DEVELOPMENT OF TOOLS FOR THE EVALUATION OF RISK OF BIAS IN NUTRITION STUDIES

Shannon Kelly

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for the
Doctorate of Philosophy

School of Epidemiology and Public Health

Faculty of Medicine, University of Ottawa

Abstract

Dietary exposure assessment is an essential concern for evaluating risk of bias in human nutrition studies, yet existing tools do not consistently incorporate nutrition specific criteria. While the principles and methodologies of evidence-based medicine globally apply to nutrition topics, diet-related challenges in nutrition studies necessitate special consideration. These include the intricacy in obtaining comprehensive and accurate estimates of exposure, the potential for complex confounding and interactive effects of food substances, and the impact of individual baseline nutritional status on the relationship between exposure to a particular food substance and any health outcome. This thesis comprises three main sections: 1) To address the limitations of existing bias assessment tools for dietary exposure assessment, an eight step process was used to develop and test a set of valid and reliable nutrition-specific risk of bias tools for randomized controlled trials, cohort studies and case-control studies, including detailed guidance and a worksheet; 2) As a continuation of this research, foundational work for the development of a risk of bias tool in cross-sectional studies, a scoping review of tools, methodology and bias concepts was conducted to inform the future development of nutrition-specific tools for cross-sectional studies; and, 3) To apply the findings of the scoping review, the rationale and scientific approach to be used for the development, testing and establishment of valid, reliable and nutrition-specific tools applicable to descriptive and analytic cross-sectional research designs is described. The intent is to assemble a suite of risk of bias tools referred to as the NUtrition QUality Evaluation Strengthening Tools (NUQUEST) with integrated consideration for dietary exposure assessment. The availability of these new tools will foster consistent and transparent assessment of potential biases commonly encountered in human nutrition studies. Future research will expand NUQUEST to develop tools applicable to other study designs and extend testing to inform systematic refinement if needed.

Contribution of the Authors

The following publications are included as part of this thesis, three of which have been published.

1. **Kelly SE**, Greene-Finestone LS, Yetley EA, Benkhedda K, Brooks SPJ, Wells GA, MacFarlane AJ. *NUQUEST-NUtrition QUality Evaluation Strengthening Tools: development of tools for the evaluation of risk of bias in nutrition studies.* Am J Clin Nutr 2022;115(1):256-71. doi: 10.1093/ajcn/nqab335.

2. **Kelly SE**, Benkhedda K, Brooks SP, MacFarlane AJ, Greene-Finestone LS, Skidmore B, Clifford TJ, Wells G. *Risk of Bias in Cross-Sectional Studies: Protocol for a Scoping Review of Concepts and Tools.* MethodsX 2024;12:102610. doi: 10.1016/j.mex.2024.102610

3. **Kelly SE**, Brooks SPJ, Benkhedda K, MacFarlane AJ, Greene-Finestone LS, Skidmore B, Clifford TJ, Wells GA. *A scoping review shows that no single existing risk of bias assessment tool considers all sources of bias for cross-sectional studies.* Journal of Clinical Epidemiology 2024;172. doi: 10.1016/j.jclinepi.2024.111408.

A fourth manuscript is planned for research still in progress when final results are available.

4**.** **Kelly SE** et al. *NUQUEST-NUtrition QUality Evaluation Strengthening Tools: development of tools for the evaluation of risk of bias in cross-sectional nutrition studies.* [Foundation and strategy are presented in this thesis]

All manuscripts were led by Shannon Kelly and were co-authored by at least one supervisor (George Wells; Tammy Clifford). Shannon is the first author on all publications, having had primary responsibility for study design, data collection, analysis, interpretation and writing of the manuscripts. The authorship criteria of the International Committee of Medical Journal Editors (ICMJE) were followed to identify the authors' specific contribution for each of the publications. All research was designed in consultation with members of the thesis advisory committee (Doug Coyle, Janet Martin, Anthony Tang).

Acknowledgments

George, I've struggled to find the appropriate words to say thank you. It doesn't seem enough. You have received every compliment, prize, and accolade over the years - each one fitting, none exaggerated. We seemed to connect effortlessly from the very beginning, and you recognized potential in me that I had not yet seen in myself. Your wisdom, kindness and humble brilliance amazes me every day. As a mentor, you bring out the best in me. You are always present with an ear to listen, a brain to pick, or a push in the right direction. You never let me give up on this work or myself, even when I wanted to many times. Thank you. Most of all, *thank you for becoming a friend*. Long may we share "tai chi" lattes, oysters, cat stories and memories. The best is yet to come.

To Tammy: You have been there for me, and *with* me, as a mentor since the start of my academic journey. You gave me opportunity, advice, and your time when I most needed it, and your support has never wavered as we have both navigated many life and career changes. I am grateful for each and every time you offered a different perspective, encouragement, ideas, or simply shared your passion and enthusiasm for whatever topic we were discussing. Our meetings never feel like meetings. I'm glad we have had this journey together and I hope it continues.

To Patrick and Erin: We have come together, and sacrificed much as a family, so that I could see this work through to completion. Thank you for your endless love, patience, and support. You mean the world to me. You picked me up every time I fell, through sleepless nights and stressful days, and constantly gave me the motivation to believe in myself. You did not knowingly sign up for this academic marathon, but you were always there on the sidelines cheering me on and keeping me running. I know it was not always easy, but I also know now that we can make it through anything *together*. Without you clearing the cats out of my office every day, this PhD would never have been finished!

Much love and an enormous thank you to my parents for always being there with nothing but encouragement, interest, and love (and warm muffins!). You were steadfast in your support and always provided a great sounding board for me... and thanks for only occasionally asking if I was "done yet."

I am deeply grateful to the members of the Nutrition Quality Evaluation Strengthening Tools (NUQUEST) working group, without whom this thesis would not have been possible. Drs. Amanda MacFarlane, Linda Greene-Finestone, Steve Brooks, Karima Benkhedda and Elizabeth Yetley (retired) - Thank you for your invaluable contributions to this thesis and for being such inspiring colleagues, mentors, and advisors. I'm thankful that a seemingly random meeting all those years ago ended up being the start of such a meaningful connection. You are all fantastic for different reasons and continue to support me in your own unique ways. I am honoured to have had the opportunity to work with each of you.

Thank you to my thesis advisory committee, Drs. Doug Coyle, Anthony Tang and Janet Martin. You helped me pivot when I needed to pivot, challenged me and provided insightful and valuable feedback as I planned and carried out this work. Every time we were together, I learned. I am a better researcher, thinker and communicator because of you, and for that I am grateful.

Thank you for keeping me out of the weeds and helping me find the right part of the pasture to focus on. I hope we are able to continue to collaborate in the future.

To each and every individual who contributed to the development and testing of this research, no matter how small the role, THANK YOU! Your input, time, feedback, resources and wisdom were valued and appreciated.

To all my colleagues and peers in the Cardiovascular Research Methods Centre and the University of Ottawa— Your dedication behind the scenes ensured that I could focus on my research. Thank you for helping me celebrate every milestone, for always encouraging me and for your friendship.

I would like to extend my heartfelt gratitude to Anne Gray who supported me behind the scenes. Anne - you do so much, for so many, and are the 'silent' member of every research project. Thank you for all that you do!

Finally, thank you to those who helped to fund this journey - The University of Ottawa, the Province of Ontario, the School of Epidemiology and Public Health, Health Canada and the Agriculture, Food, and Nutrition Evidence Center at Texas A&M.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

CCHS	Canadian Community Health Survey
CHMS	Canadian Health Measures Survey
NHANES	National Health and Nutrition Examination Survey
NRSI	non-randomized study of interventions
NUQUEST	NUtrition QUality Evaluation Strengthening Tools
PROSPERO	International prospective register of systematic reviews
RCT	randomized controlled trial
RoB	risk of bias
SR	systematic review

Chapter 1
Introduction

1.1. Statement of the Problem

Nutrition studies inform health policy decisions and recommendations related to diet and health. Still, the utility for knowledge users is limited without a fulsome assessment of the challenges associated with measuring and assessing diet in nutrition studies. These unique and complex nutrition-related considerations include baseline nutrient exposure, nutrient status, bioequivalence of bioactive compounds, bioavailability, multiple and interrelated biological functions, undefined nature of some interventions, and uncertainties in intake assessment (1). Currently available risk of bias assessment tools inadequately account for a number of key nutrition-specific concepts (2). Research has shown that risk of bias assessments in systematic reviews of nutrition studies are not appropriate or comprehensive and that commonly used tools neglected to address all important sources of biases (3). The same systematic reviews frequently misapply reporting constructs unrelated to risk of bias (3, 4). When there is insufficient consideration for important biases, this can result in misleading interpretations and erroneous conclusions (1, 5-10).

The underlying hypothesis of this thesis is that the development of nutrition-specific risk of bias tools will lead to an assessment approach for nutrition studies that integrates the judgement of methodological conduct and empirical evidence of bias for a study with nutrition-specific criteria. This will aid both non-nutrition scientists (e.g., research methodologists, epidemiologists) who may be unfamiliar with nutritional issues, and nutrition scientists who may lack in-depth methodological expertise in human study design and conduct. Nutrition-specific

risk of bias tools will permit consistent and transparent assessment of risk of bias in nutrition studies and will contribute to decreasing the uncertainty and increasing our confidence and ability to establish nutrition - health outcome associations (1, 4, 7, 9). This will be valuable to individuals, groups or organizations responsible for developing science-based recommendations and policies.

1.2. Objective

The overall objective of this thesis is to develop a set of reliable and valid risk of bias (RoB) tools specifically tailored for nutrition research with the aim to enhance the rigor and reliability of RoB assessment by addressing unique methodological challenges inherent to the nutrition field. These tools will be developed for study designs commonly used in the field of nutrition which will provide nutrition research users with robust, user-friendly tools to appraise studies. The tools will be applicable to single research studies or as part of an evidence synthesis, and this will contribute to an improved understanding of the evidence base for nutrition or dietary recommendations and health policy decisions.

1.3. Research Questions

In order to meet this objective, this thesis addresses two research questions:

Question 1*:* How can nutrition-specific risk of bias assessment tools be developed to address the unique methodological challenges posed by randomized controlled trials (RCTs), cohort studies, case-control studies, and cross-sectional studies in nutritional epidemiology?

Question 2: To what extent are the newly developed nutrition-specific risk of bias tools valid and reliable when applied across different study designs in nutritional research,

and do any additional components developed alongside the nutrition-specific risk of bias tools enhance reliability or usability?

1.4. Chapter Outlines

This thesis comprises 6 chapters. Chapter 1 provides an overall introduction and the rationale for the thesis, while Chapter 2 provides an overview of risk of bias and related issues in human nutrition research. Chapters 3, 4 and 5 present manuscripts that address the research questions. Of note, Chapter 4 contains two related manuscripts. The final chapter of the thesis, Chapter 6, integrates the major findings from the thesis and considers remaining knowledge gaps and implications for practice and policy.

The following sections present a brief overview of each chapter:

1.4.1. Chapter 1: Introduction

This introductory chapter provides a summary of the rationale for the thesis, the overall objectives of the thesis, and a brief description of each subsequent chapter.

1.4.2. Chapter 2: Background

This chapter provides relevant background information about key topics relevant to the thesis, including an overview of the study of diet and nutrition, broad concepts of quality and risk of bias in research, the challenges and complexities with nutrition research and a rationale for why nutrition-specific risk of bias tools are needed.

1.4.3. Chapter 3: Nutrition-specific RoB tools for RCTs, cohort studies and case-control studies

This chapter contains one manuscript that describes the development and testing of a series of nutrition-specific risk of bias tools intended for randomized controlled trials, cohort studies and

case-control studies. The title of the manuscript is *NUQUEST - NUtrition QUality Evaluation Strengthening Tools: development of tools for the evaluation of risk of bias in nutrition studies*.

1.4.4. Chapter 4: Risk of Bias in Cross-Sectional Studies: A Scoping Review of Concepts and Tools

This chapter contains two related manuscripts (Sections 4.2 and 4.3) that describe a scoping review of concepts and tools relevant to the assessment of risk of bias in cross-sectional studies and the related protocol. The title of each manuscript is listed below:

Section 4.2: *Risk of bias in cross-sectional studies: Protocol for a scoping review of concepts and tools*

Section 4.3: *A scoping review shows that no single existing risk of bias assessment tool considers all sources of bias for cross-sectional studies*

1.4.5. Chapter 5: Nutrition-specific RoB tools for cross-sectional studies

The chapter contains the background, rationale and methods for a strategy which describes the development of two nutrition-specific risk of bias tools for cross-sectional studies. The first is intended for analytic cross-sectional studies and the second for descriptive or prevalence studies. The anticipated title of the anticipated manuscript in draft is: *NUQUEST - NUtrition QUality Evaluation Strengthening Tools: development of tools for the evaluation of risk of bias in cross-sectional studies.* The methods for tool development have been established and a framework informed by Chapter 4 will allow for the creation and testing of the two planned nutrition-specific tools for cross-sectional studies. Completion of the tools is expected in the first half of 2025 with testing and a final version of each tool available later in 2025.

1.4.6. Chapter 6: Discussion

This summary chapter integrates the findings of the individual chapters/manuscripts and considers the implications for policy and practice. Limitations of the thesis are discussed and recommendations for future studies are proposed.

1.5. References

1. Lichtenstein AH, Yetley EA, Lau J. *Application of Systematic Review Methodology to the Field of Nutrition*. The Journal of Nutrition. 2008;138(12):2297-2306. doi:10.3945/jn.108.097154.

2. Chung M. Translation and application of evidence-based methodologies to the field of nutrition: ProQuest Dissertations & Theses; 2010.

3. Zeraatkar D, Bhasin A, Morassut RE, Churchill I, Gupta A, Lawson DO, Miroshnychenko A, Sirotich E, Aryal K, Mikhail D, Khan TA, Ha V, Sievenpiper JL, Hanna SE, Beyene J, de Souza RJ. *Characteristics and quality of systematic reviews and meta-analyses of observational nutritional epidemiology: a cross-sectional study*. The American Journal of Clinical Nutrition. 2021;113(6):1578-1592. doi:10.1093/ajcn/nqab002.

4. Zeraatkar D, Kohut A, Bhasin A, Morassut RE, Churchill I, Gupta A, Lawson D, Miroshnychenko A, Sirotich E, Aryal K. *Assessments of risk of bias in systematic reviews of observational nutritional epidemiologic studies are often not appropriate or comprehensive: a methodological study*. BMJ nutrition, prevention & health. 2021;4(2):487.

5. Chung M, Balk EM, Ip S, Raman G, Yu WW, Trikalinos TA, Lichtenstein AH, Yetley EA, Lau J. *Reporting of systematic reviews of micronutrients and health: a critical appraisal*. The American journal of clinical nutrition. 2009;89(4):1099-1113.

6. National Academies of Sciences E, Medicine. *Guiding principles for developing Dietary Reference Intakes based on chronic disease*. 2017.

7. National Academies of Sciences E, Medicine. *Dietary Reference Intakes for sodium and potassium*. 2019.

8. Trepanowski JF, Ioannidis JP. *Perspective: limiting dependence on nonrandomized studies and improving randomized trials in human nutrition research: why and how*. Advances in Nutrition. 2018;9(4):367-377.

9. Yetley EA, MacFarlane AJ, Greene-Finestone LS, Garza C, Ard JD, Atkinson SA, Bier DM, Carriquiry AL, Harlan WR, Hattis D. *Options for basing Dietary Reference Intakes (DRIs) on chronic disease endpoints: report from a joint US-/Canadian-sponsored working group*. The American journal of clinical nutrition. 2017;105(1):249S-285S.

10. Zeraatkar D, Johnston BC, Guyatt G. *Evidence Collection and Evaluation for the Development of Dietary Guidelines and Public Policy on Nutrition*. Annu Rev Nutr. 2019 Aug 21;39:227-247. doi:10.1146/annurev-nutr-082018-124610. Cited in: Pubmed; PMID 31433741.

Chapter 2
Background

2.1. Introduction

Relationships between nutrition and diet and human health have been formally and informally observed for hundreds of years (1). Nutritional epidemiology permits a more structured approach to the science of studying the relationships between dietary intake and human health and disease, which is used to inform policy and practice globally (2). Although the first formal clinical trial documented in history was in fact a nutrition study, the science of nutrition is still considered young (3, 4). In 1753, naval surgeon James Lind published his now famous *A treatise of the scurvy* in which he reported results from his experimentation cures for scurvy at sea (4). Lind was lauded when he was able to definitively show in a clinical trial that oranges and lemons could protect against the disease ravaging seamen. Despite this, most of what we know about nutrition has been discovered in the past century, over 175 years after Lind's discovery (1, 5-7). Developments over the last century were intrinsic in shaping the foundations for modern nutritional science. During World War II, Elsie Widdowson and Robert McCance pioneered efforts to prevent nutritional deficiencies related to food rationing by improving our understanding of population health, food composition and disease prevention using nutrition science (5). By providing evidence that rations must be adequately nutritious to maintain health and prevent disease and deficiency, Widdowson and McCance influenced both politics and policy to support provision and rationing decisions. In Canada, such research led to the original 1942 version of Canada's Food Guide as a national dietary recommendation (8).

Nutrition as a science continues to advance (9, 10). Scientific advancements, notably in the last 50 years, have led to the isolation, identification and synthesis of all essential known vitamins and minerals, and pushed forward knowledge of how micro- and macronutrients may lead to diseases of deficiency in animals and humans which has been foundational to informing policy and practice (3). As a science, nutrition considers some food components as essential (and without which humans may develop diseases of deficiency) and other food and food components that may be associated with disease risk in humans (11). While historical definitions of nutrition science considered it a biological life science "*concerned with how food and nutrients support human metabolism in health, prevention and management of disease*" (9), expansion of this definition to a conceptual framework including greater breadth was proposed by the Giessen Declaration in 2005 (12, 13). These guiding principles around the science of nutrition now more fulsomely consider the personal, population and planetary health and the intricate complexity of food and drink, food systems, nutrients alongside other constituents and their connections with other relevant domains including biological, social and environmental systems (12, 13).

In Canada, the evolution of nutrition science is also founded in pioneering scientific work by several scientists and a progressing regulatory and public health structure supported by all levels of government and academia. Nutrition and diet were famously an intrinsic component of Charles Best and Frederick G. Banting's scientific research in the 1920s establishing insulin as an effective treatment for diabetes. In his 1923 Nobel Lecture, Banting highlighted the pivotal role of dietary intake and timing, how macro- and micronutrients influenced insulin response, and how marked differences in individual risk and response vary based on facets of food intake or patterns (14). At the time this was a significant advancement in not only treatment of disease, but disease management using nutrition science. In the 1970s, the federal government created the Nutrition Canada survey with a mandate to address nutrition-specific knowledge gaps through

research and education with the idea that structured, organized science and education in the field would lead to public health and regulatory initiatives nationally (15). This evolved into what Canada now has federally housed as part of the Health Products and Food Branch at Health Canada.

Advancement in our understanding of nutrition science was also followed by development of large-scale national surveys and other epidemiological studies to capture critical data into diet, lifestyle and health in humans. One example of this is National Health and Nutrition Examination Survey (NHANES) designed to assess the health and nutritional status of children and adults in the United States (16, 17). NHANES began its program in 1960 and continues as a series of cross-sectional surveys focused on different health topics or population groups. As of 1999, NHANES began to change focus with every survey cycle to reflect emerging needs and priorities. Each year they selected a nationally representative sample of about 5,000 citizens. In Canada, the Canadian Community Health Survey (CCHS) was initiated in 2001 to collect cross-sectional data on health status, utilization of health services and determinants (18). The CCHS has a two-year data collection cycle, with certain cycles focusing on specific populations or topics of interest. Although the annual CCHS component is limited in nutritional data collection (i.e., consideration only for fruit and vegetable intake thresholds), it does consider some measures of nutrition status such as food insecurity, dietary supplement use and infant feeding alongside other key indicators of health (19). To date, two cycles of the CCHS, in 2004 and in 2015, have focused on food and nutrition surveillance in populations in all ten Canadian provinces, excluding populations who are institutionalized, active Canadian Forces members and persons living on reserves or other Indigenous settlements (20, 21). Data collected included food and dietary supplement consumption and anthropometric measures (self-reported weight and height) using a representative sample of Canadians at national and provincial levels. Information

collected is used by a variety of knowledge-users including federal and provincial policy-makers, health providers, researchers and the commercial food, beverage and supplement industry (22, 23). Historical shifts in nutrition science have seen the evolution from single nutrient research into the study of overall diet (24). In turn, cross-sectional surveys such as CCHS have likewise moved away from measurement of individual nutrient and energy intake towards better understanding of nutritional status and intake patterns, and population-level changes over time. In 2007, the first cycle of the Canadian Health Measures Survey (CHMS) was initiated (25, 26). The CHMS was designed similarly to NHANES and includes key information relevant to the health of Canadians, including nutritional status. Unlike CCHS, CHMS utilizes direct physical measurements for attributes of health in respondents, including blood pressure, height, weight and physical fitness. These are especially important for accurate measurements of adiposity. The survey also collects blood and urine samples to test for chronic and infectious diseases, nutrition and environment markers and limited data on socioeconomic status (25, 26). CHMS allows for assessment of the nutritional status of various nutrients (e.g. vitamin D, iron, vitamin B12) and contains a limited food frequency questionnaire. It continues to survey Canadians every two years (27). Surveys and research programs such as NHANES, CCHS and CHMS are critical to monitoring nutritional adequacy, distribution and diet quality which are used to estimate prevalence and risk factors of disease and to understand how diet and nutritional status are associated.

2.2. Why Study Diet and Nutrition?

Diet refers to the total amount of food and beverages consumed by individuals to obtain energy and nutrients (i.e., intake); while *nutrition* considers the utilization of dietary intake for growth, metabolism and repair of tissues (28). Both can have profound impact on the wellbeing and health of humans and health systems. As such, nutrition research has the potential to

substantially change human health, care systems and global economies in a positive way (29). Clinical nutrition as a discipline encompasses prevention, diagnosis and management of acute or chronic disease and conditions attributable to deficiency or excess of energy and nutrients (30). Individuals with acute or long-term health conditions or disease require adequate nutritional support, and this can be an essential component of treatment, recovery, rehabilitation or palliative care in any professional or home-based care setting. Malnutrition in acute or chronically ill individuals can lead to additional burden of disease, further exacerbation or progression of disease, or delay of recovery, which impacts both individuals and health systems. Nutritional excess, either in volume or type of energy, nutrient, food, or food pattern can lead to chronic inflammatory conditions, cardiovascular or metabolic disease, imbalances and even toxicity.

Modern nutrition science revolves around determining relationships between food, nutrition and human health, and involves intricately understanding the complexity that may influence those relationships. *Optimal nutrition* is a broad term which conceptually represents a strategy to mitigate the burden of poor health status and disease in humans and is pivotal to informing the underlying causes of major non-communicable diseases such as obesity, cardiovascular disease, diabetes and cancers (31, 32). There is no consensus as to what specifically defines optimum nutrition. Ideally, humans should consume enough essential nutrients to avoid any related deficiencies, but adequate intake may be insufficient to reduce risk for chronic disease. In the same way, humans do not become deficient in non-essential food substances, rather, intakes at varying levels may contribute to reduced disease risk or increase risk of related toxicity(33). Overarching components of nutrition and health research focus on nutrition to maximize health, acute or long-term intrinsic or extrinsic exposure across all life stages, and potential disorder due to disease, malnutrition or obesity (29). This may encompass study of disease risk factors,

identifying optimal nutritional intake and understanding how nutritional status, intake or exposure or dietary patterns influence human health. This knowledge is used to inform national dietary and intake guidelines and recommendations, program delivery, regulatory and food policy, public health campaigns, clinical management, health service delivery and individual behaviour (29).

A range of intake - health associations, including human growth and development, reduced risk for chronic disease and the positive or negative influence on overall health throughout one's life, are of interest to policy and practice (29). Commonly studied exposures can range from energy, essential food substances (i.e., micronutrients, macronutrients), non-essential food substances (e.g., bioactive compounds), whole foods (e.g., foods, beverages), or food groups, dietary patterns and the non-nutritive components of food or beverages (e.g. artificial sweeteners) (5-7, 34). Examples of each are provided in Table 2-1.

Table 2-1: Common nutrition exposures and examples

Nutrition Exposure:	**Examples:**
Micronutrients	Essential vitamins and minerals (examples: folate, vitamin C, calcium, zinc, iron)
Macronutrients	Carbohydrate (examples glucose, fructose), protein, fat
Bioactive compounds	Flavonoids, polyphenols
Foods or beverages	Milk, beef, eggs
Food groups	Fruits, vegetables, meats and alternatives, grain
Dietary patterns	Western dietary pattern, low carbohydrate diet, Mediterranean diet
Non-nutritive components of foods and beverages	Caffeine, flavour additives, food additives such as dyes or preservatives

Note: Table adapted from Zeraatkar, 2021 (35)

Micronutrients are essential vitamins and minerals required by humans to sustain healthy body function (36). These nutrients influence and perform a range of bodily functions used for normal human growth and development, including hormone and enzyme production. A deficiency of any one essential micronutrient can result in a disease of deficiency, including severe or life-threatening diseases. The World Health Organization (WHO) notes that micronutrient deficiencies disproportionately affect low and middle income countries internationally, and the burden of these deficiencies often rests on pregnant women and children (37). Micronutrient deficiencies are also common in high-income countries where they contribute to risk of chronic disease or poor health status (e.g., osteoporosis, anemia, pregnancy complications). This is still a common issue in the very young and very old and other sectors of the population (29, 37).

Macronutrients are nutrients that provide energy. Examples include carbohydrates, fats and proteins. Nutrition recommendations focus on achieving a range of macronutrient intakes that supports healthy growth and development and to prevent chronic disease while achieving recommended energy intakes (36, 38). Recommended macronutrient intake ranges differ by life stage and health status. Global differences may be attributable to differences in refined and unrefined food intake, traditional diets and divergent views as to how macronutrients effect human health risk and outcomes. Obesity is increasing in the general population year after year and it is thought that excess energy intake specifically in processed or energy-dense macronutrients is fueling this problem. Increases in metabolic disease are tied to obesity as is the risk of other diseases (e.g., type 2 diabetes, heart disease, a number of cancers) and premature mortality.

Bioactive compounds are substances that occur in nature, can be derived from foods and that have an effect on human health (39). These chemical compounds are thought to have health

benefits beyond that of the nutritional value of the edible plant, food or beverage in which they are found and may play a role in reducing risk of disease in humans. Research on bioactive compounds is broadly focused around their potential for anti-inflammatory, anti-carcinogenic, anti-oxidant, antimicrobial and cardio- or neuroprotective properties (36).

Foods and beverages are individual items that humans may ingest such as bread, beef, cereal, juice or milk. Dietary intake is generally estimated by assessing the usual intake of foods (vs. individual nutrients). From an assessment of usual food intake, one can estimate the usual nutrient intakes based on data from food composition databases. Often, the quantity, quality or avoidance of particular foods are considered in addition to nutrients, but this is complicated by *combination foods* such as pizza, lasagna or casseroles which include many different foods, which makes it difficult to accurately estimate the totality of exposure, or to explore adequacy of intake or complications or risk of deficiency (40).However, relative intakes (i.e., higher or lower) can also be useful for examining associations. Historically, dietary intake of beverages, which are not part of any food group per se, has focused on milk, 100% fruit juice and alcohol (41, 42). However, the proliferation of energy drinks, sport drinks and coffee culture has expanded interest into related human health effects in recent years (43-45). *Food groups* are categories of foods (e.g., grains, milk products, meat poultry and fish, fruit and vegetables) studied and often used to provide population level healthy eating recommendations by countries around the world (46). The global explosion in the production and intake of processed and snack foods that do not fit into these component food groups fit into a class of '*other´* foods as a group (47, 48).

Dietary patterns are defined as the "quantities, proportions, variety, or combination of different foods, drinks, and nutrients in diets, and the frequency with which they are habitually consumed" (49). There has been an emphasis in the last 20 years in designing dietary recommendations

reflecting dietary patterns that achieve nutrient adequacy and reduce the risk of chronic disease (50, 51) and it is essential to consider this and the related complexity of overall diet when considering health outcomes. Dietary patterns are seen as a more comprehensive, and potentially stronger, way of understanding how nutrients and foods modify health or risks of chronic disease and facilitate the development of guidance that can be easily communicated at the population level. It may also provide a complimentary or alternative way for scientists to understand nutrient and food intakes as they can examine nutrient content of constituent foods and beverages, related quality of the diet, macronutrient and food distribution (34, 52). Scientists are also exploring "*precision nutrition*" using dietary patterns to predict health outcomes or response in susceptible subgroups of the population while also considering genetics, microbiome, metabolic profile, health status, physical activity, food environment or related socioeconomic and psychosocial characteristics (53) "*Personal nutrition*" focuses on similar considerations at the individual level (54).

2.3. The role of research in nutrition

Nutrition research has an essential and multifaceted role in advancing our understanding of human health and how various nutrients and dietary patterns effect health status and quality of life (55-57). Since all humans must eat, and what we eat can influence our health, nutrition research is of critical importance. Nutrition can also be considered as an outcome of exogenous influences since nutritional status may be connected to education, hygiene, economic growth, sanitation and gender roles globally (58). As in broader healthcare policy and practice, decision-making in the field of nutrition relies heavily on the available scientific evidence, much of which is carried out using epidemiological study designs (59). In nutritional epidemiology, many types of relationships may be hypothesized or explored with the aim of understanding how dietary factors and human health outcomes are associated (1, 10). These hypothesized relationships

influence the design of experimental studies and guide how these associations are investigated. By exploring these relationships, epidemiological research contributes valuable evidence to inform public health policies, dietary guidelines and personalized nutrition recommendations aimed at improving population health and reducing the burden of chronic diseases (57, 60, 61).

As the field has evolved, so too has the amount of evidence available for knowledge users to translate into policy and practice (62). Owing to efficiencies and advancements in recent decades which permit and facilitate collaboration, sharing and digital access to resources, the evidence available for policy and practice is often plentiful and accessible. However, it is not always obvious which evidence is inaccurate, flawed or when health information communicated is based on misinformation, misinterpretation or pseudoscience (i.e., statements, beliefs or practices that claim to be both scientific and factual but are incompatible with the scientific method) (63). So, translating evidence into practice is arguably a difficult task in the field of nutrition.

The various approaches used to study epidemiological relationships in the field of human nutrition are summarized in Table 2-2.

Table 2-2: Approaches to the study of diet-health relationships

Approach	Hypothesis
Causal relationships	The consumption of a specific food, macro- or micronutrient or dietary pattern causes changes in human health outcomes.
Dose-response	There is a graded relationship between the amount or frequency of dietary exposure and health outcomes.
Protective relationships	Certain dietary patterns or foods confer protection against specific diseases or health conditions.
Risk relationships	Certain dietary factors increase the risk of developing specific diseases or health problems.

Approach	Hypothesis
Synergistic effects	The combined effect of multiple dietary factors or interactions between diet and other variables (e.g., genetics, lifestyle) influences health outcomes.
Mediation of effects	Dietary factors mediate the relationship between certain exposures (e.g., environmental factors, lifestyle behaviors) and health outcomes.
Moderation of effects	The effect of dietary factors on health outcomes varies based on individual characteristics (e.g., age, sex, genetic background).

A nutrition research study will broadly focus on the intake of a food substance, a whole food, dietary pattern or biomarker(s) of nutritional status or intake. In the context of a nutrition study, the interventions in a randomized controlled trial (RCT) may be based on intake of a nutrient, food or supplement, a specific dietary pattern or a combination (64). For example, an intervention group may consume a fibre supplement of 25g per day and the control group a placebo or alternative fibre supplement. Likewise, researchers may study the effects of the intervention group consuming a Mediterranean diet and the control group could consume a 'usual' or other named dietary pattern (e.g.,, a keto diet). Some RCTs may have nutritional status or intake as an outcome rather than as the intervention. For example, the intervention group may receive an educational program on the health benefits of fibre and the control group would receive no program, with the outcome being fibre intake. In an observational study, where a nutritional exposure is the focus, researchers may examine higher versus lower intakes or nutritional status, supplemented versus non-supplemented diets, Mediterranean versus Western diets, or low or high intake of a certain food group, food or nutrient. For example, researchers may explore the association of fibre intakes and the risk of diverticulitis in a cohort study. As with RCTs, the outcome in an observational cohort study may also be a change in intake or nutritional status associated with a particular behaviour intervention.

Nutrition science has been widely criticized for generating implausible claims, being overly reliant on small trials and observational studies and it has been stated that nutrition science is 'methodologically inferior' to other fields (24, 65-67). In reality, nutrition science is no less reliable than other disciplines but there are specific, complex, nuances related to dietary intake which, if not adequately considered in primary research, result in potentially biased estimation of human health effects, and likewise, misleading interpretation (2). Assessment of latent bias is an essential function of epidemiological science when the objective is to understand a relationship, or type of relationship, between an intervention or exposure and a human health outcome (68). Moreover, to understand certainty in these relationships, assessment of risk of bias is a foundational part of any well-conducted systematic review (SR) (69). SRs aim to transparently and rigorously approach the synthesis of evidence for any given research question using methods which minimize bias. They can be informative and independent components of many decision-making processes for organizations responsible for developing evidence-based guidelines, policies and programs (55). However, when nutrition studies are considered in meta-analyses by authors of SRs, bias in the individual studies included may be exacerbated or perpetuated if the nutrition-specific biases are not appropriately considered. This can lead to dissemination of misleading or false conclusions about the diet health relationships under study (70-72).

2.4. Defining critical appraisal, study quality and risk of bias

Detecting and understanding potential limitations or bias in the planning, conduct and dissemination of research allows evidence users to critically evaluate the validity of scientific research and more effectively translate evidence into policy and practice. This is an essential component of the SR process (69). Critical appraisal, study quality and risk of bias may be broadly used interchangeably (73) yet there are some important nuances in how each may be conceptually interpreted and some have argued that it is important to distinguish between the

three accordingly (74). *Critical appraisal* was first proposed within the paradigm of evidence-based medicine by Sackett et al. as a structured approach to evaluate the medical research for trustworthiness, value and relevance for any topic of interest (75, 76). In this context, the medical literature is assessed based on a clearly focused research question, validity of the methods, importance of the findings and their applicability to a target patient population. This approach historically has been a mainstay for medical journal clubs and clinical educational settings (77). Related to this, s*tudy methodological quality* involves assessment of whether a research study has been designed and executed according to the highest possible standards. More broadly, the concept of *study quality* has also evolved over time to often include consideration for items not related to the validity of the study, including sample size and power, analytic or statistical approaches, reporting quality or ethical components (i.e., approval by an institutional research ethics board).

While grassroots use of critical appraisal and study quality assessment was founded in clinical education and practice in the medical field (78, 79), there has been continual development of methods and practice in this area across fields. Conceptually, although critical appraisal and study quality can be informative to any synthesis process, even when a study is designed and implemented to the highest possible methodological standards for a given research design, it still may be at risk for bias. In 1996, Moher et al. first suggested bias as a concept and suggested that assessment of bias was required to identify trial characteristics directly related to bias (80). *Risk of bias (RoB)* assessment was conceptually proposed by the Cochrane Collaboration as a systematic way to understand both the intrinsic systematic error associated with a research study and how this error may influence the findings, both in direction (i.e., towards or away from the hypothetical *true* effect) and magnitude. The group encouraged formal and structured consideration for bias and related methods groups began to produce empirical related methods

groups began to produce empirical evidence on the influence of various study elements on study findings (81-84).

Study reporting is a concept that is often confused with study quality and risk of bias. A study that is reported according to the highest standards or guidelines is not bias free or high quality for simply meeting these requirements (85, 86). However, a study that is reported well often permits a more complete and transparent evaluation of potential for risk of bias and so therefore should be encouraged (74).

Bias can arise at many different points during the design and execution of a research study, and researchers must consider how any issues may contribute to problems with the internal validity of the results. In fact, bias may arise due to actions from study investigators or it may be unavoidable due to the practical restrictions on how a research project could be implemented in any particular setting or clinical area (87). Having some type of empirical evidence of the potential for systematic bias assists researchers to understand how the effects from that study may be over- or under-estimated and the magnitude of the bias accordingly (69). Risk of bias can be evaluated for a single study independently, or for all studies forming the evidence base within a SR. Assessment of risk of bias in the individual studies is an essential component in a SR, regardless of the study designs considered (69). This is an established process and is ideally conducted using a reliable and valid risk of bias tool which considers all applicable potential biases that may impact the validity of the study findings. Risk of bias may additionally be influenced by conflicts of interest from funders or researchers affiliated with the study (87).

In order to determine causal relationships between dietary intake and a health outcome, ideally there is evidence from RCTs, the gold standard for establishing causality as well as for establishing intake-response relationships. Alternatively, and in an even more ideal situation,

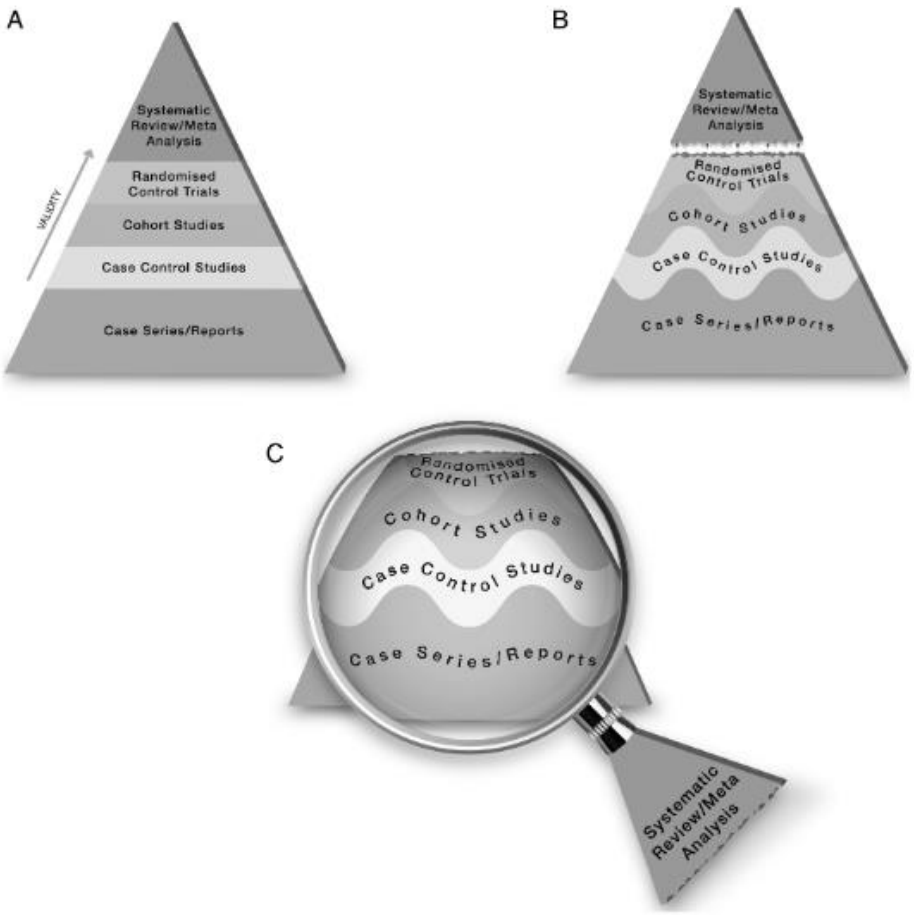
there would be several RCTs which would then inform a SR and meta-analysis of the evidence. Where evidence is sufficient for any outcome, where quantitative pooling is appropriate and where we the risk for potential bias is assessed to be low, then we can have a high level of confidence that you have a relationship between whatever the nutritional intake is under study and health outcomes.

Often however, the nature of evidence for the health outcomes we are interested in relates to chronic disease or other areas which require a study of association which is observational by design. Since observational studies are subject to inherit systematic errors and are prone to biases that do not impact randomized trials, they are often thought of as being lower value evidence.

In epidemiology, we often consider and rank the commonly used study designs using a so-called, hierarchy of evidence which has long-informed clinicians and their consideration for appraisal of evidence to inform practice (88). Levels of evidence were first described in a report by Canadian Task Force on the Periodic Health Examination in 1979 (89). Conceptually, a SR and meta-analysis is at the top of a pyramid of evidence, indicating the highest caliber of evidence, followed by the gold standard for causal relationships in primary study designs, the experimental RCT. This is followed by cohort studies and then case-control studies and so on, which infers that an RCT is always better than a cohort study (Figure 2-1A). Theoretically, this figure is an effective way to visualize the concept, but this does not hold true for any primary or secondary research design. There will always be variability in the design and conduct of all primary studies, and therefore, in the direction and magnitude of the biases that impact them (Figure 2-1B). In practice, within each study design label, there will be different specific types of studies, and varying degrees of methodological quality and RoB. Across study designs, you may have a good

quality study at a lower level of the evidence hierarchy which may provide less-biased estimates of causation or association than a study design placed higher on the hierarchy which may be less robust or at a higher risk of bias (Figure 2-1C). For example, a well-designed and rigorously conducted prospective cohort study with a low risk of bias may be more informative to policy and practice than an RCT on the same topic with limitations in these areas. What constitutes 'better' evidence can vary, and therefore, we must evaluate this in a transparent, systematic, comprehensive and valid way. Moreover, certain biases may occur more frequently or in greater magnitude in nutrition studies due to the nature of intake and exposure assessment methods, which are often self-reported.

Figure 2-1: Evidence hierarchy pyramid and related considerations



Reproduced from: Murad et al. (2016)(88)(*See Appendix 1 for permissions)*

2.4.1. Key concepts of study bias

Bias in a research study of any design may lead to systematic errors and therefore to erroneous results of different magnitudes. Potential biases must be considered when a study is designed and accounted for in the design and analysis where feasible with the goal of protecting against bias and minimizing the degree to which the study effects may be impacted. Biases may be considered broadly based on what occurs at the start of study as participants are recruited or selected into study groups, what happens after the intervention is assigned or exposures are ascertained, and then how participants are followed up and outcomes are ascertained. The focus of an RoB assessment is the internal validity of the study (or studies) under review. This may be differentiated from the external validity of a study which focuses on the extent to which the results of the study can generalize to contexts outside the confines of the research environment (i.e., applicability or generalizability) (90).

Selection of participants into a study

In a RCT, randomization is used to safeguard against differences in the groups under study at baseline. Appropriately concealing allocation also ensures that researchers are not aware of which group study participants are being assigned to. Concealment of the allocation sequence is always possible, regardless of the clinical area (91). Lastly, the intervention and control groups need to be similar in terms of composition at the start of the study and differences may guide both the analysis strategy and serve as an indicator of successful randomization. When randomization is used, allocation procedures can be concealed, and group composition is balanced, and the selection of participants may be considered robust. When allocation is not concealed, research has shown that study investigators may overestimate the effects of the

intervention by as much as 40% (92). If a baseline difference is detected, the variable must be accounted for in the planned analyses. If there are differences in several variables across the intervention and control groups, randomization may not have been successful.

Bias in the selection of study participants can be particularly confusing in the context of observational studies owing to variations in the terminology and definitions used, however, what is important for selection of study participants for an observational study is not that different from that of an RCT conceptually (93). What researchers should consider is whether the groups being studied are selected from source populations that are comparable in all respects other than the exposure under investigation, if participation rates are acceptable, whether eligible subjects could have had the outcomes at the time of enrollment, and whether the exposure ascertainment is sufficient to permit the accurate assignment of participants into their study groups. There are some other key nuances by study design with regards to selection of study groups. For example, when groups are created in a case-control study, cases and controls should similarly be taken from comparable populations, but it is also important from a bias perspective to define both cases and controls in detail (and establish that cases and controls differ) and to apply the same exclusion criteria for both cases and controls. Although it is possible to plan and account for these items, it is important to note that even the most well-designed and executed observational study may still be subject to bias of varying direction and magnitude.

Comparability of the study groups

To minimize bias in an RCT it is important that participants, investigators or research staff, and even outcome assessors are unaware of the intervention allocations to minimize bias. Ideally, participants, research staff and investigators are kept ‘blind’ after the intervention allocation to minimize risk of bias (91). Once the interventions in a trial are allocated, knowledge of the study

goals should not influence behaviour changes in the study participants. For example, in an RCT looking at fibre and diverticulitis flares, a participant who starts taking fibre supplements may confound the study and invalidate the results. Likewise, dropouts of participants before the study is completed should not be of large magnitude for the entire study (i.e., ≤20%) or have a large differential between the study groups (i.e., ≤10%). These thresholds represent working rules only, with no empirical evidence for their use. In an RCT, randomization must be maintained, and the analysis of participant outcomes must be according to which they were originally allocated, regardless of whether the intervention was actually received (i.e., intention-to-treat).

There are different aspects to consider in observational studies with regards to the comparability of the study groups. Reporting participant follow-up and attrition is a good indicator of a robust study. While attrition consideration overall and by group should be similar to that for an RCT, in longer term cohort studies dropout rates may be high, and vary depending on the exposure and or outcome of interest. To investigate potential for bias, it is important to know whether those who dropped out differ in any significant way from those who remained in the study. Exposure or prognostics factors should be assessed more than once over the course of the study using robust methods and this is true for randomized and non-randomized evidence. Most importantly for an observational cohort study is that the potential confounders are defined and considered in the design and across groups and are appropriately accounted for in any analyses for a study. Otherwise, the study groups may differ on variables unrelated to the nutrition exposure under study and the validity of the study is at risk. Unmeasured or unknown confounders are a core concern for observational studies, and a reason why risk of bias must be carefully considered. Although there is no formal approach to assessing RoB for unmeasured or uncontrolled confounding, measures have been proposed to help appraise how robust or susceptible an association potentially subject to confounding is to these risks. For example, the "*E-value*" was

can inform researchers about the range of potential confounding that is required to 'explain away' a given association (conditional on the measured covariates) and this could be reported in a study alongside findings as a sensitivity analysis (94). Although measures are available and may be informative, the *E-value* requires strong assumptions, may oversimplify multifaceted epidemiological relationships, and broadly may not be sufficient or comparable to findings from more complex bias modelling when feasible (95).

For a case-control study, similarity of the cases and controls, participation rates and confounders are also important. However, bias may also be introduced if participants differ significantly in important characteristics from non-participants or if knowledge of the exposure of interest has potential to influence how cases were ascertained. The identification and handling of potentially confounding variables is an essential part of study design and analysis; however, balanced study groups may not account for all potentially confounding factors and does not eliminate the need to consider confounders factors more broadly.

Ascertainment of the Outcome(s)

When there are errors related to how study outcomes are measured, this can bias the effects of in intervention in an RCT or observational study (91). Measurement, misclassification and under/over-ascertainment errors can be avoided by ensuring that the method(s) for measuring the outcome are appropriate (i.e., reliable and valid) and do not differ between intervention and control groups. Ideally, assessment of the study outcomes is done by individuals blind to the intervention allocation or who do not have knowledge of the exposure that may influence their measurement.

Selective Reporting of Study Outcomes

Bias can arise when a reported study result is chosen in a way that supports the vested interests of the researchers and/or to increase the likelihood of publication for academic interests. Notably, it should be differentiated from incomplete reporting of study results. All study outcomes should be analyzed based on a plan which was pre-specified and fulsomely reported. In an RCT, this includes all measured timepoints, instruments and domains for an outcome, and all pre-specified analyses of intervention effects. In observational studies, selection of the outcome measure or analysis based on the nature and direction of the results and selective reporting of any subgroup from the larger cohort may indicate bias. As with RCTs, statistical analysis plans should be used in observational research to increase transparency and validity of the findings (96).

2.4.2. Tools for Risk of Bias Assessment

There are many tools available for assessing RoB in research studies which have been designed for one or more study designs and research fields or contexts. There is no widely-accepted gold-standard, although many different tools have been proposed or used over time and have become popular based on the accepted state-of-the-art (97). For RCTs, tools commonly focus on biases related to randomization and allocation processes, blinding of participants, investigators or outcome assessors, and deviations from intended interventions (91). For quasi-experimental and observational studies, focus in the available tools is broadly on selection of the study groups, confounding, comparability and exposure or outcome ascertainment (98, 99). Users may select a RoB tool based on the study design they need to assess, the type of intervention or exposure under study, the specific research context or field of study, and how well a tool may evaluate biases specific to their discipline.

Tools available for all study designs are crucial for determining the internal validity of the study and evaluating whether the findings may reflect a true estimate of the intervention(s) being tested. In 2008, the first Cochrane RoB tool was first developed for use in RCTs as an essential component along the critical path of SR processes outlined in the Handbook of the Collaboration (91, 100). It has been evaluated and updated since to address limitations (101) and both the original Cochrane RoB tool and the updated version (RoB 2.0) tools are widely used internationally (101, 102). Despite the popularity of these tools, a number of other tools are available to assess the RoB in RCTs. An online repository tracking quality appraisal and RoB tools lists at least 13 available tools, some aimed at assessment of RoB in RCTs only, and others for simultaneous assessment of multiple study designs, including RCTs (103). A 2019 methodological study (102) identified a total of 40 RoB tools from a prospective sample of 500 PROSPERO registrations. A trend analysis also showed that several tools are more established in their use, while others have fallen out of favour or increased in use over time. Some tools were presented as part of a suite of RoB tools applicable to multiple study designs or were developed for application to studies in a given field (e.g., physiotherapy). The tools vary in their focus, applicability and complexity (103). For example, the Jadad Scale offers a tool to assess RCTs using three simple criteria (randomization, blinding and withdrawals/dropouts) and the Cochrane RoB 2.0 offers comprehensive coverage for domains of bias for parallel RCTs with additional specialty guidance to consider both cluster or crossover RCTs (101, 104).

For observational studies of exposures and non-randomized studies of interventions, there are alternative risk of bias tools are used to address biases that differ from those in randomized controlled trials (105). The Newcastle-Ottawa Scale (106) is a widely-used tool for cohort and case-control studies, and suites of tools by the JBI (107), the Scottish Intercollegiate Guidelines Network (SIGN)(108), and the National Heart, Lung and Blood Institute are commonly used

across a variety of study designs (93, 106, 109). The Risk Of Bias In Non-randomized Studies - of Interventions (ROBINS-I) tool has been gaining in popularity for the assessment of RoB in NRSI (93), and more recently, a Risk Of Bias In Non-randomized Studies - of Exposures (ROBINS-E) tool has been proposed for RoB evaluation involving exposures (including environmental, occupational and behavioural exposures)(110). Both tools share characteristics with the RoB 2.0 tools for RCTs, including a structure based on domains of bias, the use of signaling questions and the prediction of domain-specific and overall RoB for a study based on the direction and magnitude of bias and how this may threaten conclusions regarding the effect of the intervention or exposure.

Assessments are broadly made using a structured or systematic process which permits a judgement or consideration for the presence or absence of a study feature where bias may occur Tools may be structured in various ways as scales, checklists or domains to organize and facilitate an assessment of study bias (111). *Scales* offer a way to assess bias by scoring various facets of a study related to bias for study and then combining them to produce a composite summary score (91). While a popular and often straightforward way to quantify bias in a research study, use of a single quantitative summary score as a measure of RoB has been discouraged by Cochrane and other groups. Although not true for all instruments, historically many scales have combined aspects of study reporting, quality and bias, and used weights for scoring which are difficult to empirically validate or interpret (112). In particular, equal weighting of items is a concern. *Checklists* permit assessment of whether a specific list of relevant items related to the internal validity have been addressed in the study under review but may not always employ a rating system to summarize the results (113). *Domain-based instruments* use thematic or structured categories of items or questions related to the internal validity of a study and permit assessments on the items and for the domain as a whole to

contribute to judgment of the study's internal validity (111). Tools may be developed to assess a study as a whole, or they may be structured around assessment for each specific study outcome (101). This is because within the same research study, some outcomes may be more influenced by bias than others (i.e., subjective versus objective outcome measures) and the direction and magnitude of bias may also vary accordingly.

2.4.3. Considerations for existing tools

When considering which RoB tool should be used for a study design and particular research context, users may have to consider trade-offs based on the strengths and limitations of the available tools and what would best suit their requirements. There are several relevant aspects that an individual requiring a risk of bias tool may consider when selecting which tool to use. For example, if the tool is intended to assess the internal validity of a study, if it is a scale, checklist or domain-based tool, whether it is specialized to one study design or it can be applied to several different study designs, or whether the measurement properties have been established (i.e., the tool reliable and valid). There are several published studies which may support users in a more formal consideration of these factors for some of the available tools. However, information or evidence to inform these considerations is not available for all tools, and many tools have benefits and limitations which can be weighed by users for their own research context. Published studies evaluating the use and metrics for various RoB tools are more readily available for tools which are more popular in use currently, but the absence of formal critique or evaluations for other available tools does not imply they are not subject to similar, or different, challenges.

In an online repository of quality assessment and RoB tools, only 38.7% were assessed to have published validity or reliability information for the tool available. Several tools for assessment of RCTs are known to include items which do not address RoB, including reporting quality,

imprecision or applicability (114). From a total of 17 available instruments, half of the unique items were irrelevant or clearly did not address RoB, and the authors noted that the experts in methodology and epidemiology engaged could not agree on the relevance of a large proportion of the remaining items. Of the tools assessed, the Cochrane RoB 2.0 was the only tool without items not clearly addressing RoB (101, 114). Most instruments did not consider outcome-specific RoB assessment.

RoB tools may use a range of formats to consider key biases and structure the resulting assessments. Evidence suggests that overly-complex methods may broadly impede the pragmatic use of recognized evidence-based medicine principles, and that some ROB instruments for both RCTs and observational studies are contributing to this (115). Older instruments such as the NOS, the 2011 version of the Cochrane RoB tool for RCTs and others are described as following a *simple-as-possible* principle due to their limited number of intuitively labelled items with straightforward response options. This, in turn, resulted in widespread adoption by users internationally and high levels of comfort in authors with a wide range of experience applying the tools (115).

Despite being comprehensive in the biases assessed and providing a structured, outcome-specific framework and many implements to assist with standardized ratings for RCTs, the revised Cochrane tool (RoB 2.0) has been criticized for removing justifications for ratings (115) introducing complexity and non-intuitive content which has resulted in a loss of reliability and user friendliness (116, 117). The tool was assessed to have low interrater reliability for the assessment of parallel group RCTs (117). In a sample of 70 outcomes from 70 RCTs rated independently by four raters, interrater reliability (IRR) was assessed to be only slight for the overall study rating (IRR 0.16, 95% confidence interval (CI) 0.08 to 0.24), and the IRR for five

individual domains varied from *slight* (IRR 0.04; n= 1 domain), to *fair* (IRR 0.21 / 0.22 /0.30; n= 3 domains), to *moderate* (IRR 0.45; n =1 domain) (117). Reasons for disagreement were related to the complexity of the tool and lack of understanding of some signaling questions, limited experience in using the RoB 2.0 tool and inadequate knowledge of the subject matter in some studies assessed. In particular, raters were unable to decide if deviations from the intended intervention differed from clinical practice and if those deviations were likely to have affected the outcome. A second study flagged difficulty in the same domain of the RoB 2.0 tool (116). Raters also critiqued the guidance as unclear in places, despite being extensive (117) and reported difficulties implementing the guidance for certain domains and signaling questions which required operationalization of thresholds or review-specific decisions that may not be as the tool intended. Raters also reported inadequate skill or expertise to decide whether a bias had been corrected appropriately or sufficiently using statistical approaches. None of the difficulties encountered in applying the tool were attributed to a specific lack of skill or experience in the raters, rather, they were seen as difficult to reconcile the RoB 2.0 rating system with classical definitions of bias in RCTs used in clinical epidemiology or the 2011 version of the tool (100, 101, 117). These challenges in its application likely contribute to the low IRR assessed.

Substantial resources may be required to implement some RoB tools as part of a SR. The RoB 2.0 tool has a mean time to completion *per outcome* of between 28 minutes (117) and 47 minutes when applied to RCTs (116). In one study, the mean time to completion by staff resource, *per study* of 358 minutes (SD 183 minutes) (116). The mean time *per study* to conduct an individual assessment (*per reviewer*) was 127 minutes (SD 67 minutes) with an additional 54 minutes (SD 43 minutes) required for consensus amongst independent reviewers. These times were influenced by the number of outcomes to be assessed and the number of reports per RCT which increased time to assess using the RoB 2.0 tool. Time was reduced when more experienced reviewers

applied the tool (116). This SR was a network-meta-analysis involving complex interventions which was reported by study authors as particularly challenging, and these time-to-completion and challenges encountered may not be reflective of the usual application case for RoB 2.0. They do however, demonstrate some practical limitations for users to consider, particularly in resource-limited settings (117). As a potential proxy indicator of hesitancy to use the RoB 2.0 tool for RCTs in evidence syntheses, a 2023 meta-research study concluded that most Cochrane SRs and protocols sampled did not adhere to the Cochrane's risk of bias 2.0 tool (118). Despite being a recommended tool within Cochrane SRs and some evidence of increasing use, a small percentage of protocols and complete reviews and protocols used or planned to use the improved RoB 2.0 for assessing RoB of the included RCTs (118) and low uptake of the tool outside of Cochrane (115). The 2011 Cochrane RoB tool is still widely used despite some identified limitations (119).

Similarly, several challenges related to the use and application of tools for observational or non-randomized studies have also been documented (115, 120-123). While ROBINS-I implements such as the worksheet used to identify confounding domains and cointerventions exist to support users in their assessments, the completion of these tables are burdensome and resource-intensive (115). Growing evidence suggests that ROBINS-I is frequently misapplied (122), and is subject to the same limitations in usability, complexity and lengthy time-to-completion per study as the ROB 2.0 (117, 120, 122-124). Reliability may improve with experience of the rater, and some domains have demonstrated higher IRR than others (range of kappa from 0.50 to 1.0) but these are inconsistent across SRs (120). The NOS has been compared to ROBINS-I in the published literature (125) and while they demonstrate broadly similar inter-rater reliability (*ROBINS-I* IRR 0.67, 95% CI 0.50 to 0.83 and *NOS* IRR 0.73, 95% CI 0.65 to 0.81), they vary greatly in usability. The NOS was reported to have a mean time to assess a single study of 30 minutes,

while ROBINS-I had a mean time to assess a single study of 3 hours) (125). In a comparison between the NOS and ROBINS-I ratings for the same 28 observational studies, there was no agreement among tools (126). Authors attributed differences to ROBIN-I being more conservative, and studies rated as high ROB by ROBINS-I may be low RoB using the NOS. Another assessment of natural experiment studies using ROBINS-I as part of a SR of domestic energy efficiency interventions found that while the tool was able to help systematically articulate sources of bias in the included studies, there was a lack of consensus in all studies for all seven bias domains which may further demonstrate concerns over ROBINS-I's reliability and applicability in this research context. ROBINS-E was proposed in 2018 and was been published in 2024 (110). Despite its applicability to studies of dietary exposure, a published report of early use from 2018 concludes that ROBINS-E did not discriminate between studies with a single RoB or multiple ROBs and omitted questions related to funding source and exposure (127). It is not clear whether the 2024 version of ROBINS-E addresses these shortcomings. As such, users of these tools must consider which tool to select based on these published evaluations, while factoring in their own use case, experience, research context and resource availability.

2.5. Why is a nutrition-specific risk of bias tool needed?

Accurate assessment of RoB for nutrition studies is critical because these studies are necessary to advance our scientific knowledge and they provide essential input to health guidance and policy. Dietary Reference Intakes (DRIs), for example, are international nutrient reference value standards, used to inform clinical practices, nutrition policies, and dietary recommendations in Canada and the US (61, 128-130). These are primarily related to nutrient adequacy but increasingly chronic disease is considered. The strength of the related policy and guidelines is highly reliant on the scientific rigour of the studies behind them. When we have policies and guidance based on high quality nutrition studies, they maintain relevance over time and arguably

have greater impact. Policies based on sound evidence which are stable over time minimize the need for frequent changes which can undermine public trust and contribute to scientific uncertainty. This, in turn, minimizes related resource consumption and research waste. Uncertainty in nutrition research often arises from conflicting study outcomes or avoidable methodological flaws, making it challenging to develop lasting policy and guidance.

SRs are often a foundational evidence resource when setting nutrient standards, nutrition policies and dietary guidance. Low quality SRs may be ineffective, misleading or wasteful, and this is a particular problem in syntheses of nutrition research. RoB is an essential component of a SR, and nutritional reviews have important limitations in their assessment and reporting of RoB, which limits the utility and uptake of their findings (131). A cross-sectional study exploring the characteristics and quality of SRs of observational nutrition research (including and without RCTs) showed that 87.3% of SRs assessed RoB but within this group, the method of RoB assessment was not reported in almost 50% of SRs which is consistent with reviews in other topic areas (35, 132). Over 18% of SRs in the study used a tool developed de novo by the SR authors to examine internal validity of the studies included instead of applying an existing RoB tool. This was twice as common in non-therapeutic SRs (36%). Risk of bias or quality assessment of included studies was performed in 70% of SRs but was infrequently incorporated into any meta-analysis (16%) and reviews which considered non-RCT evidence were much less likely to formally assess RoB (59%). Authors concluded that the quality of conduct and reporting in nutritional SRs has not improved in the decade since they last examined a similar sample of nutrition SRs. Results from a second meta-epidemiological study based on a random sample conveyed similar concerns that assessments of RoB in SRs of observational nutritional epidemiologic studies are often inappropriate or not comprehensive (131). In addition to a wide variety of ad-hoc criteria being added on to existing risk of bias tools as part of the assessments,

RoB assessments were seldom reported for included studies individually, or tools not intended for RoB were misapplied (e.g., reporting guidelines were used to report RoB). Conclusions noted that RoB tools and criteria often had serious limitations, but that existing tools did not provide sufficient guidance to assist with application, particularly for nutritional epidemiologic studies. None provided sufficient guidance to facilitate judgements regarding whether tools for measuring dietary exposures are sufficiently valid and reliable. Further to this, study authors found that review authors often disregarded biases that may be broadly considered to be inherent to the design of nutritional studies, such as the method of exposure assessment (131). This reinforces the need for advancement of the science around the assessment of RoB in nutrition research. There is a need for nutrition-specific tools and guidance for applying RoB concepts to the nutrition context.

As noted, there are many tools available to assess research studies using approaches which generically consider items related to quality and/or RoB without being specific to a particular clinical area or discipline. As with research studies, these RoB tools may vary in design or coverage, how tools are applied and in guidance provided to users. Generic RoB tools may not be specific enough to assess the nuances in some topic areas. For this reason, there are a number of formal discipline-specific tools (133, 134), however there are no current RoB tools specific to nutrition. A modified version of the ROBINS-I tool proposed by the Nutrition Evidence Systematic Review group was used in the development of the Dietary Guidelines for Americans to consider bias in nutrition observational studies (ROB-NOBS) but is not available for external use (135). The Academy of Nutrition and Dietetics methodology for conducting SRs provides an example of a nutrition-specific quality criteria checklist for critically appraising research articles which is founded on constructs and domains identified by the Agency for Healthcare Research and Quality (AHRQ) report on Systems to Rate the Strength of Scientific Evidence (136).

Factors considered are almost all generic risk of bias or quality items, although there are specific questions pertaining to nutrition measures as outcomes (item 7.2) and broad dose-response as part of analysis (item 8.4) which may be additionally important in nutrition studies. This tool is intended to be applied to any type of nutrition or dietary research study design (RCTs, cohort studies and cross-sectional studies are named within the tool) yet items are unstructured and there is no guidance to provide clarity, context or examples for the items being assessed (136). International calls for improvement in methods and quality of research in the field of nutrition have resulted in other nutrition-specific endeavors, including extensions to the Consolidated Standards of Reporting Trials for Nutrition (CONSORT-nut)(48) and the Strengthening the Reporting of Observational Studies in Epidemiology-Nutritional Epidemiology (STROBE-nut)(137) and a meta-evidence scoring system (NutriGrade)(138).

Why are nutrition studies so different from those in any other field or discipline? In a seminal commentary from Dr. John Ioannidis, several unique challenges were identified that are common to dietary and nutrition research and, related to this, in understanding the true epidemiologic associations between dietary-health relationships (139). First, some aspects of nutrition research are not so different from other fields and research contexts. As an example, it is somewhat intuitive to understand how inadequate intakes of an essential nutrient leads to deficiency or that excessive consumption of high-calorie foods may contribute to obesity or excess mortality in humans. What is different, and difficult to capture, is how relatively small changes in human intake of foods, nutrients, dietary patterns or other nutritional interventions contribute to changes in overall health, risk for chronic disease or life-expectancy (56, 57, 140). Given that these associations are correlated with, and further influenced by numerous endogenous and exogenous factors such as a person’s age, metabolic profile, genetic factors, comorbidities, and even environmental factors, it is very complicated to disentangle the effect of a single dietary

component from numerous other variables/factors. Therefore, results from nutritional research studies can be sensitive to the choices of investigators related to study design, implementation and analysis and these must be fulsomely considered when evaluating the confidence that a study's results. Developing effective dietary guidance and policy requires a rigorous evidence base. While there have been many initiatives to globally improve the rigour of nutrition studies (74), their reporting, and the uptake of such evidence into policy and guidelines, there is a lack of adequate consideration for key but nuanced issues such as baseline nutritional status, dietary exposure and other important aspects of nutrition science in tools that are used to evaluate the validity of the findings of existing studies. The Nordic Nutrition Recommendations (NNR; 2022) for the structure and conduct of high quality SRs in the field (141, 142) dictates that the unique and complex methodological facets of nutrition research should be considered as they factor into important related implications for synthesizing and interpreting the nutrition research literature (2, 67, 137, 143). This updated guidance builds off guidance from the Cochrane Collaboration and AHRQ. These important and specific considerations defined by the NNR include: (1) the nutritional status and background diet of the study population; (2) the validity of the dietary assessment methods; (3) the bioavailability of nutrients and other food substances; (4) biological interactions of food components; and, (5) the many known or unknown factors that may confound or mediate relationships between foods/diets and health outcomes (141, 142).

In nutrition studies, if particular aspects of nutritional intake were not comprehensively documented or considered, then major uncertainty can be added to judgments around an intake-health relationship (55, 57, 67, 140, 144-146). This can have a cascading influence in a research study. Intake exposure assessment is essential because it directly impacts the estimation regarding the accuracy of the intake. This, in turn, will impact the assessment of an intake-outcome dose-response relationship and, as an example, can further impact the assessment of

prevalence of population groups with adequate or inadequate intakes or higher or lower intakes of a particular food or nutrient (67, 128, 147). Within a nutrition study population, there may be groups with higher or lower intake and these are essential for assessing the existence, or direction of, and strength of the relationships between specific nutritional components, foods or dietary patterns and a human health outcome. Different methods of intake assessment have different strengths and weaknesses, and those strengths and weaknesses can lead to increased or reduced potential for RoB in any particular study (67). Ultimately, this impacts how we interpret the study and how we use the findings of the study, especially in downstream settings for policy and guidance. Differences in methods for intake exposure assessment are also nutrient-dependent. One method may be accurate for a certain nutrient or food and may not apply well to other nutrients or foods.

Likewise, dietary intake is often assessed using self-reported data (e.g., diet records) and these are often misreported. Individuals may actually alter their behaviour when they are maintaining a diet record, intentionally or unintentionally as a subconscious action or behaviour (67, 148). An individual for a portion of the diet record, may change what they are eating or not accurately document their intake, which can also occur subconsciously or on purpose. In nutrition research, when intake of nutrients or foods is considered, scientists often study what is referred to as a 'usual intake' (29, 149). To obtain this data, research participants are asked to recall their diet retrospectively over a 24-hour period with multiple passes required to estimate usual intake. Although widely-used, intake measured through a 24-hour recall is often under-reported (149). To accurately estimate usual intake, at least two 24-hour recalls are needed to permit adjustments needed for any intra-individual variability (and ideally more) (67, 148). Food frequency questionnaires (FFQs) are also widely used in nutrition research and, by design, tend to increase potential for RoB even more than a 24-hour recall. Researchers use FFQs to obtain information

about consumption of foods and beverages over a specified time, and this data can often be linked to other databases to yield information about total dietary intake (148). A FFQ is also subject to underreporting as they are structured to document a record for specific items (e.g., a question may ask 'How often do you eat a ½ cup of rice') versus everything an individual had to eat or drink on a particular day (148, 149). Also, self-reported intake has other limitations which impact potential for RoB. It is essential to know how an instrument was validated, in what population and, if applicable, for what intake (nutrient, food, dietary pattern etc.) as this often varies (150). A new FFQ is often validated against an existing FFQ and when (which is usually the case) the existing tool is not a gold standard or when it has existing flaws that the new tool is measured against, these flaws are then carried over or propagated in the new tools (67, 150). When an FFQ is validated in a specific study population (e.g., individuals who are pregnant or of a certain age, individuals from specific ethnic groups), it is important to understand whether the assessed validity will be maintained in applications to different study populations. If populations are not comparable, there could be reason for concern. The FFQ data outputs (i.e. could be continuous or categorical) should also be considered when dietary intake exposure is assessed, as data manipulation and standardization, and use of natural groupings, cut-offs or clustering approaches may influence the resulting intake estimates (151). Depending on the methods used to manipulate or standardize data, bias may be introduced.

Another aspect which is often overlooked in nutrition research is the accuracy of the food composition databases used to estimate nutrient and other component intakes from foods (57, 144, 146). Many of the existing databases were not developed using gold standard methodologies or are outdated, or both, which adds to the uncertainty of the data. Food composition databases are widely used to estimate both individual and population-level food and nutrient intakes for use in epidemiological research, mainly due to the impossibility of

chemically analyzing all foods consumed by study participants (151). Although many countries maintain their own unique food composition databases, and despite international standards, the majority keep incomplete, outdated or unreliable food composition datasets, or none at all (151). Databases may be in a specific language and the absence of an English version has been documented as a limiting factor to disclosure of particular diets, food items and nutrients (151, 152). These databases may also fail to capture details relevant to intake assessment of some nutrients such as the extent of the food processing, the food matrix (for example, solid, semi-solid or liquids, in the context of other foods) or other food properties. They may also have missing data for key nutrients or foods which are more difficult to capture. A food is conceptually a complex matrix which can contribute to health effects in humans as a whole complex food or through the actions of individual food components, and these complexities are near impossible to capture in a standardized database. Properties of foods interact synergistically within the food and with other ingested foods and nutrients which is complex to capture or simplify. Additional considerations relate to the *bioavailability* of various foods. This refers to the proportion of the nutrient or food that can be absorbed or used by the human body after it is ingested (39). Not all of the nutrients in a food are bioavailable. For example, the stated amount of Calcium in an individually portioned yogurt cup could be stated as 100 mg, but only 30 mg of that Calcium may be bioavailable to the human body. It is critical to assess how a database would consider this nuance and whether the output would reflect this. These fundamental properties are complex beyond simple nutritional composition.

Use of biomarkers of exposure in nutrition research may contribute to improvement in the accuracy of exposure estimates (153-155). Metabolites recovered from human urine or blood may be used as biomarkers of intake (recent exposure) and status (long-term exposure) (153). It is important to emphasize that it is possible to improve the accuracy of exposure estimates using

validated biomarker assessment methods, but not all nutrients or foods have these. Practically, there are a limited number of biomarkers for a limited number of nutrients and, therefore, not every nutrient and food has a validated biomarker of exposure (155). When a biomarker is available to assess intake exposure, there are several considerations relevant to the study methods that must be considered, including whether it has been validated to accurately reflect the intake, whether the methodology used was accurate and reliable, if the method used was gold standard and whether the standard reference material, quality assessment or quality control data are collected and reported (155). From a population standpoint, you must consider if the biomarker of exposure is accurate across subgroups within the population. For example, some biomarkers may be accurate in adults, but not in children, or similarly, accurate for individuals who are not pregnant but not accurate or not validated in individuals who are pregnant. Biomarkers must be valid and reliable, and must clearly elucidate which settings and conditions they may be best applied to (156). Biomarker validation results should be similar across population subgroups as biomarkers may be subject to confounding. For example, when red blood cell folate is used as a marker for human folate status, age, pregnancy, vitamin B-12 deficiency, medication use, smoking and intake of alcohol may influence results (156, 157). Selection of biomarkers, regardless of use, may be influenced or constrained by setting as environment (temperature, hygiene), training of technical staff and collection variables (equipment, storage, sample collection procedures, facilities) and, in turn, limitations in these settings may compromise the precision of a measured biomarker (156). Biomarker cut-offs are commonly used in research and clinical settings. They provide benchmarks for risk classification (e.g., deficient compared with excess status and are informative for the interpretation of related effect measures (e.g., relative risk or odds ratios) but may introduce bias when cut-offs are applied inappropriately or without justification (158). These are essential concepts to consider in a RoB assessment. For some

biomarkers, additional considerations are warranted (67, 156). For example, if a urinary biomarker is considered, the timing and completeness of the collection should be considered, along with whether the collections were repeated in multiple passes, or multiple days, or done using a spot collection (155). Multiple passes or days are required to estimate usual intakes to account for day-to-day variability. The metabolite of interest must have a good recovery in urine.

Existing RoB tools do not comprehensively consider the unique aspects related to causal or intake-response relationships in nutrition research yet are still used widely out of necessity. In an analysis of 50 randomly-selected nutrition-specific SRs of RCTs (18% Cochrane reviews), a total of 70% applied the original Cochrane ROB assessment tool (74) but the tool does not include any consideration for important factors such as dietary adherence (74, 100). The updated RoB 2.0 tool (101) permits assessment of non-adherence to a specific dietary intervention and can be evaluated within the bias domain "assessing deviations from intended interventions", but this has shown to be a difficult and complex domain to assess with very limited IRR (117). This has been attributed to complexity in the tool, gaps in the accompanying guidance, and difficulty considering whether deviations from the intended intervention reflect usual practice and whether they were likely to have affected study outcomes (117). Additionally, there is choice required around whether raters are interested in potential biases related to quantifying the effect of assignment to intervention (intention-to-treat effect) or with the effect of adhering to intervention (per-protocol effect). This can be difficult to interpret in general as research has shown the signalling questions are challenging to interpret and users find it hard to decipher cointerventions and how these may have influenced the groups to produce imbalances (159). This could be additionally challenging in the dietary or nutritional intervention context, as both the intervention and the outcome may be dietary intake, and non-adherence may not be easy to disentangle when dietary patterns are considered. The NOS has also been criticized for its inability to appropriately

discriminate low to high levels of bias based on empirical evidence of its use within nutrition SRs (121).

When existing tools do not offer a comprehensive way to assess the nuances in nutrition studies, users will often adapt tools to meet their needs or develop their own de novo approaches to assess RoB. A meta-research study of Cochrane nutrition reviews examining how dietary adherence to a specific dietary regimen was assessed found considerable variation (35). Several of the reviews included adapted existing RoB tools by adding a new risk of bias domain for dietary adherence. This is problematic as dietary adherence is not a domain of bias *per se*, and ad hoc modifications to tools may impact the reliability and the validity of the assessments and contribute to a lack of standardization across assessments and studies. Similar adaptations to include ad hoc nutrition-specific risk of bias criteria have been used by Evidence-based Practice Centers for the AHRQ in the United States in their comparative effectiveness reviews (160). As an example, in a review of human health effects related to sodium and potassium, both the Cochrane RoB tool and the NOS were modified to "cover concerns" in the types of RCTs and observational nutrition studies, respectively, considered in the review. Specifically, additional criteria were used to consider the methods used to determine sodium and potassium intake assessment, adherence and compliance, and similarities at baseline related to absence of the outcome and measured sodium excretion. Related to these RoB criteria additions, the methods for rating individual studies were adapted to consider the additional RoB criteria. A separate RoB rating was applied to observational studies based specifically on the assessment of sodium or potassium intake, meaning that the same study could receive different overall RoB ratings for sodium- and potassium-outcome pairs. Modifications to existing RoB tools within the context of a nutrition-specific SR are not always reported transparently, or at all. Often, modifications are noted in the methods but it is difficult to ascertain how the tools were adapted in order to

facilitate appropriate uptake or reproducibility of the evidence synthesized in the review. While there may not be any nutrition-specific RoB tools, there is empirical evidence based on amendments to existing tools that further justifies the need for these types of tools, if only to provide a standardized, transparent and consistent way to evaluate RoB in nutrition research.

There are several nutrition-specific items that must be given special attention, and which are not adequately covered, in existing RoB tools. The particular nuances of dietary intake and exposure assessment introduce significant variability and uncertainty into research studies, and these issues may not be as prominent in other research disciplines. Existing tools, old and new, are not well-suited to assess biases common in nutrition studies, which require a nuanced or unique RoB criteria along with detailed guidance to fulsomely capture the issues. Furthermore, the complexity of nutrient interactions and lifestyle factors adds another layer of potential bias that general tools do not fully capture. Therefore, a tailored approach to address these gaps is needed to ensure more accurate and meaningful evaluations of bias in nutrition research. At the 2022 International Conference on Nutrition in Tokyo, Japan, several opportunities for improving the rigour of research in the field of nutrition were chosen by an international group of 40 scientists and other interested individuals and parties, with a nutrition-specific RoB rating system being one of six prioritized options (74). The development of nutrition-specific RoB tools is therefore proposed to address the limitations and gaps in existing instruments which relate to the unique challenges present in dietary and nutrition research.

2.6. Summary

Robust nutrition studies are necessary for developing high quality dietary recommendations and health policies which, in turn, contribute to clear guidance and policies which optimize resource use and reduce public misperceptions of nutrition evidence. The assessment of RoB in these

studies is critical, as flaws and limitations of nutrition study design and conduct undermine the validity and impact of any resulting health policy and guidance. Nutrition research has unique methodological challenges, particularly around the assessment of dietary exposure and intake, and existing RoB tools do not comprehensively consider or facilitate appropriate assessment of these factors. This results in inconsistent and incomplete RoB evaluations for nutrition studies. While several discipline-specific RoB tools exist, none are tailored to nutrition research. Nutrition-specific RoB tools are necessary to standardize appraisal of nutrition evidence and to fulsomely consider the biases that can influence the estimation of causal and intake-response relationships to human health, including nutritional status of study participants, the variety of intake sources, the validity of dietary intake measurement and methods to consider nutrient interactions.

In this chapter, background information about the importance of diet and nutrition for human health, the role of research in nutrition, and concepts on bias in research are presented. Additionally, the rationale for why nutrition-specific RoB tools are required is described. The following three chapters (Chapters 3 to 5) present a series of manuscripts that each address the thesis research questions building on the context and concepts introduced in this chapter.

2.7. References

1. Byers T. *The role of epidemiology in developing nutritional recommendations: past, present, and future*. The American Journal of Clinical Nutrition. 1999 1999/06/01/;69(6):1304S-1308S. doi:https://doi.org/10.1093/ajcn/69.6.1304S.

2. Satija A, Yu E, Willett WC, Hu FB. *Understanding nutritional epidemiology and its role in policy*. Advances in nutrition. 2015;6(1):5-18.

3. Mozaffarian D, Rosenberg I, Uauy R. *History of modern nutrition science—implications for current research, dietary guidelines, and food policy*. BMJ. 2018;361:k2392. doi:10.1136/bmj.k2392.

4. Wyatt HV. *James Lind and the prevention of scurvy*. Medical History. 1976;20(4):433-438. Epub 2012/08/16. doi:10.1017/S0025727300031264.

5. Carpenter KJ. *A short history of nutritional science: part 3 (1912–1944)*. The Journal of nutrition. 2003;133(10):3023-3032.

6. Carpenter KJ. *A short history of nutritional science: part 2 (1885–1912)*. The Journal of nutrition. 2003;133(4):975-984.

7. Carpenter KJ. *A short history of nutritional science: part 4 (1945–1985)*. The Journal of nutrition. 2003;133(11):3331-3342.

8. Pett L. *Food makes a difference*. Canadian Public Health Journal. 1942;33(12):565-570.

9. *Transitions of nutritional sciences*. Nat Food. 2021 Mar;2(3):129. doi:10.1038/s43016-021-00254-4. Cited in: Pubmed; PMID 37117455.

10. Ohlhorst SD, Russell R, Bier D, Klurfeld DM, Li Z, Mein JR, Milner J, Ross AC, Stover P, Konopka E. *Nutrition research to affect food and a healthy life span*. J Nutr. 2013 Aug;143(8):1349-54. Epub 20130619. doi:10.3945/jn.113.180638. Cited in: Pubmed; PMID 23784071.

11. Ross AC, Caballero B, Cousins RJ, Tucker KL. Modern nutrition in health and disease. Jones & Bartlett Learning; 2020. ISBN: 1284229858.

12. *The Giessen Declaration*. Public Health Nutr. 2005 Sep;8(6A):783-6. doi:10.1079/phn2005768. Cited in: Pubmed; PMID 16236217.

13. Beauman C, Cannon G, Elmadfa I, Glasauer P, Hoffmann I, Keller M, Krawinkel M, Lang T, Leitzmann C, Lotsch B, Margetts BM, McMichael AJ, Meyer-Abich K, Oltersdorf U, Pettoello-Mantovani M, Sabate J, Shetty P, Soria M, Spiekermann U, Tudge C, Vorster HH, Wahlqvist M, Zerilli-Marimo M. *The principles, definition and dimensions of the new nutrition science*. Public Health Nutr. 2005 Sep;8(6A):695-8. doi:10.1079/phn2005820. Cited in: Pubmed; PMID 16236202.

14. Banting FG. *Frederick G. Banting - Nobel Lecture.* . The Nobel Prize in Physiology or Medicine. . Amsterdam: Elsevier Publishing Company; 1923.

15. McHenry E. *Nutrition in Canada*. Canadian Public Health Journal. 1939;30(9):431-434.

16. Harris T, Woteki C, Briefel RR, Kleinman JC. *NHANES III for older persons: nutrition content and methodological considerations*. Am J Clin Nutr. 1989 Nov;50(5 Suppl):1145-9; discussion 1231-5. doi:10.1093/ajcn/50.5.1145. Cited in: Pubmed; PMID 2816808.

17. Johnson CL. The U.S. National Health and Nutrition Examination Survey: What Forty Years of Experience Has Taught Us. In: Methodological Issues in Measuring Population Health. Statistics Canada Symposium 2006; 2006 Gatineau, Quebec. Statistics Canada; 2006. Available from: https://www150.statcan.gc.ca/n1/en/pub/11-522-x/2006001/article/10413-eng.pdf?st=jYu0YoNH.

18. Canadian Community Health Survey (CCHS) - Detailed information for 2000-2001 (Cycle 1.1). Statistics Canada - Surveys and Statistical Programs; 2002 April 2, 2023. Available from: https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=3359.

19. *Canadian Community Health Survey*. In: Canada S, editor. Public Use Microdata File. Ottawa2024.

20. *Canadian Community Health Survey*. In: Canada S, editor. Public Use Microdata File. Ottawa2004.

21. *Canadian Community Health Survey*. In: Canada S, editor. Public Use Microdata File. Ottawa2015.

22. Sacks G, Riesenberg D, Mialon M, Dean S, Cameron AJ. *The characteristics and extent of food industry involvement in peer-reviewed research articles from 10 leading nutrition-related journals in 2018*. PLoS One. 2020;15(12):e0243144. eng. Epub 20201216. doi:10.1371/journal.pone.0243144. Cited in: Pubmed; PMID 33326431.

23. Vandenbrink D, Pauzé E, Potvin Kent M. *Strategies used by the Canadian food and beverage industry to influence food and nutrition policies*. Int J Behav Nutr Phys Act. 2020 Jan 29;17(1):3. eng. Epub 20200129. doi:10.1186/s12966-019-0900-8. Cited in: Pubmed; PMID 31992304.

24. Mozaffarian D, Forouhi NG. *Dietary guidelines and health-is nutrition science up to the task?* BMJ. 2018 Mar 16;360:k822. Epub 20180316. doi:10.1136/bmj.k822. Cited in: Pubmed; PMID 29549076.

25. Tremblay M, Wolfson M, Connor Gorber S. *Canadian Health Measures Survey: rationale, background and overview*. Health Rep. 2007;18 Suppl:7-20. Cited in: Pubmed; PMID 18210866.

26. Tremblay MS, Connor Gorber S. *Canadian health measures survey: brief overview*. Can J Public Health. 2007 Nov-Dec;98(6):453-6. doi:10.1007/BF03405437. Cited in: Pubmed; PMID 19039881.

27. Yousef S, Colman I, Papadimitropoulos M, Manuel D, Hossain A, Faris M, Wells GA. *Vitamin D and chronic diseases among first-generation immigrants: A large-scale study using Canadian Health Measures Survey (CHMS) Data*. Nutrients. 2022;14(9):1760.

28. Zohoori FV. *Chapter 1: Nutrition and Diet*. Monogr Oral Sci. 2020;28:1-13. eng. Epub 20191107. doi:10.1159/000455365. Cited in: Pubmed; PMID 31940634.

29. Medical Research Council & National Institute of Health Research. (2017) (on behalf of OSCHR) Review of Nutrition and Human Health Research. Accessed 19th April 2024].

30. Cardenas D. *What is clinical nutrition?* Clinical Nutrition. 2017;36(6):1725-1726. doi:10.1016/j.clnu.2017.08.030.

31. Mozaffarian D, Rosenberg I, Uauy R. *History of modern nutrition science-implications for current research, dietary guidelines, and food policy*. BMJ. 2018 Jun 13;361:k2392. s. Epub 20180613. doi:10.1136/bmj.k2392. Cited in: Pubmed; PMID 29899124.

32. Collaborators. GBoDRF. *Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019*. Lancet. 2020 Oct 17;396(10258):1223-1249. doi:10.1016/S0140-6736(20)30752-2. Cited in: Pubmed; PMID 33069327.

33. Shao A, Drewnowski A, Willcox D, Krämer L, Lausted C, Eggersdorfer M, Mathers J, Bell J, Randolph R, Witkamp R. *Optimal nutrition and the ever-changing dietary landscape: a conference report*. European journal of nutrition. 2017;56(1):1-21.

34. Hu FB. *Dietary pattern analysis: a new direction in nutritional epidemiology*. Curr Opin Lipidol. 2002 Feb;13(1):3-9. doi:10.1097/00041433-200202000-00002. Cited in: Pubmed; PMID 11790957.

35. Zeraatkar D, Bhasin A, Morassut RE, Churchill I, Gupta A, Lawson DO, Miroshnychenko A, Sirotich E, Aryal K, Mikhail D, Khan TA, Ha V, Sievenpiper JL, Hanna SE, Beyene J, de Souza RJ. *Characteristics and quality of systematic reviews and meta-analyses of observational nutritional epidemiology: a cross-sectional study*. Am J Clin Nutr. 2021 Jun 1;113(6):1578-1592. eng. Epub 2021/03/20. doi:10.1093/ajcn/nqab002. Cited in: Pubmed; PMID 33740039.

36. Kussmann M, Abe Cunha DH, Berciano S. *Bioactive compounds for human and planetary health*. Front Nutr. 2023;10:1193848. Epub 20230717. doi:10.3389/fnut.2023.1193848. Cited in: Pubmed; PMID 37545571.

37. World Health O. Global nutrition targets 2025: Stunting policy brief. 2014. World Health Organization.

38. Venn BJ. *Macronutrients and Human Health for the 21st Century*. Nutrients. 2020;12(8):2363. Cited in: Pubmed; PMID doi:10.3390/nu12082363.

39. Biesalski HK, Dragsted LO, Elmadfa I, Grossklaus R, Muller M, Schrenk D, Walter P, Weber P. *Bioactive compounds: definition and assessment of activity*. Nutrition. 2009 Nov-Dec;25(11-12):1202-5. Epub 20090819. doi:10.1016/j.nut.2009.04.023. Cited in: Pubmed; PMID 19695833.

40. Phillips JA. *Dietary guidelines for Americans, 2020–2025*. Workplace health & safety. 2021;69(8):395-395.

41. Ferruzzi MG, Tanprasertsuk J, Kris-Etherton P, Weaver CM, Johnson EJ. *Perspective: The Role of Beverages as a Source of Nutrients and Phytonutrients*. Adv Nutr. 2020 May 1;11(3):507-523. eng. doi:10.1093/advances/nmz115. Cited in: Pubmed; PMID 31755901.

42. DeSalvo KB, Olson R, Casavale KO. *Dietary guidelines for Americans*. Jama. 2016;315(5):457-458.

43. Abalo R. *Coffee and caffeine consumption for human health*. MDPI; 2021. p. 2918.

44. Alsunni AA. *Energy drink consumption: beneficial and adverse health effects*. International journal of health sciences. 2015;9(4):468.

45. Muñoz-Urtubia N, Vega-Muñoz A, Estrada-Muñoz C, Salazar-Sepúlveda G, Contreras-Barraza N, Castillo D. *Healthy behavior and sports drinks: a systematic review*. Nutrients. 2023;15(13):2915.

46. Herforth A, Arimond M, Álvarez-Sánchez C, Coates J, Christianson K, Muehlhoff E. *A global review of food-based dietary guidelines*. Advances in Nutrition. 2019;10(4):590-605.

47. Chen X, Zhang Z, Yang H, Qiu P, Wang H, Wang F, Zhao Q, Fang J, Nie J. *Consumption of ultra-processed foods and health outcomes: a systematic review of epidemiological studies*. Nutrition journal. 2020;19:1-10.

48. Campos S, Doxey J, Hammond D. *Nutrition labels on pre-packaged foods: a systematic review*. Public health nutrition. 2011;14(8):1496-1506.

49. Snetselaar LG, de Jesus JM, DeSilva DM, Stoody EE. *Dietary guidelines for Americans, 2020–2025: understanding the scientific process, guidelines, and key recommendations*. Nutrition today. 2021;56(6):287-295.

50. Liese AD, Krebs-Smith SM, Subar AF, George SM, Harmon BE, Neuhouser ML, Boushey CJ, Schap TE, Reedy J. *The Dietary Patterns Methods Project: synthesis of findings across cohorts and relevance to dietary guidance*. J Nutr. 2015 Mar;145(3):393-402. Epub 20150121. doi:10.3945/jn.114.205336. Cited in: Pubmed; PMID 25733454.

51. Liese AD, Wambogo E, Lerman JL, Boushey CJ, Neuhouser ML, Wang S, Harmon BE, Tinker LF. *Variations in Dietary Patterns Defined by the Healthy Eating Index 2015 and Associations with Mortality: Findings from the Dietary Patterns Methods Project*. J Nutr. 2022 Mar 3;152(3):796-804. doi:10.1093/jn/nxab383. Cited in: Pubmed; PMID 34755860.

52. Collier R. *Health Canada proposes nutrition label changes*. CMAJ. 2015 Aug 11;187(11):E331. doi:10.1503/cmaj.109-5099. Cited in: Pubmed; PMID 26261276.

53. Berciano S, Figueiredo J, Brisbois TD, Alford S, Koecher K, Eckhouse S, Ciati R, Kussmann M, Ordovas JM, Stebbins K, Blumberg JB. *Precision nutrition: Maintaining scientific integrity while realizing market potential*. Front Nutr. 2022;9:979665. eng. Epub 20220902. doi:10.3389/fnut.2022.979665. Cited in: Pubmed; PMID 36118748.

54. Prajapati D, Dodiya T. *Personalized Nutrition—A trend gaining momentum: Hope or Hype?* 2024.

55. Lichtenstein AH, Yetley EA, Lau J. *Application of Systematic Review Methodology to the Field of Nutrition*. The Journal of Nutrition. 2008;138(12):2297-2306. doi:10.3945/jn.108.097154.

56. Lichtenstein AH, Yetley EA, Lau J. *Nutritional Research Series*. 2009.

57. Zeraatkar D, Johnston BC, Guyatt G. *Evidence Collection and Evaluation for the Development of Dietary Guidelines and Public Policy on Nutrition*. Annu Rev Nutr. 2019 Aug 21;39:227-247. doi:10.1146/annurev-nutr-082018-124610. Cited in: Pubmed; PMID 31433741.

58. Webb P. *Nutrition and the post-2015 sustainable development goals*. A policy brief Geneva: UNSCN Secretariat. 2014.

59. Jacobs DR, Temple NJ. Methods in Nutrition Research. In: Temple NJ, Wilson T, Jacobs JDR, Bray GA, editors. Nutritional Health: Strategies for Disease Prevention. Cham: Springer International Publishing; 2023. p. 3-20.

60. Authority EGfTCoSREFS. *Application of systematic review methodology to food and feed safety assessments to support decision making*. EFSA journal. 2010;8(6):1637.

61. National Academies of Sciences E, Medicine. *Guiding principles for developing Dietary Reference Intakes based on chronic disease*. 2017.

62. Collier R. *Containing health myths in the age of viral misinformation*. Can Med Assoc; 2018. p. E578-E578.

63. Neale EP, Tapsell LC. *Perspective: The Evidence-Based Framework in Nutrition and Dietetics: Implementation, Challenges, and Future Directions*. Adv Nutr. 2019 Jan 1;10(1):1-8. doi:10.1093/advances/nmy113. Cited in: Pubmed; PMID 30649173.

64. Lichtenstein AH, Petersen K, Barger K, Hansen KE, Anderson CAM, Baer DJ, Lampe JW, Rasmussen H, Matthan NR. *Perspective: Design and Conduct of Human Nutrition Randomized Controlled Trials*. Adv Nutr. 2021 Feb 1;12(1):4-20. Epub 2020/11/18. doi:10.1093/advances/nmaa109. Cited in: Pubmed; PMID 33200182.

65. Ioannidis JP. *Implausible results in human nutrition research*. BMJ. 2013 Nov 14;347:f6698. Epub 20131114. doi:10.1136/bmj.f6698. Cited in: Pubmed; PMID 24231028.

66. Taubes G. *Epidemiology faces its limits*. Science. 1995 Jul 14;269(5221):164-9. doi:10.1126/science.7618077. Cited in: Pubmed; PMID 7618077.

67. Yetley EA, MacFarlane AJ, Greene-Finestone LS, Garza C, Ard JD, Atkinson SA, Bier DM, Carriquiry AL, Harlan WR, Hattis D. *Options for basing Dietary Reference Intakes (DRIs) on chronic disease endpoints: report from a joint US-/Canadian-sponsored working group*. The American journal of clinical nutrition. 2017;105(1):249S-285S.

68. Rothman KJ. *Modern epidemiology*. Lippincott Williams & Wilkins; 2008.

69. Cumpston M, Li T, Page MJ, Chandler J, Welch VA, Higgins JP, Thomas J. *Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane Database Syst Rev. 2019 Oct 3;10:ED000142. Epub 2019/10/24. doi:10.1002/14651858.ED000142. Cited in: Pubmed; PMID 31643080.

70. Chowdhury R, Warnakula S, Kunutsor S, Crowe F, Ward HA, Johnson L, Franco OH, Butterworth AS, Forouhi NG, Thompson SG, Khaw KT, Mozaffarian D, Danesh J, Di

Angelantonio E. *Association of dietary, circulating, and supplement fatty acids with coronary risk: a systematic review and meta-analysis*. Ann Intern Med. 2014 Mar 18;160(6):398-406. doi:10.7326/M13-1788. Cited in: Pubmed; PMID 24723079.

71. Flegal KM, Kalantar-Zadeh K. *Overweight, mortality and survival*. Obesity (Silver Spring). 2013 Sep;21(9):1744-5. doi:10.1002/oby.20588. Cited in: Pubmed; PMID 23929522.

72. Bittman M. Butter is back. . The New York Times 25 March 2014.

73. Stone JC, Barker TH, Aromataris E, Ritskes-Hoitinga M, Sears K, Klugar M, Leonardi-Bee J, Munn Z. *From critical appraisal to risk of bias assessment: clarifying the terminology for study evaluation in JBI systematic reviews*. JBI Evid Synth. 2023 Mar 1;21(3):472-477. Epub 20230301. doi:10.11124/JBIES-22-00434. Cited in: Pubmed; PMID 36882947.

74. Weaver C, Ahles S, Murphy KJ, Shyam S, Cade J, Plat J, Schwingshackl L, Roche HM, Lachat C, Minihane A-M. *Perspective: peer evaluation of recommendations for CONSORT guidelines for randomized controlled trials in nutrition*. Advances in Nutrition. 2023:100154.

75. *How to read clinical journals: I. why to read them and how to start reading them critically*. Can Med Assoc J. 1981 Mar 1;124(5):555-8. eng. Cited in: Pubmed; PMID 7471000.

76. Sackett DL. *Evidence-based medicine*. Semin Perinatol. 1997 Feb;21(1):3-5. doi:10.1016/s0146-0005(97)80013-4. Cited in: Pubmed; PMID 9190027.

77. Fowkes F, Fulton P. *Critical appraisal of published research: introductory guidelines*. BMJ: British Medical Journal. 1991;302(6785):1136.

78. Chalmers TC, Smith H, Jr., Blackburn B, Silverman B, Schroeder B, Reitman D, Ambroz A. *A method for assessing the quality of a randomized control trial*. Control Clin Trials. 1981 May;2(1):31-49. doi:10.1016/0197-2456(81)90056-8. Cited in: Pubmed; PMID 7261638.

79. Kramer MS, Shapiro SH. *Scientific challenges in the application of randomized trials*. JAMA. 1984 Nov 16;252(19):2739-45. Cited in: Pubmed; PMID 6492351.

80. Moher D, Jadad AR, Tugwell P. *Assessing the quality of randomized controlled trials. Current issues and future directions*. Int J Technol Assess Health Care. 1996 Spring;12(2):195-208. doi:10.1017/s0266462300009570. Cited in: Pubmed; PMID 8707495.

81. Naylor CD. *Meta-analysis and the meta-epidemiology of clinical research*. BMJ. 1997 Sep 13;315(7109):617-9. doi:10.1136/bmj.315.7109.617. Cited in: Pubmed; PMID 9310553.

82. Schulz KF. *Subverting randomization in controlled trials*. JAMA. 1995 Nov 8;274(18):1456-8. Cited in: Pubmed; PMID 7474192.

83. Schulz KF, Chalmers I, Hayes RJ, Altman DG. *Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials*. JAMA. 1995 Feb 1;273(5):408-12. doi:10.1001/jama.273.5.408. Cited in: Pubmed; PMID 7823387.

84. Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. *Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological'*

research. Stat Med. 2002 Jun 15;21(11):1513-24. doi:10.1002/sim.1184. Cited in: Pubmed; PMID 12111917.

85. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG, Consolidated Standards of Reporting Trials G. *CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials*. J Clin Epidemiol. 2010 Aug;63(8):e1-37. Epub 20100325. doi:10.1016/j.jclinepi.2010.03.004. Cited in: Pubmed; PMID 20346624.

86. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, Initiative S. *The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies*. J Clin Epidemiol. 2008 Apr;61(4):344-9. doi:10.1016/j.jclinepi.2007.11.008. Cited in: Pubmed; PMID 18313558.

87. Boutron I, Page MJ, Higgins JPT, Altman DG, Lundh A, Hróbjartsson A, on behalf of the Cochrane Bias Methods G. Considering bias and conflicts of interest among the included studies. Cochrane Handbook for Systematic Reviews of Interventions. 2019. p. 177-204.

88. Murad MH, Asi N, Alsawas M, Alahdab F. *New evidence pyramid*. Evidence Based Medicine. 2016;21(4):125. doi:10.1136/ebmed-2016-110401.

89. Burns PB, Rohrich RJ, Chung KC. *The levels of evidence and their role in evidence-based medicine*. Plast Reconstr Surg. 2011 Jul;128(1):305-310. eng. doi:10.1097/PRS.0b013e318219c171. Cited in: Pubmed; PMID 21701348.

90. Murad MH, Katabi A, Benkhadra R, Montori VM. *External validity, generalisability, applicability and directness: a brief primer*. BMJ Evid Based Med. 2018 Feb;23(1):17-19. doi:10.1136/ebmed-2017-110800. Cited in: Pubmed; PMID 29367319.

91. Higgins JPT, Savović J, Page MJ, Elbers RG, Sterne JAC. *Assessing risk of bias in a randomized trial*. Cochrane handbook for systematic reviews of interventions. 2019:205-228.

92. Schulz KF, Grimes DA. *Allocation concealment in randomised trials: defending against deciphering*. Lancet. 2002 Feb 16;359(9306):614-8. doi:10.1016/S0140-6736(02)07750-4. Cited in: Pubmed; PMID 11867132.

93. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, Henry D, Altman DG, Ansari MT, Boutron I, Carpenter JR, Chan AW, Churchill R, Deeks JJ, Hrobjartsson A, Kirkham J, Juni P, Loke YK, Pigott TD, Ramsay CR, Regidor D, Rothstein HR, Sandhu L, Santaguida PL, Schunemann HJ, Shea B, Shrier I, Tugwell P, Turner L, Valentine JC, Waddington H, Waters E, Wells GA, Whiting PF, Higgins JP. *ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions*. BMJ. 2016 Oct 12;355:i4919. Epub 20161012. doi:10.1136/bmj.i4919. Cited in: Pubmed; PMID 27733354.

94. VanderWeele TJ, Ding P. *Sensitivity Analysis in Observational Research: Introducing the E-Value*. Ann Intern Med. 2017 Aug 15;167(4):268-274. Epub 20170711. doi:10.7326/M16-2607. Cited in: Pubmed; PMID 28693043.

95. Greenland S. *Multiple-bias modelling for analysis of observational data*. Journal of the Royal Statistical Society Series A: Statistics in Society. 2005;168(2):267-306.

96. Hiemstra B, Keus F, Wetterslev J, Gluud C, van der Horst ICC. *DEBATE-statistical analysis plans for observational studies*. BMC Medical Research Methodology. 2019 2019/12/09;19(1):233. doi:10.1186/s12874-019-0879-5.

97. Seehra J, Pandis N, Koletsi D, Fleming PS. *Use of quality assessment tools in systematic reviews was varied and inconsistent*. J Clin Epidemiol. 2016 Jan;69:179-84 e5. Epub 20150704. doi:10.1016/j.jclinepi.2015.06.023. Cited in: Pubmed; PMID 26151664.

98. Waddington H, Aloe AM, Becker BJ, Djimeu EW, Hombrados JG, Tugwell P, Wells G, Reeves B. *Quasi-experimental study designs series—paper 6: risk of bias assessment*. Journal of clinical epidemiology. 2017;89:43-52. doi:10.1016/j.jclinepi.2017.02.015.

99. Nguyen VTE, Mishelle ; Davison, Mauricia ; Ravaud, Philippe ; Porcher, Raphael ; Boutron, Isabelle. *Risk of bias in observational studies using routinely collected data of comparative effectiveness research: a meta-research study*. 2021;19(1):279-294. Epub 2021-11.

100. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, Savovic J, Schulz KF, Weeks L, Sterne JA, Cochrane Bias Methods G, Cochrane Statistical Methods G. *The Cochrane Collaboration's tool for assessing risk of bias in randomised trials*. BMJ. 2011 Oct 18;343:d5928. Epub 20111018. doi:10.1136/bmj.d5928. Cited in: Pubmed; PMID 22008217.

101. Sterne JAC, Savovic J, Page MJ, Elbers RG, Blencowe NS, Boutron I, Cates CJ, Cheng HY, Corbett MS, Eldridge SM, Emberson JR, Hernan MA, Hopewell S, Hrobjartsson A, Junqueira DR, Juni P, Kirkham JJ, Lasserson T, Li T, McAleenan A, Reeves BC, Shepperd S, Shrier I, Stewart LA, Tilling K, White IR, Whiting PF, Higgins JPT. *RoB 2: a revised tool for assessing risk of bias in randomised trials*. BMJ. 2019 Aug 28;366:l4898. Epub 20190828. doi:10.1136/bmj.l4898. Cited in: Pubmed; PMID 31462531.

102. Farrah K, Young K, Tunis MC, Zhao L. *Risk of bias tools in systematic reviews of health interventions: an analysis of PROSPERO-registered protocols*. Syst Rev. 2019 Nov 15;8(1):280. Epub 20191115. doi:10.1186/s13643-019-1172-8. Cited in: Pubmed; PMID 31730014.

103. *Quality Assessment and Risk of Bias Tool Repository © 2021 Leila Ledbetter and Stephanie Hendren. Licensed under Attribution-NonCommercial 4.0 International. Available at: https://doi.org/10.17605/OSF.IO/WS824*.

104. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. *Assessing the quality of reports of randomized clinical trials: is blinding necessary?* Control Clin Trials. 1996 Feb;17(1):1-12. eng. doi:10.1016/0197-2456(95)00134-4. Cited in: Pubmed; PMID 8721797.

105. Bai A, Shukla VK, Bak G, G W. Quality Assessment Tools Project Report. Ottawa, Ontario, Canada; 2012. Canadian Agency for Drugs and Technologies in Health

106. Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses*. 2000.

107. Aromataris E MZ, (Editors). JBI Manual for Evidence Synthesis [internet]. Adelaide:: JBI; 2020 June 20, 2022. Available from: https://synthesismanual.jbi.global.

108. *Scottish Intercollegiate Guidelines Network (SIGN) 50 Methodology: Critical Appraisal notes and checklists.* . version. 2014

109. Porritt K, Gomersall J, Lockwood C. *JBI's systematic reviews: study selection and critical appraisal*. AJN The American Journal of Nursing. 2014;114(6):47-52.

110. Higgins JPT, Morgan RL, Rooney AA, Taylor KW, Thayer KA, Silva RA, Lemeris C, Akl EA, Bateson TF, Berkman ND, Glenn BS, Hrobjartsson A, LaKind JS, McAleenan A, Meerpohl JJ, Nachman RM, Obbagy JE, O'Connor A, Radke EG, Savovic J, Schunemann HJ, Shea B, Tilling K, Verbeek J, Viswanathan M, Sterne JAC. *A tool to assess risk of bias in non-randomized follow-up studies of exposure effects (ROBINS-E)*. Environ Int. 2024 Apr;186:108602. Epub 20240324. doi:10.1016/j.envint.2024.108602. Cited in: Pubmed; PMID 38555664.

111. Stone J. *A unified framework for bias assessment of primary analytic research in evidence synthesis*. 2022.

112. Drucker AM, Fleming P, Chan A-W. *Research Techniques Made Simple: Assessing Risk of Bias in Systematic Reviews*. Journal of Investigative Dermatology. 2016 2016/11/01/;136(11):e109-e114. doi:https://doi.org/10.1016/j.jid.2016.08.021.

113. Juni P, Altman DG, Egger M. *Systematic reviews in health care: Assessing the quality of controlled clinical trials*. BMJ. 2001 Jul 7;323(7303):42-6. doi:10.1136/bmj.323.7303.42. Cited in: Pubmed; PMID 11440947.

114. Wang Y, Ghadimi M, Wang Q, Hou L, Zeraatkar D, Iqbal A, Ho C, Yao L, Hu M, Ye Z, Couban R, Armijo-Olivo S, Bassler D, Briel M, Gluud LL, Glasziou P, Jackson R, Keitz SA, Letelier LM, Ravaud P, Schulz KF, Siemieniuk RAC, Brignardello-Petersen R, Guyatt GH. *Instruments assessing risk of bias of randomized trials frequently included items that are not addressing risk of bias issues*. Journal of Clinical Epidemiology. 2022;152:218-225. doi:10.1016/j.jclinepi.2022.10.018.

115. Kuehn R, Wang Y, Guyatt G. *Overly complex methods may impair pragmatic use of core evidence-based medicine principles*. BMJ Evid Based Med. 2024 May 22;29(3):139-141. Epub 20240522. doi:10.1136/bmjebm-2024-112868. Cited in: Pubmed; PMID 38453419.

116. Crocker TF, Lam N, Jordao M, Brundle C, Prescott M, Forster A, Ensor J, Gladman J, Clegg A. *Risk-of-bias assessment using Cochrane's revised tool for randomized trials (RoB 2) was useful but challenging and resource-intensive: observations from a systematic review*. J Clin Epidemiol. 2023 Sep;161:39-45. Epub 20230624. doi:10.1016/j.jclinepi.2023.06.015. Cited in: Pubmed; PMID 37364620.

117. Minozzi S, Cinquini M, Gianola S, Gonzalez-Lorenzo M, Banzi R. *The revised Cochrane risk of bias tool for randomized trials (RoB 2) showed low interrater reliability and challenges in its application*. J Clin Epidemiol. 2020 Oct;126:37-44. Epub 20200618. doi:10.1016/j.jclinepi.2020.06.015. Cited in: Pubmed; PMID 32562833.

118. Martimbianco ALC, Sá KMM, Santos GM, Santos EM, Pacheco RL, Riera R. *Most Cochrane systematic reviews and protocols did not adhere to the Cochrane's risk of bias 2.0 tool*. Rev Assoc Med Bras (1992). 2023;69(3):469-472. eng. Epub 20230220. doi:10.1590/1806-9282.20221593. Cited in: Pubmed; PMID 36820779.

119. Puljak L, Ramic I, Arriola Naharro C, Brezova J, Lin YC, Surdila AA, Tomajkova E, Farias Medeiros I, Nikolovska M, Poklepovic Pericic T, Barcot O, Suarez Salvado M. *Cochrane risk of bias tool was used inadequately in the majority of non-Cochrane systematic reviews*. J Clin Epidemiol. 2020 Jul;123:114-119. eng. Epub 20200401. doi:10.1016/j.jclinepi.2020.03.019. Cited in: Pubmed; PMID 32247026.

120. Bilandzic A, Fitzpatrick T, Rosella L, Henry D. *Risk of Bias in Systematic Reviews of Non-Randomized Studies of Adverse Cardiovascular Effects of Thiazolidinediones and Cyclooxygenase-2 Inhibitors: Application of a New Cochrane Risk of Bias Tool*. PLoS medicine. 2016;13(4):e1001987-e1001987. doi:10.1371/journal.pmed.1001987.

121. Schwingshackl L, Schunemann HJ, Meerpohl JJ. *Improving the trustworthiness of findings from nutrition evidence syntheses: assessing risk of bias and rating the certainty of evidence*. Eur J Nutr. 2021 Sep;60(6):2893-2903. Epub 20201230. doi:10.1007/s00394-020-02464-1. Cited in: Pubmed; PMID 33377996.

122. Igelström E, Campbell M, Craig P, Katikireddi SV. *Cochrane's risk of bias tool for non-randomized studies (ROBINS-I) is frequently misapplied: A methodological systematic review*. Journal of clinical epidemiology. 2021;140:22-32.

123. Thomson H, Craig P, Hilton-Boon M, Campbell M, Katikireddi SV. *Applying the ROBINS-I tool to natural experiments: an example from public health*. Syst Rev. 2018 Jan 24;7(1):15. Epub 20180124. doi:10.1186/s13643-017-0659-4. Cited in: Pubmed; PMID 29368630.

124. Jeyaraman MM, Rabbani R, Copstein L, Robson RC, Al-Yousif N, Pollock M, Xia J, Balijepalli C, Hofer K, Mansour S, Fazeli MS, Ansari MT, Tricco AC, Abou-Setta AM. *Methodologically rigorous risk of bias tools for nonrandomized studies had low reliability and high evaluator burden*. Journal of Clinical Epidemiology. 2020 2020/12/01/;128:140-147. doi:https://doi.org/10.1016/j.jclinepi.2020.09.033.

125. Zhang Y, Huang L, Wang D, Ren P, Hong Q, Kang D. *The ROBINS-I and the NOS had similar reliability but differed in applicability: A random sampling observational studies of systematic reviews/meta-analysis*. J Evid Based Med. 2021 May;14(2):112-122. Epub 20210518. doi:10.1111/jebm.12427. Cited in: Pubmed; PMID 34002466.

126. Losilla J-M, Oliveras I, Marin-Garcia JA, Vives J. *Three risk of bias tools lead to opposite conclusions in observational research synthesis*. Journal of clinical epidemiology. 2018;101:61-72.

127. Bero L, Chartres N, Diong J, Fabbri A, Ghersi D, Lam J, Lau A, McDonald S, Mintzes B, Sutton P, Turton JL, Woodruff TJ. *The risk of bias in observational studies of exposures (ROBINS-E) tool: concerns arising from application to observational studies of exposures*. Systematic reviews. 2018;7(1):242-242. doi:10.1186/s13643-018-0915-2.

128. Del Valle HB, Yaktine AL, Taylor CL, Ross AC. Dietary reference intakes for calcium and vitamin D. National Academies Press; 2011. ISBN: 0309163943.

129. European Food Safety A. *Outcome of public consultations on the Scientific Opinions of the EFSA Panel on Nutrition, Novel Foods and Food Allergens (NDA) on Dietary Reference Values for sodium and chloride*. EFSA Supporting Publications. 2019;16(9):1679E.

130. Yetley EA, MacFarlane AJ, Greene-Finestone LS, Garza C, Ard JD, Atkinson SA, Bier DM, Carriquiry AL, Harlan WR, Hattis D, King JC, Krewski D, O'Connor DL, Prentice RL, Rodricks JV, Wells GA. *Options for basing Dietary Reference Intakes (DRIs) on chronic disease endpoints: report from a joint US-/Canadian-sponsored working group*. Am J Clin Nutr. 2017 Jan;105(1):249S-285S. Epub 2016/12/09. doi:10.3945/ajcn.116.139097. Cited in: Pubmed; PMID 27927637.

131. Zeraatkar D, Kohut A, Bhasin A, Morassut RE, Churchill I, Gupta A, Lawson D, Miroshnychenko A, Sirotich E, Aryal K. *Assessments of risk of bias in systematic reviews of observational nutritional epidemiologic studies are often not appropriate or comprehensive: a methodological study*. BMJ nutrition, prevention & health. 2021;4(2):487.

132. Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, Catalá-López F, Li L, Reid EK, Sarkis-Onofre R, Moher D. *Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study*. PLoS Med. 2016 May;13(5):e1002028. eng. Epub 20160524. doi:10.1371/journal.pmed.1002028. Cited in: Pubmed; PMID 27218655.

133. WHO. (Do formal reference) Air Quality WHO - https://www.euro.who.int/__data/assets/pdf_file/0007/425518/Risk-of-bias-assessment-instrument-for-systematic-reviews-informing-who-global-air-quality-guidelines.pdf. 2020.

134. Tonia T, Buitrago-Garcia D, Peter NL, Mesa-Vieira C, Li T, Furukawa TA, Cipriani A, Leucht S, Low N, Salanti G. *Tool to assess risk of bias in studies estimating the prevalence of mental health disorders (RoB-PrevMH)*. BMJ Ment Health. 2023 Oct;26(1). eng. doi:10.1136/bmjment-2023-300694. Cited in: Pubmed; PMID 37899074.

135. Van Elswyk ME, Teo L, Lau CS, Shanahan CJ. *Dietary Patterns and the Risk of Sarcopenia: A Systematic Review and Meta-Analysis*. Current Developments in Nutrition. 2022 2022/05/01/;6(5):nzac001. doi:https://doi.org/10.1093/cdn/nzac001.

136. Handu D, Moloney L, Wolfram T, Ziegler P, Acosta A, Steiber A. *Academy of Nutrition and Dietetics Methodology for Conducting Systematic Reviews for the Evidence Analysis Library*. Journal of the Academy of Nutrition and Dietetics. 2016 2016/02/01/;116(2):311-318. doi:https://doi.org/10.1016/j.jand.2015.11.008.

137. Lachat C, Hawwash D, Ocke MC, Berg C, Forsum E, Hornell A, Larsson C, Sonestedt E, Wirfalt E, Akesson A, Kolsteren P, Byrnes G, De Keyzer W, Van Camp J, Cade JE, Slimani N, Cevallos M, Egger M, Huybrechts I. *Strengthening the Reporting of Observational Studies in Epidemiology-Nutritional Epidemiology (STROBE-nut): An Extension of the STROBE Statement*. PLoS Med. 2016 Jun;13(6):e1002036. Epub 2016/06/09. doi:10.1371/journal.pmed.1002036. Cited in: Pubmed; PMID 27270749.

138. Schwingshackl L, Knüppel S, Schwedhelm C, Hoffmann G, Missbach B, Stelmach-Mardas M, Dietrich S, Eichelmann F, Kontopantelis E, Iqbal K, Aleksandrova K, Lorkowski S, Leitzmann MF, Kroke A, Boeing H. *Perspective: NutriGrade: A Scoring System to Assess and Judge the Meta-Evidence of Randomized Controlled Trials and Cohort Studies in Nutrition Research*. Adv Nutr. 2016 Nov;7(6):994-1004. eng. Epub 20161115. doi:10.3945/an.116.013052. Cited in: Pubmed; PMID 28140319.

139. Ioannidis JPA. *The Challenge of Reforming Nutritional Epidemiologic Research*. JAMA. 2018 Sep 11;320(10):969-970. doi:10.1001/jama.2018.11025. Cited in: Pubmed; PMID 30422271.

140. Trepanowski JF, Ioannidis JP. *Perspective: limiting dependence on nonrandomized studies and improving randomized trials in human nutrition research: why and how*. Advances in Nutrition. 2018;9(4):367-377.

141. Christensen JJ, Arnesen EK, Andersen R, Eneroth H, Erkkola M, Hoyer A, Lemming EW, Meltzer HM, Halldorsson Thorn I, Thornorsdottir I, Schwab U, Trolle E, Blomhoff R. *The Nordic Nutrition Recommendations 2022 - principles and methodologies*. Food Nutr Res. 2020;64. Epub 20200618. doi:10.29219/fnr.v64.4402. Cited in: Pubmed; PMID 32612489.

142. Arnesen EK, Christensen JJ, Andersen R, Eneroth H, Erkkola M, Hoyer A, Lemming EW, Meltzer HM, Halldorsson Thorn I, Thornorsdottir I, Schwab U, Trolle E, Blomhoff R. *The Nordic Nutrition Recommendations 2022 - structure and rationale of qualified systematic reviews*. Food Nutr Res. 2020;64. Epub 20200618. doi:10.29219/fnr.v64.4403. Cited in: Pubmed; PMID 32612488.

143. Balk EM, Adams GP, Langberg V, Halladay C, Chung M, Lin L, Robertson S, Yip A, Steele D, Smith BT, Lau J, Lichtenstein AH, Trikalinos TA. *Omega-3 Fatty Acids and Cardiovascular Disease: An Updated Systematic Review*. Evid Rep Technol Assess (Full Rep). 2016 Aug;(223):1-1252. Epub 2016/08/01. doi:10.23970/AHRQEPCERTA223. Cited in: Pubmed; PMID 30307737.

144. Chung M, Balk EM, Ip S, Raman G, Yu WW, Trikalinos TA, Lichtenstein AH, Yetley EA, Lau J. *Reporting of systematic reviews of micronutrients and health: a critical appraisal*. The American journal of clinical nutrition. 2009;89(4):1099-1113.

145. National Academies of Sciences E, Medicine. *Dietary Reference Intakes for sodium and potassium*. 2019.

146. Oria MP, Kumanyika S. *Guiding principles for developing dietary reference intakes based on chronic disease*. 2017.

147. National Academies of Sciences E, Medicine. Guiding Principles for Developing Dietary Reference Intakes Based on Chronic Disease. Washington, DC: The National Academies Press; 2017. (Shiriki K, Maria PO, editors.). English. ISBN: 978-0-309-46256-3. https://www.nap.edu/catalog/24828/guiding-principles-for-developing-dietary-reference-intakes-based-on-chronic-disease.

148. Bailey RL. *Overview of dietary assessment methods for measuring intakes of foods, beverages, and dietary supplements in research studies*. Curr Opin Biotechnol. 2021 Aug;70:91-96. eng. Epub 20210311. doi:10.1016/j.copbio.2021.02.007. Cited in: Pubmed; PMID 33714006.

149. Okada E, Nakade M, Hanzawa F, Murakami K, Matsumoto M, Sasaki S, Takimoto H. *National Nutrition Surveys Applying Dietary Records or 24-h Dietary Recalls with Questionnaires: A Scoping Review*. Nutrients. 2023 Nov 9;15(22). eng. Epub 20231109. doi:10.3390/nu15224739. Cited in: Pubmed; PMID 38004132.

150. Cade J, Thompson R, Burley V, Warm D. *Development, validation and utilisation of food-frequency questionnaires - a review*. Public Health Nutr. 2002 Aug;5(4):567-87. eng. doi:10.1079/phn2001318. Cited in: Pubmed; PMID 12186666.

151. Balakrishna Y, Manda S, Mwambi H, van Graan A. *Statistical methods for the analysis of food composition databases: a review*. Nutrients. 2022;14(11):2193.

152. Lucarini M, Durazzo A. Food Composition and Dedicated Databases : Key Tools for Human Health and Public Nutrition. Place of publication not identified: MDPI - Multidisciplinary Digital Publishing Institute; 2023. (Food Composition and Dedicated Databases). ISBN: 3-0365-6851-4.

153. Kristensen M, Pelletier X, Ross AB, Thielecke F. *A High Rate of Non-Compliance Confounds the Study of Whole Grains and Weight Maintenance in a Randomised Intervention Trial-The Case for Greater Use of Dietary Biomarkers in Nutrition Intervention Studies*. Nutrients. 2017;9(1):55. doi:10.3390/nu9010055.

154. Tasevska N, Midthune D, Tinker LF, Potischman N, Lampe JW, Neuhouser ML, Beasley JM, Van Horn L, Prentice RL, Kipnis V. *Use of a urinary sugars biomarker to assess measurement error in self-reported sugars intake in the nutrition and physical activity assessment study (NPAAS)*. Cancer epidemiology, biomarkers & prevention. 2014;23(12):2874-2883. doi:10.1158/1055-9965.EPI-14-0594.

155. Picó C, Serra F, Rodríguez AM, Keijer J, Palou A. Biomarkers of Nutrition and Health: New Tools for New Approaches. Nutrients. 2019;11(5). doi: 10.3390/nu11051092

156. Raiten DJ, Namasté S, Brabin B, Combs G, L'Abbe MR, Wasantwisut E, Darnton-Hill I. *Executive summary—Biomarkers of Nutrition for Development: Building a Consensus123*. The American Journal of Clinical Nutrition. 2011 2011/08/01/;94(2):633S-650S. doi:https://doi.org/10.3945/ajcn.110.008227.

157. Bailey LB, Stover PJ, McNulty H, Fenech MF, Gregory Iii JF, Mills JL, Pfeiffer CM, Fazili Z, Zhang M, Ueland PM. *Biomarkers of nutrition for development—folate review*. The Journal of nutrition. 2015;145(7):1636S-1680S.

158. Raghavan R, Ashour FS, Bailey R. *A review of cutoffs for nutritional biomarkers*. Advances in Nutrition. 2016;7(1):112-120.

159. Tomlinson E, Cooper C, Davenport C, Rutjes AWS, Leeflang M, Mallett S, Whiting P. *Common challenges and suggestions for risk of bias tool development: a systematic review of methodological studies*. Journal of Clinical Epidemiology. 2024;171. doi:10.1016/j.jclinepi.2024.111370.

160. Newberry SJ, Chung M, Anderson CA, Chen C, Fu Z, Tang A, Zhao N, Booth M, Marks J, Hollands S. *Sodium and potassium intake: effects on chronic disease outcomes and risks*. 2018.

Chapter 3
Nutrition-specific risk of bias tools for randomized controlled trials, cohort studies and case-control studies

3.1. Chapter Overview

This chapter contains one manuscript that describes the development and testing of a series of nutrition-specific risk of bias tools intended for randomized controlled trials, cohort studies and case-control studies. Chapter 3 addresses both thesis research questions:

How can nutrition-specific risk of bias assessment tools be developed to address the unique methodological challenges posed by studies in nutritional epidemiology?

To what extent are the newly developed nutrition-specific risk of bias tools valid and reliable when applied across different study designs in nutritional research, and do any additional components developed alongside the nutrition-specific risk of bias tools enhance reliability or usability?

To address these questions, the common sources of bias in nutrition research and their potential to impact related human health outcomes are investigated. Next, existing risk of bias tools and how they account for, or facilitate, in-depth consideration for biases specific to nutrition studies are considered. Third, as a precursor to the development of a set of nutrition-specific risk of bias tools, the constructs needed for risk of bias tools to effectively address the unique methodological challenges inherent to nutrition research are summarized. Following the development and testing of nutrition-specific risk of bias tools for randomized controlled trials, cohort studies and case-control studies, detailed nutrition-specific guidance and a worksheet to facilitate user ratings was created, along with a worked example for demonstration purposes.

3.2. Manuscript Status

Chapter 3 contains one manuscript which is published.:

> **Kelly SE**, Greene-Finestone LS, Yetley EA, Benkhedda K, Brooks SPJ, Wells GA, MacFarlane AJ. *NUQUEST-NUtrition QUality Evaluation Strengthening Tools: development of tools for the evaluation of risk of bias in nutrition studies.* Am J Clin Nutr. 2022 Jan 11;115(1):256-271. doi: 10.1093/ajcn/nqab335. PMID: 34605544; PMCID: PMC8755056.

3.3. Author Contributions

AJM, EAY, GAW, LSGF conceptualized the project. All authors developed the research plan, conducted the research and analyzed the data. SEK drafted the manuscript and all authors contributed to the final version. AJM and GAW had primary responsibility for final content.

3.4. Related Thesis Appendices

The PDF of the published manuscript is presented in APPENDIX 2 along with associated permissions from Elsevier. Additional Materials related to the published manuscript are also located in APPENDIX 2, including a summary of dissemination, knowledge translation and impact.

3.5. Manuscript

NUQUEST – NUtrition QUality Evaluation Strengthening Tools:

Development of tools for the evaluation of risk of bias in nutrition studies

Shannon E. Kelly, Linda S. Greene-Finestone, Elizabeth A. Yetley, Karima Benkhedda, Stephen P.J. Brooks, George A. Wells*, and Amanda J. MacFarlane*

Cardiovascular Research Methods Centre, University of Ottawa Heart Institute, Ottawa, ON, Canada (SEK and GAW)

Public Health Agency of Canada, Ottawa, ON, Canada (LSGF)

U.S. National Institutes of Health (Retired), Bethesda, MD, USA (EAY)

Bureau of Nutritional Sciences, Health Canada, Ottawa, ON, Canada (AJM, KB, SPJB)

Department of Biology, Carleton University, Ottawa, ON, Canada (AJM, SPJB)

*indicates shared senior authorship

Author last names: Kelly, Greene-Finestone, Yetley, Benkhedda, Brooks, Wells, MacFarlane

Corresponding author: Amanda MacFarlane.

Disclaimer: Amanda MacFarlane is an Associate Editor for the American Journal of Clinical Nutrition and played no role in the Journal's evaluation of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Support: This work was supported by Health Canada.

Short title: Nutrition-specific RoB tools

Abbreviations: AHRQ, Agency for Healthcare Research and Quality; CVD, cardiovascular disease; NUQUEST, Nutrition QUality Evaluation Strengthening Tools; PICO, Population, Intervention (or exposure), Comparison and Outcome; QAI, quality assessment instrument; RCT, randomized controlled trial; RoB, Risk of Bias; SIGN, Scottish Intercollegiate Guidelines Network;

Online supplementary material: YES

Data availability: Data described in the manuscript will be made publicly and freely available without restriction in the Online Supplementary Material.

Keywords: Quality assessment instrument, nutrition, randomized controlled trial, cohort study, case-control study, risk of bias tool

ABSTRACT

Background: Dietary exposure assessments are a critical issue in evaluating human nutrition studies; however, nutrition-specific criteria are not consistently included in existing bias assessment tools.

Objective: Our objective was to develop a set of Risk of Bias (RoB) tools that integrated nutrition-specific criteria into validated generic assessment tools to address RoB issues, including those specific to dietary exposure assessment.

Methods: Nutrition QUality Evaluation Strengthening Tools (NUQUEST) development and validation process included eight steps. The first steps identified: 1) a development strategy; 2) generic assessment tools with demonstrated validity; and 3) nutrition-specific appraisal issues. This was followed by: 4) generation of nutrition-specific items, and 5) development of guidance to aid users of NUQUEST. The final steps used established ratings of selected studies and feedback from independent raters to: 6) assess reliability and validity; 7) assess formatting and usability; and 8) finalize NUQUEST.

Results: NUQUEST is based on the Scottish Intercollegiate Guidelines Network checklists for randomized controlled trials, cohort studies and case-control studies. Using a purposive sample of 45 studies representing the three study designs, inter-rater reliability was high (Cohen's kappa 0.73, 95% confidence interval (CI) 0.52, 0.93) across all tools and at least moderate for individual tools (range 0.57 to 1.00). The use of a worksheet improved usability and consistency of overall inter-rater agreement across all study designs (40% without worksheet, 80-100% with worksheet). When compared to published ratings, NUQUEST ratings for evaluated studies demonstrated high concurrent validity (93% perfect or near-perfect agreement). Where there was

disagreement, the nutrition-specific component was a contributing factor in discerning exposure methodological issues.

Conclusion: NUQUEST integrates nutrition-specific criteria with generic criteria from assessment tools with demonstrated reliability and validity. NUQUEST represents a consistent and transparent approach for evaluating RoB issues related to dietary exposure assessment commonly encountered in human nutrition studies.

Introduction

Scientific consensus committees tasked with evaluating diet and health/disease relations for policy, clinical, and educational purposes are increasingly applying evidence-based methodologies to the evaluation of nutrition studies (1). However, a universal observation is that while the concepts and methods of evidence-based medicine are applicable to nutrition topics, diet-related challenges in nutrition studies require consideration (1-7). These include the difficulty in obtaining accurate and comprehensive exposure estimates (i.e., intake or status estimates), the potential for confounding and interactive effects of food substances, and the impact of participant baseline nutritional status on the relation between exposure to a particular food substance and health outcome. If not properly accounted for, these challenges can lead to bias, misleading interpretation, and erroneous conclusions. Thus, the development of an assessment approach that integrates nutrition-specific concepts with accepted general concepts of study design and conduct is of paramount importance to users of nutrition evidence (8-10).

Our objective was to develop a set of reliable and valid tools - Nutrition QUality Evaluation Strengthening Tools (NUQUEST) - to aid in evaluating the types of study designs commonly used for nutrition studies while retaining generic assessment components from existing tools. We found inconsistencies among publications regarding the appropriate terminology as to whether our goal was best described by the term ‘Quality’ or ‘Risk of Bias’ (RoB) (11). These terms, often applied interchangeably, describe conditions related to study design and conduct associated with the validity of study results (12). Quality Assessment Instruments (QAIs) assess the quality of a study from conception to interpretation whereas RoB tools assess the accuracy of estimates of benefit and risk. A QAI, an older term than RoB, includes the assessment of precision and generalizability, as well as RoB. Existing QAIs and RoB tools (assessment instruments) are usually generic and focus on the evaluation of general

study design and conduct issues that are universally important. Examples of generic tools for assessing RoB previously used in the nutrition literature include Cochrane RoB for RCTs and ROBINS-I for observational studies (10, 13). However, they do not address specific nutrition-related challenges that contribute to uncertainty in nutrition studies, but, on occasion, are adapted to address them. For example, the Agency for Health Research Quality (AHRQ) added several nutrient-specific questions to the Cochrane RoB tool for RCTs and the Newcastle-Ottawa Scale for observational studies to address uncertainties associated with dietary assessment measures (10, 14-16). While modifications can improve the sensitivity of these assessment instruments for nutrition studies, inconsistent adaptation and application limit the validity and comparability of review outcomes.

Here we present the development and validation of NUQUEST, a suite of RoB tools for evaluating nutrition RCTs, cohort studies and case control studies. They combine RoB components from an existing assessment instrument with nutrition-specific criteria, detailed guidance, and worksheets that address RoB issues related to dietary exposure assessment. We intended NUQUEST to be helpful to both non-nutrition scientists (e.g., research methodologists, epidemiologists) who may be unfamiliar with nutritional issues, and nutrition scientists who may lack in-depth expertise in human study design and conduct. NUQUEST complements nutrition reporting guidelines, guidelines for conducting nutrition systematic reviews (SRs), and guidelines for grading nutrition evidence, and it can be used in conjunction with these tools when appropriate.

Methods

The co-authors constituted a multidisciplinary working group of seven US and Canadian experts in the areas of nutrition science and epidemiology with the goal to develop a set of nutrition-

specific instruments to be based on validated generic instruments for the assessment of RoB in nutrition research studies. For this work, the term "nutrition studies" would broadly apply to studies in which the focus is on the intake of a food substance (nutrient or other food component), food, dietary pattern or biomarker(s) of nutritional intake/status. The development of the nutrition-specific RoB tools for RCTs, cohort studies, and case-control studies followed an 8-step approach: 1) identify development strategy; 2) identify existing and appropriate assessment instruments for adaptation; 3) identify nutrition-specific appraisal issues; 4) generate nutrition-specific items; 5) adapt and/or develop guidance for use; 6) assess reliability and validity; 7) assess formatting and usability; and 8) finalize NUQUEST (**Figure 1**).

(1) Identify development strategy

The group's overall strategy was to develop a nutrition-specific RoB tool that captured the generic RoB concepts addressed in existing instruments while adding specificity for assessing nutrition issues in nutrition studies. The working group defined the goals, focus, and general approach for the project.

(2) Identify existing and appropriate assessment instruments for adaptation

A rapid, systematic process was used to identify instruments from systematic or comparative review articles that assessed existing QAIs and RoB tools that could serves as a starting point for developing nutrition-specific RoB tools. The rationale for starting with existing instruments was to leverage their demonstrated properties (i.e., validity) and experience of use in the field. We performed a literature search in December 2018 of PubMed, MEDLINE and Google Scholar using a combination of keywords and MeSH syntax relevant to critical appraisal, quality assessment, or RoB to locate potentially relevant articles (**Supplementary File 1**). We included reviews that formally evaluated the appraisal domains and properties of different types of study

designs (e.g., RCTs or observational studies). A working group member screened titles and abstracts and then the full text of potentially relevant articles. Included reviews were vetted by the working group until agreement on inclusion status was achieved. In addition, we searched bibliographic databases to identify instruments published after the search date of the most recent review. Individual generic instruments were identified from recommendations made by included reviews, and generic instruments were assessed by the working group using a process that considered the value and feasibility of adding nutrition-specific items in context with the existing RoB items.

(3) Identify nutrition-specific appraisal issues

In order to develop appropriate nutrition-specific items to add to the generic instruments, the working group identified nutrition-specific appraisal issues. Nutrition-specific issues were defined as those related to nutrition exposure measurement and the interpretation of nutrition data. We considered 'dietary exposure assessment' broadly to encompass multiple components of exposure assessment including accuracy and completeness of exposure assessment methodologies, assurances of adherence to interventions, and useful context information for interpreting dietary exposure data. We drew on published literature describing the experiences of scientists who had either conducted or used nutrition-based SRs (7-10), recommendations of expert groups (4, 5, 17-21), articles that had identified nutrition-specific appraisal issues (2, 3), and considered the experiences among the working group members in evaluating nutrition studies. Preliminary nutrition-specific appraisal issues were compiled from all available sources and then vetted within the working group.

(4) Generate nutrition-specific items

The working group drafted a set of nutrition-specific RoB items for each study type (RCT, cohort studies, case-control studies) based on the identified nutrition-specific appraisal issues

with the intention of "bolting" them onto the selected generic instruments as an additional checklist section.

(5) *Adapt and/or develop guidance for use*

The working group reviewed the existing item-specific guidance for the generic RoB items and edited or revised the text as needed to include context and examples relevant to nutrition studies and the specific study design. We developed detailed guidance *de novo* for each nutrition-specific item using a structured, iterative process. Two additional experts (non-working group members) in clinical and population nutrition reviewed NUQUEST, including the generic and nutrition-specific items and guidance. The proposed guidance was refined and finalized using an iterative approach.

(6) Assess reliability and validity

To assess the impact of the nutrition-specific items on overall study rating, we used a purposive sample of 45 nutrition-based studies representing three study designs (**Supplementary File 2**). Studies were selected based on two criteria. First, the study had to have been included in a SR that had examined the association between a nutrition exposure and a health outcome in humans and assigned an overall study rating based on an assessment instrument. Second, the studies as a whole had to cover a diverse range of study ratings, as determined in the original SR, which allowed for comparisons between the original SR overall rating and the NUQUEST overall rating across a range of ratings. The sample consisted of 15 RCTs, 15 cohort studies and 15 case-control studies on a range of topics. Reliability testing was performed in two rounds; the first round consisted of 10 studies for each study design and the second consisted of 5 additional studies for each design. Two raters (not members of the working group) with knowledge of

research methodology and critical appraisal or nutrition science, independently applied NUQUEST to the selected studies for each set of studies/study designs.

Reliability: We assessed reliability using the Cohen's kappa statistic for inter-rater agreement comparing the overall rating of 'poor' to 'neutral' or 'good'. A kappa value was determined for each checklist and for all checklists combined (22). The working group regarded kappa values less than or equal to 0 as *less than chance agreement*; 0.01–0.20, *slight agreement*; 0.21–0.40, *fair agreement*; 0.41–0.60, *moderate agreement*; 0.61–0.80, *substantial agreement*; and 0.81–1.00, *almost perfect agreement* (23).

In the first round of reliability testing, the working group assessed the proportion of agreement between raters across 10 studies for each of the three types of study design. Inter-rater reliability was assessed using the proportion of agreement between raters (i.e., raters' assessment of good, neutral or poor were exactly the same) and the proportion of near agreement (i.e., raters' assessment of good, neutral or poor were either exactly the same or differed by only one category of good, neutral or poor) were calculated. After assessing the first 10 studies for each checklist, raters met with the working group to discuss sources of disagreement and strategies for improving agreement. The working group made modifications, as needed, before raters completed the remaining 5 studies for each checklist. Consensus for the 45 sets of NUQUEST ratings was reached by discussion between the two raters with adjudication by a third rater.

Validity: As there is no accepted 'gold standard' nutrition-specific assessment instrument, it was not possible to assess criterion validity (i.e., how well NUQUEST is in alignment with a gold standard comparator). A degree of validity was established by identifying a generic tool on which to build NUQUEST that had a long history of application and has performed well in identifying studies of different RoB to the satisfaction of authors and users provides a degree of

validity (Step 2). In addition, content and concurrent validity of NUQUEST was assessed. Content validity was established by including issues specifically related to dietary exposure and other key aspects needed when evaluating nutrition studies (Step 3 and 4), which provided specific content that was in alignment with the important issues related to nutrition studies, over and above the general methods items of the selected generic tool on which NUQUEST was based. Concurrent validity, how a new tool compares to an existing tool, was assessed by comparing the NUQUEST overall rating for each study (poor, neutral and good) to the overall rating reported in the original SR from which the study was selected. The ratings from the original SR assessment were recoded into categories of poor, neutral and good to allow for comparison to NUQUEST.

(7) Assess formatting and usability

Verbal and written feedback was sought from the raters pertaining to the usability of NUQUEST, clarity of the appraisal items and corresponding guidance, and their overall experience applying the tools. They also tracked the time it took to complete each NUQUEST assessment, and documented instances where they perceived scoring to be challenging or guidance unclear.

(8) Finalize NUQUEST

Taking into consideration final feedback from raters, the wording of the NUQUEST guidance was edited for clarity and made consistent across the three study design-specific checklists.

Results

(1) Identify development strategy

The working group identified their initial goals for the development of NUQUEST: (1) to integrate accepted generic criteria for assessing study design and conduct with nutrition-specific methodological issues relating primarily to nutritional intake and exposure; and, (2) to have the

resultant NUQUEST be useful to both epidemiological experts who may have limited knowledge of nutritional methodologies and to nutritional scientists who may have limited knowledge of human study design and conduct issues. To achieve these goals, the group decided to adapt an existing generic assessment instrument to maintain the scientific concepts and qualities associated with the items included in the instrument, with some limited modifications to enhance the usability for nutrition contexts. In addition, the group decided that "bolt-on" nutrition-specific items could be added to the generic assessment instruments when those concepts were not addressed by items in the generic instrument. The group agreed that NUQUEST should initially focus on common study designs employed in nutrition science.

(2) Identify existing and appropriate assessment instruments for adaptation

The rapid review (24, 25) located six SRs (26-31) comparing various assessment instruments following a review of 367 titles and abstracts and 14 full-text articles. Six potentially relevant records (5, 9, 32-35) identified through additional top-up searching were also considered. One unpublished environmental scan was identified (36, 37), but eventually excluded as it identified, but did not evaluate, a number of available assessment instruments.

The working group used the most recent, comprehensive and rigorously conducted SR by Bai et al. to select an assessment instrument for adaptation (26). The Bai et al. SR (26) is an agency-sponsored report completed by the Canadian Agency for Drugs and Technologies in Health that updated the findings of a comprehensive 2002 AHRQ Report (38) that identified and evaluated assessment instruments. Briefly, a total of 267 unique assessment instruments (57 for SRs, 94 for RCTs, 99 for observational studies, 17 for multiple designs, and 60 evidence grading systems) were included following a comprehensive multi-database search of published and unpublished sources and applying a rigorous approach to comparisons and evaluation. Each assessment

instrument was assessed independently by two or more review authors and external content experts based on the adequacy of addressing five study-level domains important for generic assessment instruments: comparability of subjects, exposure ascertainment/intervention, outcome measure/ascertainment, statistical analysis and funding. Bai et al. noted that a variety of instruments for appraising studies exist; yet, no gold standard has been identified. They recommended three generic assessment instruments produced by the Scottish Intercollegiate Guidelines Network (SIGN) (39, 40) (one checklist each for RCTs, cohort studies and case-control studies) based on internal validity and the potential for feasible, consistent and systematic application of the checklists during evidence appraisal.

There are many assessment instruments but only a few that have been developed to assess with consistency a range of study designs including randomized controlled trials (RCTs) and observational studies (cohort studies and case-control studies), such as those developed by SIGN (39, 40) and the Joanna Briggs Institute (41-43). Other advantages of the SIGN instruments are that they are simple to use and include key criteria for RoB, clear guidance for their application and interpretation, and have been used extensively. Also, SIGN instruments are not outcome-specific so they can be used for a wide range of health outcomes. The working group concluded that the advantages of the SIGN approach supported its use as the basis for the concurrent development of nutrition-specific RoB tools across three commonly used nutrition study designs (i.e., RCTs, cohort studies and case-control studies) for which a suite of SIGN tools were available. Because a SIGN checklist for cross-sectional studies was not available, a de novo process will be required for the future development of NUQUEST for cross-sectional studies.

(3) Identify nutrition-specific appraisal issues

Based on findings from a broad literature review (1-5, 17, 19, 34, 44-47) and building on the experiential knowledge of the working group members in performing assessments of nutrition studies, the group identified five essential nutrition-specific issues to consider when evaluating studies. These issues are explained in detail in **Table 1** and broadly relate to: 1) accuracy of exposure/status estimates; 2) baseline exposure/status; 3) exposure/status throughout the study; 4) consideration of the dietary context and 5) the duration of nutrition exposure.

The first issue identified was the challenge involved in obtaining accurate estimates of exposure and status – an issue that has received considerable attention (1, 4, 5, 17, 44, 45, 47). Several authoritative reports have offered guidance on how to deal with these issues in reporting and evaluating study RoB, and in analyzing study results (4, 5, 17, 19, 34). The second issue – the need for accurate baseline exposure or status measurements - has been recognized as an important issue for cohort studies (34) but is also important for all nutrition study designs (2, 3, 46). The third issue - the need to assess exposure/status throughout the study - impacts all study designs, and is especially relevant to the issue of adherence to study protocols in intervention studies (46). The fourth issue - the need to consider dietary context when evaluating the effect of specific food substances or dietary patterns on health outcomes – is important because it can impact the exposure-outcome relation. Diets are complex mixtures containing multiple biologically active components that affect food substance bioavailability and metabolic utilization, and any food substance of interest may be correlated with other food substances in the diet (1, 3, 4). Energy intakes are important for understanding the full effect of an addition or deletion of a food, or substitution of one food for another, for identifying potential errors in dietary intake reporting (e.g., under- or- over reporting of dietary intake) (44, 45, 48). The various dietary and biological complexities and their effects on food substance and outcome relations of interest may vary systematically among subpopulation groups. The last nutrition

issue – the duration of exposure relative to outcome – is important because the duration of exposure must be of sufficient duration to reasonably expect the full effect of the exposure on biomarkers of status on the development of the outcome from both an effectiveness and safety perspective (46, 49-51).

(4) Generate nutrition-specific items

Reflecting on the identified nutrition-specific issues from Step 3, the working group developed nutrition-specific items to address the issues when generic items of the existing assessment instrument either addressed them insufficiently or not at all. In the case of the former, the nutrition-specific items were designed to build on concepts addressed by generic RoB items, thereby emphasizing specific important details related to nutrition exposure. The items were focused on measurement of exposure, assessment of baseline exposure, exposure assessment throughout the study, the dietary context of the exposure and potential for confounding, and the duration of exposure relative to the outcome. They were "bolted-on" to the SIGN checklists, as appropriate for each study design. An iterative process among the working group and raters was used to refine the wording of the nutrition-specific items. The final nutrition-specific items vary in number and wording according to the study design and consider methodological rigour, complexity and user-friendliness (**Table 2**). We added five nutrition-specific items to the RCT checklist and four to the cohort checklist. For the case-control checklist, we retained the original SIGN item for exposure assessment and added two additional nutrition-specific items.

(5) Adapt and/or develop guidance for use

Key nutrition concepts applied to both generic and nutrition-specific items. As such, we added nutrition-specific guidance to the existing guidance for the generic SIGN items, including contextual examples pertinent to nutrition studies (examples of nutrition-specific guidance for generic items included in NUQUEST for cohorts are found in **Table 3**). We also developed d*e*

novo guidance for each nutrition-specific item (examples of guidance for nutrition-specific items included in NUQUEST for cohorts are found in **Table 4**). While exposure was assessed by generic items, we emphasized the importance of accurate exposure assessment in the context of nutrition studies (NUQUEST for RCTs, cohort studies and case-control studies with complete guidance are in **Supplementary files 3, 4 and 5**, respectively). In NUQUEST for case-control studies, the generic item on exposure assessment received greatly enhanced nutrition-specific guidance, similar to that of NUQUEST for RCTs and cohort studies.

Guidance pertaining to the assessment of exposure and intake was expanded with each round of validation to include discussion of accuracy and validity in a nutrition context. Topics covered include baseline assessments, adherence to intervention, self-report, supplement and food product composition, dietary patterns or exposures, biomarkers and appropriate documentation, and the appropriateness of dietary assessment methodologies for different research purposes. Guidance was edited to include examples from nutrition studies to illustrate specific issues and facilitate tool application and decision-making.

(6) Assess reliability and validity

Raters completed the assessment of 45 nutrition studies (i.e., 15 RCTs, 15 cohort studies, and 15 case-control studies) selected from SRs in the nutrition literature in two separate rounds (ten of each study design in round one without a worksheet; five in a second round with a worksheet; **Supplementary file 6**). The working group developed the NUQUEST worksheet based on rater feedback after the first round of testing. This worksheet is intended to provide, as needed, a framework for raters to *a priori* learn how to assess each item as it relates to the research question, the PICO, and methods of assessing intake/nutritional status, outcomes and confounders. Each of the NUQUEST sections, based on items grouped according to selection,

comparability, ascertainment of outcomes (or exposure) and the overall assessment, were rated independently by each rater as ‘good’, ‘neutral’ or ‘poor’.

Reliability: Overall inter-rater reliability for NUQUEST across all study designs (n=45 studies) was *substantial* (0.73, 95% confidence interval (CI) 0.52, 0.93) and *moderate* to *almost perfect* for the individual instruments for RCTs (0.59, 95% CI 0.19, 1.00; n=15 studies), cohort studies (0.57, 95% CI 0.15, 1.00; n=15 studies) and case-control studies (1.00, 95% CI 1.00, 1.00; n=15 studies). Kappa values were not robust when considering only the first round of reliability testing due to the limited sample size (n=30 studies) and the corresponding number of nominal categories, therefore only the combined Kappa values are presented (n=45 studies).

The proportion of agreement of the two raters was generally satisfactory for the individual sections and NUQUEST overall rating (**Table 5**). The proportion of ‘near agreement’ (i.e., the raters’ assessment of good, neutral or poor were either exactly the same or differed by only one category of good, neutral or poor) on the selection, comparability, ascertainment, and nutrition sections of NUQUEST, as well as the overall rating, ranged from 80 to 100% for RCTs, 80 to 100% for cohort studies, and 87 to 100% for case-control studies. The proportion for ‘agreement’ (i.e., raters’ assessment of good, neutral, or poor were exactly the same) was lower. In particular, for the overall rating, agreement was only 40% based on the first 10 studies assessed for each of the three study designs. For the assessment of the remaining five studies for each study design, the agreement improved to 100% for RCTs, 80% for cohort studies, and 100% for case-control studies when raters used the completed worksheets.

Rater feedback indicated that it would have been helpful to meet to review their consensus on NUQUEST items after assessment of the first two to three studies before completing the assessment for all of the studies. They felt that this would have increased their agreement for

items for which ratings were inconsistent. The multidisciplinary backgrounds of the raters underscored the need to understand the nature and sources of discrepancies between their ratings, which was generally dependent on the expertise of each individual rater (study appraisal methodology or nutrition).

Validity: As noted, the SIGN instruments on which NUQUEST was built demonstrates high content validity (26) by virtue of its original development process and extensive history of use. When the working group compared the overall study ratings from NUQUEST (based on consensus from two independent raters) to the overall study rating from the original SR (**Figure 2**), there was agreement for 21 of the 45 studies (46.7%; 10 RCTs, 9 cohort studies, 2 case-control studies), and near-agreement for 42 studies (93.3%; 14 RCTs, 15 cohort studies, 13 case-control studies). When the overall ratings differed between NUQUEST and the original SR, the rating was lower 79% of the time (19 of 24 studies) and higher 21% of the time (5 of 24 studies). For the three studies for which there was not near-agreement (Study ID 1, 32, 39; Figure 2), the nutrition bolt-on section impacted the NUQUEST overall study rating negatively and led to greater disagreement between the NUQUEST overall rating and SR overall rating. In summary, the overall study ratings generally demonstrated agreement between NUQUEST and the original SR ratings and, in instances of major disagreement, the nutrition-specific section was a key contributor to the difference.

(7) Assess formatting and usability

Based on feedback from raters and consensus discussions amongst the working group members, we modified the format and organization of NUQUEST to improve the logical structure and usability while maintaining the scientific concepts outlined in the original SIGN instruments (Supplementary files 1, 2 and 3). We edited the original SIGN items and guidance slightly to improve clarity and, where appropriate, to highlight their relevance to nutrition studies.

NUQUEST was re-organized into five sections that varied in title depending on the study design: (1) *Selection* - Selection of participants (RCT)/Selection of cohorts (cohort) /Creation of study groups (case-control); (2) *Comparability* - Comparability of study groups (RCT, case-control)/Comparability of cohorts (cohort) (3) *Ascertainment* - Ascertainment of outcomes (RCT, cohort)/Exposure ascertainment (case-control); (4) *Nutrition* - Nutrition-specific issues (RCT, cohort and case-control); and (5) *Overall study rating*. A table in the guidance section of each checklist describes the individual sources of bias that NUQUEST are designed to address (i.e., performance, attrition, detection, selection, misclassification and dietary exposure assessment biases). The revised structure and improved guidance brought focus to the generic concepts of selection, comparability and ascertainment and the *de novo* developed bolt-on nutrition concepts, and facilitated the determination of the overall rating for a specific concept (e.g.; selection) that in turn assisted the decision-making required for the overall study rating.

The original SIGN rating scheme for each item included the following response options: *yes*/*no*/*can't say*/*does not apply*. In the first round of validation (30 studies), the raters frequently gave 'fence-sitting' responses where they used the response option of "*can't say*"; however, the working group determined that this rating was not always justified. In response, we modified the rating system for individual items to include a broader range of response options (*yes/probably yes/probably no/no)* and removed the "*can't say*" option. The re-named response option "*Not applicable"* was included only for items where a response would not be pertinent.

The overall rating options for each section are: *good* (+), where almost all criteria are met, there is little or no concern, and low RoB; *neutral* (0), where most criteria are met but there are some flaws with an associated concern, and a moderate RoB; and *poor* (-), where either most or all criteria are not met, there are significant flaws, and a high RoB. The overall ratings for each section are carried forward to the *Overall Study Rating* section, and used when considering the

final assessment of the study (response options*: poor*, *neutral*, or *good*). Guidance for determining overall ratings, including the consideration of the relative importance of specific items within the context of the study design, was edited to reflect the updated format and response options.

Based on qualitative feedback, the raters felt that NUQUEST was easy to apply and feasible to complete within a reasonable time by raters with general knowledge of epidemiology or nutrition science (average 16.5 minutes/study, range 10 to 33 minutes). NUQUEST for cohort studies generally took more time for raters to complete than the others, but time to completion improved with each completed study assessment. To aid the future use of NUQUEST, we identified points to consider when developing a review (**Text Box 1**) and when training raters to use the worksheet (**Text Box 2**).

The raters indicated that they felt more comfortable rating items within their area of expertise, than rating items that did not. Specifically, raters with expertise in epidemiology/study design methodology felt confident to assign stricter ratings to the generic items, whereas raters with expertise in nutrition felt confident to assign stricter ratings for the nutrition-specific items. In addition, raters were more likely to use the "*no information*" option for items outside of their expertise. Often other information in the article could have permitted the rater to assess whether something likely occurred or not. With the addition of a completed worksheet in the second validation round (completed example in **Supplementary file 7**), the raters indicated that they were more confident in applying a rating for items for which they had less expertise and overall more likely to be stricter when applying ratings to all items. This resulted in a tendency to downgrade individual items and overall study ratings. The addition of the worksheet also decreased the use of the "*no information*" rating, however, raters still tended to use this option in areas for which they had less expertise. As such, the working group decided to remove this

option in the final checklists to avoid uninformative answers. Users are now forced to make a judgement on each item, even in cases where details are reported poorly, while taking into account stated accounts of study design, conduct and analysis, and items that the rater judges are likely to have (or have not) occurred in the study.

(8) Finalize NUQUEST

NUQUEST for RCTs, cohort studies and case-control studies, and guidance for each instrument are available in the Online Supplementary Materials, where we have also included an example of a completed worksheet (Supplementary file 7**)**. The wording of the guidance is critical for the application of NUQUEST and refinement is expected to continue with future use.

Discussion

With an emphasis on using the best available evidence to inform public health and nutrition initiatives (52), several evidence appraisal instruments have been developed; however, current instruments do not adequately address methodological challenges inherent to nutrition studies. These challenges contribute to uncertainty and lower our ability to establish nutrition-health outcome associations with confidence. Here we present a series of developed and tested RoB tools that retain generic study appraisal components but with the addition of nutrition-specific appraisal items to address challenges common to nutrition studies. NUQUEST presents an integrated approach for combining generic and nutrition-specific issues in a single tool for assessing the RoB of individual nutrition studies. Other complementary tools include guidelines for designing and conducting nutrition studies (46), guidelines for reporting nutrition studies (2, 34, 35), procedures for conducting nutrition SRs and meta-analyses (53), and guidelines for grading the evidence (32, 54-57). The rigorous evidentiary evaluation and transparency resulting from the use of RoB tools for nutritional epidemiology can identify problematic areas for which

more research or creative statistical approaches for dealing with limitations of dietary exposure data are needed (17).

We built on the existing properties of the SIGN instruments, which are valid, reliable, and have extensive use experience (26). The SIGN instruments have a long history of application and have performed well identifying studies of different quality over the years to the satisfaction of authors and users. The SIGN instruments provided a complimentary suite of tools for evaluating three different study designs commonly used in nutrition studies on which a nutrition-specific instrument could be developed. We enhanced the clarity of the SIGN items by reorganizing them into sections focused on selection, comparability, and ascertainment, and supplemented the items by adding nutrition-related guidance. Nutrition-specific RoB items were added to recognize the nature and impact of the challenges when considering nutrition exposures. We offer suggestions in Text Boxes 1 and 2 and in the guidance and a worksheet template to elucidate topic-specific issues important for study assessment (See Supplementary File 6). We provide an example of a completed worksheet to illustrate its use in the context of a specific research question. We also include reminders in the guidance to consider both general and topic-specific criteria when rating studies. NUQUEST focuses on the extent to which nutrition studies were designed, conducted, analysed, and reported to the highest possible methodological standards, but are not intended to quantify the magnitude of bias that may be attributable to methodological flaws.

Our nutrition-specific RoB tools can be used for assessing single studies or multiple studies included in an SR, the latter being similar to the adaptation of ROBINS-I for evaluation of non-randomized studies of exposures (58). Realizing that some existing assessment instruments are challenging to use (59, 60), we aimed to make NUQUEST user-friendly for both epidemiology and nutrition users via an improved rating system, organization, and detailed nutrition-specific guidance. Our multidisciplinary raters felt that NUQUEST was feasible: assessments could be

completed in a reasonable amount of time, time for completion improved with repeated use, and the inclusion of a completed worksheet facilitated study evaluation. Importantly, we wanted NUQUEST to be educational for users and to promote improvement in nutrition study design, conduct and reporting.

While SRs and meta-analyses are important for informing evidence-based nutrition policy, some nutrition scientists have expressed concern that there is potential for misuse if they are poorly conducted (61). Indeed, a recent SR and meta-analysis revealed that the vast majority of a large sample of nutrition SRs had serious methodological issues, including no RoB assessment of included studies and a lack of formal evaluation of the certainty of evidence (62). Given that most nutrition evidence comes from observational studies where exposure is based on self-reported intakes with systematic biases and random errors that can produce inaccurate and misleading results (4), it is imperative that nutrition-focused SRs are rigorous and that RoB is assessed in a consistent and transparent fashion. Our results showed that the effect of nutrition-specific criteria on overall ratings varied, with some ratings increasing and others decreasing when comparing NUQUEST to other instruments. The use of NUQUEST has several advantages for supporting SR review development and nutrition guidance decisions, including clarity of the key research question, flexibility and transparency in developing the PICO and worksheet, and transparency of the strengths and weaknesses of the relied upon evidence. Additionally, policy and guidance decisions are not solely dependent on the strength of the evidence but also consider other factors (e.g., public health impact, economics) (52). We believe that NUQUEST, with its emphasis on assessment of exposure data, can support future nutrition guidance initiatives.

A common challenge in assessing the RoB in nutrition studies, particularly for observational studies, is the failure of study authors to adequately describe the procedure used to establish the accuracy and appropriateness of their dietary exposure assessment. Researchers completing the

NUQUEST worksheet will be responsible for establishing rating criteria for assessing the accuracy, reliability and appropriateness of dietary intake in selected studies, as well as the appropriateness of the dietary methodology for meeting research objectives. This need can be partially addressed if journal editors require that authors follow reporting guidelines for dietary studies (2, 34, 35). Resources exist to aid in these decisions, such as the NCI's *Dietary Assessment Primer* (17), which provides a comprehensive description of the available dietary assessment instruments, the key concerns in dietary assessment (i.e., measurement error and validation), and recommendations for choosing an approach for dietary assessment for different research objectives. Additionally, the evolving use of statistical procedures to improve the ability to estimate 'true' intakes in nutritional epidemiology research (47) can also provide useful information for establishing NUQUEST worksheet criteria. The failure to document validation procedures or the use of validation procedures generally considered to increase the potential for RoB (e.g., repeatability properties of an exposure methodology) would likely result in lower nutrition-specific ratings; documentation of validation against an external 'gold standard' (e.g., use of a recovery biomarker or controlled feeding trials), particularly when also combined with statistical adjustments for other dietary components and personal characteristics, would warrant higher ratings.

Although NUQUEST strengthens the ability to assess nutrition studies, these instruments have limitations. As is common practice, we recommend training on a set of studies to test topic-specific guidance and use of the worksheet, and discussing consensus on items among raters. The working group did not address issues of outcome specificity. However, in practice, any tool can be made outcome-specific, as can be done with the use of specific criteria in the outcome section of the worksheet, keeping in mind potential biases related to the ascertainment of each outcome of interest (63). Ultimately, even with the guidance and assessment criteria defined, user

judgement will be required. We found that the inclusion of a worksheet improved inter-rater reliability and increased discussion among raters to achieve consensus. These limitations underscore the need for an interdisciplinary team, including experts with knowledge of study rating systems, study design and nutrition. The effective application of NUQUEST will require the team to develop *a priori* rating criteria and estimate their relative impact on study RoB. Since there is no current gold standard approach for the assessment of RoB in nutrition studies, the working group could not assess the criterion validity of NUQUEST. Instead, the exercise here represents an initial assessment of content, construct and concurrent validity. Our results suggest NUQUEST performed well in comparison to other RoB tools and that where there was disagreement, the nutrition-specific items influenced the overall rating. Finally, the absence of a SIGN checklist for cross-sectional studies precluded the development of a NUQUEST for cross-sectional studies, at least for the moment. We recognize the need for such a tool given the plethora of cross-sectional nutrition studies. As such, NUQUEST for cross-sectional studies is being developed *de novo* and will be reported separately.

Conclusion

We developed a set of validated nutrition-specific RoB tools to assess nutrition studies of commonly used study designs. NUQUEST is grounded in the scientific concepts reported in a broadly used generic assessment instrument, while improving the usability for application to nutrition studies. These nutrition-specific RoB tools for RCTs, cohort studies, and case-control studies are intended for the evaluation of individual nutrition studies or studies included in SRs. In addition, NUQUEST will help researchers improve the design, conduct and reporting of future nutrition research studies through its focus on nutrition-specific appraisal items. Finally, the availability of these new tools will foster consistency among future reviews of nutrition studies while enabling future systematic refinement, as warranted, by experience and as science evolves.

Acknowledgments: The authors gratefully acknowledge the Scottish Intercollegiate Guidelines Network (SIGN) for their permission to adapt the SIGN checklists in the development of the NUQUEST checklists. The authors would like to thank Joan Peterson, Dr. Zemin Bai, Amy Johnston, Dr. Shu-Ching Hsieh, Dr. Matthew Parrot, Alexandre Martel, Dr. Jesse Elliott and Cynthia Boudrault for their contributions as raters for the validation of the QAIs. We greatly appreciate Drs. Stephanie Atkinson, Alice Lichtenstein and Maria Makrides for their time and insight on the nutrition-specific items and guidance, and/or individual study RoB.

References

1. Yetley EA, MacFarlane AJ, Greene-Finestone LS, Garza C, Ard JD, Atkinson SA, Bier DM, Carriquiry AL, Harlan WR, Hattis D, King JC, Krewski D, O'Connor DL, Prentice RL, Rodricks JV, Wells GA. *Options for basing Dietary Reference Intakes (DRIs) on chronic disease endpoints: report from a joint US-/Canadian-sponsored working group*. Am J Clin Nutr. 2017 Jan;105(1):249S-285S. Epub 2016/12/09. doi:10.3945/ajcn.116.139097. Cited in: Pubmed; PMID 27927637.

2. Chung M, Balk EM, Ip S, Raman G, Yu WW, Trikalinos TA, Lichtenstein AH, Yetley EA, Lau J. *Reporting of systematic reviews of micronutrients and health: a critical appraisal*. Am J Clin Nutr. 2009 Apr;89(4):1099-113. Epub 2009/02/27. doi:10.3945/ajcn.2008.26821. Cited in: Pubmed; PMID 19244363.

3. Lichtenstein AH, Yetley EA, Lau J. *Application of systematic review methodology to the field of nutrition*. J Nutr. 2008 Dec;138(12):2297-306. Epub 2008/11/22. doi:10.3945/jn.108.097154. Cited in: Pubmed; PMID 19022948.

4. National Academies of Sciences E, Medicine. Guiding Principles for Developing Dietary Reference Intakes Based on Chronic Disease. Washington, DC: The National Academies Press; 2017. (Shiriki K, Maria PO, editors.). English. ISBN: 978-0-309-46256-3. https://www.nap.edu/catalog/24828/guiding-principles-for-developing-dietary-reference-intakes-based-on-chronic-disease.

5. National Academies of Sciences E, Medicine. Dietary Reference Intakes for Sodium and Potassium. Washington, DC: The National Academies Press; 2019. (Virginia AS, Meghan H, Maria O, editors.). English. ISBN: 978-0-309-48834-1. https://www.nap.edu/catalog/25353/dietary-reference-intakes-for-sodium-and-potassium.

6. Trepanowski JF, Ioannidis JPA. *Perspective: Limiting Dependence on Nonrandomized Studies and Improving Randomized Trials in Human Nutrition Research: Why and How*. Adv Nutr. 2018 Jul 1;9(4):367-377. Epub 2018/07/23. doi:10.1093/advances/nmy014. Cited in: Pubmed; PMID 30032218.

7. Zeraatkar D, Johnston BC, Guyatt G. *Evidence Collection and Evaluation for the Development of Dietary Guidelines and Public Policy on Nutrition*. Annu Rev Nutr. 2019 Aug 21;39:227-247. doi:10.1146/annurev-nutr-082018-124610. Cited in: Pubmed; PMID 31433741.

8. Suh JR, Herbig AK, Stover PJ. *New perspectives on folate catabolism*. Annu Rev Nutr. 2001;21:255-82. Cited in: Pubmed; PMID 11375437.

9. Evidence Analysis Manual: Steps in the Academy Evidence Analysis Process. Chicago, Illinois: Dietetics AoNa; April 2016. 1-106 p. Evidence Analysis Library®.

10. Newberry SJ, Chung M, Anderson CAM, Chen C, Fu Z, Tang A, Zhao N, Booth M, Marks J, Hollands S, Motala A, Larkin J, Shanman R, Hempel S. Sodium and Potassium Intake: Effects on Chronic Disease Outcomes and Risks. Rockville (MD): 2018. (AHRQ Comparative Effectiveness Reviews). eng.

11. Guyatt G, Busse J. Methods Commentary: Risk of Bias a Better Term Than Alternatives [Online Commentary]. Evidence Partners for the CLARITY Group at McMaster University; 2021. Available from: https://www.evidencepartners.com/resources/methodological-resources/risk-of-bias-commentary.

12. Hartling L, Ospina M, Liang Y, Dryden DM, Hooton N, Krebs Seida J, Klassen TP. *Risk of bias versus quality assessment of randomised controlled trials: cross sectional study*. BMJ. 2009 Oct 19;339:b4012. Epub 2009/10/21. doi:10.1136/bmj.b4012. Cited in: Pubmed; PMID 19841007.

13. Handel MN, Rohde JF, Jacobsen R, Nielsen SM, Christensen R, Alexander DD, Frederiksen P, Heitmann BL. *Processed meat intake and incidence of colorectal cancer: a systematic review and meta-analysis of prospective observational studies*. Eur J Clin Nutr. 2020 Aug;74(8):1132-1148. Epub 2020/02/08. doi:10.1038/s41430-020-0576-9. Cited in: Pubmed; PMID 32029911.

14. Balk EM, Adams GP, Langberg V, Halladay C, Chung M, Lin L, Robertson S, Yip A, Steele D, Smith BT, Lau J, Lichtenstein AH, Trikalinos TA. *Omega-3 Fatty Acids and Cardiovascular Disease: An Updated Systematic Review*. Evid Rep Technol Assess (Full Rep). 2016 Aug;(223):1-1252. Epub 2016/08/01. doi:10.23970/AHRQEPCERTA223. Cited in: Pubmed; PMID 30307737.

15. Newberry SJ, Chung M, Shekelle PG, Booth MS, Liu JL, Maher AR, Motala A, Cui M, Perry T, Shanman R, Balk EM. *Vitamin D and Calcium: A Systematic Review of Health Outcomes (Update)*. Evid Rep Technol Assess (Full Rep). 2014 Sep;(217):1-929. Epub 2014/09/01. doi:10.23970/AHRQEPCERTA217. Cited in: Pubmed; PMID 30313003.

16. Newberry SJ, Chung M, Booth M, Maglione MA, Tang AM, O'Hanlon CE, Wang DD, Okunogbe A, Huang C, Motala A, Trimmer M, Dudley W, Shanman R, Coker TR, Shekelle PG. *Omega-3 Fatty Acids and Maternal and Child Health: An Updated Systematic Review*. Evid Rep Technol Assess (Full Rep). 2016 Oct;(224):1-826. Epub 2016/10/01. doi:10.23970/AHRQEPCERTA224. Cited in: Pubmed; PMID 30307735.

17. Thompson FE, Kirkpatrick SI, Subar AF, Reedy J, Schap TE, Wilson MM, Krebs-Smith SM. *The national cancer institute's dietary assessment primer: A resource for diet research*. Journal of the Academy of Nutrition and Dietetics. 2015;115(12):1986-1995.

18. Del Valle HB, Yaktine AL, Taylor CL, Ross AC. Dietary reference intakes for calcium and vitamin D. National Academies Press; 2011. ISBN: 0309163943.

19. European Food Safety A. *Outcome of public consultations on the Scientific Opinions of the EFSA Panel on Nutrition, Novel Foods and Food Allergens (NDA) on Dietary Reference Values for sodium and chloride*. EFSA Supporting Publications. 2019;16(9):1679E.

20. National Heart L, Blood I. *Lifestyle interventions to reduce cardiovascular risk: Systematic evidence review from the Lifestyle Work Group, 2013*. 2015.

21. World Health O. Evidence-informed guideline development for World Health Organization nutrition-related normative work: continuous quality improvement for efficiency and impact. 2018. World Health Organization.

22. Cohen J. *A Coefficient of Agreement for Nominal Scales*. Educational and Psychological Measurement. 1960 1960/04/01;20(1):37-46. doi:10.1177/001316446002000104.

23. Cohen J. *Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit*. Psychol Bull. 1968 Oct;70(4):213-20. Epub 1968/10/01. doi:10.1037/h0026256. Cited in: Pubmed; PMID 19673146.

24. Garritty C, Gartlehner G, Nussbaumer-Streit B, King VJ, Hamel C, Kamel C, Affengruber L, Stevens A. *Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews*. J Clin Epidemiol. 2021 Feb;130:13-22. Epub 2020/10/18. doi:10.1016/j.jclinepi.2020.10.007. Cited in: Pubmed; PMID 33068715.

25. Tricco AC, Antony J, Zarin W, Strifler L, Ghassemi M, Ivory J, Perrier L, Hutton B, Moher D, Straus SE. *A scoping review of rapid review methods*. BMC Med. 2015 Sep 16;13:224. Epub 2015/09/18. doi:10.1186/s12916-015-0465-6. Cited in: Pubmed; PMID 26377409.

26. Bai A, Shukla VK, Bak G, G W. Quality Assessment Tools Project Report. Ottawa, Ontario, Canada; 2012. Canadian Agency for Drugs and Technologies in Health

27. Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar VS, Grimmer KA. *A systematic review of the content of critical appraisal tools*. BMC medical research methodology. 2004;4(1):1-11.

28. Lohr KN. *Rating the strength of scientific evidence: relevance for quality improvement programs*. International Journal for Quality in Health Care. 2004;16(1):9-18.

29. Sanderson S, Tatt ID, Higgins J. *Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography*. International journal of epidemiology. 2007;36(3):666-676.

30. Penny Whiting PD, Jelena Savović, Deborah Caldwell, Rachel Churchill. Evidence to inform the development of ROBIS, a new tool to assess the risk of bias in systematic reviews. Bristol, UK; September 2013. 160 p. School of Social and Community Medicine, University of Bristol.

31. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, Petticrew M, Altman DG. *Evaluating non-randomised intervention studies*. Health Technol Assess. 2003;7(27):iii-x, 1-173. eng. Epub 2003/09/23. doi:10.3310/hta7270. Cited in: Pubmed; PMID 14499048.

32. Schwingshackl L, Knuppel S, Schwedhelm C, Hoffmann G, Missbach B, Stelmach-Mardas M, Dietrich S, Eichelmann F, Kontopantelis E, Iqbal K, Aleksandrova K, Lorkowski S, Leitzmann MF, Kroke A, Boeing H. *Perspective: NutriGrade: A Scoring System to Assess and Judge the Meta-Evidence of Randomized Controlled Trials and Cohort Studies in Nutrition Research*. Adv Nutr. 2016 Nov;7(6):994-1004. Epub 2017/02/01. doi:10.3945/an.116.013052. Cited in: Pubmed; PMID 28140319.

33. Yang C, Pinart M, Kolsteren P, Van Camp J, De Cock N, Nimptsch K, Pischon T, Laird E, Perozzi G, Canali R, Hoge A, Stelmach-Mardas M, Dragsted LO, Palombi SM, Dobre I, Bouwman J, Clarys P, Minervini F, De Angelis M, Gobbetti M, Tafforeau J, Coltell O, Corella D, De Ruyck H, Walton J, Kehoe L, Matthys C, De Baets B, De Tre G, Bronselaer A, Rivellese A, Giacco R, Lombardo R, De Clercq S, Hulstaert N, Lachat C. *Perspective: Essential Study Quality Descriptors for Data from Nutritional Epidemiologic Research*. Adv Nutr. 2017 Sep;8(5):639-651. Epub 2017/09/17. doi:10.3945/an.117.015651. Cited in: Pubmed; PMID 28916566.

34. Hornell A, Berg C, Forsum E, Larsson C, Sonestedt E, Akesson A, Lachat C, Hawwash D, Kolsteren P, Byrnes G, De Keyzer W, Van Camp J, Cade JE, Greenwood DC, Slimani N, Cevallos M, Egger M, Huybrechts I, Wirfalt E. *Perspective: An Extension of the STROBE Statement for Observational Studies in Nutritional Epidemiology (STROBE-nut): Explanation and Elaboration*. Adv Nutr. 2017 Sep;8(5):652-678. Epub 2017/09/17. doi:10.3945/an.117.015941. Cited in: Pubmed; PMID 28916567.

35. Lachat C, Hawwash D, Ocke MC, Berg C, Forsum E, Hornell A, Larsson C, Sonestedt E, Wirfalt E, Akesson A, Kolsteren P, Byrnes G, De Keyzer W, Van Camp J, Cade JE, Slimani N, Cevallos M, Egger M, Huybrechts I. *Strengthening the Reporting of Observational Studies in Epidemiology-Nutritional Epidemiology (STROBE-nut): An Extension of the STROBE Statement*. PLoS Med. 2016 Jun;13(6):e1002036. Epub 2016/06/09. doi:10.1371/journal.pmed.1002036. Cited in: Pubmed; PMID 27270749.

36. Farrah K, Young K, Tunis MC, Zhao L. *Risk of bias tools in systematic reviews of health interventions: an analysis of PROSPERO-registered protocols*. Systematic Reviews. 2019 2019/11/15;8(1):280. doi:10.1186/s13643-019-1172-8.

37. Charlton P, Doucet S, Azar R, Nagel DA, Boulos L, Luke A, Mears K, Kelly KJ, Montelpare WJ. *The use of the environmental scan in health services delivery research: a scoping review protocol*. BMJ Open. 2019;9(9):e029805. doi:10.1136/bmjopen-2019-029805.

38. West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, Lux L. *Systems to rate the strength of scientific evidence*. Evid Rep Technol Assess (Summ). 2002 Mar;(47):1-11. Epub 2002/05/01. Cited in: Pubmed; PMID 11979732.

39. *Scottish Intercollegiate Guidelines Network (SIGN) 50 Methodology. Critical appraisal: Notes and checklists. Checklist 2: Randomised controlled trials, methodology checklist 3: Cohort studies, methodology checklist 4: Case-control studies.* version. 2014.

40. *Scottish Intercollegiate Guidelines Network (SIGN) 50 Methodology: Critical Appraisal notes and checklists.* . version. 2014

41. Joanna Briggs Institute. *Critical Appraisal Checklist for Randomized Controlled Trials*. 2020.

42. Joanna Briggs Institute. *Critical Appraisal Checklist for Cohort Studies*. 2020.

43. Joanna Briggs Institute. *Critical Appraisal Checklist for Case Control Studies*. 2020.

44. Freedman LS, Commins JM, Moler JE, Arab L, Baer DJ, Kipnis V, Midthune D, Moshfegh AJ, Neuhouser ML, Prentice RL, Schatzkin A, Spiegelman D, Subar AF, Tinker LF, Willett W. *Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake*. Am J Epidemiol. 2014 Jul 15;180(2):172-88. Epub 2014/06/12. doi:10.1093/aje/kwu116. Cited in: Pubmed; PMID 24918187.

45. Freedman LS, Commins JM, Moler JE, Willett W, Tinker LF, Subar AF, Spiegelman D, Rhodes D, Potischman N, Neuhouser ML, Moshfegh AJ, Kipnis V, Arab L, Prentice RL. *Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for potassium and sodium intake*. Am J Epidemiol. 2015 Apr 1;181(7):473-87. Epub 2015/03/20. doi:10.1093/aje/kwu325. Cited in: Pubmed; PMID 25787264.

46. Lichtenstein AH, Petersen K, Barger K, Hansen KE, Anderson CAM, Baer DJ, Lampe JW, Rasmussen H, Matthan NR. *Perspective: Design and Conduct of Human Nutrition Randomized Controlled Trials*. Adv Nutr. 2021 Feb 1;12(1):4-20. Epub 2020/11/18. doi:10.1093/advances/nmaa109. Cited in: Pubmed; PMID 33200182.

47. Zheng C, Beresford SA, Van Horn L, Tinker LF, Thomson CA, Neuhouser ML, Di C, Manson JE, Mossavar-Rahmani Y, Seguin R, Manini T, LaCroix AZ, Prentice RL. *Simultaneous association of total energy consumption and activity-related energy expenditure with risks of cardiovascular disease, cancer, and diabetes among postmenopausal women*. Am J Epidemiol. 2014 Sep 1;180(5):526-35. Epub 2014/07/14. doi:10.1093/aje/kwu152. Cited in: Pubmed; PMID 25016533.

48. Dietary Assessment Primer, Energy Adjustment. National Institutes of Health, National Cancer Institute. (https://dietassessmentprimer.cancer.gov/). [September 15, 2021].

49. Houghton LA, Gray AR, Rose MC, Miller JC, Hurthouse NA, Gregory JF, 3rd. *Long-term effect of low-dose folic acid intake: potential effect of mandatory fortification on the prevention of neural tube defects*. Am J Clin Nutr. 2011 Jul;94(1):136-41. Epub 2011/05/20. doi:10.3945/ajcn.110.004549. Cited in: Pubmed; PMID 21593499.

50. Nguyen P, Tam C, O'Connor DL, Kapur B, Koren G. *Steady state folate concentrations achieved with 5 compared with 1.1 mg folic acid supplementation among women of childbearing*

age. Am J Clin Nutr. 2009 Mar;89(3):844-52. Epub 2009/01/23. doi:10.3945/ajcn.2008.26878. Cited in: Pubmed; PMID 19158211.

51. Tighe P, Ward M, McNulty H, Finnegan O, Dunne A, Strain J, Molloy AM, Duffy M, Pentieva K, Scott JM. *A dose-finding trial of the effect of long-term folic acid intervention: implications for food fortification policy*. Am J Clin Nutr. 2011 Jan;93(1):11-8. Epub 2010/10/29. doi:10.3945/ajcn.2010.29427. Cited in: Pubmed; PMID 20980493.

52. Johnston BC, Seivenpiper JL, Vernooij RWM, de Souza RJ, Jenkins DJA, Zeraatkar D, Bier DM, Guyatt GH. *The Philosophy of Evidence-Based Principles and Practice in Nutrition*. Mayo Clin Proc Innov Qual Outcomes. 2019 Jun;3(2):189-199. Epub 2019/06/14. doi:10.1016/j.mayocpiqo.2019.02.005. Cited in: Pubmed; PMID 31193887.

53. Cumpston M, Li T, Page MJ, Chandler J, Welch VA, Higgins JP, Thomas J. *Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane Database Syst Rev. 2019 Oct 3;10:ED000142. Epub 2019/10/24. doi:10.1002/14651858.ED000142. Cited in: Pubmed; PMID 31643080.

54. Johnston BC, Alonso-Coello P, Bala MM, Zeraatkar D, Rabassa M, Valli C, Marshall C, El Dib R, Vernooij RWM, Vandvik PO, Guyatt GH. *Methods for trustworthy nutritional recommendations NutriRECS (Nutritional Recommendations and accessible Evidence summaries Composed of Systematic reviews): a protocol*. BMC Med Res Methodol. 2018 Dec 5;18(1):162. Epub 2018/12/07. doi:10.1186/s12874-018-0621-8. Cited in: Pubmed; PMID 30518328.

55. Committee ES, Hardy A, Benford D, Halldorsson T, Jeger MJ, Knutsen HK, More S, Naegeli H, Noteborn H, Ockleford C, Ricci A, Rychen G, Schlatter JR, Silano V, Solecki R, Turck D, Benfenati E, Chaudhry QM, Craig P, Frampton G, Greiner M, Hart A, Hogstrand C, Lambre C, Luttik R, Makowski D, Siani A, Wahlstroem H, Aguilera J, Dorne J-L, Fernandez Dumont A, Hempen M, Valtueña Martínez S, Martino L, Smeraldi C, Terron A, Georgiadis N, Younes M. *Guidance on the use of the weight of evidence approach in scientific assessments* [https://doi.org/10.2903/j.efsa.2017.4971]. EFSA Journal. 2017 2017/08/01;15(8):e04971. doi:https://doi.org/10.2903/j.efsa.2017.4971.

56. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ. *GRADE: an emerging consensus on rating quality of evidence and strength of recommendations*. Bmj. 2008 Apr 26;336(7650):924-6. eng. Epub 2008/04/26. doi:10.1136/bmj.39489.470347.AD. Cited in: Pubmed; PMID 18436948.

57. Katz DL, Karlsen MC, Chung M, Shams-White MM, Green LW, Fielding J, Saito A, Willett W. *Hierarchies of evidence applied to lifestyle Medicine (HEALM): introduction of a strength-of-evidence approach based on a methodological systematic review*. BMC Med Res Methodol. 2019 Aug 20;19(1):178. Epub 2019/08/21. doi:10.1186/s12874-019-0811-z. Cited in: Pubmed; PMID 31429718.

58. Morgan RL, Thayer KA, Santesso N, Holloway AC, Blain R, Eftim SE, Goldstone AE, Ross P, Ansari M, Akl EA, Filippini T, Hansell A, Meerpohl JJ, Mustafa RA, Verbeek J, Vinceti M, Whaley P, Schunemann HJ, Group GW. *A risk of bias instrument for non-randomized studies of exposures: A users' guide to its application in the context of GRADE*. Environ Int. 2019

Jan;122:168-184. Epub 2018/11/27. doi:10.1016/j.envint.2018.11.004. Cited in: Pubmed; PMID 30473382.

59. Minozzi S, Cinquini M, Gianola S, Castellini G, Gerardi C, Banzi R. *Risk of bias in nonrandomized studies of interventions showed low inter-rater reliability and challenges in its application*. J Clin Epidemiol. 2019 Aug;112:28-35. Epub 2019/04/15. doi:10.1016/j.jclinepi.2019.04.001. Cited in: Pubmed; PMID 30981833.

60. Thomson H, Craig P, Hilton-Boon M, Campbell M, Katikireddi SV. *Applying the ROBINS-I tool to natural experiments: an example from public health*. Syst Rev. 2018 Jan 24;7(1):15. Epub 20180124. doi:10.1186/s13643-017-0659-4. Cited in: Pubmed; PMID 29368630.

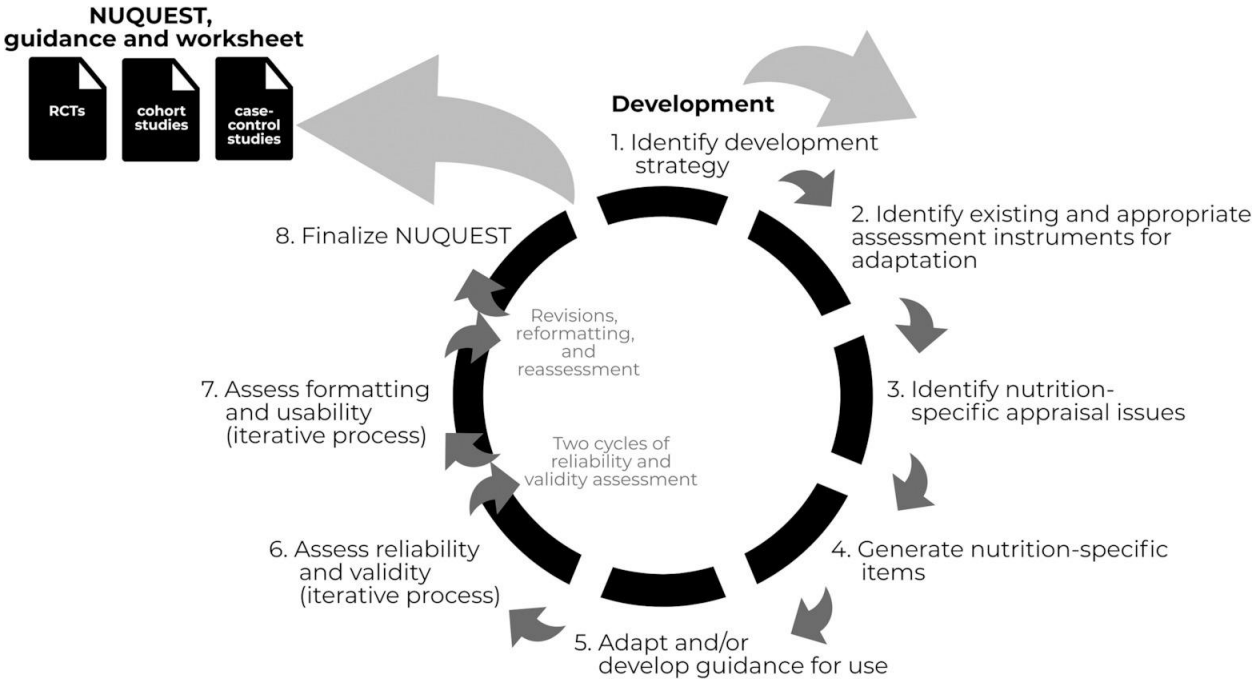
61. Tobias DK, Wittenbecher C, Hu FB. *Grading nutrition evidence: where to go from here?* The American Journal of Clinical Nutrition. 2021;113(6):1385-1387. doi:10.1093/ajcn/nqab124.

62. Zeraatkar D, Bhasin A, Morassut RE, Churchill I, Gupta A, Lawson DO, Miroshnychenko A, Sirotich E, Aryal K, Mikhail D, Khan TA, Ha V, Sievenpiper JL, Hanna SE, Beyene J, de Souza RJ. *Characteristics and quality of systematic reviews and meta-analyses of observational nutritional epidemiology: a cross-sectional study*. Am J Clin Nutr. 2021 Jun 1;113(6):1578-1592. eng. Epub 2021/03/20. doi:10.1093/ajcn/nqab002. Cited in: Pubmed; PMID 33740039.

63. Viswanathan M, Patnode CD, Berkman ND, Bass EB, Chang S, Hartling L, Murad MH, Treadwell JR, Kane RL. *Recommendations for assessing the risk of bias in systematic reviews of health-care interventions*. Journal of Clinical Epidemiology. 2018 2018/05/01/;97:26-34. doi:https://doi.org/10.1016/j.jclinepi.2017.12.004.

FIGURES

Figure 1: Approach for the development of NUQUEST, guidance and worksheet.



NUQUEST = NUtrition QUality Evaluation Strengthening Tools; RCT = randomized controlled trials.

Figure 2: NUQUEST ratings for individual sections and overall and the overall rating from the original systematic review.

STUDY DESIGN	STUDY ID	NUQUEST ASSESSMENT[1]: Selection[2]	Comparability[3]	Ascertainment[4]	Nutrition[5]	Overall[6]	SR: Overall
RCT	1	◐	○	○	○	○[7]	●[8]
	2	◐	◐	●	●	◐	○[8]
	3	●	◐	●	◐	◐	◐[9]
	4	●	◐	●	●	●[7]	●[8]
	5	○	○	●	◐	◐	◐[9]
	6	○	◐	○	○	○[7]	○[8]
	7	◐	○	●	●	◐	◐[9]
	8	●	●	●	●	●	●[8]
	9	◐	◐	●	●	◐	◐[9]
	10	●	●	●	●	●	●[8]
	11	●	◐	○	○	◐[7]	●[8]
	12	○	○	●	○	○	◐[9]
	13	●	●	●	●	●	◐[8]
	14	○	◐	○	○	○[7]	○[8]
	15	○	○	●	○	○	○[9]
Cohort study	16	●	◐	◐	◐	◐	◐[10]
	17	◐	○	◐	◐	◐	○[10]
	18	●	○	◐	◐	○	○[10]
	19	◐	○	○	◐	○	○[10]
	20	○	○	◐	○	○[7]	○[10]
	21	●	○	◐	●	◐	◐[10]
	22	●	○	◐	◐	◐	◐[10]
	23	◐	◐	●	◐	◐	●[10]
	24	●	●	●	●	●	●[10]
	25	◐	◐	◐	◐	◐[7]	●[10]
	26	◐	○	◐	◐	◐	○[10]
	27	○	○	○	○	○[7]	○[10]
	28	○	○	◐	○	○[7]	○[10]
	29	●	○	○	●	◐	○[10]
	30	○	○	○	○	○[7]	◐[10]
Case-control study	31	●	○	○	○	○[7]	◐[11]
	32	●	◐	○	○	○	●[11]
	33	●	○	○	○	○[7]	◐[11]
	34	●	◐	●	◐	◐	◐[11]
	35	●	◐	◐	●	◐	●[11]
	36	●	◐	●	◐	◐	●[11]
	37	●	◐	◐	◐	◐	●[12]
	38	◐	○	○	○	○[7]	◐[11]
	39	◐	○	◐	○	○	●[12]
	40	◐	○	○	○	○[7]	◐[11]
	41	◐	○	●	○	○	○[13]
	42	○	○	◐	○	○[7]	◐[11]
	43	●	◐	●	◐	◐	●[11]
	44	●	○	●	◐	◐	●[11]
	45	●	◐	○	●	◐	●[11]

[1]NUQUEST assessments are by section and overall by study: ● *good*, where almost all criteria are met, little or no concern and low RoB; ● *neutral*, where most criteria are met, there are some flaws and moderate RoB; and ○ *poor*, where either most or all criteria are not met, there are significant flaws and high RoB.

[2]Selection = Selection of participants (RCT)/Selection of cohorts (cohort) /Creation of study groups (case-control).

[3]Comparability = Comparability of study groups (RCT, case-control)/Comparability of cohorts (cohort).

[4]Ascertainment = Ascertainment of outcomes (RCT, cohort)/Exposure ascertainment (case-control).

[5]Nutrition = Nutrition-specific (RCT, cohort, case-control).

[6]Overall = Overall NUQUEST rating.

[7]Assessed using the NUQUEST worksheet.

[8]Evaluated by the Cochrane RoB Tool modified for nutrition studies.

[9]Evaluated by the Cochrane RoB Tool.

[10]Evaluated by NOS modified for nutrition cohort studies.

[11]Evaluated by the NOS.

[12]Evaluated by the JBI critical appraisal tool for case-control studies.

[13]Evaluated by the JBI critical appraisal tool for cross-sectional studies. Because the original systematic reviews used a variety of tools to assess RoB, the original scores were re-coded to "poor, neutral and good" to allow for comparison to NUQUEST scores. Recoding for Cochrane RoB Tool or the NOS modified for nutrition cohort studies was: Low risk of bias, it was assigned good ●; Moderate Risk of bias, it was assigned neutral ●; High risk of bias, it was assigned poor ○. Recoding for NOS was: 1-3, it was assigned ○; 4-6, it was assigned ●; 7-9, it was assigned ●. Recoding for JBI cross-sectional tool was: 1-4, it was assigned ○; 5-8, it was assigned ●. Recoding for JBI case-control tool was: 1-3, it was assigned ○; 4-7, it was assigned ●; 8-10, it was assigned ●.Where the authors of the original systematic review did not make an overall RoB judgement, section-specific judgements and related text within the report were examined to assign an overall RoB assessment by two members of the working group.

JBI, Joanna Briggs Institute; NOS, Newcastle-Ottawa Scale; NUQUEST, NUtrition QUality Evaluation Strengthening Tools. RCT, randomized controlled trials; RoB, Risk of Bias.

Tables

Table 1. Nutrition-specific issues.[1,2]

Nutrition-specific issue	Rationale for consideration of the issue in the NUQUEST checklists
Assessing exposure	This issue refers to the ability to assess nutritional exposures accurately and with minimum error. • The two main issues are: measurement error and validation of the assessment methodology. • Measurement error is the difference between the measured value and the true value. It has two forms - random error and systematic bias. • Random error (day-to-day or within-person variation) is related to an individual's measured intake on a specific day of testing vs. an individual's long term average intake ("usual" intake) based on multiple administrations of the same instrument. o These data are imprecise but not biased. o With repeated measures on a subsample of a population, statistical modeling can be used to adjust for these effects (i.e., to estimate "usual" intakes). o Failure to deal with random error results in overestimates of tail probabilities, attenuated relationships, and loss of power. • Systematic error (bias) results in measurements that depart from the 'true intake" in the same direction. o This can result from under- or over-reporting or mis-reporting intakes and may also be related to individual characteristics (e.g., BMI). Some biomarker methodologies can also exhibit bias. o This type of error cannot be reduced or eliminated by taking repeated measurements. o Bias can result in loss of power to detect diet/health relationships or in errors in establishing accurate intake/response curves. o The magnitude of bias varies by type of methodology used, among different nutrients, and with participant characteristics. • Validation studies are conducted to determine how accurately exposure instruments measure true exposures o One type, rarely used, assesses validation by collecting reference measures by direct observation or feeding studies for a time period that is exactly consistent with when each exposure measurement is taken. o A second type collects reference measures such as recovery biomarkers or less biased assessment instruments for a time period that is not exactly the same as the measured exposure

Nutrition-specific issue	Rationale for consideration of the issue in the NUQUEST checklists
	assessment. Bias is the difference between the average reported intake and average true intake at the group level. o Correlation coefficients between measured and true usual intakes are related to the loss of power to detect exposure/outcome relations. o Unbiased reference measures are limited, but include some recovery biomarkers and accurately collected data from feeding studies or direct observation. o Validation of one method using a comparison to a second method that has a known bias or imprecision (i.e., a weak or flawed reference instrument) can propagate the bias. A common example of this is when one Food Frequency Questionnaire with a known bias is used as the comparator for another new Food Frequency Questionnaire. • Sources of error and lack of validation for exposure assessments can result in misleading results for assessing diet/health relations, intake/response curves, and population prevalence estimates. • Differences in research objectives can affect the appropriateness of different dietary assessment approaches. • Failure to consider measurement error and validation can lead to erroneous and misleading study conclusions.
Baseline exposure	This issue refers to the need for consideration of baseline exposure in assessing exposure-outcome relations. • Allows for correct description/identification of study groups. • Allows for comparison of nutritional and dietary similarities and differences among study groups • Facilitates the evaluation of generalizability of study populations to user target populations • Facilitates the identification of potential nutrition-based confounders and nutrient-nutrient interactions that can influence the relations between the exposure of interest and the outcome, and therefore should be accounted for in the analysis • Observational studies: Allows for appropriate assignment of participants to study groups • Intervention trials: Baseline exposures can inform on total exposure (from both diet/supplements and intervention, for example) allowing for the determination of accurate intake-response curves.
Exposure throughout study	This issue refers to the need to document changes to exposure throughout the study that could affect exposure-outcome relations. • Identifies exposure changes during the study that can occur due to self-initiated changes in participant dietary patterns and supplement use or compositional changes in marketed food products used by study participants that occur after baseline assessment. These

Nutrition-specific issue	Rationale for consideration of the issue in the NUQUEST checklists
	changes can affect exposure-outcome relations in unpredictable ways and lead to erroneous conclusions. Accounting for these changes in analyses is necessary to prevent misleading and erroneous results. • Intervention studies: Adherence to intervention products should be monitored as it can impact the actual exposure and thus the exposure-outcome relation.
Dietary context	This issue refers to the need to consider the complexity of the dietary exposure to determine its impact, if any, on the exposure-outcome relation. • Food substances may correlate with other food substances in the diet thereby confounding the attribution of an exposure-outcome relation to a single specific food substance • Bioavailability differences among food substance sources (e.g., naturally occurring or synthetic) and/or the varying effects of different dietary patterns on the bioavailability of the food substance of interest can affect the exposure-outcome relation. • Direct or indirect interactions between the food substance and other dietary components can alter its effect on the outcome. • Dietary context provides information on total intakes or exposure of the food substance (e.g., background diet plus supplements, bioavailable vs less bioavailable formats) thus more accurately capturing the effect of the food substance on the outcome. • Dietary context can provide information on energy intakes. This information is important in adjusting for potential biases in nutrient exposures related to energy intakes, when evaluating equivalencies of food or dietary pattern substitutions, or in identifying the potential for under- or over- reporting of food substances of interest. • Dietary context provides information to calculate ranges or distributions of the food substance thus enhancing the ability to create intake-response curves • Dosing or eating conditions (e.g., single bolus vs. multiple exposures/day; with meals vs. between meals; chemical forms of the supplement) can affect the intake-response relation • Intervention studies: the ability to "blind" a study may be affected when foods or dietary patterns are the intervention; in studies where there is an addition or deletion of a food, or a food substitution is made, energy equivalency between comparison groups must be considered.
Duration of exposure relative to outcome	This issue refers to the need for the exposure to the food substance to be of sufficient duration to see the full effect on the outcome. • The time needed to show a significant and non-transient change in the outcome may vary by food substance and outcome assessed

Nutrition-specific issue	Rationale for consideration of the issue in the NUQUEST checklists
	• Baseline exposure of study participants can affect the length of time of exposure needed to see an effect of the food substance on the outcome • The dose of the exposure may affect the duration required to observe a change in the outcome. For example, low doses consumed for a longer duration can have the same effect on the outcome as high doses consumed for a shorter time. • Similarly, the format of the exposure may affect the duration required to observe a change in the outcome. For example, food substance in supplement form may be more bioavailable than as part of a dietary intervention. • Outcomes based on a surrogate marker might reasonably be expected to occur in a shorter time period than the occurrence of a disease outcome.

Table 2. Nutrition-specific items in NUQUEST.

Study design	Nutrition-specific items
Randomized controlled trial	• The frequency and quantity of the intervention exposure under study are accurately and reliably measured.
	• The relevant baseline exposure is measured and taken into account.
	• The baseline exposures of all groups have been maintained over the course of the study.
	• Adherence to the intervention and control intakes is monitored in a reliable and accurate manner.
	• The interval between the intervention and outcome is of sufficient duration to observe an effect, if there is one.
Cohort study	• The frequency and quantity of the exposure under study are accurately and reliably measured.
	• The relevant exposure at baseline is measured and taken into account.
	• The baseline exposure differences between the groups have been maintained over the course of the study.
	• The interval between the exposure and outcome is of sufficient duration to observe an effect, if there is one.
	• The exposure under study is accurately and reliably measured.

Case-control study[1]	• The baseline exposure differences between the groups have been maintained over the course of the study.
	• The interval between the exposure and outcome is of sufficient duration to observe an effect, if there is one.

[1]All items except for the exposure assessment in NUQUEST for case-control studies are found in the nutrition-specific section of NUQUEST. An exception was made for NUQUEST for case-control studies because the generic assessment items for this study design included an exposure ascertainment item. Nutrition-specific guidance was developed for this generic item to highlight the specific issues in assessing nutrition exposures, and to highlight the fact that retrospective case-control study designs are particularly vulnerable to recall bias.

Table 3. Examples of nutrition-specific guidance developed for generic items of NUQUEST for cohort studies.

Concept	Item	Sample guidance
Selection of participants/ creation of study groups	1.1 The groups being studied are selected from source populations that are comparable in all respects other than the exposure under investigation.	…groups that are formed from two different source populations or are selected from the same source population using different approaches are at higher risk for selection bias. For example, comparing a group of women of childbearing age from an urban center who took folic acid during pregnancy to those from a rural area who did not would be problematic…
	1.2 The study indicates how many of the people asked to take part did so in each of the groups being studied.	…Groups selected from different source populations or from the same source population using different approaches, may have differential participation rates and should be evaluated more closely. For example, one would be less concerned about differential participation rates in a cohort of women of childbearing age who were selected from the same urban area and groups were formed based on whether or not they took folic acid during their pregnancy. In contrast, if populations from rural and urban areas are compared, we would have more concern that participation rates might differ because, for instance, access to the study centers may be more difficult for the rural participants.
Comparability of groups	1.8 The main potential confounders are identified and/or taken into account	…There may be fundamental differences between the groups that may impact nutrient intake or exposure and these should be considered and accounted for in a study (e.g., health-related behavior or personal/lifestyle variables). These factors may be known or unknown. For example, in a cohort of

Concept	Item	Sample guidance
	in the design and analysis.	women of childbearing age where folic acid intake (compared to no intake) is being studied, we may want to consider potential confounders that may be related to the nutrient intake such as education, physical activity and race or ethnicity. In addition, and particularly when we have different source populations (e.g., different geographies), a variety of known and unknown differences between the groups may confound the nutrient intake or exposure, including but not limited to population structure, culture, genetic diversity and health or personal behaviors.

Table 4. Examples of nutrition-specific items and guidance developed for the nutrition-specific items of NUQUEST for cohort studies.

Concept	Item	Sample guidance
Exposure of interest/ Intervention intake	2.1 The frequency and quantity of the exposure under study are accurately and reliably measured.	…**The study should report the method used to assess exposures or status and the validity of such methods.** Validity refers to the degree to which the exposure assessment methodology accurately measures the aspect of exposure it was intended to measure. Exposure assessment methods are validated in different ways. Comparing an exposure method to a "gold standard' or reference method thought to capture all the food and supplement sources containing the exposure of interest provides the best validation approach. Comparing a self-reported method to another self-reported method requires caution as the two methods may be affected by the same types of errors and biases. An exposure assessment method could be modified to reflect the population under study if the food/supplement supply differs for the new study population. Validation of a new questionnaire against a previous questionnaire targeting the exposure under study allows for comparability among similarly validated studies… Currently available assessment methods are often subject to random errors and systematic biases that can result in misleading and erroneous conclusions. Optimal methodologies as well as the nature and severity of potential errors and biases vary among food substances and, therefore, must be considered on a case-by-case basis. Methods used in the study under evaluation need to be evaluated for their potential for **risk** of bias (RoB) that includes both general and topic-specific criteria. Topic-specific RoB criteria are issues related specifically to the exposure of interest. For example, issues related to

Concept	Item	Sample guidance
		assessing sodium intake will be different from those of other nutrients or total energy… **Food substance intakes from self-reported diets** (e.g., food frequency questionnaire, food record, or 24-hour recall). Methods have different strengths and weaknesses (known errors and biases)… The frequency of evaluation of the exposure will often depend on the food substance, the dosage conditions and the outcome. **Food substance intakes from supplements or specific food products** (e.g., bars, liquid drinks, fortified foods). The study should report product composition details and the methods used to confirm the composition of the product (i.e., assay procedure), and to how adherence to the use of these products was monitored (e.g., supplement count, assessment of changes in nutritional status, etc.). **For dietary pattern or food intakes**, the study should include ranges or distributions of any pre-specified food substance or status exposures of interest. The study should also include exposure measurement characteristics and errors, of which some can be controlled for statistically (e.g., 'usual' intakes). This can be accomplished by intake methodologies that represent averages over a specified time period or by statistically adjusting for within-person-variabilities in 24-hour recalls. Some intake methodologies cannot adjust for usual intakes (e.g., if they are unknown or unmeasured; or subject to a systematic bias such as underreporting) and this can introduce bias. **Biomarkers of intake and exposure** are usually metabolites recovered from the blood or urine that objectively and accurately assess intake over a period of time. However, biomarker assays can vary in accuracy and reliability. If a biomarker has been accurately measured and appropriately qualified for its intended purpose, it is more reliable than exposure assessed by dietary intake estimation methods. Note that use of non-standardized methodologies could affect status measures, even when the method itself includes standards and controls. For example, 25-hydroxyvitamin D values differ depending on the assay platform and the laboratory performing the measure. A measure from a laboratory that has been certified for the method will be more reliable than one from an uncertified laboratory and standardized procedures are more reliable than non-standardized procedures. **Other important considerations for assessing exposure include the composition, the form, any potential**

Concept	Item	Sample guidance
		interactions with other exposure components, supplement brand name, and food types or recipes. The importance and weight of these issues (and other relevant characteristics of the exposure) must be determined *a priori* by the reviewer. Knowledge about the exposure form (e.g., naturally occurring vs. synthetic food substance) and intake conditions (e.g., taken with or without meals; bolus vs. multiple daily exposures; preparation practices for meal components) can be important especially when bioavailability and/or bioactivity are issues. Ideally, product composition would be confirmed by analysis because food/supplement label values and/or composition databases do not always accurately reflect actual composition. **If the exposure is not accurately measured or adequately characterized, confidence in the degree of exposure is reduced. A value on a product label could be acceptable but might not accurately reflect composition and should be interpreted with caution.** Energy intakes are another consideration. They are important for understanding the full effect of an addition or deletion of a food, or substitution of one food for another, and for identifying potential errors in dietary intake reporting (e.g., under- or- over reporting of dietary intake). **Rating:** Documentation of exposure should be sufficient to determine its validity. For example, intake/nutritional status based on a comparison to an accurate external method would have the highest rating. Self-reported intake or use of a biomarker assay procedure with accuracy or bias concerns that have been described/evaluated would have higher uncertainty. Self-reported intake with no description of accuracy or bias would have the lowest rating. In addition, because of large day-to-day variabilities in intakes by a given individual, higher rating should be given to results expressed as "usual" or long-term intakes. If adherence to the exposure is assessed by a supplement count then there is greater confidence in the measurement than if the assessment is achieved by recall. The tolerance for uncertainty will vary depending on the purpose of your evaluation and will be reflected in the study rating. The rationale for the rating system used, including a description of tolerance for uncertainty, should be defined *a priori*.
Baseline nutrient	2.2 The relevant exposure at	Accurate baseline intake/status data are necessary to ensure appropriate group assignments; to determine whether changes in intake occurred during the study period (which

Concept	Item	Sample guidance
intake or exposure	baseline is measured and taken into account.	could confound study results); and to compare results across studies in different populations. If supplement use is the exposure, baseline intakes can be used to ascertain the nutritional similarity of exposed and unexposed groups at baseline and estimate total intake. If expressed as distributions, baseline intakes of study groups can aid in the interpretation of results. For example, if a wide range of baseline statuses is present among groups of a supplement study, the overall effect of the exposure may be null; this could happen if nutrient-depleted participants experience benefit but replete participants experience no effect or even harm…
Intake adherence and maintenance	2.3 The baseline exposure differences between the groups have been maintained over the course of the study.	…All participants are exposed to diets that include food substances, foods or dietary patterns of interest. Changes in intake over the duration of the study for any or all of the groups could make it more difficult to detect differences in outcome due to the effect of the baseline exposures. Has the potential for this been monitored? If a change in intake has occurred, does the study have appropriate methods to deal with it?...
Intake/outcome interval	2.4 The interval between the exposure and outcome is of sufficient duration to observe an effect, if there is one.	In order to determine whether a 'true' relationship exists, the duration of the interval between the exposure and the outcome must be sufficient to allow the outcome to occur. False positives can occur in studies of inadequate duration. For example, transient decreases in LDL-cholesterol may be observed in short-term lipid studies but may not persist in long-term studies. Confounding factors such as seasonal changes can transiently affect status or intake. Even study enrolment can transiently affect intake or status in the short-term…

PICO = population, intervention/exposure, comparator, outcome.

Table 5. ***Rater agreement on individual sections and overall rating using NUQUEST.***

Study Design	No. of Studies	Degree of Agreement	Percentage in Agreement[1] - NUQUEST Assessment (%)				
			Selection[4]	Comparability[5]	Ascertainment[6]	Nutrition[7]	Overall[8]
Randomized controlled trial	15	Near agreement[2]	87	80	93	100	87
		Agreement[3]	40	47	93	80	60
	10	*Agreement without worksheet*	*40*	*40*	*80*	*100*	*40*
	5	*Agreement with worksheet*	*40*	*60*	*80*	*80*	*100*
Cohort	15	Near agreement	100	80	80	93	93
		Agreement	67	47	33	87	53
	10	*Agreement without worksheet*	*60*	*30*	*30*	*80*	*40*
	5	*Agreement with worksheet*	*80*	*80*	*40*	*100*	*80*
Case-control	15	Near agreement	93	87	93	100	100
		Agreement	60	47	40	60	60
	10	*Agreement without worksheet*	*70*	*20*	*40*	*40*	*40*
	5	*Agreement with worksheet*	*40*	*100*	*40*	*100*	*100*

1. Agreement/near agreement was determined section-by-section and overall. Percentage in agreement was calculated for each section/overall using number of studies with rater agreement or near agreement out of the total number of studies assessed.
2. Near agreement: raters' assessment of good, neutral or poor were either exactly the same or within one category of good, neutral or poor
3. Agreement: raters' assessment of good, neutral or poor were exactly the same.
4. Selection = Selection of participants (RCT)/Selection of cohorts (cohort) /Creation of study groups (case-control).
5. Comparability = Comparability of study groups (RCT, case-control)/Comparability of cohorts (cohort).
6. Ascertainment = Ascertainment of outcomes (RCT, cohort)/Exposure ascertainment (case-control).
7. Nutrition = Nutrition-specific (RCT, cohort and case-control).
8. Overall = NUQUEST Overall rating.

BOX 1

If Initiating a Review, Points to Consider When Developing the Worksheet[1]

- Is there a clearly stated key research question for the review?
    - In cases where the study's purpose differs from the key research question, does the study design provide information that can still be used to evaluate the key research question?
- Has a PICO been prepared before the identification of articles to assist in study selection/exclusion decisions?
    - Are the selection/exclusion criteria sufficiently broad to include a range of acceptable study methodologies that can subsequently be subjected to assessment with application of the NUQUEST criteria?
- Have the rating criteria to assist reviewers in using NUQUEST been prepared before initiation of the reviews?
    - Does the worksheet include criteria for evaluating a range of exposure methodologies that are specific for the food substance and for the analysis and expression of results?
    - Does the worksheet differentiate between key confounders and other potential confounders of lesser importance?

[1]NUQUEST, NUtrition QUality Evaluation Strengthening Tools; PICO, population, intervention/exposure, comparator, outcome.

BOX 2
For the Initiators of the Review, Points to Consider When Preparing Raters to Use the Worksheet in Study Assessments[1]

- Do raters understand the worksheet's purpose for the review, the PICO, and the rating criteria?
- If possible, do raters represent multidisciplinary expertise (e.g., epidemiology, nutrition)?
- If there is >1 rater, are differences adjudicated?
- Do raters recognize the types of situations that may require judgment?

For example:

- If detailed study design and methods information is primarily available in companion articles, have these articles been accessed?
- Where double blinding may not be possible (e.g., 1 of the groups receives an educational program or a whole food substitution), have the study investigators taken steps to reduce the potential biases (e.g., detection bias)?
- If randomization failed to some degree, have the study investigators taken steps to reduce potential biases by treating important variables as covariates in analyses?
- If comparability of groups is based on a broad range of potential confounders or limited to commonly used demographic factors, how well have study investigators dealt with the key critical confounders?
- If intention-to-treat analysis was not specifically mentioned, was there other information (e.g., analysis descriptions) to suggest that it was or was not the approach used?
- When evaluating nutrient exposures, have the complexities and nuances of exposure assessments for different nutrients, different purposes, different study designs, and different statistical analyses been taken into account?
- If nutrient exposure is an intermediate factor (e.g., when education is the intervention), is it still possible to evaluate the exposure–outcome relation of interest?
- Has the rater maintained focus on the evaluation of the scientific quality of reviewed studies and not been distracted by the potentially trivial nature of a study topic? Trivial or not, does the study provide information that can be used to evaluate the key research question?

[1]In many cases, ≥2 raters will perform the assessment. However, we recognize that in some cases, a single rater will conduct the review. In this latter case, adaptation of the "Points to Consider" may be appropriate.

PICO, population, intervention/exposure, comparator, outcome.

NUQUEST – NUtrition QUality Evaluation Strengthening Tool: Development of tools for the evaluation of risk of bias in nutrition studies

Shannon E. Kelly, Linda S. Greene-Finestone, Elizabeth A. Yetley, Karima Benkhedda, Stephen P.J. Brooks, George A. Wells, and Amanda J. MacFarlane

Online supplementary file 1: Search syntax for identifying generic study assessment instruments

"quality assessment tool" [tiab] OR "quality assessment tools" [tiab] OR "quality assessment instrument" [tiab] OR "quality assessment instruments" [tiab] OR "quality assessment checklists" [tiab] OR "quality assessment check list" [tiab] OR "quality assessment check lists" [tiab] OR "quality assessment checklist" [tiab] OR "quality assessment guidelines" [tiab] OR "quality assessment guideline" [tiab] OR "quality assessment scale" [tiab] OR "quality assessment scales" [tiab] OR “risk of bias” OR “internal validity”

OR

(meta analysis[MeSH Major Topic] OR review literature [mesh major topic] OR cohort studies[MeSH Major Topic] OR case control studies[MeSH Major Topic] OR randomized controlled trials/st[MeSH Major Topic] OR randomized controlled trials/mt[MeSH Major Topic]) AND quality [tiab] OR bias [tiab]

OR

(in process [filter] OR publisher[filter]) AND "quality assessment" [tiab] AND (tool [tiab] OR tools [tiab] OR instrument [tiab] OR instruments [tiab] OR checklists [tiab] OR checklist [tiab] OR "check list" [tiab] OR "check lists" [tiab] OR guidelines [tiab] OR guideline [tiab] OR scale [tiab] OR scales[tiab] OR tool[tiab])

OR

(evidence based medicine/cl [mesh major topic] OR evidence based medicine/st [mesh major topic]) AND quality [tiab]

OR

((rating [tiab] OR grade [tiab] OR grading [tiab] OR score [tiab] OR scoring [tiab] OR checklist [tiab] OR checklists [tiab] OR measure [tiab] OR measuring [tiab] OR assessing [tiab] OR assess [tiab]) AND ("level of evidence" [tiab] OR "levels of evidence" [tiab] OR "hierarchy of evidence" [tiab] OR "hierarchies of evidence" [tiab])) AND quality [tiab]

NUQUEST – NUtrition QUality Evaluation Strengthening Tool: Development of tools for the evaluation of risk of bias in nutrition studies

Shannon E. Kelly, Linda S. Greene-Finestone, Elizabeth A. Yetley, Karima Benkhedda, Stephen P.J. Brooks, George A. Wells, and Amanda J. MacFarlane

Online supplementary file 2: Studies included in the evaluation of NUQUEST

Adzersen K-H, Jess P, Freivogel KW, Gerhard I, Bastert G. Raw and cooked vegetables, fruits, selected micronutrients, and breast cancer risk: a case-control study in Germany. Nutrition and cancer 2003;46(2):131-7.

Alderman MH, Cohen H, Madhavan S. Dietary sodium intake and mortality: the National Health and Nutrition Examination Survey (NHANES I). The Lancet 1998;351(9105):781-5.

Baird IM, Shephard N, Merritt R, Hildick-Smith G. Repeated dose study of sucralose tolerance in human subjects. Food and Chemical Toxicology 2000;38:123-9.

Blackburn GL, Kanders BS, Lavin PT, Keller SD, Whatley J. The effect of aspartame as part of a multidisciplinary weight-control program on short-and long-term control of body weight. The American journal of clinical nutrition 1997;65(2):409-18.

Browne JC, Scott KM, Silvers KM. Fish consumption in pregnancy and omega-3 status after birth are not associated with postnatal depression. Journal of affective disorders 2006;90(2-3):131-9.

Burkievcz CJC, Skare TL, Malafaia O, Nassif PAN, Ribas CSG, Santos LRP. Vitamin D deficiency in patients with chronic venous ulcers. Revista do Colegio Brasileiro de Cirurgioes 2012;39:60-3.

Centonze S, Boeing H, Leoci C, Guerra V, Misciagna G. Dietary habits and colorectal cancer in a low-risk area. Results from a population-based case-control study in southern Italy. Nutrition and Cancer 1994;21(3):233-46. doi: 10.1080/01635589409514322.

Chien K-L, Hsu H-C, Chen P-C, Su T-C, Chang W-T, Chen M-F, Lee Y-T. Urinary sodium and potassium excretion and risk of hypertension in Chinese: report from a community-based cohort study in Taiwan. Journal of hypertension 2008;26(9):1750-6.

Chun YJ, Sohn S-K, Song HK, Lee SM, Youn YH, Lee S, Park H. Associations of colorectal cancer incidence with nutrient and food group intakes in Korean adults: a case-control study. Clinical nutrition research 2015;4(2):110.

Cohen HW, Hailpern SM, Fang J, Alderman MH. Sodium intake and mortality in the NHANES II follow-up study. The American journal of medicine 2006;119(3):275. e7-. e14.

Cook NR, Appel LJ, Whelton PK. Lower levels of sodium intake and reduced cardiovascular risk. Circulation 2014;129(9):981-9. *This study was included in both rounds of evaluation (with and without the worksheet. The two rounds of validation were performed by different raters therefore this paper is treated as an independent study in our assessments.

de Franciscis S, De Sarro G, Longo P, Buffone G, Molinari V, Stillitano DM, Gallelli L, Serra R. Hyperhomocysteinaemia and chronic venous ulcers. International wound journal 2015;12(1):22-6.

de Ruyter JC, Katan MB, Kuijper LD, Liem DG, Olthof MR. The effect of sugar-free versus sugar-sweetened beverages on satiety, liking and wanting: an 18 month randomized double-blind trial in children. PLoS One 2013;8(10):e78039.

Dubbert PM, Cushman WC, Meydrech EF, Rowland AK, Maury P. Effects of dietary instruction and sodium excretion feedback in hypertension clinic patients. Behavior therapy 1995;26(4):721-32.

Ellison RC, Capper AL, Stephenson WP, Goldberg RJ, Hosmer Jr DW, Humphrey KF, Ockene JK, Gamble WJ, Witschi JC, Stare FJ. Effects on blood pressure of a decrease in sodium use in institutional food preparation: the Exeter-Andover Project. Journal of clinical epidemiology 1989;42(3):201-8.

Graham U, McCance D, Young I, Mullan K. A randomised controlled trial evaluating the effect of potassium supplementation on vascular function and the renin–angiotensin–aldosterone system. Journal of human hypertension 2014;28(5):333-9.

Hamazaki K, Harauma A, Otaka Y, Moriguchi T, Inadera H. Serum n-3 polyunsaturated fatty acids and psychological distress in early pregnancy: Adjunct Study of Japan Environment and Children's Study. Translational psychiatry 2016;6(2):e737-e.

He J, Ogden LG, Vupputuri S, Bazzano LA, Loria C, Whelton PK. Dietary sodium intake and subsequent risk of cardiovascular disease in overweight adults. Jama 1999;282(21):2027-34.

Hwang JH, Chin HJ, Kim S, Kim DK, Kim S, Park JH, Shin SJ, Lee SH, Choi BS, Lim CS. Effects of intensive low-salt diet education on albuminuria among nondiabetic patients with hypertension treated with olmesartan: a single-blinded randomized, controlled trial. Clinical Journal of the American Society of Nephrology 2014;9(12):2059-69.

Kartal Durmazlar SP, Akgul A, Eskioglu F. Hyperhomocysteinemia in patients with stasis dermatitis and ulcer: a novel finding with important therapeutic implications. Journal of dermatological treatment 2009;20(6):336-9.

Kuzma JN, Cromer G, Hagman DK, Breymeyer KL, Roth CL, Foster-Schubert KE, Holte SE, Callahan HS, Weigle DS, Kratz M. No difference in ad libitum energy intake in healthy men and women consuming beverages sweetened with fructose, glucose, or high-fructose corn syrup: a randomized trial. The American journal of clinical nutrition 2015;102(6):1373-80.

Lamelas PM, Mente A, Diaz R, Orlandini A, Avezum A, Oliveira G, Lanas F, Seron P, Lopez-Jaramillo P, Camacho-Lopez P. Association of urinary sodium excretion with blood pressure and

cardiovascular clinical events in 17,033 Latin Americans. American Journal of Hypertension 2016;29(7):796-805.

Lee HP, Gourley L, Duffy SW, Esteve J, Lee J, Day NE. Colorectal cancer and diet in an Asian population—a case-control study among Singapore Chinese. International journal of cancer 1989;43(6):1007-16.

Legendre C, Debure C, Meaume S, Lok C, Golmard JL, Senet P. Impact of protein deficiency on venous ulcer healing. Journal of vascular surgery 2008;48(3):688-93.

Levi F, La Vecchia C, Gulie C, Negri E. Dietary factors and breast cancer risk in Vaud, Switzerland. Nutrition and Cancer 1993;19(3):327-35. doi: 10.1080/01635589309514263.

Maersk M, Belza A, Stødkilde-Jørgensen H, Ringgaard S, Chabanova E, Thomsen H, Pedersen SB, Astrup A, Richelsen B. Sucrose-sweetened beverages increase fat storage in the liver, muscle, and visceral fat depot: a 6-mo randomized intervention study. The American journal of clinical nutrition 2012;95(2):283-9.

Mizoue T, Kimura Y, Toyomura K, Nagano J, Kono S, Mibu R, Tanaka M, Kakeji Y, Maehara Y, Okamura T. Calcium, dairy foods, vitamin D, and colorectal cancer risk: the Fukuoka Colorectal Cancer Study. Cancer Epidemiology and Prevention Biomarkers 2008;17(10):2800-7.

Morikawa N, Yamasue K, Tochikubo O, Mizushima S. Effect of salt reduction intervention program using an electronic salt sensor and cellular phone on blood pressure among hypertensive workers. Clinical and experimental hypertension 2011;33(4):216-22.

Mourouti N, Kontogianni MD, Papavagelis C, Psaltopoulou T, Kapetanstrataki MG, Plytzanopoulou P, Vassilakou T, Malamos N, Linos A, Panagiotakos DB. Whole grain consumption and breast cancer: a case-control study in women. Journal of the American College of Nutrition 2016;35(2):143-9.

Nakano M, Eguchi K, Sato T, Onoguchi A, Hoshide S, Kario K. Effect of intensive salt-restriction education on clinic, home, and ambulatory blood pressure levels in treated hypertensive patients during a 3-month education period. The Journal of Clinical Hypertension 2016;18(5):385-92.

Nerbass FB, Pecoits-Filho R, McIntyre NJ, Shardlow A, McIntyre CW, Taal MW. Reduction in sodium intake is independently associated with improved blood pressure control in people with chronic kidney disease in primary care. British Journal of Nutrition 2015;114(6):936-42.

O'Donnell M, Mente A, Rangarajan S, McQueen MJ, Wang X, Liu L, Yan H, Lee SF, Mony P, Devanath A. Urinary sodium and potassium excretion, mortality, and cardiovascular events. New England Journal of Medicine 2014;371(7):612-23.

Ohta Y, Tsuchihashi T, Kiyohara K, Oniki H. High salt intake promotes a decline in renal function in hypertensive patients: a 10-year observational study. Hypertension Research 2013;36(2):172-6.

Rees A-M, Austin M-P, Owen C, Parker G. Omega-3 deficiency associated with perinatal depression: case control study. Psychiatry research 2009;166(2-3):254-9.

Reid M, Hammersley R, Duffy M, Ballantyne C. Effects on obese women of the sugar sucrose added to the diet over 28 d: a quasi-randomised, single-blind, controlled trial. British journal of nutrition 2014;111(3):563-70.

Serra R, Grande R, Butrico L, Buffone G, Caliò FG, Squillace A, Rizzo BA, Massara M, Spinelli F, Ferrarese AG. Effects of a new nutraceutical substance on clinical and molecular parameters in patients with chronic venous ulceration. International Wound Journal 2016;13(1):88-96.

Shiraishi M, Matsuzaki M, Yatsuki Y, Murayama R, Severinsson E, Haruna M. Associations of dietary intake and plasma concentrations of eicosapentaenoic and docosahexaenoic acid with prenatal depressive symptoms in J apan. Nursing & health sciences 2015;17(2):257-62.

Singer P, Cohen H, Alderman M. Assessing the associations of sodium intake with long-term all-cause and cardiovascular mortality in a hypertensive cohort. American journal of hypertension 2015;28(3):335-42.

Stolarz-Skrzypek K, Kuznetsova T, Thijs L, Tikhonoff V, Seidlerová J, Richart T, Jin Y, Olszanecka A, Malyutina S, Casiglia E. Fatal and nonfatal outcomes, incidence of hypertension, and blood pressure changes in relation to urinary sodium excretion. Jama 2011;305(17):1777-85.

Sun J, Jiang T, Qiu Z, Cen G, Cao J, Huang K, Pu Y, Liang H, Huang R, Chen S. Short-term and medium-term clinical outcomes of laparoscopic-assisted and open surgery for colorectal cancer: a single center retrospective case-control study. BMC gastroenterology 2011;11(1):1-12.

Trials of Hypertension Prevention Collaborative Research Group. Effects of weight loss and sodium reduction intervention on blood pressure and hypertension incidence in over-weight people with high normal blood pressure: the Trials of Hypertension Prevention, Phase II. Arch Intern Med 1997;157:657-67.

Tuomilehto J, Jousilahti P, Rastenyte D, Moltchanov V, Tanskanen A, Pietinen P, Nissinen A. Urinary sodium excretion and cardiovascular mortality in Finland: a prospective study. The Lancet 2001;357(9259):848-51.

Umesawa M, Yamagishi K, Noda H, Ikeda A, Sawachi S, Muraki I, Chei C-L, Cui R, Nagao M, Ohira T. The relationship between sodium concentrations in spot urine and blood pressure increases: a prospective study of Japanese general population: the Circulatory Risk in Communities Study (CIRCS). BMC cardiovascular disorders 2016;16(1):1-9.

Yang Q, Liu T, Kuklina EV, Flanders WD, Hong Y, Gillespie C, Chang M-H, Gwinn M, Dowling N, Khoury MJ. Sodium and potassium intake and mortality among US adults: prospective data from the Third National Health and Nutrition Examination Survey. Archives of internal medicine 2011;171(13):1183-91.

NUQUEST – NUtrition QUality Evaluation Strengthening Tool: Development of tools for the evaluation of risk of bias in nutrition studies

Shannon E. Kelly, Linda S. Greene-Finestone, Elizabeth A. Yetley, Karima Benkhedda, Stephen P.J. Brooks, George A. Wells, and Amanda J. MacFarlane

Online supplementary file 3: NUQUEST Randomized Controlled Trials Checklist for Nutrition Studies

NUQUEST: RANDOMIZED CONTROLLED TRIALS CHECKLIST FOR NUTRITION STUDIES

We recommend reviewing the Guidance and the Worksheet provided before starting.

NUTRITION STUDY[i] INFORMATION
CITATION:
REVIEWER NAME:
REVIEW DATE:
RESEARCH QUESTION:
REVIEWER NOTES:
THE STUDY UNDER EVALUATION ADDRESSES AN APPROPRIATE AND CLEARLY FOCUSED QUESTION: ☐ YES ☐ NO

		YES	PROBABLY YES	PROBABLY NO	NO	NOT APPLICABLE
SELECTION OF PARTICIPANTS						
1.1	The assignment of participants to intervention[ii] groups is randomized.	☐	☐	☐	☐	-
1.2	An adequate concealment method is used.	☐	☐	☐	☐	-
1.3	The intervention and control groups are similar at the start of the trial.	☐	☐	☐	☐	-
Taking the ratings of 1.1, 1.2, and 1.3 into consideration, provide an overall rating for Selection of Participants using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Selection of Participants. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Selection of Participants. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Selection of Participants. High RoB.			☐ Good (+) ☐ Neutral (0) ☐ Poor (-)			
COMPARABILITY OF STUDY GROUPS						
1.4	Participants and investigators are kept 'blind' about intervention allocation.	☐	☐	☐	☐	-
1.5	The only difference between groups is the intervention under investigation.	☐	☐	☐	☐	-
1.6	The percentage of the individuals or clusters recruited into each arm of the study that dropped out before the study was completed is reasonable.	☐	☐	☐	☐	-
1.7	All the participants are analyzed in the groups to which they were randomly allocated (often referred to as intention to treat analysis).	☐	☐	☐	☐	☐
1.8	Where the study is carried out at more than one site, results are comparable for all sites.	☐	☐	☐	☐	☐
Taking the ratings of 1.4, 1.5, 1.6, 1.7, and 1.8 into consideration, provide an overall rating for Comparability of Groups using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Comparability of Study Groups. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Comparability of Study Groups. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Comparability of Study Groups. High RoB.		**OVERALL RATING FOR COMPARABILITY** ☐ Good (+) ☐ Neutral (0) ☐ Poor (-)			

NUQUEST: RANDOMIZED CONTROLLED TRIALS CHECKLIST FOR NUTRITION STUDIES

We recommend reviewing the Guidance and the Worksheet provided before starting.

		YES	PROBABLY YES	PROBABLY NO	NO	NOT APPLICABLE
ASCERTAINMENT OF OUTCOMES						
1.9	All relevant outcomes are measured in a valid and reliable way.	☐	☐	☐	☐	-
Taking the rating of 1.9 into consideration, provide an overall rating for Ascertainment of Outcomes using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Ascertainment of Outcomes. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Ascertainment of Outcomes. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Ascertainment of Outcomes. High RoB.		**OVERALL RATING FOR ASCERTAINMENT** ☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**				
NUTRITION-SPECIFIC						
2.1	The frequency and quantity of the intervention under study are accurately and reliably measured.	☐	☐	☐	☐	-
2.2	The relevant baseline exposure is measured and taken into account.	☐	☐	☐	☐	-
2.3	The baseline exposures of all groups have been maintained over the course of the study.	☐	☐	☐	☐	-
2.4	Adherence to the intervention and control intakes is monitored in a reliable and accurate manner.	☐	☐	☐	☐	-
2.5	The interval between the intervention and outcome is of sufficient duration to observe an effect, if there is one.	☐	☐	☐	☐	-
Taking the ratings of 2.1, 2.2, 2.3, 2.4 and 2.5 into consideration, provide an overall rating for Nutrition-Specific using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Nutrition-specific issues. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Nutrition-specific issues. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Nutrition-specific issues. High RoB.		**OVERALL RATING FOR NUTRITION** ☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**				

OVERALL STUDY RATING

ASSESSMENT SUMMARY: Record the overall ratings from each section above in the boxes below:

SELECTION OF PARTICIPANTS:	☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**	COMPARABILITY OF STUDY GROUPS:	☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**	ASCERTAINMENT OF OUTCOMES:	☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**	NUTRITION-SPECIFIC:	☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**

Taking the overall ratings for each section into consideration, provide an overall study rating:

Good (+): Almost all criteria met. Little or no concern over study design. Low RoB.

Neutral (0): Most criteria met. Some flaws in the study with an associated concern over study design. Moderate RoB

Poor (-): Most or all criteria not met, or significant flaws related to key aspects of the study design. High RoB.

OVERALL STUDY RATING

☐ **GOOD (+)**
☐ **NEUTRAL (0)**
☐ **POOR (-)**

NUQUEST: RANDOMIZED CONTROLLED TRIALS CHECKLIST FOR NUTRITION STUDIES

We recommend reviewing the Guidance and the Worksheet provided before starting.

ASSESSMENT OF APPLICABILITY:
How applicable is the study under evaluation to your research question?

This section is optional.
Although it is not necessarily related to RoB or study quality, it may help interpret the study results in the context of the research question.

Taking into account the study design and methodology used, are the results directly applicable to your research question?	YES ☐	PROBABLY YES ☐	PROBABLY NO ☐	NO ☐
Taking into account the study design and methodology used, do you have confidence that there is clear evidence of an association between the intervention and the outcome?	YES ☐	PROBABLY YES ☐	PROBABLY NO ☐	NO ☐
Are the results of this study directly applicable to the population of interest?	YES ☐	PROBABLY YES ☐	PROBABLY NO ☐	NO ☐

NOTES:

[i]For this checklist, the term "nutrition study" is used broadly to apply to a study in which the focus is on **intake of a food substance, food, dietary pattern or biomarker(s) of nutritional status.**

Food substances are defined as essential or conditionally essential nutrients, energy nutrients, or other naturally occurring bioactive food components. Food substance intakes can be derived from the intake of food sources (naturally occurring foods, fortified/supplemented food) and dietary supplements. For some food substances, exposure may come from other sources (e.g., from industrial minerals; vitamin D from UV-B rays), which should be taken into account. Studies that focus on **foods** will generally require documentation of food composition and factors that alter composition, such as processing and fortification/supplementation. Studies of **dietary patterns** often also require documentation of food composition and/or prevalence of the use of specific types of foods.

[ii]For this checklist, in the context of a nutrition study, **'intervention' often refers to intake or status.** The intervention group consumes the intervention that is being tested and the control group consumes a placebo, a different intake level, or a 'usual' intake. For example, an intervention group could consume a vitamin D supplement of 600 IU/day and the control group a placebo or a lower dose vitamin D supplement. If the trial is a dietary pattern study, the intervention group could consume a Mediterranean diet and the control group could consume a 'usual' dietary pattern or a 'Western-style' diet.

In some randomized controlled trial nutrition studies, the intervention is not an intake, but the outcome is an intake or nutritional status. For example, the intervention group could receive an intensive educational program and the control group the usual program, with the outcome being sodium intake. The checklist is flexible enough to be used for these types of studies. The guidance addresses nutritional assessment whether intake is an exposure, an outcome or both.

NUQUEST: CHECKLIST GUIDANCE FOR RANDOMIZED CONTROLLED NUTRITION TRIALS

1. BEFORE YOU START

Prior to completing the NUQUEST randomized controlled trial checklist, ensure that:

(1). The study under evaluation is a randomized controlled trial.
If in doubt, make sure you have the correct checklist for your study design.

(2). The study is relevant to the key research question of the review.
Inclusion or exclusion of individual studies into an evaluation will depend on the key question(s) and the *a priori* criteria, including the **PICO** (Participants/Population, Intervention/Exposure, Comparator, and Outcome(s)) components that are guiding the evaluation. Decisions should be documented and in place prior to beginning the checklist. Documenting definitions of exclusion/inclusion criteria prior to initiation of the evaluation is important for establishing the applicability of the evaluation. The worksheet can be used to document these criteria.

(3). The worksheet can be used to identify important study criteria and evaluate the risk of bias associated with them. The worksheet can be referred to, as appropriate, to facilitate the study assessment.
One purpose of an assessment tool is to identify and take into account the potential for risk of bias (RoB). The worksheet can be used to identify the importance of each component of the study (intervention, confounding factors, co-intervention factors) for the study assessment.

(4). General risk of bias sources are considered.
While the type of bias is not specified within the NUQUEST checklist for each item, items in each section identify potential issues associated with facets of a study that are related to different sources of bias.

Below are examples of important general bias domains that may be commonly observed in nutrition randomized controlled trials. Note that this is not an exhaustive list of potential bias domains and others may apply to your specific study assessment (see Resources following guidance).

Performance bias	Systematic differences between groups in the care provided or exposure to factors other than the interventions of interest.
Attrition bias	Systematic differences between groups in withdrawals from the study.
Detection bias	Systematic differences between groups in how exposure/status and outcomes are determined.
Selection bias	Systematic differences between groups on baseline characteristics.
Dietary exposure assessment bias	Error associated with the use of methodologies for assessing dietary intakes. A subcategory for self-reporting methodologies is **recall bias**, which refers to systematic error due to differences in completeness or accuracy of recall. Self-reported dietary intakes are at risk of this bias.
Misclassification bias	Systematic error due to inaccurate measurements or classifications of participants' exposure or outcome; error may be related to the risk of outcome. If the error is unrelated to the risk of outcome, the effect is usually biased to the null.

2. INSTRUCTIONS FOR COMPLETING THE NUQUEST CHECKLIST

This NUQUEST checklist is intended to be applied to individual epidemiological studies in nutrition with a randomized controlled trial design. Individual items may be answered using:

Yes	Probably Yes	Probably No	No	Not Applicable[a]
Based on something factual or known from evidence in the study publication.	Based on something the assessor judges likely to have occurred in the study.	Based on something the assessor judges likely to have not occurred in the study.	Based on something factual or known from evidence in the study publication.	This should be determined and recorded *a priori*. It would not be considered in the overall assessment.

a – the NOT APPLICABLE rating is only relevant for certain items in the checklist.

All items within a section should be rated prior to assessing the overall rating for that section (see guidance by individual section below).

All of the overall section ratings should be completed prior to completing the Overall Study Rating.

The Assessment of Applicability is optional.

3. GUIDANCE FOR COMPLETING THE NUQUEST RANDOMIZED CONTROLLED TRIALS CHECKLIST

NUTRITION STUDY INFORMATION

The study under evaluation addresses an appropriate and clearly focused question.

Unless a clear and well-defined research question is specified in the study report, it will be difficult to assess how well it has met its objectives and how relevant it is to the question you are trying to answer. Implementation of the pre-study PICO and other inclusion/exclusion criteria will often result in a "Yes" answer for this item.

SELECTION OF PARTICIPANTS

1.1	**The assignment of participants to intervention or control groups is randomized.**
	Random allocation of participants by a valid randomization method to receive one or other of the interventions under investigation, or to receive either intervention or placebo, is fundamental to this type of study.
1.2	**An adequate concealment method is used.**
	Allocation concealment refers to the process used to ensure that researchers are unaware which group participants are being allocated to at the time they enter the study. Research has shown that where allocation concealment is inadequate, investigators can significantly overestimate the effect of interventions (e.g., by up to 40%).
1.3	**The intervention and control groups are similar at the start of the trial.**
	Participants selected for inclusion in a trial must be as similar as possible. The study should report any significant differences in the composition of the study groups including factors such as gender, age, social background, ethnic origin, stage of disease (if appropriate), clinical biomarkers, and co-morbid conditions. Baseline intakes and nutritional status should also be described (see 2.2). These factors may be covered by inclusion and exclusion criteria. **Failure to address this item could lead to the study being downgraded.** Reviewers should identify *a priori* the most important factors for evaluation of the study. If a baseline difference exists, the variable should be considered as a covariate in any of the planned analyses. **If this has not been done, then the study could be downgraded.** **If several relevant variables at baseline are different, then a randomization failure may be an issue and the study could be downgraded.** Relevant variables should be identified a *priori* and used as a guide for your assessment of baseline differences and/or the likelihood of randomization failure.

COMPARABILITY OF STUDY GROUPS

1.4	**Participants and investigators are kept 'blind' about intervention allocation.**
	Blinding refers to the process whereby participants and investigators are kept unaware of intervention allocation. Blinding is an important tool in reducing the risk of bias. Single blinding refers to the case when either participants or investigators, but not both, are aware of the intervention allocation. In double blind studies, neither the investigator nor the participant knows the intervention allocation. The higher the level of blinding, the lower the risk of bias in the study.

	When blinding is either not possible or is not done, bias may occur. Note, however, that the possible levels of blinding are dependent on the study intervention. In nutrition studies, double blinding can be achieved when the intervention is a supplement or liquid diet. However, participant and investigator blinding may not be possible when foods and/or dietary patterns are the intervention; it may be similarly difficult when a behavioral or educational intervention is employed. In these cases, those conducting the outcome assessment (see 1.9) and/or data analysis can be blinded to the intervention allocation. It is important to note whether the study under review dealt with blinding challenges to mitigate potential bias and whether the approach was likely to be effective. Failure of the study to do so requires caution in interpretation of study results and should be considered when determining the rating.
1.5	**The only difference between groups is the intervention under investigation.**
	If some participants received intervention beyond the original protocol and/or changed their behavior because of knowledge of study goals (e.g., started taking calcium supplements if enrolled in a calcium and fracture study), even if of a minor nature or consisting of advice and counselling rather than a physical intervention, this is a potential confounding factor that could invalidate the results. **If there is a concern, the study could be downgraded.**
1.6	**The percentage of the individuals or clusters recruited into each arm of the study that dropped out before the study was completed is reasonable.**
	The number of participants that drop out of a study should give concern if the number is high. Conventionally, a 20% dropout rate is regarded as acceptable, but this may vary. Determining the acceptable dropout rate may vary depending on the intervention, the outcome or the population. It should be noted why participants dropped out, how many dropped out, and whether the dropout rate differed among study groups. The dropout rate may be expected to be higher in studies conducted over a long period of time**. A higher dropout rate could lead to downgrading of a study depending on the purpose of the evaluation.**
1.7	**All the participants are analyzed in the groups to which they were randomly allocated (often referred to as intention to treat analysis).**
	In practice, it is rarely the case that all participants allocated to the intervention group receive the intervention throughout the trial, or that all those in the comparison group do not. Participants may refuse treatment, or contra-indications may arise that lead them to be switched to another group. If the comparability of groups through randomization is to be maintained, participant outcomes must be analyzed according to the group to which they were originally allocated irrespective of the intervention they actually received; this is known as intention to treat analysis**. If the analysis was not performed on an intention to treat basis, the study could be downgraded;** if it is used, it should be evaluated as if it were a non-randomized cohort study.
1.8	**Where the study is carried out at more than one site, results are comparable for all sites.**
	In multi-site studies, confidence in the results should be increased if it can be shown that similar results were obtained at the different participating centers. **Conversely, confidence in the results could be downgraded and noted if there appears to be a site/outcome interaction.**
ASCERTAINMENT OF OUTCOMES	
1.9	**All relevant outcomes are measured in a valid and reliable way.**
	The outcome measures should be clearly stated in the study and be consistent with the outcome definitions identified in the PICO documentation**. If the outcome measures are not clearly stated, the study could be downgraded.** Where outcome measures require any degree of subjectivity, evidence should be provided showing

	that the measures are reliable and have been validated prior to the study. Appropriate steps should be taken to reduce the risk of outcome detection bias e.g. by having an independent assessment performed by assessors who are blinded (masked) to the intervention. **If the outcome measures are not measured in a valid or reliable way, the study could be downgraded.** Whether and how secondary outcomes are to be used, need to be determined in conjunction with your PICO documentation. If the outcome is an intake or nutritional status, refer to the guidance in section 2. Nutrition-Specific.
NUTRITION-SPECIFIC	
2.1	**The frequency and quantity of the intervention exposure under study are accurately[1] and reliably[2] measured.**
	The accuracy and reliability of the exposure assessment, including an intervention, is of paramount importance for all nutrition-related studies. In general, PICO criteria need not be overly restrictive, can include relevant, and commonly accepted exposure assessment methods**.** A study should indicate how the degree of exposure or presence of prognostic factors or markers were assessed. The methods must be sufficient to accurately quantify the exposure under investigation. **The study should report the methods used to assess exposures or status and the validity of such methods.** Validity refers to the degree to which the exposure assessment methodology accurately measures the aspect of exposure it was intended to measure. Exposure assessment methods are validated in different ways. Comparing an exposure method to a "gold standard' or reference method thought to capture all the food and supplement sources containing the exposure of interest provides the best validation approach. Comparing a self-reported method to another self-reported method requires caution as the two methods may be affected by the same types of errors and biases. An exposure assessment method or tool could be modified to reflect the population under study if the food/supplement supply differs for the new study population. Validation of a new questionnaire against a previous questionnaire targeting the exposure under study allows for comparability among similarly validated studies. **If the exposure measures are not clearly stated, the study could be downgraded.** Currently available assessment methods are often subject to random errors and systematic biases that can result in misleading and erroneous conclusions. Optimal methodologies, as well as the nature and severity of potential errors and biases, vary among food substances and, therefore, must be considered on a case-by-case basis. Methods used in the study under evaluation need to be evaluated for their potential for **risk of bias (RoB)** that includes both general and topic-specific criteria. Topic-specific criteria are issues related specifically to the exposure of interest. For example, issues related to assessing sodium intake will be different from those of other nutrients or total energy. **Topic-specific RoB criteria** assist in rating the exposure methodologies included in the review as higher to lower RoB, which need to be considered when rating the nutrition-specific checklist items described below. **The worksheet can be used to define and evaluate topic-specific criteria.** **Food substance intakes from self-reported diets** (e.g., food frequency questionnaire, food record, or 24-hour recall). Methods have different strengths and weaknesses (known errors and biases). The best method depends on the research question[1] and generally needs to be determined as part of the *a priori* PICO documentation process. The frequency of evaluation of the exposure will often depend on the food substance, the dosage conditions and the outcome.

[1] Accuracy: The extent to which a measured value is close to that of the true value. Low accuracy can be a result of bias or systematic error in measurement (https://www.nutritools.org/glossary)

[2] Reliability: The degree of stability of a measurement repeated under the same conditions (https://www.nutritools.org/glossary)

	Food substance intakes from supplements or specific food products (e.g., bars, liquid drinks, fortified foods). The study should report product composition details and the methods used to confirm the composition of the product (i.e., assay procedure), and how adherence to the use of these products was monitored (e.g., supplement count, assessment of changes in nutritional status, etc.). **For dietary pattern or food intakes**, the study should include ranges or distributions of any pre-specified food substance or status exposures of interest. The study should also include exposure measurement characteristics and errors, of which some can be controlled for statistically (e.g., 'usual' intakes). This can be accomplished by intake methodologies that represent averages over a specified time period or by statistically adjusting for within-person-variabilities in 24-hour recalls. Some intake methodologies cannot adjust for usual intakes (e.g., if they are unknown or unmeasured; or subject to a systematic bias such as underreporting) and this can introduce bias. **Biomarkers of intake and exposure** are usually metabolites recovered from the blood or urine that objectively and accurately assess intake over a period of time. However, biomarker assays can vary in accuracy and reliability. If a biomarker has been accurately measured and appropriately qualified for its intended purpose, it is more reliable than exposure assessed by dietary intake estimation methods. Note that use of non-standardized methodologies could affect status measures, even when the method itself includes standards and controls. For example, 25-hydroxyvitamin D values differ depending on the assay platform and the laboratory performing the measure. A measure from a laboratory that has been certified for the method will be more reliable than one from an uncertified laboratory and standardized procedures are more reliable than non-standardized procedures. **Other important considerations for assessing exposure include the composition, the form, any potential interactions with other exposure components, supplement brand name, and food types or recipes. The importance and weight of these issues (and other relevant characteristics of the exposure) must be determined *a priori* by the reviewer.** Knowledge about the exposure form (e.g., naturally occurring vs. synthetic food substance) and intake conditions (e.g., taken with or without meals; bolus vs. multiple daily exposures; preparation practices for meal components) can be important especially when bioavailability and/or bioactivity are issues. Ideally, product composition would be confirmed by analysis because food/supplement label values and/or composition databases do not always accurately reflect actual composition. **If the exposure is not accurately measured or adequately characterized, confidence in the degree of exposure is reduced. A value on a product label could be acceptable but might not accurately reflect composition and should be interpreted with caution.** **Energy intakes** are another consideration. They are important for understanding the full effect of an addition or deletion of a food, or substitution of one food for another, and for identifying potential errors in dietary intake reporting (e.g., under- or- over reporting of dietary intake). **Rating:** Documentation of the intervention exposure should be sufficient to determine its validity. For example, intake/nutritional status based on a comparison to an accurate external method would have the highest rating. Self-reported intake or use of a biomarker assay procedure with accuracy or bias concerns that have been described/evaluated would have higher uncertainty. Self-reported intake with no description of accuracy or bias would have the lowest rating. In addition, because of large day-to-day variabilities in intakes by a given individual, higher rating should be given to results expressed as "usual" or long-term intakes. If adherence to the exposure is assessed by a supplement count then there is greater confidence in the measurement than if the assessment is achieved by recall. The tolerance for uncertainty will vary depending on the purpose of your evaluation and will be

	reflected in the study rating. The rationale for the rating system used, including a description of tolerance for uncertainty, should be defined a priori.
2.2	**The relevant baseline exposure is measured and taken into account.**
	Exposure to the intervention under investigation, from all sources, must be taken into account in the methods, results and/or interpretation. Accurate baseline intake level or level of biomarker of intake (or status) are necessary to ensure ascertainment of the nutritional similarity of intervention and control comparator groups at baseline (see 1.3), estimation of total intake (e.g., baseline intake plus intervention intake) and comparison of results across studies in different populations. Baseline intake or status can affect responsiveness to dietary intervention. For example, if a wide range of baseline statuses is present in a study group, the overall effect of the intervention may be null; this could happen if nutrient-depleted participants experience benefit, but replete participants experience no effect or even harm. Section 2.1 offers guidance on how to assess the RoB of the method to monitor intake/status or supplement consumption.
2.3	**The baseline exposures of all groups have been maintained over the course of the study.**
	All participants have baseline intakes of food substances, foods or dietary patterns of interest. Changes in the baseline intake over the duration of the study for any or all of the groups could make it more difficult to detect differences in outcome due to the effect of the intervention. Has the potential for this been monitored? If a change in baseline intake has occurred, does the study have appropriate methods to deal with it? **If it is possible/likely that this has occurred, and has not been addressed, then the study could be downgraded.** A higher rating would be given to studies where there are multiple measures of intake over time and especially where this is confirmed/assessed with biomarker data. A higher rating would also be given if baseline intake changes over time have been addressed in the analysis and/or interpretation. Section 2.1 offers guidance on how to assess the RoB of the method to monitor intake/status or supplement consumption.
2.4	**Adherence to the intervention and control intakes is monitored in a reliable and accurate manner.**
	The study should report the method used to assess/monitor adherence to the intervention and control intake. The method and validation characteristics should be pre-specified. Methods may include a dietary assessment (e.g., food frequency questionnaire, food record, 24-hour recall, or supplement pill counts) as well as changes in biomarkers of intake or nutritional status. **Failure to document the method used to monitor adherence to the intervention and control intake could result in the study being downgraded.** The method and validation characteristics are context specific and therefore the reviewer should pre-specify what is acceptable. Section 2.1 offers guidance on how to assess the RoB of the method to monitor intake/status or supplement consumption.
2.5	**The interval between the intervention and outcome is of sufficient duration to observe an effect, if there is one.**
	In order to determine whether a 'true' relationship exists, the duration of the interval between the intervention and the outcome must be sufficient for the outcome to occur; the minimum acceptable duration for the outcome of interest should be pre-specified by the reviewer. False positives can occur in studies of inadequate duration e.g., transient decreases that do not persist in long-term studies (e.g., LDL-cholesterol on short-term lipid studies that

	returns to normal over time). Confounding factors such as seasonal changes can transiently affect status or intake. Even study enrolment can transiently affect intake or status in the short-term. **Identify a minimum follow-up time for each outcome in the PICO documentation step. If this minimum is not met, the study could be downgraded.**

4. GUIDANCE FOR COMPLETING THE OVERALL STUDY RATING

Taking the overall ratings for the selection of participants, comparability of study groups, ascertainment of outcomes and nutrition-specific aspects into consideration, the rating of the overall risk of bias for the study could consider the following as a guide:

- All criteria may not have equal importance.
- The importance of a particular criterion may vary depending on your research question (PICO). In this regard, it may be helpful to consider the criteria in 3 parts, which could be differentially weighted:
  - (a) creation of the study groups (Items 1.1 to 1.4, and 2.2);
  - (b) comparability of the study groups (Items 1.5 to 1.8, 2.1, 2.3, and 2.4) ; and
  - (c) outcome ascertainment (Items 1.9 and 2.5).

For example, if the randomization scheme is not clear or concealment is difficult, selection bias could result and you may want to put more weight on 'creation of study groups'. If assessment of exposure is prone to error including bias or if adherence to the intervention is problematic, then more emphasis may be placed on 'comparability of study groups'. If a study involves an assessment of outcome that is prone to bias, you may want to put more weight on criteria related to 'outcome ascertainment'.

RESOURCES

- Chung M, Balk EM, Ip S, Raman G, Yu WW, Trikalinos TA, Lichtenstein AH, Yetley EA, Lau J. *Reporting of systematic reviews of micronutrients and health: a critical appraisal.* Am J Clin Nutr. 2009 Apr;89(4):1099-113. doi: https://doi.org/10.3945/ajcn.2008.2 6821.
- Delgado-Rodríguez M, Llorca J. *Bias.* Journal of Epidemiology & Community Health 2004;58:635-641.
- Institute of Medicine (IOM) Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease Board on Health Care Services Board on Health Sciences Policy Food and Nutrition Board, *Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease*, C.M. Micheel and J.R. Ball, Editors. Washington (DC): National Academies Press; 2010. Available from: https://iom.nationalacademies.org/Reports/2010/Evaluation-of-Biomarkers-and-Surrogate-Endpoints-in-Chronic-Disease.aspx
- National Academies of Sciences, Engineering, and Medicine. *Guiding principles for developing Dietary Reference Intakes based on chronic disease.* Washington, DC: The National Academies Press; 2017. doi: https://doi.org/10.17226/2482.
- National Academies of Sciences, Engineering, and Medicine. Methodological Considerations. In: *Dietary Reference Intakes for sodium and potassium.* Washington, DC: The National Academies Press, 2019:61-98. doi: https://doi.org/10.17226/25353.
- Dietary Assessment Primer, Summary Tables: Recommendations on Potential Approaches to Dietary Assessment for Different Research Objectives Requiring Group-level Estimates. National Institutes of Health, National Cancer Institute. Internet: https://dietassessmentprimer.cancer.gov/approach/table.html (accessed September 20, 2020).
- Scottish Intercollegiate Guidelines Network *(SIGN). A guideline developer's handbook.* Edinburgh: SIGN; 2019. (SIGN publication no. 50). [November 2019]. Available from URL: http://www.sign.ac.uk.

NUQUEST – NUtrition QUality Evaluation Strengthening Tool: Development of tools for the evaluation of risk of bias in nutrition studies

Shannon E. Kelly, Linda S. Greene-Finestone, Elizabeth A. Yetley, Karima Benkhedda, Stephen P.J. Brooks, George A. Wells, and Amanda J. MacFarlane

Online supplementary file 4: NUQUEST Cohort Checklist for Nutrition Studies

NUQUEST: COHORT CHECKLIST FOR NUTRITION STUDIES

We recommend reviewing the Guidance and the Worksheet provided before starting.

NUTRITION STUDY[i] INFORMATION
CITATION:
REVIEWER NAME:
REVIEW DATE:
RESEARCH QUESTION:
REVIEWER NOTES:
THE STUDY UNDER EVALUATION ADDRESSES AN APPROPRIATE AND CLEARLY FOCUSED QUESTION: ☐ YES ☐ NO

		YES	PROBABLY YES	PROBABLY NO	NO	NOT APPLICABLE
SELECTION OF COHORTS						
1.1	The groups being studied are selected from source populations that are comparable in all respects other than the exposure[ii] under investigation.	☐	☐	☐	☐	-
1.2	The study indicates how many of the people asked to take part did so in each of the groups being studied.	☐	☐	☐	☐	-
1.3	The likelihood that some eligible subjects might have the outcome at the time of enrolment is assessed and taken into account in the analysis.	☐	☐	☐	☐	☐
1.4	The exposure assessment method is adequate to differentiate exposure among study groups.	☐	☐	☐	☐	-
Taking the ratings of 1.1, 1.2, 1.3, and 1.4 into consideration, provide an overall rating for Selection of Cohorts using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Selection of Cohorts. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Selection of Cohorts. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Selection of Cohorts. High RoB.		**OVERALL RATING FOR SELECTION** ☐ Good (+) ☐ Neutral (0) ☐ Poor (-)				
COMPARABILITY OF COHORTS						
1.5	The percentage of individuals or clusters recruited into each arm of the study that dropped out before the study was completed is reasonable.	☐	☐	☐	☐	-
1.6	Comparison is made between full participants and those lost to follow up, by exposure.	☐	☐	☐	☐	-
1.7	Exposure level or prognostic factor is assessed more than once.	☐	☐	☐	☐	☐
1.8	The main potential confounders are identified and taken into account in the design and/or analysis.	☐	☐	☐	☐	-
Taking the ratings of 1.5, 1.6, 1.7 and 1.8 into consideration, provide an overall rating for Comparability of Cohorts using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Comparability of Cohorts. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Comparability of Cohorts. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Comparability of Cohorts. High RoB.		**OVERALL RATING FOR COMPARABILITY** ☐ Good (+) ☐ Neutral (0) ☐ Poor (-)				

NUQUEST: COHORT CHECKLIST FOR NUTRITION STUDIES

We recommend reviewing the Guidance and the Worksheet provided before starting.

		YES	PROBABLY YES	PROBABLY NO	NO	NOT APPLICABLE
ASCERTAINMENT OF OUTCOMES						
1.9	The outcomes are clearly defined.	☐	☐	☐	☐	-
1.10	The assessment of outcome is made blind to exposure. If the study is retrospective, this may not be applicable.	☐	☐	☐	☐	☐
1.11	Where blinding was not possible, there is some recognition that knowledge of exposure could have influenced the assessment of outcome.	☐	☐	☐	☐	☐
1.12	Evidence from other sources is used to demonstrate that the method of outcome assessment is valid and reliable.	☐	☐	☐	☐	☐
Taking the rating of 1.9, 1.10, 1.11, and 1.12 into consideration, provide an overall rating for Ascertainment of Outcomes using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Ascertainment of Outcomes. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Ascertainment of Outcomes. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Ascertainment of Outcomes. High RoB.		**OVERALL RATING FOR ASCERTAINMENT** ☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**			
NUTRITION-SPECIFIC						
2.1	The frequency and quantity of the exposure under study are accurately and reliably measured.	☐	☐	☐	☐	-
2.2	The relevant exposure at baseline is measured and taken into account.	☐	☐	☐	☐	-
2.3	The baseline exposure differences between the groups have been maintained over the course of the study.	☐	☐	☐	☐	-
2.4	The interval between the exposure and outcome is of sufficient duration to observe an effect, if there is one.	☐	☐	☐	☐	-
Taking the ratings of 2.1, 2.2, 2,.3 and 2.4 into consideration, provide an overall rating for Nutrition-Specific using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Nutrition-specific issues. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Nutrition-specific issues. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Nutrition-specific issues. High RoB.		**OVERALL RATING FOR NUTRITION** ☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**			

NUQUEST: COHORT CHECKLIST FOR NUTRITION STUDIES

We recommend reviewing the Guidance and the Worksheet provided before starting.

OVERALL STUDY RATING

ASSESSMENT SUMMARY: Record the overall ratings from each sections above in the boxes below:

CREATION OF COHORTS:	☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**	COMPARABILITY OF COHORTS:	☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**	ASCERTAINMENT OF OUTCOMES:	☐ **Good (+)** ☐ **Neutral(0)** ☐ **Poor (-)**	NUTRITION-SPECIFIC:	☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**

Taking the overall ratings for each section into consideration, provide an overall study rating:

Good (+): Almost all criteria met. Little or no concern over study design. Low RoB.

Neutral (0): Most criteria met. Some flaws in the study with an associated concern over study design. Moderate RoB

Poor (-): Most or all criteria not met, or significant flaws related to key aspects of the study design. High RoB.

OVERALL STUDY RATING

☐ **GOOD (+)**

☐ **NEUTRAL (0)**

☐ **POOR (-)**

NUQUEST: COHORT CHECKLIST FOR NUTRITION STUDIES

We recommend reviewing the Guidance and the Worksheet provided before starting.

ASSESSMENT OF APPLICABILITY: **How applicable is the study under evaluation to your research question?**				
This section is optional. Although it is not necessarily related to RoB or study quality, it may help interpret the study results in the context of the research question.				
Taking into account the study design and methodology used, are the results directly applicable to your research question?	YES ☐	PROBABLY YES ☐	PROBABLY NO ☐	NO ☐
Taking into account the study design and methodology used, do you have confidence that there is clear evidence of an association between the intervention and the outcome?	YES ☐	PROBABLY YES ☐	PROBABLY NO ☐	NO ☐
Are the results of this study directly applicable to the population of interest?	YES ☐	PROBABLY YES ☐	PROBABLY NO ☐	NO ☐

NOTES:

[i]For this checklist, the term 'nutrition study' is used broadly to apply to studies in which the focus is on **intake of a food substance, food, dietary pattern, or biomarker(s) of nutritional status.**

'**Food substances**' are defined as essential or conditionally essential nutrients, energy nutrients, or other naturally-occurring bioactive food components. Food substance intakes can be derived from the intake of food sources (naturally-occurring foods, fortified/supplemented food) and dietary supplements. For some food substances, exposure may come from other sources (e.g., from industrial minerals; vitamin D from UV-B rays), which should be taken into account Studies that focus on **foods** will generally require documentation of food composition and factors that alter composition, such as processing and fortification/supplementation. Studies of **dietary patterns** often also require documentation of food composition and prevalence of the use of specific types of foods.

[ii]For this checklist, in the context of a nutrition study, **'exposure' refers to intake or nutritional status of the food substance or dietary pattern of interest** (e.g., folate, Mediterranean diet). As noted above, exposure may also derive from other sources that may need to be taken into account (e.g., from industrial minerals, from UV-B rays in the case of vitamin D). A study may compare the intake/status between groups (e.g., higher vs lower intakes or status, supplemented vs. non-supplemented, diets high vs low in fruits and vegetables, Mediterranean- vs Western- style diets). For example, to assess the association of saturated fat intakes and risk of coronary heart disease, the comparator groups would be participants with higher saturated fat intakes and participants with lower saturated fat intakes. In a second example, to evaluate the association of folic acid intakes and the risk of birth defects, the comparator groups could be women who consume folic acid supplements and those who do not.

In some cohort studies, the outcome is a difference/change in intake or nutritional status and exposure is a behavior (e.g., education is the 'exposure' and changes in dietary patterns are the 'outcome'). The checklist is flexible enough to be used for these types of studies. The guidance addresses nutritional assessment whether intake is an exposure, an outcome or both.

NUQUEST: CHECKLIST GUIDANCE FOR COHORT NUTRITION STUDIES

1. BEFORE YOU START

Prior to completing the NUQUEST cohort checklist, ensure that:

(1). The study under evaluation is a cohort study.
If in doubt, make sure you have the correct checklist for your study design.

(2). The study is relevant to the key research question of the review.
Inclusion or exclusion of individual studies into an evaluation will depend on the key question(s) and the *a priori* criteria, including the **PICO** (Participants/Population, Intervention/Exposure, Comparator, and Outcome(s)) components that are guiding the evaluation. Decisions should be documented and in place prior to beginning the checklist. Documenting definitions of exclusion/inclusion criteria prior to initiation of the evaluation is important to establishing the applicability of the evaluation. The worksheet can be used to document these criteria.

(3). The worksheet can be used to identify important study criteria and evaluate the risk of bias associated with them. The worksheet can be referred to, as appropriate, to facilitate the study assessment.
One purpose of an assessment tool is to identify and take into account the potential for risk of bias (RoB).The worksheet can be used to identify the importance of each component of the study (exposure, confounding factors, co-exposure factors) for the study assessment.

(4). General risk of bias sources are considered.
While the type of bias is not specified within the NUQUEST checklist for each item, items in each section identify potential issues associated with facets of a study that are related to different sources of bias.

Below are examples of important general bias domains that may be commonly observed in nutrition cohort studies. Note that this is not an exhaustive list of potential bias domains and others may apply to your specific study assessment (see Resources following guidance).

Performance bias	Systematic differences between groups in the care provided or exposure to factors other than the exposure of interest.
Attrition bias	Systematic differences between groups in withdrawals from the study.
Detection bias	Systematic differences between groups in how exposure/status and outcomes are determined.
Selection bias	Systematic differences between groups on baseline characteristics.
Dietary exposure assessment bias	Error associated with the use of methodologies for assessing dietary intakes. A subcategory for self-reporting methodologies is **recall bias**, which refers to systematic error due to differences in completeness or accuracy of recall. Self-reported dietary intakes are at risk of this bias.
Misclassification bias	Systematic error due to inaccurate measurements or classifications of participants' exposure or outcome; error may be related to the risk of outcome. If the error is unrelated to the risk of outcome, the effect is usually biased to the null.

2. INSTRUCTIONS FOR COMPLETING THE NUQUEST CHECKLIST

This NUQUEST checklist is intended to be applied to individual epidemiological studies in nutrition with a cohort study design. Individual items may be answered using:

Yes	Probably Yes	Probably No	No	Not Applicable[a]
Based on something factual or known from evidence in the study publication.	Based on something the assessor judges likely to have occurred in the study.	Based on something the assessor judges likely to have not occurred in the study.	Based on something factual or known from evidence in the study publication.	This should be determined and recorded *a priori*. It would not be considered in the overall assessment.

a – the NOT APPLICABLE rating is only relevant for certain items in the checklist.

All items within a section should be rated prior to assessing the overall rating for that section (see guidance by individual section below).

All of the overall section ratings should be completed prior to completing the Overall Study Rating.

The Assessment of Applicability is optional.

3. GUIDANCE FOR COMPLETING THE NUQUEST COHORT CHECKLIST

NUTRITION STUDY INFORMATION

The study under evaluation addresses an appropriate and clearly focused question.

Unless a clear and well-defined research question is specified in the study report, it will be difficult to assess how well it has met its objectives and how relevant it is to the question you are trying to answer. Implementation of the pre-study PICO and other inclusion/exclusion criteria will often result in a "Yes" answer for this item.

SELECTION OF COHORTS

1.1	**The groups being studied are selected from source populations that are comparable in all respects other than the exposure under investigation.**
	It is important that the groups selected for comparison are from source populations that are as similar as possible in all characteristics except for their exposure, or the presence of specific prognostic factors. This is because measured, unmeasured and unknown factors are more likely to be similar in groups from similar source populations. Note that this will not eliminate all confounding factors and does not remove the need to consider confounding factors as discussed in section 1.8 below. The corollary condition also applies: groups that are formed from two different source populations or are selected from the same source population using different approaches are at higher risk for selection bias. For example, comparing a group of women of childbearing age from an urban center who took folic acid during pregnancy to those from a rural area who did not would be problematic. Although known and measured factors that may affect the comparison of the groups may be adjusted for in the analysis, unmeasured and unknown factors could still be unbalanced. ***If there is a concern, the study could be downgraded.***
1.2	**The study indicates how many of the people asked to take part did so in each of the groups being studied.**
	The participation rate is defined as the number of study participants divided by the number of eligible subjects, and should be calculated separately for each group of the study. ***A large difference in participation rate between the groups of the study indicates that selection bias may be present, and the study results should be treated with caution.*** Groups selected from different source populations or from the same source population using different approaches, may have differential participation rates and should be evaluated more closely. For example, one would be less concerned about differential participation rates in a cohort of women of childbearing age who were selected from the same urban area and groups were formed based on whether or not they took folic acid during their pregnancy. In contrast, if populations from rural and urban areas are compared, we would have more concern that participation rates might differ because, for instance, access to the study centers may be more difficult for the rural participants.
1.3	**The likelihood that some eligible subjects might have the outcome at the time of enrolment is assessed and taken into account in the analysis.**
	If some of the eligible subjects, particularly those in the unexposed group, already have the outcome at baseline, the final result will be biased (See guidance for item 1.11). A study should attempt to estimate the likelihood of this occurring, and take it into account in the analysis through the use of sensitivity analyses or other methods. For example, if the outcome is a disease, e.g., a form of cancer, the study should ensure that all participants are cancer-free at the start of the study. This may require an analysis that includes the provision for a delay prior to recording

	new cases of cancer. On the other hand, if the outcome is a continuous variable such as blood pressure, this item may not be applicable if changes in blood pressure measurements are the study outcome.
1.4	**The exposure assessment method is adequate to differentiate exposure among study groups.**
	The method used to assess exposure allows for correct assignment of participants to study groups. Further details are provided in Section 2.1.
COMPARABILITY OF COHORTS	
1.5	**The percentage of individuals or clusters recruited into each arm of the study that dropped out before the study was completed is reasonable.**
	The number of participants that drop out of a study should give concern if the number is high. Conventionally, a 20% dropout rate overall is regarded as acceptable, but in observational studies conducted over a lengthy period of time a higher dropout rate is possible. The dropout rates may vary depending on the exposure or outcome of interest. The overall and group dropout rate should be reported and an acceptable dropout rate for both should be determined *a priori*. A decision on whether to downgrade a study because of a high dropout rate is a matter of judgement based on the reasons why people dropped out, and whether dropout rates were comparable in the study groups. **A study could be downgraded if it does not report attrition from all study groups.** Reporting of efforts to follow up participants that dropped out may be regarded as an indicator of a well-conducted study. A working rule is that less than a 10% difference in the dropout rate between study groups is acceptable. **Otherwise, critical differences between groups may be a concern and the study could be downgraded.**
1.6	**Comparison is made between full participants and those lost to follow up, by exposure.**
	For valid study results, it is essential that the study participants are truly representative of the source population. It is always possible that participants who dropped out of the study will differ in some significant way from those who remained part of the study throughout. A study should attempt to identify any such differences between full and partial participants to assess whether they are comparable among study groups**. Failure to make the comparison, or to consider any unexplained differences, could lead to the study results being treated with caution.**
1.7	**Exposure level or prognostic factor is assessed more than once.**
	Confidence in data quality should be increased if exposure level is measured more than once in the course of the study. Independent assessment by more than one investigator is preferable.
1.8	**The main potential confounders are identified and taken into account in the design and analysis.**
	Confounding is the distortion of a link between exposure and outcome by another factor that is associated with both exposure and outcome. The possible presence of confounding factors, particularly unmeasured or unknown factors, is one of the principal reasons why observational studies are not more highly rated as a source of evidence. Potential confounders should be defined *a priori* in the PICO. The study should indicate which potential confounders have been considered. Judgement should be applied to consider whether all likely confounders have been considered. .If there are differences in potential confounders in the source group(s), the differences should have been accounted for in the design and/or adjusted for in the analysis **If the measures used to address confounding are considered inadequate, the study could be downgraded, depending on how serious the risk of confounding is**

	considered to be. A study that has not measured or analyzed for the effects of likely confounders and has not addressed the likelihood of confounding could be downgraded. The worksheet can be used to define and evaluate potential confounders. There may be fundamental differences between the groups that may impact nutrient intake or exposure and these should be considered and accounted for in a study (e.g., health-related behavior or personal/lifestyle variables). These factors may be known or unknown. For example, in a cohort of women of childbearing age where folic acid intake (compared to no intake) is being studied, we may want to consider potential confounders that may be related to the nutrient intake such as education, physical activity and race or ethnicity. In addition, and particularly when we have different source populations (e.g., different geographies), a variety of known and unknown differences between the groups may confound the nutrient intake or exposure, including but not limited to population structure, culture, genetic diversity and health or personal behaviors.
ASCERTAINMENT OF OUTCOMES	
1.9	**The outcomes are clearly defined.**
	Once enrolled in the study, participants should be followed until specified endpoints or outcomes are reached. In a study of the effect of exercise on the death rates from heart disease in middle-aged men, for example, participants might be followed up until death, or until reaching a predefined age**. If *a priori* outcomes and the criteria used for measuring them are not clearly defined, the study could be downgraded.** Consideration needs to be given to whether secondary or post-hoc outcomes will be acceptable.
1.10	**The assessment of outcome is made blind to exposure. If the study is retrospective this may not be applicable.**
	This relates to the risk of detection bias. If the assessor is blinded to which participants received the exposure, and which did not, the prospects of unbiased results are significantly increased. Studies in which this is done should be rated more highly than those where it is not done, or not done adequately. If the outcome is objective, this is less of a problem. **Subjective or self-reported outcomes or those that are assessed without blinding are of particular concern and the study could be downgraded.**
1.11	**Where blinding was not possible, there is some recognition that knowledge of exposure could have influenced the assessment of outcome.**
	Blinding may not be possible in some cohort studies. In order to assess the extent of any bias that may be present, it may be helpful to compare process measures used on the participant groups - e.g., frequency of observations, who carried out the observations, the degree of detail and completeness of observations. If these process measures are comparable between the groups, the results may be regarded with more confidence.
1.12	**Evidence from other sources is used to demonstrate that the method of outcome assessment is valid and reliable.**
	Acceptable methods should be defined *a priori* in the PICO. Evidence should be provided showing that the outcome measures used are reliable and validated prior to their use in the study. This is particularly important where outcome measures require any degree of subjectivity. For example, whether Type 2 diabetes is diagnosed by a specific chemical test or is self-reported, or whether an evaluation of cognitive function is based on a validated and standardized instrument or by a non-validated 'in-house' measurement tool. This issue becomes important when biomarker concentrations are compared against cut points for classification of participants (e.g., disease/non-disease, adequacy/deficiency) if the cut-point was determined by different or non-standardized methods**. If the**

	outcome measures are not measured in a valid or reliable way, the study could be downgraded. Whether and how secondary outcomes are to be used needs to be determined in conjunction with the PICO documentation.
NUTRITION-SPECIFIC	
2.1	**The frequency and quantity of the exposure under study are accurately[1] and reliably[2] measured.**
	The accuracy and reliability of exposure assessments is of paramount importance for all nutrition-related studies. In general, PICO criteria need not be overly restrictive and can include relevant and commonly accepted exposure assessment methods**.** A study should indicate how the degree of exposure or presence of prognostic factors or markers were assessed. The methods must be sufficient to accurately quantify the exposure under investigation. **The study should report the method used to assess exposures or status and the validity of such methods.** Validity refers to the degree to which the exposure assessment methodology accurately measures the aspect of exposure it was intended to measure. Exposure assessment methods are validated in different ways. Comparing an exposure method to a "gold standard" or reference method thought to capture all the food and supplement sources containing the exposure of interest provides the best validation approach. Comparing a self-reported method to another self-reported method requires caution as the two methods may be affected by the same types of errors and biases. An exposure assessment method of tool could be modified to reflect the population under study if the food/supplement supply differs for the new study population. Validation of a new questionnaire against a previous questionnaire targeting the exposure under study allows for comparability among similarly validated studies. **If the exposure measures are not clearly stated, the study could be downgraded.** Currently available assessment methods are often subject to random errors and systematic biases that can result in misleading and erroneous conclusions. Optimal methodologies as well as the nature and severity of potential errors and biases vary among food substances and, therefore, must be considered on a case-by-case basis. Methods used in the study under evaluation need to be evaluated for their potential for **risk of bias (RoB)** that includes both general and topic-specific criteria. Topic-specific criteria are issues related specifically to the exposure of interest. For example, issues related to assessing sodium intake will be different from those of other nutrients or total energy. **Topic-specific RoB criteria** assist in rating the exposure methodologies included in the review as higher to lower-risk of bias, which need to be considered when rating the nutrition-specific checklist items described below. **The worksheet can be used to define and evaluate topic-specific criteria.** **Food substance intakes from self-reported diets** (e.g., food frequency questionnaire, food record, or 24-hour recall). Methods have different strengths and weaknesses (known errors and biases). The best method depends on the research question[1] and generally needs to be determined as part of the *a priori* PICO documentation process. The frequency of evaluation of the exposure will often depend on the food substance, the dosage conditions and the outcome. **Food substance intakes from supplements or specific food products** (e.g., bars, liquid drinks, fortified foods). The study should report product composition details and the methods used to confirm the composition of the product (i.e., assay procedure), and to how adherence to the use of these products was monitored (e.g., supplement count, assessment of changes in nutritional status, etc.).

[1] Accuracy: The extent to which a measured value is close to that of the true value. Low accuracy can be a result of bias or systematic error in measurement (https://www.nutritools.org/glossary)

[2] Reliability: The degree of stability of a measurement repeated under the same conditions (https://www.nutritools.org/glossary)

For dietary pattern or food intakes, the study should include ranges or distributions of any pre-specified food substance or status exposures of interest. The study should also include exposure measurement characteristics and errors, of which some can be controlled for statistically (e.g., 'usual' intakes). This can be accomplished by intake methodologies that represent averages over a specified time period or by statistically adjusting for within-person-variabilities in 24-hour recalls. Some intake methodologies cannot adjust for usual intakes (e.g., if they are unknown or unmeasured; or subject to a systematic bias such as underreporting) and this can introduce bias.

Biomarkers of intake and exposure are usually metabolites recovered from the blood or urine that objectively and accurately assess intake over a period of time. However, biomarker assays can vary in accuracy and reliability. If a biomarker has been accurately measured and appropriately qualified for its intended purpose, it is more reliable than exposure assessed by dietary intake estimation methods.

Note that use of non-standardized methodologies could affect status measures, even when the method itself includes standards and controls. For example, 25-hydroxyvitamin D values differ depending on the assay platform and the laboratory performing the measure. A measure from a laboratory that has been certified for the method will be more reliable than one from an uncertified laboratory and standardized procedures are more reliable than non-standardized procedures.

Other important considerations for assessing exposure include the composition, the form, any potential interactions with other exposure components, supplement brand name, and food types or recipes. The importance and weight of these issues (and other relevant characteristics of the exposure) must be determined *a priori* by the reviewer. Knowledge about the exposure form (e.g., naturally occurring vs. synthetic food substance) and intake conditions (e.g., taken with or without meals; bolus vs. multiple daily exposures; preparation practices for meal components) can be important especially when bioavailability and/or bioactivity are issues.

Ideally, product composition would be confirmed by analysis because food/supplement label values and/or composition databases do not always accurately reflect actual composition. **If the exposure is not accurately measured or adequately characterized, confidence in the degree of exposure is reduced. A value on a product label could be acceptable but might not accurately reflect composition and should be interpreted with caution.**

Energy intakes are another consideration. They are important for understanding the full effect of an addition or deletion of a food, or substitution of one food for another, and for identifying potential errors in dietary intake reporting (e.g., under- or- over reporting of dietary intake).

Rating: Documentation of exposure should be sufficient to determine its validity. For example, intake/nutritional status based on a comparison to an accurate external method would have the highest rating. Self-reported intake or use of a biomarker assay procedure with accuracy or bias concerns that have been described/evaluated would have higher uncertainty. Self-reported intake with no description of accuracy or bias would have the lowest rating. In addition, because of large day-to-day variabilities in intakes by a given individual, higher rating should be given to results expressed as "usual" or long-term intakes. If adherence to the exposure is assessed by a supplement count then there is greater confidence in the measurement than if the assessment is achieved by recall. The tolerance for uncertainty will vary depending on the purpose of your evaluation and will be reflected in the study rating. The rationale for the rating system used, including a description of tolerance for uncertainty, should be defined *a priori*.

2.2	**The relevant exposure at baseline is measured and taken into account.**
	Accurate baseline intake/status data are necessary to ensure appropriate group assignments; to determine whether changes in intake occurred during the study period (which could confound study results); and to compare results across studies in different populations. If supplement use is the exposure, baseline intakes can be used to ascertain the nutritional similarity of exposed and unexposed groups at baseline and estimate total intake. If expressed as distributions, baseline intakes of study groups can aid in the interpretation of results. For example, if a wide range of baseline statuses is present among groups of a supplement study, the overall effect of the exposure may be null; this could happen if nutrient-depleted participants experience benefit but replete participants experience no effect or even harm. The relevant exposure at baseline is distinct from the concept of similarity to a source population (See 1.1). Section 2.1 offers guidance on how to assess the RoB of the method to monitor intake/status or supplement consumption.
2.3	**The baseline exposure differences between the groups have been maintained over the course of the study.**
	This item follows on 1.7 and refers to monitoring of exposure over time. All participants are exposed to diets that include food substances, foods or dietary patterns of interest. Changes in intake over the duration of the study for any or all of the groups could make it more difficult to detect differences in outcome due to the effect of the exposure of interest. Has the potential for this been monitored? If a change in intake has occurred, does the study have appropriate methods to deal with it? **If it is possible/likely that this has occurred, and has not been addressed, then the study could be downgraded.** Higher rating would be given to studies where there are multiple measures of exposure over time and especially where this is confirmed/assessed with biomarker data. A higher rating would also be given if baseline intake changes over time have been addressed in the analysis and/or interpretation. Section 2.1 offers guidance on how to assess the RoB of the method to monitor intake/status or supplement consumption.
2.4	**The interval between the exposure and outcome is of sufficient duration to observe an effect, if there is one.**
	In order to determine whether a 'true' relationship exists, the duration of the interval between the exposure and the outcome must be sufficient to allow the outcome to occur; the minimum acceptable duration for the outcome of interest should be pre-specified by the reviewer False positives can occur in studies of inadequate duration. For example, transient decreases in LDL-cholesterol may be observed in short-term lipid studies but may not persist in long-term studies. Confounding factors such as seasonal changes can transiently affect status or intake. Even study enrolment can transiently affect intake or status in the short-term. **Identify a minimum follow-up time for each outcome during the PICO documentation step. If this minimum cannot be met, the study could be downgraded.**

4. GUIDANCE FOR COMPLETING THE OVERALL STUDY RATING

Taking the overall ratings for selection of cohorts, comparability of cohorts, ascertainment of outcomes and nutrition-specific aspects into consideration, the rating of the overall risk of bias for the study could consider the following as a guide:

- All criteria may not have equal importance.
- The importance of a particular criterion may vary depending on your research question (PICO). In this regard, it may be helpful to consider the criteria in 3 parts which could be differentially weighted:
  - (a) creation of the study groups (Items 1.1 to 1.4, and 2.2);
  - (b) comparability of the study groups (Items 1.5 to 1.8, 2.1and 2.3) ; and
  - (c) outcome ascertainment (Items 1.9 to 1.12, and 2.4).

For example, if there is a concern that the formation of the cohorts is highly susceptible to selection bias you may want to put more weight on 'creation of study groups'. If assessment of exposure is prone to error including bias or if adherence to the exposure is problematic, then more emphasis may be placed on 'comparability of study groups'. If a study involves an assessment of outcome that is prone to bias, you may want to put more weight on criteria related to 'outcome ascertainment'.

RESOURCES

- Chung M, Balk EM, Ip S, Raman G, Yu WW, Trikalinos TA, Lichtenstein AH, Yetley EA, Lau J. *Reporting of systematic reviews of micronutrients and health: a critical appraisal.* Am J Clin Nutr. 2009 Apr;89(4):1099-113. doi: https://doi.org/10.3945/ajcn.2008.2 6821.
- Delgado-Rodríguez M, Llorca J. *Bias.* Journal of Epidemiology & Community Health 2004;58:635-641.
- Institute of Medicine (IOM) Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease Board on Health Care Services Board on Health Sciences Policy Food and Nutrition Board, *Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease*, C.M. Micheel and J.R. Ball, Editors. Washington (DC): National Academies Press; 2010. Available from: https://iom.nationalacademies.org/Reports/2010/Evaluation-of-Biomarkers-and-Surrogate-Endpoints-in-Chronic-Disease.aspx
- National Academies of Sciences, Engineering, and Medicine. *Guiding principles for developing Dietary Reference Intakes based on chronic disease.* Washington, DC: The National Academies Press; 2017. doi: https://doi.org/10.17226/2482.
- National Academies of Sciences, Engineering, and Medicine. Methodological Considerations. In: *Dietary Reference Intakes for sodium and potassium.* Washington, DC: The National Academies Press, 2019:61-98. doi: https://doi.org/10.17226/25353.
- Dietary Assessment Primer, Summary Tables: Recommendations on Potential Approaches to Dietary Assessment for Different Research Objectives Requiring Group-level Estimates. National Institutes of Health, National Cancer Institute. Internet: https://dietassessmentprimer.cancer.gov/approach/table.html (accessed September 20, 2020).
- Scottish Intercollegiate Guidelines Network *(SIGN). A guideline developer's handbook.* Edinburgh: SIGN; 2019. (SIGN publication no. 50). [November 2019]. Available from URL: http://www.sign.ac.uk.

NUQUEST – NUtrition QUality Evaluation Strengthening Tool: Development of tools for the evaluation of risk of bias in nutrition studies

Shannon E. Kelly, Linda S. Greene-Finestone, Elizabeth A. Yetley, Karima Benkhedda, Stephen P.J. Brooks, George A. Wells, and Amanda J. MacFarlane

Online supplementary file 5: NUQUEST Case-Control Checklist for Nutrition Studies

NUQUEST: CASE-CONTROL CHECKLIST FOR NUTRITION STUDIES

We recommend reviewing the Guidance and Worksheet provided before starting.

NUTRITION STUDY[i] INFORMATION
CITATION:
REVIEWER NAME:
REVIEW DATE:
RESEARCH QUESTION:
REVIEWER NOTES:
THE STUDY UNDER EVALUATION ADDRESSES AN APPROPRIATE AND CLEARLY FOCUSED QUESTION: ☐ YES ☐ NO

		YES	PROBABLY YES	PROBABLY NO	NO	NOT APPLICABLE
CREATION OF STUDY GROUPS						
1.1	The cases and controls are taken from comparable populations.	☐	☐	☐	☐	-
1.2	The same exclusion criteria are used for both cases and controls.	☐	☐	☐	☐	-
1.3	Cases are clearly defined and differentiated from controls.	☐	☐	☐	☐	-
1.4	It is clearly established that controls are non-cases.	☐	☐	☐	☐	-
Taking the ratings of 1.1, 1.2, 1.3, and 1.4 into consideration, provide an overall rating for Creation of Study Groups using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Creation of Study Groups. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Creation of Study Groups. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Creation of Study Groups. High RoB.		**OVERALL RATING FOR CREATION OF STUDY GROUPS** ☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**				
COMPARABILITY OF STUDY GROUPS						
1.5	The percentage of individuals in the study who participated as cases and controls is reasonable.	☐	☐	☐	☐	-
1.6	Measures will have been taken to prevent knowledge of primary exposure[ii] influencing case ascertainment.	☐	☐	☐	☐	☐
1.7	Comparison is made between participants and non-participants to establish their similarities or differences.	☐	☐	☐	☐	-
1.8	The main potential confounders are identified and taken into account in the design and/or analysis.	☐	☐	☐	☐	-
Taking the ratings of 1.5, 1.6, 1.7, and 1.8 into consideration, provide an overall rating for Comparability of Groups using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Comparability of Study Groups. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Comparability of Study Groups. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Comparability of Study Groups. High RoB.		**OVERALL RATING FOR COMPARABILITY** ☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**				

NUQUEST: CASE-CONTROL CHECKLIST FOR NUTRITION STUDIES

We recommend reviewing the Guidance and Worksheet provided before starting.

		YES	PROBABLY YES	PROBABLY NO	NO	NOT APPLICABLE
EXPOSURE ASCERTAINMENT						
1.9	The exposure under study is accurately and reliably measured.	☐	☐	☐	☐	-
Taking the rating of 1.9 into consideration, provide an overall rating for Exposure Ascertainment using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Exposure Ascertainment. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Exposure Ascertainment. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Exposure Ascertainment. High RoB.		**OVERALL RATING FOR ASCERTAINMENT** ☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**				
NUTRITION-SPECIFIC						
2.1	The baseline exposure differences between the groups have been maintained over the course of the study.	☐	☐	☐	☐	-
2.2	The interval between the exposure and outcome is of sufficient duration to observe an effect, if there is one.	☐	☐	☐	☐	-
Taking the ratings of 2.1 and 2.2 into consideration, provide an overall rating for Nutrition-Specific using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Nutrition-specific issues. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Nutrition-specific issues. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Nutrition-specific issues. High RoB.		**OVERALL RATING FOR NUTRITION** ☐ **GOOD (+)** ☐ **NEUTRAL (0)** ☐ **POOR (-)**				

OVERALL STUDY RATING

ASSESSMENT SUMMARY: Record the overall ratings from each sections above in the boxes below:

CREATION OF STUDY GROUPS:	☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**	COMPARABILITY OF STUDY GROUPS:	☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**	EXPOSURE ASCERTAINMENT:	☐ **Good (+)** ☐ **Neutral(0)** ☐ **Poor (-)**	NUTRITION-SPECIFIC:	☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**

Taking the overall ratings for each section into consideration, provide an overall study rating: ***Good (+):*** Almost all criteria met. Little or no concern over study design. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern over study design. Moderate RoB ***Poor (-):*** Most or all criteria not met, or significant flaws related to key aspects of the study design. High RoB.	**OVERALL STUDY RATING** ☐ **GOOD (+)** ☐ **NEUTRAL (0)** ☐ **POOR (-)**

NUQUEST: CASE-CONTROL CHECKLIST FOR NUTRITION STUDIES

We recommend reviewing the Guidance and Worksheet provided before starting.

ASSESSMENT OF APPLICABILITY:
How applicable is the study under evaluation to your research question?

This section is optional.
Although it is not necessarily related to RoB or study quality, it may help interpret the study results in the context of the research question.

Taking into account the study design and methodology used, are the results directly applicable to your research question?	YES ☐	PROBABLY YES ☐	PROBABLY NO ☐	NO ☐
Taking into account the study design and methodology used, do you have confidence that there is clear evidence of an association between the intervention and the outcome?	YES ☐	PROBABLY YES ☐	PROBABLY NO ☐	NO ☐
Are the results of this study directly applicable to the population of interest?	YES ☐	PROBABLY YES ☐	PROBABLY NO ☐	NO ☐

NOTES:

[i]For this checklist, the term "nutrition study" is used broadly to apply to studies in which the focus is on **intake of a food substance, food, dietary pattern or biomarker(s) of nutritional status.**

'**Food substances**' are defined as essential or conditionally essential nutrients, energy nutrients, or other naturally-occurring bioactive food components. Food substance intakes can be derived from the intake of food sources (naturally occurring foods, fortified/supplemented food) and dietary supplements. For some food substances, exposure may come from other sources (e.g., from industrial minerals; vitamin D from UV-B rays), which should be taken into account. Studies that focus on **foods** will generally require documentation of food composition and factors that alter composition, such as processing and fortification/supplementation. Studies of **dietary patterns** often also require documentation of food composition and/or prevalence of the use of specific types of foods.

[ii]For this checklist, in the context of a nutrition study, **'exposure' refers to history of intake or nutritional status of the food substance or dietary pattern of interest** (e.g., folate, Mediterranean diet). As noted above, exposure may also derive from other sources (e.g., from industrial minerals; vitamin D from UV-B rays). A study compares the intake/status between groups (e.g., higher vs. lower intakes or status, supplemented vs. non-supplemented, diets high vs. low in fruits and vegetables, Mediterranean- vs. Western-style diets). For example, to assess the association of coronary heart disease and saturated fat intakes, the comparator groups would be participants with higher saturated fat intakes and participants with lower saturated fat intakes. In a second example, to evaluate the association of birth defects and folic acid intakes, the comparator groups would be women who consumed folic acid supplements and those who did not.

In some case-control nutrition studies, the historical exposure is not an intake or nutritional status, but the outcome is an intake or nutritional status. For example, the case group could have received an income supplement and the control group no supplement with the outcome being fruit and vegetable intake. The checklist is flexible enough to be used for these types of studies. The guidance addresses nutritional assessment whether intake is an exposure, an outcome or both.

NUQUEST: CHECKLIST GUIDANCE FOR CASE-CONTROL NUTRITION STUDIES

1. BEFORE YOU START

Prior to completing the NUQUEST case-control study checklist, ensure that:

(1). The study under evaluation is a case-control study.
If in doubt, make sure you have the correct checklist for your study design.

(2). The study is relevant to the key research question of the review.
Inclusion or exclusion of individual studies into an evaluation will depend on the key question(s) and the *a priori* criteria, including the **PICO** (Participants/Population, Intervention/Exposure, Comparator, and Outcome(s)) components that are guiding the evaluation. Decisions should be documented and in place prior to beginning the checklist. Documenting definitions of exclusion/inclusion criteria prior to initiation of the evaluation is important for establishing the applicability of the evaluation. The worksheet can be used to document these criteria.

(3). The worksheet can be used to identify important study criteria and evaluate the risk of bias associated with them. The worksheet can be referred to, as appropriate, to facilitate the study assessment.
One purpose of an assessment tool is to identify and take into account the potential for risk of bias (RoB).The worksheet can be used to identify the importance of each component of the study (exposure, confounding factors, co-exposure factors) for the study assessment.

(4). General risk of bias sources are considered.
While the type of bias is not specified within the NUQUEST checklist for each item, items in each section identify potential issues associated with facets of a study that are related to different sources of bias

Below are examples of important general bias domains that may be commonly observed in nutrition studies. Note that this is not an exhaustive list of potential bias domains and others may apply to your specific study assessment (see Resources following guidance).

Performance bias	Systematic differences between groups in the care provided or exposure to factors other than the exposure of interest.
Attrition bias	Systematic differences between groups in withdrawals from the study.
Detection bias	Systematic differences between groups in how exposure/status and outcomes are determined.
Selection bias	Systematic differences between groups on baseline characteristics.
Dietary exposure assessment bias	Error associated with the use of self-reporting methodologies for assessing dietary intakes. A subcategory for self-reporting methodologies is **recall bias**, which refers to systematic error due to differences in completeness or accuracy of recall. Self-reported dietary intakes are at risk of this bias. **Case-control studies** are particularly at risk of recall bias because cases (those with the outcome) and controls (those without) may recall their dietary exposures differently.
Misclassification bias	Systematic error due to inaccurate measurements or classifications of participants' exposure or outcome; error may be related to the risk of outcome. If the error is unrelated to the risk of outcome, the effect is usually biased to the null.

2. INSTRUCTIONS FOR COMPLETING THE NUQUEST CHECKLIST

This NUQUEST checklist is intended to be applied to individual epidemiological studies in nutrition with a case-control design. Individual items may be answered using:

Yes	Probably Yes	Probably No	No	Not Applicable[a]
Based on something factual or known from evidence in the study publication.	Based on something the assessor judges likely to have occurred in the study.	Based on something the assessor judges likely to have not occurred in the study.	Based on something factual or known from evidence in the study publication.	This should be determined and recorded *a priori*. It would not be considered in the overall assessment.

a – the NOT APPLICABLE rating is only relevant for certain items in the checklist.

All items within a section should be rated prior to assessing the overall rating for that section (see guidance by individual section below).

All of the overall section ratings should be completed prior to completing the Overall Study Rating.

The Assessment of Applicability is optional.

3. GUIDANCE FOR COMPLETING THE NUQUEST CASE-CONTROL CHECKLIST

NUTRITION STUDY INFORMATION

The study under evaluation addresses an appropriate and clearly focused question.

Unless a clear and well-defined research question is specified in the study report, it will be difficult to assess how well it has met its objectives and how relevant it is to the question you are trying to answer. Implementation of the pre-study PICO and other inclusion/exclusion criteria will often result in a "Yes" answer for this item.

SELECTION OF STUDY GROUPS

1.1	**The cases and controls are taken from comparable populations.**
	Study participants may be selected from the target population (all individuals to which the results of the study could be applied), the source population (a defined subset of the target population from which participants are selected), or from a pool of eligible subjects (a clearly defined and counted group selected from the source population). It is important that the cases and controls selected for comparison are from source populations that are as similar as possible in all characteristics except for their exposure, (or) the presence of specific prognostic factors. This is because measured, unmeasured and unknown factors are more likely to be similar in groups from similar source populations. Cases and controls should be selected from a defined and comparable period of time so that environmental and medical exposures are more likely to be similar. Note that this will not eliminate all confounding factors and does not remove the need to consider confounding factors. **If the study does not include clear definitions of the source population, it could be downgraded.**
1.2	**The same exclusion criteria are used for both cases and controls.**
	All selection and exclusion criteria should be applied equally to cases and controls. Failure to do so may introduce a significant degree of bias into the results of the study.
1.3	**Cases are clearly defined and differentiated from controls.**
	The method of selection of cases is of critical importance to the validity of the study. Investigators have to be certain that cases are truly cases, but must balance this with the need to ensure that the cases admitted into the study are representative of the eligible population. **The issues involved in case selection are complex, and should ideally be evaluated by someone with a good understanding of the design of case-control studies. If the study does not include information on how cases were selected, it is probably safest to downgrade it.**
1.4	**It is clearly established that controls are non-cases.**
	Just as it is important to be sure that cases are true cases, it is important to be sure that controls do not have the outcome under investigation. It is possible that a control is asymptomatic and/or may not have been assessed for eligibility in a similar manner as a case, which could be a limitation of the study. Control subjects should be chosen so that information on exposure can be obtained or assessed in a similar way to that used for the selection of cases. **If the methods of control selection are not described or differ between cases and controls, the study could be downgraded. If different methods of selection are used for cases and controls the study should be evaluated by someone with a good understanding of the design of case-control studies.**

COMPARABILITY OF STUDY GROUPS	
1.5	The percentage of individuals in the study who participated as cases and controls is reasonable.
	Differences between the eligible population and the participants are important, as they may influence the validity of the study. Participation rates can be calculated by dividing the number of study participants by the number of eligible subjects. If the participation rates are low, or there is a large difference between the two groups, the study results may well be invalid due to differences between participants and non-participants. **If the participation rates are low or the differences between the groups are large, the study could be downgraded.** One working rule is that an overall loss of 10 to 20%, or a difference in participation of 5 to 10%, is often regarded as acceptable for group formation. In particular, for a nested case-control, consideration should be given to: • Losses when cases and/or controls were determined from the cohort; and, • Losses of included cases and/or controls when exposure could not be determined.
1.6	**Measures will have been taken to prevent knowledge of primary exposure influencing case ascertainment.**
	If there is a possibility that case ascertainment can be influenced by knowledge of exposure, assessment of any association is likely to be biased. A well-conducted study should take this into account in the design of the study. Section 1.9 provides guidance offers guidance on how to assess exposure.
1.7	**Comparison is made between participants and non-participants to establish their similarities or differences.**
	Even if participation rates were comparable and acceptable, it is still possible that the participants selected to act as cases or controls may have differed from other members of the source population in some significant way. A well-conducted case-control study will look at samples of the non-participants among the source population to ensure that the participants are a truly representative sample. **Failure to make the comparison, or to consider any unexplained differences, could lead to the study being downgraded.**
1.8	**The main potential confounders are identified and taken into account in the design and/or analysis.**
	Confounding is the distortion of a link between exposure and outcome by another factor that is associated with both exposure and outcome. The possible presence of confounding factors, particularly unmeasured or unknown factors, is one of the principal reasons why case-control studies are not more highly rated as a source of evidence. Potential confounding variables should be identified *a priori* in the PICO. The study should indicate which potential confounders have been considered. Judgement should be applied to consider whether all likely confounders have been considered. If there were differences in potential confounders in the source group(s), the differences should have been accounted for in the design and/or adjusted for in the analysis. **If the measures used to address confounding are considered inadequate, the study could be downgraded, depending on how serious the risk of confounding is considered to be. A study that did not measure or analyze for the effects of likely confounders and has not addressed the likelihood of confounding could be downgraded. The worksheet can be used to define and evaluate potential confounders.** There may be fundamental differences between cases and controls that may impact the exposure and these should be considered and accounted for in a study (e.g., health-related behavior or personal/lifestyle variables). These factors may be known or unknown. For example, in women of childbearing age where folic acid intake (compared to no intake) is being studied, one may want to consider potential confounders that may be related to the nutrient intake such as education, physical activity and race or ethnicity.

	In addition, and particularly when there are different source populations (e.g., different geographies), a variety of known and unknown differences between the groups may confound the nutrient intake or exposure, including but not limited to population structure, culture, genetic diversity, and health or personal behaviors.
EXPOSURE ASCERTAINMENT	
1.9	**The exposure under study is accurately[1] and reliably[2] measured.**
	The accuracy and reliability of exposure assessments is of paramount importance for all nutrition-related studies. In general, PICO criteria need not be overly restrictive, and can include relevant and commonly accepted exposure assessment methods**.** A study should indicate how the degree of exposure or presence of prognostic factors or markers was assessed. The methods must be sufficient to accurately quantify the exposure under investigation. **Study design.** Nutritional assessment methods may vary depending on whether the case-control study is **retrospective** or **nested in a prospective study**. In retrospective case-control studies, assessment is usually based on historical food use reflecting recalls of food consumed at a designated time prior to study enrollment. They typically assess historical exposures using food frequency questionnaires. In nested case-control studies, intake assessments reflecting a past time are taken from assessments obtained at the past time. The assessment methods for nested case-control studies may include FFQs, 24-hr recalls or food records, and exposure biomarkers. The accuracy and reliability of the nutritional assessment in case-control studies will largely depend on two factors 1) the accuracy of the assessment method used, including how well quantitation was assessed, and 2) how well the exposure can be assessed for the historical time of interest (i.e., before initiation of the disease of interest). **The study should report the methods used to assess exposures or status and the validity of such methods.** Validity refers to the degree to which the exposure assessment methodology accurately measures the aspect of exposure it was intended to measure. Exposure assessment methods are validated in different ways. Comparing an exposure method to a "gold standard" or reference method thought to capture all the food and supplement sources containing the exposure of interest provides the best validation approach. Comparing a self-reported method to another self-reported method requires caution as the two methods may be affected by the same types of errors and biases. An exposure assessment method of tool could be modified to reflect the population under study if the food/supplement supply differs for the new study population. Validation of a new questionnaire against a previous questionnaire targeting the exposure under study allows for comparability among similarly validated studies. **If the exposure measures are not clearly stated, the study could be downgraded.** Currently available assessment methods are often subject to random errors and systematic biases that can result in misleading and erroneous conclusions. Optimal methodologies as well as the nature and severity of potential errors and biases vary among food substances and, therefore, must be considered on a case-by-case basis. Methods used in the study under evaluation need to be evaluated for their potential for **risk of bias (RoB)** that includes both general and topic-specific criteria. **Ideally, the ascertainment of exposure should be blinded to the case or control grouping - otherwise the study could be downgraded.** Topic-specific criteria are issues related specifically to the exposure of interest. For example, issues related to assessing sodium intake will be different from those of other nutrients or total energy. **Topic-specific RoB criteria** assist in rating the exposure methodologies included in the review as higher to lower-risk of bias, which need to be considered when rating the nutrition-specific checklist items described below. **The worksheet can be used to define and evaluate topic-specific criteria**.

[1] Accuracy: The extent to which a measured value is close to that of the true value. Low accuracy can be a result of bias or systematic error in measurement (https://www.nutritools.org/glossary).

[2] Reliability: The degree of stability of a measurement repeated under the same conditions (https://www.nutritools.org/glossary).

	Historical food substance intakes from self-reported diets. The study should report the method used to assess/monitor the intakes of the comparator groups (e.g., type of dietary assessment such as food frequency questionnaire, food record or 24-hour recall; type of assay used to measure the intake or status biomarker) and the approach used to validate the method. Methods have different strengths and weaknesses (known errors). The best method depends on the research question and generally needs to be determined as part of the *a priori* PICO documentation process. The frequency of evaluation of the intake/status will often depend on the food substance, the outcome and availability of historical data. **Historical food substance intakes from supplements or specific food products** (e.g., bars, liquid drinks, and fortified foods)**.** The study should report product composition details and how adherence to the use of these products was monitored (e.g., supplement count, assessment of changes in nutritional status, etc.). In the case of nested case control studies, it should also report the methods used to confirm the composition of the product (i.e., assay procedure). **For historical dietary pattern or food intakes**, the study should include ranges or distributions of any pre-specified food substance intakes of interest. The study should also include exposure measurement characteristics and errors, of which some can be controlled for statistically (e.g., by known or measured variables or adjusted usual intakes) while others cannot (e.g., if they are unknown or unmeasured; or subject to a systematic bias such as underreporting) which can introduce bias. **Biomarkers of intake and exposure** are usually metabolites recovered from the blood or urine that objectively and accurately assess intake over a period of time. However, biomarker assays can vary in accuracy and reliability. If a biomarker has been accurately measured and appropriately qualified for its intended purpose, it is more reliable than intake assessed by dietary intake estimation methods. To attain historic intake or status data, biomarkers could have been measured prior to the study or measured in biobanked samples. Note that use of non-standardized methodologies could affect status measures, even when the method itself includes standards and controls. For example, 25-hydroxyvitamin D values differ depending on the assay platform and the laboratory performing the measure. A measure from a laboratory that has been certified for the method will be more reliable than one from an uncertified laboratory and standardized procedures are more reliable than non-standardized procedures. **Other important considerations for assessing exposure include the composition, the form, any potential interactions with other exposure components, supplement brand name, and food types or recipes. The importance and weight of these issues (and other relevant characteristics of the exposure) must be determined *a priori* by the reviewer.** Knowledge about the exposure form (e.g., naturally occurring vs. synthetic food substance) and intake conditions (e.g., taken with or without meals; bolus vs. multiple daily exposures; preparation practices for meal components) can be important especially when bioavailability and/or bioactivity are issues. For nested case control studies, product composition would ideally be confirmed by analysis because food/supplement label values and/or composition databases do not always accurately reflect actual composition. **If the exposure is not accurately measured or adequately characterized, confidence in the degree of exposure is reduced. A value on a product label could be acceptable but might not accurately reflect composition and should be interpreted with caution.** **Energy intakes** are another consideration. They are important for understanding the full effect of an addition or deletion of a food, or substitution of one food for another, and for identifying potential errors in dietary intake reporting (e.g., under- or- over reporting of dietary intake).

	Rating: Documentation of exposure should be sufficient to determine its validity. For example, intake/nutritional status based on a comparison to an accurate external method would have the highest rating. Self-reported intake or use of a biomarker assay procedure with accuracy or bias concerns that have been described/evaluated would have higher uncertainty. Case-control studies are particularly susceptible to **recall bias** in which cases and controls may recall their self-reported intakes differently. For this reason, self-reported intake with no description of accuracy or bias would have the lowest rating. In addition, because of large day-to-day variabilities in intakes by a given individual, higher rating should be given to results expressed as "usual" or long-term intakes. If adherence to the exposure is assessed by a supplement count then there is greater confidence in the measurement than if the assessment is achieved by recall. The tolerance for uncertainty will vary depending on the purpose of your evaluation and will be reflected in the study rating. The rationale for the rating system used, including a description of tolerance for uncertainty, should be defined *a priori*.
NUTRITION-SPECIFIC	
2.1	**The baseline exposure differences between the groups have been maintained over the course of the study.**
	All participants are exposed to diets, which include food substances, foods or dietary patterns of interest. Changes in intake over the duration of the study for any or all of the groups could make it more difficult to detect differences in outcome due to the effect of the exposure of interest. Has the potential for this been monitored? If a change in intake occurred, does the study have appropriate methods to deal with it? **If it is possible/likely that this has occurred, and has not been addressed, then the study could be downgraded**. Higher rating would be given to studies where there are multiple measures of exposure over time and especially where this is confirmed/assessed with biomarker data. A higher rating would also be given if baseline intake changes over time have been addressed in the analysis and/or interpretation. Section. 1.9 offers guidance on how to assess the RoB of the method to monitor intake/status or supplement consumption.
2.2	**The interval between the exposure and outcome is of sufficient duration to observe an effect, if there is one.**
	In order to determine whether a 'true' relationship exists, the duration of the interval between the historic exposure and the current outcome must be sufficient to allow for the outcome to occur; the minimum acceptable duration for the outcome of interest should be pre-specified by the reviewer. False positives can occur in studies of inadequate duration. For example,, transient decreases in LDL-cholesterol may be observed in short-term lipid studies but may not persist in long-term studies. Confounding factors such as seasonal changes can transiently affect status or intake. Even study enrolment can transiently affect intake or status in the short-term. **Identify a minimum interval time between exposure and outcome during the PICO documentation step. If this minimum cannot be met, the study could be downgraded**.

4. GUIDANCE FOR COMPLETING THE OVERALL STUDY RATING

Taking the overall ratings for the creation of study groups, comparability of study groups, exposure ascertainment and other nutrition-specific aspects into consideration, the rating of the overall risk of bias for the study could consider the following as a guide:

- All criteria may not have equal importance.
- The importance of a particular criterion may vary depending on your research question (PICO). In this regard, it may be helpful to consider the criteria in 3 parts which could be differentially weighted:
  - (a) creation of the study groups (Items 1.2 to 1.4, and 2.2);
  - (b) comparability of the study groups (Items 1.1, 1.5 to 1.8, and 2.1); and
  - (c) exposure ascertainment (Item 1.9).

For example, if the cases are not clearly defined and differentiated from the controls, selection bias could result and you may want to put more weight on 'creation of study groups'. If there are important differences between the eligible population and the participants then more emphasis may be placed on 'comparability of study groups'. If assessment of exposure is prone to error including bias you may want to put more weight on criteria related to 'exposure ascertainment'.

RESOURCES

- Chung M, Balk EM, Ip S, Raman G, Yu WW, Trikalinos TA, Lichtenstein AH, Yetley EA, Lau J. *Reporting of systematic reviews of micronutrients and health: a critical appraisal.* Am J Clin Nutr. 2009 Apr;89(4):1099-113. doi: https://doi.org/10.3945/ajcn.2008.2 6821.
- Delgado-Rodríguez M, Llorca J. *Bias.* Journal of Epidemiology & Community Health 2004;58:635-641.
- Institute of Medicine (IOM) Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease Board on Health Care Services Board on Health Sciences Policy Food and Nutrition Board, *Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease*, C.M. Micheel and J.R. Ball, Editors. Washington (DC): National Academies Press; 2010. Available from: https://iom.nationalacademies.org/Reports/2010/Evaluation-of-Biomarkers-and-Surrogate-Endpoints-in-Chronic-Disease.aspx
- National Academies of Sciences, Engineering, and Medicine. *Guiding principles for developing Dietary Reference Intakes based on chronic disease.* Washington, DC: The National Academies Press; 2017. doi: https://doi.org/10.17226/2482.
- National Academies of Sciences, Engineering, and Medicine. Methodological Considerations. In: *Dietary Reference Intakes for sodium and potassium.* Washington, DC: The National Academies Press, 2019:61-98. doi: https://doi.org/10.17226/25353.
- Dietary Assessment Primer, Summary Tables: Recommendations on Potential Approaches to Dietary Assessment for Different Research Objectives Requiring Group-level Estimates. National Institutes of Health, National Cancer Institute. Internet: https://dietassessmentprimer.cancer.gov/approach/table.html (accessed September 20, 2020).
- Scottish Intercollegiate Guidelines Network *(SIGN). A guideline developer's handbook.* Edinburgh: SIGN; 2019. (SIGN publication no. 50). [November 2019]. Available from URL: http://www.sign.ac.uk.

NUQUEST – NUtrition QUality Evaluation Strengthening Tool: Development of tools for the evaluation of risk of bias in nutrition studies

Shannon E. Kelly, Linda S. Greene-Finestone, Elizabeth A. Yetley, Karima Benkhedda, Stephen P.J. Brooks, George A. Wells, and Amanda J. MacFarlane

Online supplementary file 6: Blank NUQUEST worksheet

NUQUEST Worksheet

This worksheet is to be completed before using the NUQUEST checklist. There are three components: the key research question, the PICO framework, and the rating criteria to assist reviewers in assessing checklist items. These components provide focus for the reviews, guidance on selection of relevant and useful articles for review, and guidance for reviewers when using the NUQUEST checklist.

Key Research Question(s) of the review

The key research question(s) focuses the review. This serves as a basis for developing article selection inclusion/exclusion criteria via a PICO framework and rating criteria for checklist items.

Key Research Question	

PICO Framework

The PICO framework, developed by the review investigators before starting the reviews but after development of the key research question(s), provides a framework for determining the applicability of potential articles within the review context. Specifically, it guides decisions related to article selection or exclusion for subsequent review. Typically, acceptable criteria for four types of items are identified: **P**articipants/Population, **I**ntervention/Exposure, **C**omparator, and **O**utcome(s). Some reviews may add additional PICO items, such as timing (e.g., duration of acceptable studies), setting (e.g., outpatient clinic, home), or study design (e.g., parallel and cross-over trials with a specific washout period).

Participants/Population	
Intervention/Exposure	
Comparator	
Outcome(s)	

Examples of PICO characteristics:

- **Participants/Population**: Study population characteristics could include demographics (e.g., sex, age, racial/ethnic origin), economic and educational distributions, baseline nutritional status and clinical aspects (e.g., BMI, medication use, baseline chronic disease status, or biomarkers of outcomes).
- **Intervention/Exposure**: Amounts of food substances from supplements, diets, and/or nutritional status assessments; other relevant factors could include dosing conditions (e.g., with or without meals, bolus or continuous exposures), and/or form (e.g., synthetic vs. naturally-occurring).
- **Comparator**: Comparison groups could include higher vs. lower intakes or supplemented vs. non-supplemented intakes.
- **Outcome(s)**: These could include endpoints which are nutritional (e.g., biochemical markers of nutritional status, adequacy/inadequacy) or clinical (e.g. condition, risk factor or disease incidence or prevalence).

Rating criteria to assist reviewers in evaluating NUQUEST checklist items

1. List and rank the research objectives and their methods of assessing intake/nutritional status.

Research objectives	Method of assessment	Risk of Bias[1]

In addition to the elements discussed in the checklist guidance, other points to consider include:

- Evolving evidence has shown the importance of accurate and appropriate nutritional intake/status methodologies in assessing reported outcomes. When evaluating intake/status assessment methodologies, these points need to be addressed:
  - What is the accuracy for a specific nutrient? The criteria used to evaluate the accuracy of the intake of interest needs to be determined on a nutrient-by-nutrient basis. For example, for sodium intakes, a complete 24 hr urinary excretion collection with quality control measures is generally considered to be the gold standard while intakes based on self-reported methodologies tend to be less accurate (1). However, for potassium intakes, both 24-hr urine samples and 24-hr recall assessment methods provide reasonably accurate estimates of intake. The inclusion of quality controls is important because they can affect accuracy (1).
  - Data should be verified for incomplete or implausible nutritional intake data, which could lead to erroneous conclusions if not dealt with appropriately during the analysis (2).
  - How are the results for the intake/status reported and used in the study? While self-reported dietary intake assessment methodologies are commonly used for most nutrients, they are often susceptible to systematic bias and, therefore, caution is required when interpreting results. For a few nutrients, alternate methods such as 24h urinary collections provide more accurate intake measures. For example, nutrients like sodium are more accurately assessed when using a 24h urine collection (see the first bullet above and reference (1)) than when assessed by a 24 h dietary recall or FFQ. The most appropriate method and the appropriate number of measurements per participant will depend on the analysis being done. For example, for assessing the mean usual intake difference between two groups, a single 24 hr dietary recall or 24h urine collection on the whole sample (if validated for the nutrient of interest) may be appropriate. However, if the goal is to determine the change in the proportion of individuals above/below a threshold value between two points in time, then the best methodological approach would be *multiple* 24 hr recalls or 24 h urines on the whole sample at each point in time (3).
- The following adjustments may also need to be taken into account:
  - Energy intake adjustment in self-reported studies. Energy intake is often over- or under-reported and is dependent on body size and other factors. Because of the correlation between reported intake of energy and nutrients, this can result in bias. Adjustment for energy may reduce measurement error for nutrients that have been demonstrated to correlate with energy intakes. In these cases, the adjustment may be useful when evaluating diet-disease relationships;
  - Adjustment to correct for within-person variation (regression bias) using a repeated subsample of participants (i.e., in 24 hour diet recalls which estimate the proportion of a population above or below a cut-point); and
  - Adjustment for season or study population using weighting to ensure representativeness (2).

- Other factors are important when comparing reported methods. For example, has the method been validated for the nutrient of interest? If an adjustment or a weighting factor has been applied, is this reasonable and correctly applied (adjustments may not be equally valid for every nutrient)? If a measure is repeated, is there more confidence in the result? Does the method have higher/lower risk of bias when applied to an individual versus a group?

2. **List and rank the possible methods for assessing outcomes:**

Research objectives	Method of assessment	Risk of Bias[1]

In addition to the elements discussed in the checklist guidance, points to consider include:

- Consider the relative validity and reliability of the types of assessments that can be used to measure the outcome variable(s). Outcomes are commonly clinical (e.g., growth, blood pressure, LDL cholesterol) or nutritional (e.g., nutritional intake or status). When outcomes are nutrient intakes or status, consider the issues in Table 1 (intake/nutritional status).
- Similar to Table 1, it is important to consider how the outcome results are to be reported and used in the analysis of study results. For example, precise blood pressure values are important when the outcome is a measure of the percentage of individuals below/above a cut point.
- Other factors may be important when comparing methods. For example, has the method been validated for the outcome of interest? If an adjustment or a weighting factor was applied to the results, is this reasonable and correctly applied? If a measure is repeated, is there more confidence in the result? This could affect the relative weighting of the method. Does the method have a higher/lower risk of bias when applied to an individual versus a group?

3. **List key confounders (important and should be accounted for) for the specified outcome(s):**
A confounder is a variable that is related to both the outcome (e.g., systolic blood pressure) and the intake/exposure (e.g., sodium exposure) in individuals and therefore may distort (i.e., confound) the relationship. Generally, randomization in an RCT should balance confounders among the groups. Nevertheless, imbalances may arise and, if they exist, must be taken into account during analysis. List confounders that can result in higher risk of bias if not addressed in a study.

Key confounders	Method of assessment	Risk of Bias[1]

In addition to the elements discussed in the checklist guidance, points to consider include:

- If the confounder is an exposure, how often would this need to be assessed? For example, changes in background diet many occur in both the control and test group as a result of participating in the trial. These may not be consistent over the course of the trial. Risk of bias ratings for exposure in Table 1 can be useful but it is important to ensure that they apply to the confounder in question. It is important to remember that, as pointed out in the notes associated with Table 1, methods can be nutrient specific. For example, a method with a low risk of bias for assessing sodium exposure (group vs individual) may not have the same risk of bias when assessing potassium exposure.

- Has the method been validated? If a measure is repeated, is there more confidence in the result? This could affect the relative weighting of the method. Does the method have a higher/lower risk of bias when applied to an individual versus a group?
- If an adjustment or a weighting factor was applied to the results, is this reasonable and correctly applied?

Notes

[1]Risk of Bias levels for assessing methods of exposures/outcomes/confounders

Risk of Bias	Definition
Low	The method is reproducible with a high degree of accuracy.
Moderate	The method is generally reproducible with an acceptable degree of accuracy.
High	The method does not appear accurate, has a low reproducibility and/or a low degree of precision. Additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect

References

1. National Academies of Sciences E, and Medicine. Dietary Reference Intakes for Sodium and Potassium. Washington, DC, 2019.
2. Hörnell A, Berg C, Forsum E, Larsson C, Sonestedt E, Åkesson A, Lachat C, Hawwash D, Kolsteren P, Byrnes G, et al. Perspective: An Extension of the STROBE Statement for Observational Studies in Nutritional Epidemiology (STROBE-nut): Explanation and Elaboration. Advances in Nutrition 2017;8(5):652-78. doi: 10.3945/an.117.015941.
3. National Institutes of Health, National Cancer Institute. Internet: https://dietassessmentprimer.cancer.gov/approach/table.html (accessed September 20, 2020).

NUQUEST – NUtrition QUality Evaluation Strengthening Tool: Development of tools for the evaluation of risk of bias in nutrition studies

Shannon E. Kelly, Linda S. Greene-Finestone, Elizabeth A. Yetley, Karima Benkhedda, Stephen P.J. Brooks, George A. Wells, and Amanda J. MacFarlane

Online supplementary file 7: Example of completed NUQUEST worksheet

Example of Completed NUQUEST Worksheet

This worksheet is to be completed before using the NUQUEST checklist. There are three components: the key research question, the PICO framework, and the rating criteria to assist reviewers in assessing checklist items. These components provide focus for the reviews, guidance on selection of relevant and useful articles for review, and guidance for reviewers when using the NUQUEST checklist.

Key Research Question(s) of the review

The key research question(s) focuses the review. This serves as a basis for developing article selection inclusion/exclusion criteria via a PICO framework and rating criteria for checklist items.

Example: Relationship between sodium intakes and blood pressure.

Key Research Question	Among all adults, what is the relationship (benefits and harms) between dietary sodium intake and blood pressure?

PICO Framework

The PICO framework, developed by the review investigators before starting the reviews but after development of the key research question(s), provides a framework for determining the applicability of potential articles within the review context. Specifically, it guides decisions related to article selection or exclusion for subsequent review. Typically, acceptable criteria for four types of items are identified: **P**articipants/Population, **I**ntervention/Exposure, **C**omparator, and **O**utcome(s). Some reviews may add additional PICO items, such as timing (e.g., duration of acceptable studies), setting (e.g., outpatient clinic, home), or study design (e.g., parallel and cross-over trials with a specific washout period).

Example: Relationship between sodium intakes and blood pressure.

Participants/Population	All adults excluding persons with end stage renal disease, heart failure, HIV, or cancer.
Intervention/Exposure	Interventions to reduce dietary sodium (e.g., lower sodium diet or salt substitute).
Comparator	Higher dietary sodium intakes (e.g., usual diet) vs. lower sodium intakes (e.g., dietary salt reductions or use of salt substitutes).
Outcome(s)	Blood pressure (e.g., systolic and/or diastolic blood pressure, percent participants meeting a blood pressure goal, or changes in blood pressure).

Examples of PICO characteristics:

- **Participants/Population**: Study population characteristics could include demographics (e.g., sex, age, racial/ethnic origin), economic and educational distributions, baseline nutritional status and clinical aspects (e.g., BMI, medication use, baseline chronic disease status, or biomarkers of outcomes).
- **Intervention/Exposure**: Amounts of food substances from supplements, diets, and/or nutritional status assessments; other relevant factors could include dosing conditions (e.g., with or without meals, bolus or continuous exposures), and/or form (e.g., synthetic vs. naturally-occurring).
- **Comparator**: Comparison groups could include higher vs. lower intakes or supplemented vs. non-supplemented intakes.
- **Outcome(s)**: These could include endpoints which are nutritional (e.g., biochemical markers of nutritional status, adequacy/inadequacy) or clinical (e.g. condition, risk factor or disease incidence or prevalence).

Rating criteria to assist reviewers in evaluating NUQUEST checklist items

1. List and rank the research objectives and their methods of assessing intake/nutritional status.

Example: Usual Sodium Intake.

Research objectives	Method of assessment	Risk of Bias[1]
Determining difference between 2 points in time or between 2 groups	One or more 24 hr urine collections with quality control on whole sample at each point in time	Low
	One or more 24 hr urine collections without quality control on whole sample at each point in time	Moderate
	One or more 24 hr diet recalls on whole sample at each point in time	Moderate
	FFQ on whole sample at each point in time.	High
Determining change between 2 points in time or between 2 groups over time, in proportion of individuals above/below some threshold	Multiple 24 hr urine collections with quality control on whole sample at each point in time	Low
	Multiple 24 hr urine collections without quality control on whole sample at each point in time	Moderate
	Multiple 24 hr diet recalls on whole sample at each point in time	Moderate
	Single 24 hr recall on whole sample plus repeat on large subsample	Moderate
	FFQ or single 24 hr recall without repeat subsample	High
Examining the association between diet as an independent variable and a dependent variable (outcome)	Multiple 24 hr urine collections with quality control on whole sample at each point in time	Low
	Multiple 24 hr diet recalls on whole sample at each point in time	Moderate
	Single 24 hr recall on whole sample plus repeat on large subsample	Moderate

In addition to the elements discussed in the checklist guidance, other points to consider include:

- Evolving evidence has shown the importance of accurate and appropriate nutritional intake/status methodologies in assessing reported outcomes. When evaluating intake/status assessment methodologies, these points need to be addressed:
  - What is the accuracy for a specific nutrient? The criteria used to evaluate the accuracy of the intake of interest needs to be determined on a nutrient-by-nutrient basis. For example, for sodium intakes, a complete 24 hr urinary excretion collection with quality control measures is generally considered to be the gold standard while intakes based on self-reported methodologies tend to be less accurate (1). However, for potassium intakes, both 24-hr urine samples and 24-hr recall assessment methods provide reasonably accurate estimates of intake. The inclusion of quality controls is important because they can affect accuracy (1).
  - Data should be verified for incomplete or implausible nutritional intake data, which could lead to erroneous conclusions if not dealt with appropriately during the analysis (2).
  - How are the results for the intake/status reported and used in the study? While self-reported dietary intake assessment methodologies are commonly used for most nutrients, they are often susceptible to systematic bias and, therefore, caution is required when interpreting results. For a few nutrients, alternate methods such as 24h urinary collections provide more accurate intake measures. For example, nutrients like sodium are more accurately assessed when using a 24h urine collection (see the first bullet above and reference (1)) than when assessed by a 24 h dietary recall or FFQ. The most appropriate method and the appropriate number of measurements per participant will depend on the analysis being done. For example, for assessing the mean usual intake difference between two groups, a single 24 hr dietary recall or 24h urine collection on the whole sample (if validated for the nutrient of interest) may be appropriate. However, if the goal is to determine the change in the proportion of individuals above/below a threshold value between two points in time, then the

best methodological approach would be *multiple* 24 hr recalls or 24 h urines on the whole sample at each point in time (3).

- The following adjustments may also need to be taken into account:
  - Energy intake adjustment in self-reported studies. Energy intake is often over- or under-reported and is dependent on body size and other factors. Because of the correlation between reported intake of energy and nutrients, this can result in bias. Adjustment for energy may reduce measurement error for nutrients that have been demonstrated to correlate with energy intakes. In these cases, the adjustment may be useful when evaluating diet-disease relationships;
  - Adjustment to correct for within-person variation (regression bias) using a repeated subsample of participants (i.e., in 24 hour diet recalls which estimate the proportion of a population above or below a cut-point); and
  - Adjustment for season or study population using weighting to ensure representativeness (2).
- Other factors are important when comparing reported methods. For example, has the method been validated for the nutrient of interest? If an adjustment or a weighting factor has been applied, is this reasonable and correctly applied (adjustments may not be equally valid for every nutrient)? If a measure is repeated, is there more confidence in the result? Does the method have higher/lower risk of bias when applied to an individual versus a group?

2. List and rank the possible methods for assessing outcomes:

Example: Blood pressure measurements.

Research objectives	Method of assessment	Risk of Bias[1]
Systolic and diastolic blood pressure for classification of individuals	Multiple day, 3-repeat measure by trained personnel in clinic setting according to accepted and referenced guidelines.	Low
	Multiple day, 3-repeat measure at home with a validated automated sphygmomanometer with accepted guidelines (4).	Moderate
	Single measure at home with an automated sphygmomanometer without training or guidelines (5).	High

In addition to the elements discussed in the checklist guidance, points to consider include:

- Consider the relative validity and reliability of the types of assessments that can be used to measure the outcome variable(s). Outcomes are commonly clinical (e.g., growth, blood pressure, LDL cholesterol) or nutritional (e.g., nutritional intake or status). When outcomes are nutrient intakes or status, consider the issues in Table 1 (intake/nutritional status).
- Similar to Table 1, it is important to consider how the outcome results are to be reported and used in the analysis of study results. For example, precise blood pressure values are important when the outcome is a measure of the percentage of individuals below/above a cut point.
- Other factors may be important when comparing methods. For example, has the method been validated for the outcome of interest? If an adjustment or a weighting factor was applied to the results, is this reasonable and correctly applied? If a measure is repeated, is there more confidence in the result? This could affect the relative weighting of the method. Does the method have a higher/lower risk of bias when applied to an individual versus a group?

3. List key confounders (important and should be accounted for) for the specified outcome(s):

A confounder is a variable that is related to both the outcome (e.g., systolic blood pressure) and the intake/exposure (e.g., sodium exposure) in individuals and therefore may distort (i.e., confound) the relationship. Generally, randomization in an RCT should balance confounders among the groups. Nevertheless, imbalances may arise and, if they exist, must be taken into account during analysis. List confounders that can result in higher risk of bias if not addressed in a study.

Example: Confounders related to the relationship between sodium intakes and blood pressure.

Key confounders	Method of assessment	Risk of Bias[1]
Sex	Self-reported	Low
Age	Birth certificate (documented)	Low
Use of blood pressure medication	Self-reported	Moderate
Changes in sodium intakes unrelated to the intervention	Dietary assessment method conducted more than once during study (see Table 1 on intake/exposure methods)	Low to High
Pre-existing medical conditions (e.g., hypertension, conditions that affect the relationship between intakes and outcome)	Self-reported	Moderate
Potassium intake	Dietary assessment method conducted more than once during study (see Table 1 on intake/exposure methods)	Low to High
Physical activity	Self-reported	High
Physical activity	Activity monitor	Moderate

In addition to the elements discussed in the checklist guidance, points to consider include:

- If the confounder is an exposure, how often would this need to be assessed? For example, changes in background diet many occur in both the control and test group as a result of participating in the trial. These may not be consistent over the course of the trial. Risk of bias ratings for exposure in Table 1 can be useful but it is important to ensure that they apply to the confounder in question. It is important to remember that, as pointed out in the notes associated with Table 1, methods can be nutrient specific. For example, a method with a low risk of bias for assessing sodium exposure (group vs individual) may not have the same risk of bias when assessing potassium exposure.
- Has the method been validated? If a measure is repeated, is there more confidence in the result? This could affect the relative weighting of the method. Does the method have a higher/lower risk of bias when applied to an individual versus a group?
- If an adjustment or a weighting factor was applied to the results, is this reasonable and correctly applied?

Notes

[1]Risk of Bias levels for assessing methods of exposures/outcomes/confounders

Risk of Bias	Definition
Low	The method is reproducible with a high degree of accuracy.
Moderate	The method is generally reproducible with an acceptable degree of accuracy.
High	The method does not appear accurate, has a low reproducibility and/or a low degree of precision. Additional evidence is needed before concluding either that the findings are stable or that the estimate of effect is close to the true effect

References

1. National Academies of Sciences E, and Medicine. Dietary Reference Intakes for Sodium and Potassium. Washington, DC, 2019.
2. Hörnell A, Berg C, Forsum E, Larsson C, Sonestedt E, Åkesson A, Lachat C, Hawwash D, Kolsteren P, Byrnes G, et al. Perspective: An Extension of the STROBE Statement for Observational Studies in Nutritional Epidemiology (STROBE-nut): Explanation and Elaboration. Advances in Nutrition 2017;8(5):652-78. doi: 10.3945/an.117.015941.
3. National Institutes of Health, National Cancer Institute. Internet: https://dietassessmentprimer.cancer.gov/approach/table.html (accessed September 20, 2020).
4. Muntner P, Shimbo D, Carey RM, Charleston JB, Gaillard T, Misra S, Myers MG, Ogedegbe G, Schwartz JE, Townsend RR, et al. Measurement of Blood Pressure in Humans: A Scientific Statement From the American Heart Association. Hypertension 2019;73(5):e35-e66. doi: doi:10.1161/HYP.0000000000000087.

5. De León-Robert A, Gascón-Cánovas JJ, Antón-Botella JJ, Hidalgo-García IM, López-Alegría C, Pérez-Cabrera YD, Campusano-Castellanos HM. Validity of self blood pressure measurement in the control of the hypertensive patient: factors involved. BMC Cardiovascular Disorders 2019;19(1):171. doi: 10.1186/s12872-019-1145-9.

Chapter 4
Risk of bias in cross-sectional studies: A scoping review of concepts and tools

4.1. Chapter Overview

As a continuation of the development of nutrition-specific risk of bias tools intended for randomized controlled trials, cohort studies and case-control studies in Chapter 3, the focus in this chapter is on the foundational work for the development of the risk of bias tool in cross-sectional studies. This chapter contains two related manuscripts (Sections 4.2 and 4.3) that describe a scoping review of concepts and tools relevant to the assessment of risk of bias in cross-sectional studies and the related protocol. The title of each manuscript is listed below:

Section 4.2: *Risk of bias in cross-sectional studies: Protocol for a scoping review of concepts and tools*

Section 4.3: *A scoping review shows that no single existing risk of bias assessment tool considers all sources of bias for cross-sectional studies*

Chapter 4 addresses one thesis research question:

How can nutrition-specific risk of bias assessment tools be developed to address the unique methodological challenges posed by studies in nutritional epidemiology?

To address this question, tools available for risk of bias assessment in cross-sectional studies are investigated along with the common sources of bias unique to this study design. Similar to the foundational work in Chapter 3 for randomized controlled trials, cohort studies and case-control studies, this baseline review of tools, methodology and concepts is an essential first step in the

development of a valid and reliable nutrition-specific tool for cross-sectional studies. Unlike the initial NUQUEST research, there is no tool for cross-sectional studies available from the Scottish Intercollegiate Guidelines Consortium available for adaptation. As such, a scoping review was needed to identify RoB tools relevant to cross-sectional study designs that could be used to 'bolt on' nutrition-specific RoB considerations (See Chapter 5). The scoping review was also used to identify key biases thought to be relevant to cross-sectional research. An analytic framework was used to assess the comprehensiveness of each tool located in the scoping review.

4.2. Risk of bias in cross-sectional studies: Protocol for a scoping review of concepts and tools

4.2.1. Section Overview

This section presents a manuscript that describes the protocol for a scoping review aiming to identify and describe available tools to assess the risk of bias in cross-sectional studies and to compile the key bias concepts relevant to cross-sectional studies into an item bank.

4.2.2. Manuscript Status

This manuscript is published:

> **Kelly SE**, Benkhedda K, Brooks SP, MacFarlane AJ, Greene-Finestone LS, Skidmore B, Clifford TJ, Wells G.A. *Risk of Bias in Cross-Sectional Studies: Protocol for a Scoping Review of Concepts and Tools.* 2024. MethodsX. 2024 Feb 12;12:102610. doi: 10.1016/j.mex.2024.102610. PMID: 38371462; PMCID: PMC10874705.

4.2.3. Author Contributions

Author contributions for the manuscript are outlined below using the Contributor Roles Taxonomy (CRediT) author statement (1) as submitted with the manuscript.

Shannon E. Kelly: Conceptualization, Methodology, Validation, Writing – Original Draft, Writing – Review & Editing, Visualization. **Karima Benkhedda:** Conceptualization, Methodology, Writing – Review & Editing. **Stephen P. J. Brooks:** Conceptualization, Methodology, Writing – Review & Editing. **Amanda J. MacFarlane:** Conceptualization, Methodology, Writing – Review & Editing. **Linda S. Greene-Finestone:** Conceptualization, Methodology, Writing – Review & Editing. **Becky Skidmore:** Methodology (search), Validation (search), Writing – Original Draft (search), Writing – Review & Editing. **Tammy J. Clifford:** Supervision, Methodology, Writing – Review & Editing. **George A. Wells:** Conceptualization, Methodology, Supervision, Writing – Review & Editing.

4.2.4. Related Thesis Appendices

A copy of the published manuscript is presented in Appendix 3 along with associated permissions from Elsevier.

4.2.5. Manuscript

Risk of bias in cross-sectional studies: protocol for a scoping review of concepts and tools

Shannon E. Kelly*[1,2] (ORCID: 0000-0002-5520-6095)

Karima Benkhedda[3]

Stephen P. J. Brooks[3,4]

Amanda J. MacFarlane[4,5]

Linda S. Greene-Finestone[6]

Becky Skidmore[7]

Tammy J. Clifford[1]

George A. Wells[1,2]

Affiliations

1. School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada
2. Cardiovascular Research Methods Centre, University of Ottawa Heart Institute, Ottawa, Ontario, Canada
3. Bureau of Nutritional Sciences, Health Canada, Ottawa, Ontario, Canada
4. Department of Biology, Carleton University, Ottawa, Ontario, Canada
5. Agrilife Evidence Center at Texas A&M University, College Station, Texas, USA
6. Applied Research Division, Public Health Agency of Canada, Ottawa, Ontario, Canada
7. Independent contractor, Ottawa, Ontario, Canada

Corresponding author's email address: *redacted*

Keywords

risk of bias, cross-sectional, scoping review, tool, methodological quality, critical appraisal

Abstract[1]

Cross-sectional studies are commonly used to study human health and disease, but are especially susceptible to bias. This scoping review aims to identify and describe available tools to assess the risk of bias (RoB) in cross-sectional studies and to compile the key bias concepts relevant to cross-sectional studies into an item bank. Using the JBI scoping review methodology, the strategy to locate relevant RoB concepts and tools is a combination of database searches, prospective review of PROSPERO registry records; and consultation with knowledge users and content experts. English language records will be included if they describe tools, checklists, or instruments which describe or permit assessment of RoB for cross-sectional studies. Systematic reviews will be included if they consider eligible RoB tools or use RoB tools for RoB of cross-sectional studies. All records will be independently screened, selected, and extracted by one researcher and checked by a second. An analytic framework will be used to structure the extraction of data. Results for the scoping review are pending. Results from this scoping review will be used to inform future selection of RoB tools and to consider whether development of a new RoB tool for cross-sectional studies is needed.

Specifications[2]**:**

[1] Note: An unstructured abstract was required by the journal.

[2] Note: All items reported under the *Specifications* heading were required by the journal as part of the submission.

Subject area: Medicine and Dentistry

More specific subject area: Epidemiology, knowledge synthesis methods, study designs

Name of your protocol: Risk of bias in cross-sectional studies: protocol for a scoping review of concepts and tools

Reagents/tools: Not applicable

Experimental design: Not applicable. This protocol describes a scoping review

Trial registration: Not applicable. This is not a trial. This scoping review is registered with the Open Science Framework (https://osf.io/y7vst/)

Ethics: This work did not involve human subjects, animal experiments or data collected from social media platforms.

Value of the Protocol/Scientific contribution:

- Comprehensive summary of different methodological considerations for risk of bias in cross-sectional study designs.
- Maps coverage for key bias considerations in existing risk of bias tools.
- Identified gaps may be used to select tools for use and/or to inform future risk of bias tool development.

Description of protocol

A cross-sectional study is a type of observational research study that is used to examine a group of individuals at a specific point in time in order to understand their characteristics and

behaviors. The objective of this scoping review is to identify tools used to evaluate the risk of bias (RoB) in cross-sectional studies of humans; and to identify, describe, and to compile the key items, domains or concepts relevant to RoB in cross-sectional studies into an item bank.

Methods

The proposed scoping review will be conducted in accordance with the JBI methodology for scoping reviews and reported in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR)(2, 3).

Review questions

1) What tools are used to assess the risk of bias in cross-sectional studies?
2) What risk of bias items, domains or concepts are included in tools intended for cross-sectional study designs?
3) How do risk of bias concepts vary when tools are intended to be applied to analytic studies of association or prevalence?

Inclusion criteria

The Cochrane definition will be used to consider RoB as a measure of how features of the design, conduct or analysis of a study may cause systematic error in the results of a study; That is, whether the prevalence or associated effects for an outcome in a cross-sectional study may be over- or under-estimated. Since tools may be described inconsistently using language on methodological quality, quality tools will be eligible if they are judged by reviewers to contain one or more items relating to internal validity of the study.

This review will consider records for inclusion if they meet the following criteria:

- Permit the assessment of RoB in human observational studies defined as cross-sectional studies, including analytic or prevalent/survey approaches;
- Describe or permit global assessment of RoB for multiple study designs and originally developed for application to cross-sectional study designs; or,
- Describe the application of a tool not originally designed for application to cross-sectional study designs which has been adapted or modified for application to cross-sectional study designs for the assessment of RoB; and,
- Published or available in English.

“Tools” may be described as checklists, scored instruments, ‘question/item-based‘ or 'domain-based' according to groupings or categories of bias. Systematic reviews of RoB tools will also be included if they endeavoured to locate and/or include relevant tools/items/concepts for cross-sectional studies.

Records identified will be excluded if they:

- Report tools intended for other study designs and are not intended for evaluation of cross-sectional study designs;
- Report single-study or multi-study tools modified for application to cross-sectional studies in a systematic review and the modified tool is not included or cited.

Types of sources

This scoping review will consider records of any study design for inclusion. As there is a limited number of RoB tools published in peer-reviewed literature, four approaches will be used to search for and locate eligible RoB tools relevant to cross-sectional studies:

1. **Bibliographic database search:** This approach will be used to identify published RoB tools and relevant systematic reviews of RoB tools. For the search of bibliographic databases, an experienced medical information specialist developed and tested search strategies through an iterative process in consultation with the review team.

2. **Prospective scan of PROSPERO registrations:** This approach will be used to identify RoB tools identified in 1,000 prospectively registered systematic review protocols in the PROSPERO database. In order to identify currently used tools, all records will be searched for proposals including the term "cross-sectional" and systematic reviews indicating the proposed inclusion of cross-sectional studies in the analysis will be retained. Applicable RoB tools will be identified from the details in the "Risk of bias (quality) assessment" section of the registration.

3. **Outreach to key knowledge users:** This approach will be used to gather information regarding RoB tools being applied in practice for cross-sectional studies. A group of users in the field of nutrition will be selected as they commonly use or assess cross-sectional research. The email-based outreach will briefly ask participants to reply if they are willing to share a citation or the name of a relevant RoB tool they use or are aware of.

4. **Experiential knowledge of content experts:** This approach will supplement the other approaches by collecting the experiential knowledge of key content experts to identify additional tools not represented they are aware of.

Bibliographic database search strategy

An experienced medical information specialist developed and tested the search strategies through an iterative process in consultation with the review team. A full search strategy for OvidMEDLINE® ALL, including Epub Ahead of Print, In-Process & Other Non-Indexed Citations, and Embase Classic+Embase will be searched on the Ovid platform, followed by a search of Web of Science (see Appendix I). The strategies utilize a combination of controlled vocabulary (e.g.,"Cross-Sectional Studies", "Bias","Checklists") and keywords (e.g.,"cross-sectional design", "quality rating","tool"). The search strategy will be adapted for each included

information source. The search strategy will aim to locate both published and unpublished tools. Results will be limited to the publication years 2011 to the present. Results will be downloaded and deduplicated using EndNote version 20.4.1 (Clarivate Analytics, PA, USA) and uploaded to Covidence systematic review software for screening (Veritas Health Innovation, Melbourne, Australia. Available at www.covidence.org). The database search will be updated prior to completion of the scoping review.

Evidence selection

Following the search, all identified database records will be collated and uploaded into Covidence systematic review software (Veritas Health Innovation, Melbourne, Australia. Available at www.covidence.org) and EndNote and duplicates removed. Following a pilot test, titles and abstracts will then be screened by two independent reviewers for assessment against the inclusion criteria for the review. Potentially relevant records will be retrieved in full and tracked through Covidence. The full text of selected citations will be assessed in detail against the inclusion criteria by one reviewer and included records will be confirmed by a second independent reviewer.

The included study lists of all eligible systematic reviews will be screened to identify potentially relevant RoB tools. All records identified through systematic reviews, the environmental scan, expert working group and PROSPERO registry search will be retrieved in full, assessed in detail against the inclusion criteria by one reviewer and confirmed by a second independent reviewer. All non-database records will be tracked using an MS Excel spreadsheet (Microsoft Corp. (2018).

Following a detailed assessment, a super-set of unique RoB tools relevant to cross-sectional studies will be created once de-duplication of records and tools across the approaches is complete.

Reasons for exclusion of full-text records that do not meet the inclusion criteria will be recorded and reported in the scoping review. Any disagreements that arise between the reviewers at each stage of the selection process will be resolved through discussion or with a third reviewer. The results of the search will be reported in full in the final scoping review and presented in a PRISMA flow diagram (4).

Data extraction

Data will be extracted from tools included in the scoping review by one reviewer and checked by a second reviewer. The data extracted will focus on descriptive information, characteristics of the tool, items and domains used to evaluate risk of bias and any psychometric properties reported, including:

- citation, tool name;
- reported study design(s) intended for use/application (single or multiple study designs);
- tool development and testing: reported methods used to develop the tool, reported reliability, validity or usability testing, including whether the tool was developed or tested in a particular clinical area;
- characteristics of the tool including but not limited to: use of domains/items/rating probes, format for completing the item or overall rating, whether rating summaries of quality or bias are used (e.g. poor/good/moderate) or if scoring is calculated;
- discrimination between concepts of quality, risk of bias, and reporting quality;

- how and why an existing tool was modified for application to cross-sectional studies; and,
- declared organization supporting the tool development, any sponsors or reported funding.

Data extraction will be standardized using a framework reported for RoB tools by the National Toxicology Program of the United States Department of Health and Human Services (5). The extraction framework focuses on key bias domains (*selection, exposure, outcome, confounding, loss-to-follow-up, analysis, selective reporting, conflict of interest and other*), the topic(s) within each domain, the questions and available response options. The extraction framework will be adapted if necessary for considerations specific to cross-sectional studies during a pilot to assess the feasibility of the process for extracting data from each included record. Any modifications will be detailed in the full scoping review. Any disagreements that arise between the reviewers will be resolved through discussion or with a third reviewer. All extracted information will be compiled into detailed evidence tables in Microsoft Excel.

Data analysis and presentation

Results will be presented in tables and figures (if applicable) and described in detail. A narrative summary will accompany the charted results and will describe how the results relate to the scoping review objective and questions. The final evidence product will be an item bank of risk of bias concepts relevant to cross-sectional studies.

CRediT author statement

Shannon E. Kelly: Conceptualization, Methodology, Validation, Writing – Original Draft, Writing – Review & Editing, Visualization Karima Benkhedda: Conceptualization, Methodology, Writing – Review & Editing Stephen P. J. Brooks: Conceptualization,

Methodology, Writing – Review & Editing Amanda J. MacFarlane: Conceptualization, Methodology, Writing – Review & Editing Linda S. Greene-Finestone: Conceptualization, Methodology, Writing – Review & Editing Becky Skidmore: Methodology (search), Validation (search), Writing – Original Draft (search), Writing – Review & Editing Tammy J. Clifford: Supervision, Methodology, Writing – Review & Editing George A. Wells: Conceptualization, Methodology, Supervision, Writing – Review & Editing

Acknowledgments

The review team would like to acknowledge Beth Yetley for her early involvement and for help with conceptualization of the work.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Additional information[3]

[3] This journal includes all background and introductory material after the protocol methods and other content.

In a cross-sectional study, a sample of individuals is selected from a larger population and data is collected from each individual in the sample. This data can then be analyzed to draw conclusions about the population as a whole (6). Researchers may use cross-sectional studies for population-based surveys, to estimate the prevalence of outcomes and/or intervention/exposures or to measure association between the intervention/exposure and outcome by calculating an odds ratio (6, 7).

The strengths of this design are that studies can be conducted relatively more resource efficiently when compared to other study designs, they can provide information that can be used to plan subsequent research studies (hypothesis generating), as well as provide data to support planning, monitoring and evaluation of health services or public health programs (7). They often employ survey questionnaires which can be quick and simple approaches to collect prevalence data (8). At the population level, the resulting prevalence estimates can be informative and generalizable. However, there are several limitations associated with cross-sectional studies. Causal relationships cannot be derived from a one-time measurement of intervention/exposure and outcome and there are no temporal elements within the design to strengthen estimates or sufficiently determine trends (7, 8). It can be difficult to plan and conduct a high-quality cross-sectional study as this design is prone to real or potential bias related to difficulties obtaining representative samples, inaccuracies in collected information, unclear temporal relationships and the possibility of reverse causality. Like other observational study designs, a major limitation is the inability to account for unobservable confounding when associations of interest are identified (7, 9-11).

When trying to answer research questions, knowledge users may use or synthesize evidence from cross-sectional studies. Including a risk-of-bias (RoB) assessment is crucial to understanding

relationships among interventions, exposures and the parameters of interest, and interpreting study findings. Although several tools exist for assessing the RoB in cross-sectional studies, there is no consensus on a 'gold standard' tool.

Recent reviews of RoB tools have shown that existing tools may apply to one or more study design, but vary in their approach to scoring or rating, the included RoB concepts, and their response options, and variably include additional items related to generalizability, precision or reporting (12, 13). Previous research concluded that many RoB tools are developed explicitly for experimental research and do not adequately address key biases common in observational research (14, 15). Many tools developed broadly for multiple observational study designs are not specifically applicable to cross-sectional studies or fail to sufficiently assess important aspects of bias common in cross-sectional studies (15-17). While multi-study RoB tools cover a breadth of methodological bias concepts, using design-specific RoB tools can help ensure that all relevant sources of bias are considered and evaluated appropriately. These design-specific tools may be more applicable, user-friendly and easier to interpret where there is no control group present (16).

References

1. Brand A, Allen L, Altman M, Hlava M, Scott J. *Beyond authorship: attribution, contribution, collaboration, and credit*. Learned Publishing. 2015;28(2):151-155.

2. McGowan J, Straus S, Moher D, Langlois EV, O'Brien KK, Horsley T, Aldcroft A, Zarin W, Garitty CM, Hempel S, Lillie E, Tuncalp, Tricco AC. *Reporting scoping reviews-PRISMA ScR extension*. J Clin Epidemiol. 2020 Jul;123:177-179. Epub 20200327. doi:10.1016/j.jclinepi.2020.03.016. Cited in: Pubmed; PMID 32229248.

3. Aromataris E MZ, (Editors). JBI Manual for Evidence Synthesis [internet]. Adelaide:: JBI; 2020 June 20, 2022. Available from: https://synthesismanual.jbi.global.

4. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hrobjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J,

Tricco AC, Welch VA, Whiting P, Moher D. *The PRISMA 2020 statement: An updated guideline for reporting systematic reviews*. Int J Surg. 2021 Apr;88:105906. Epub 20210329. doi:10.1016/j.ijsu.2021.105906. Cited in: Pubmed; PMID 33789826.

5. Risk of Bias Tools. [internet]: National Toxicology Program of the United States Department of Health and Human Services; 2020 June 13, 2022. Available from: https://ntp.niehs.nih.gov/whatwestudy/assessments/noncancer/tools/index1.html?utm_source=direct&utm_medium=prod&utm_campaign=ntpgolinks&utm_term=ohat_tools.

6. Reeves BC, Wells GA, Waddington H. *Quasi-experimental study designs series—paper 5: a checklist for classifying studies evaluating the effects on health interventions—a taxonomy without labels*. Journal of Clinical Epidemiology. 2017 2017/09/01/;89:30-42. doi:https://doi.org/10.1016/j.jclinepi.2017.02.016.

7. Setia MS. *Methodology Series Module 3: Cross-sectional Studies*. Indian J Dermatol. 2016 May-Jun;61(3):261-4. eng. doi:10.4103/0019-5154.182410. Cited in: Pubmed; PMID 27293245.

8. Van der Stede WA. *A manipulationist view of causality in cross-sectional survey research*. Accounting, Organizations and Society. 2014;39(7):567-574.

9. Yu I, Tse S. *Workshop 6--sources of bias in cross-sectional studies; summary on sources of bias for different study designs*. Hong Kong Medical Journal= Xianggang yi xue za zhi. 2012;18(3):226-227.

10. Sedgwick P. *Bias in observational study designs: cross sectional studies*. BMJ : British Medical Journal (Online). 2015 2015 Mar 06

2016-10-31;350. English. doi:https://doi.org/10.1136/bmj.h1286. Cited in: Pubmed; PMID 1777762032.

11. Sedgwick P. *Cross sectional studies: advantages and disadvantages*. BMJ : British Medical Journal (Online). 2014 2014 Mar 26

2017-10-05;348. English. doi:https://doi.org/10.1136/bmj.g2276. Cited in: Pubmed; PMID 1777773341.

12. Downes MJ, Brennan ML, Williams HC, Dean RS. *Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS)*. BMJ Open. 2016 Dec 8;6(12):e011458. Epub 20161208. doi:10.1136/bmjopen-2016-011458. Cited in: Pubmed; PMID 27932337.

13. Wang Z, Taylor, K., Allman-Farinelli, M., Armstrong, B., Askie, L., Ghersi, D., McKenzie, J., Norris, S.L., Page, M., Rooney, A., Woodruff, T. & Bero, L. . Systematic review: Tools for assessing methodological quality of human observational studies. [Online]. Available from: osf.io/preprints/metaarxiv/pnqmy. 2019.

14. Mallen C, Peat G, Croft P. *Quality assessment of observational studies is not commonplace in systematic reviews*. J Clin Epidemiol. 2006 Aug;59(8):765-9. eng. Epub 20060530. doi:10.1016/j.jclinepi.2005.12.010. Cited in: Pubmed; PMID 16828667.

15. Shamliyan T, Kane RL, Dickinson S. *A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases*. J Clin Epidemiol. 2010 Oct;63(10):1061-70. doi:10.1016/j.jclinepi.2010.04.014. Cited in: Pubmed; PMID 20728045.

16. Glasgow MJ, Edlin R, Harding JE. *Comparison of risk-of-bias assessment approaches for selection of studies reporting prevalence for economic analyses*. BMJ Open. 2020;10(9):e037324. doi:10.1136/bmjopen-2020-037324.

17. Moskalewicz A, Oremus M. *No clear choice between Newcastle-Ottawa Scale and Appraisal Tool for Cross-Sectional Studies to assess methodological quality in cross-sectional studies of health-related quality of life and breast cancer*. J Clin Epidemiol. 2020 Apr;120:94-103. Epub 20191219. doi:10.1016/j.jclinepi.2019.12.013. Cited in: Pubmed; PMID 31866469.

18. Wang X, Cheng Z. *Cross-sectional studies: strengths, weaknesses, and recommendations*. Chest. 2020;158(1):S65-S71.

19. Barker TH, Stone JC, Sears K, Klugar M, Leonardi-Bee J, Tufanaru C, Aromataris E, Munn Z. *Revising the JBI quantitative critical appraisal tools to improve their applicability: an overview of methods and the development process*. JBI Evidence Synthesis. 2023;21(3):478-493. doi:10.11124/jbies-22-00125. Cited in: Pubmed; PMID 02174543-202303000-00004.

Online supplementary material: Search strategy for MEDLINE

Database: Ovid MEDLINE(R) ALL <1946 to November 17, 2021>

---

1 Cross-Sectional Studies/ (398684)
2 ((cross-section* or crosssection*) adj2 (study or studies or design?)).ti,kw,kf. (48153)
3 1 or 2 [CROSS-SECTIONAL STUDIES] (412208)
4 Cross-Sectional Studies/st [standards] (41)
5 Research Design/st [standards] (12411)
6 Evidence-Based Medicine/mt, st [methods,standards] (9222)
7 or/4-6 [RESEARCH DESIGN STANDARDS] (21297)
8 3 and 7 [CROSS-SECTIONAL STUDIES - RESEARCH DESIGN STANDARDS] (320)
9 exp bias/ (72351)
10 bias$2.ti,kw,kf. (38345)
11 ((assess* or apprais* or estimat* or grade? or grading) adj3 (bias* or quality)).tw,kw,kf. (118427)
12 ((susceptib* or risk?) adj3 bias*).tw,kw,kf. (33354)
13 ((rate or rated or rates or rating) adj2 quality).tw,kw,kf. (7834)
14 Research Design/ (116838)
15 (research adj3 (appropriate* or design? or method* or protocol? or strateg* or valid*)).tw,kw,kf. (132941)
16 (methodological* adj3 (appropriate* or consider* or decision? or design? or issue? or quality or valid*)).tw,kw,kf. (33428)
17 ((design? or report* or study or studies) adj3 (appropriate* or quality or valid*)).tw,kw,kf. (160543)
18 "Reproducibility of Results"/ (431267)
19 reproducib*.ti,kw,kf. (18259)
20 (reliabilit* or reliabl*).ti,kw,kf. (60000)
21 (replicable or replicat*).ti,kw,kf. (58248)
22 valid*.ti,kw,kf. (142161)
23 confound*.ti,kw,kf. (5461)
24 Data Accuracy/ (3315)
25 ((calculat* or data or measur* or statistic*) adj3 (accurat* or accurac* or appropriate* or objective)).tw,kw,kf. (141906)
26 ((calculat* or data or measur* or statistic*) adj3 (correct* or exact* or precis*)).tw,kw,kf. (50759)
27 or/9-26 [QUALITY] (1285786)
28 3 and 27 [CSS - QUALITY] (38988)
29 Checklists/ (7539)
30 (checklist* or check list*).ti,kw,kf. (8020)
31 (tool or tools).ti,kw,kf. (100645)

32 (instrument or instruments).ti,kw,kf. (37759)
33 (framework? or frame work?).ti,kw,kf. (58639)
34 ((assess* or apprais* or grade? or grading or quality) adj3 (criteri* or guide or guideline? or guidance? or standard?)).ti,kw,kf. (7910)
35 or/29-34 [TOOLS/FRAMEWORKS] (215219)
36 28 and 35 [CSS - QUALITY ASPECTS - TOOLS] (1844)
37 8 or 36 [CSS - QUALITY ASPECTS - TOOLS/STANDARDS] (2136)
38 limit 37 to yr="2010-current" (1713)

Graphical Abstract

4.3. Risk of bias in cross-sectional studies: A scoping review of concepts and tools

4.3.1. Section Overview

This section presents a manuscript that describes the findings of the scoping review describing methods, concepts or tools used to consider risk of bias in health research reported to be descriptive/prevalence survey or analytic/association (cross-sectional) study designs. The protocol described in Section 4.2 was followed to consider and describe the references included. The manuscript describes the state of the evidence current to November 21, 2021.

4.3.2. Manuscript Status

This manuscript is published:

> **Kelly SE**, Brooks SP, Benkhedda K, MacFarlane AJ, Greene-Finestone LS, Skidmore B, Clifford TJ, Wells G.A. *A scoping review shows that no single existing risk of bias assessment tool considers all sources of bias for cross-sectional studies.* J Clin Epidemiol. 2024 Aug;172:111408. doi: 10.1016/j.jclinepi.2024.111408. Epub 2024 Jun 4. PMID: 38844117.

4.3.3. Author Contributions

Author contributions for the manuscript are outlined below using the Contributor Roles Taxonomy (CRediT) author statement (1).

Shannon E. Kelly: Conceptualization, Methodology, Investigation, Validation, Project administration, Writing – Original Draft, Writing – Review & Editing, Formal analysis, Visualization. **Karima Benkhedda:** Conceptualization, Methodology, Writing – Review & Editing. **Stephen P. J. Brooks:** Formal analysis, Methodology, Investigation, Writing – Review & Editing. **Amanda J. MacFarlane:** Conceptualization, Resources, Methodology,

Writing – Review & Editing. **Linda S. Greene-Finestone:** Conceptualization, Methodology, Investigation, Writing – Review & Editing. **Becky Skidmore:** Methodology (search), Investigation (search), Validation (search), Writing – Original Draft (search), Writing – Review & Editing. **Tammy J. Clifford:** Supervision, Methodology, Writing – Review & Editing. **George A. Wells:** Conceptualization, Resources, Methodology, Supervision, Writing – Review & Editing.

4.3.4. Related Thesis Appendices

A copy of the published manuscript is presented in Appendix 3 along with associated permissions from Elsevier.

4.3.5. Manuscript

A scoping review shows that no single existing risk of bias assessment tool considers all sources of bias for cross-sectional studies

Shannon E. Kelly[1,2] (ORCID: 0000-0002-5520-6095)*Corresponding author

Stephen P. J. Brooks[3]

Karima Benkhedda[3]

Amanda J. MacFarlane[4,5]

Linda S. Greene-Finestone[6]

Becky Skidmore[7]

Tammy J. Clifford[1]

George A. Wells[1,2]

1. School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada

2. Cardiovascular Research Methods Centre, University of Ottawa Heart Institute, Ottawa, Ontario,

Canada

3. Bureau of Nutritional Sciences, Health Canada, Ottawa, Ontario, Canada, retired.

4. Department of Biology, Carleton University, Ottawa, Ontario, Canada

5. Texas A&M Agriculture, Food, and Nutrition Evidence Center, Fort Worth, Texas, USA

6. Applied Research Division, Public Health Agency of Canada, Ottawa, Ontario, Canada

7. Independent contractor, Ottawa, Ontario, Canada

Abstract

Objective

Different tools to assess the potential risk of bias (RoB) for cross-sectional studies have been developed, but it is unclear whether all pertinent bias concepts are addressed. We aimed to identify RoB concepts applicable to cross-sectional research validity and to explore coverage for each in existing appraisal tools.

Study design and setting

This scoping review followed the Joanna Briggs Institute methodology. We included records of any study design describing or reporting methods, concepts or tools used to consider RoB in health research reported to be descriptive/ prevalence survey or analytic/association (cross-sectional) study designs. Synthesis included quantitative and qualitative analysis.

Results

Of the 4,556 records screened, 90 were selected for inclusion; 67 (74%) described the development of, or validation process for, appraisal tools, 15 (17%) described methodological content or theory relevant to RoB for cross-sectional studies and 8 (9%) records of methodological systematic reviews. Review of methodological reports identified important RoB concepts for both descriptive/prevalence and analytic/association studies. Tools identified (n=64 unique tools) were either intended to appraise quality or assess RoB in multiple study designs including cross-sectional studies (n=21; 33%) or cross-sectional designs alone (n=43; 67%). Several existing tools were modified (n=17; 27%) for application to cross-sectional studies. The RoB items most frequently addressed in the RoB tools were validity and reliability of the

exposure (53%) or outcome (65%) measurement and representativeness of the study population (59%). Most tools did not consider non-response or missingness appropriately or at all.

Conclusion

Assessing cross-sectional studies involves unique risk of bias (RoB) considerations. We identified RoB tools designed for broad applicability across various study designs as well as those specifically tailored for crosssectional studies. However, none of the identified tools comprehensively address all potential biases pertinent to cross-sectional studies. Our findings indicate a need for continued improvement of RoB tools and suggest that the development of context-specific or more precise tools for this study design may be necessary.

Plain Language Summary

What is the problem?

Cross-sectional research looks at a group of people at one specific point in time to understand how common certain traits or behaviours are and to find connections between different factors, and often, health or disease. These studies provide a snapshot at one time point but do not show changes over time. Results of these studies are often used to provide initial insights that can inform important healthcare-related decisions by patients, policy makers, researchers and health providers.

We rely on these studies to be well conducted and trustworthy, but there may be problems with how these studies are designed and conducted that may influence the results. These problems are referred to as biases. There are tools designed to help researchers identify possible biases in cross-sectional studies which may influence the results, but it's not clear if these tools cover all the important issues, some of which may be unique to these types of studies.

What did we do?

We searched for and identified tools that could be used to evaluate possibles biases in cross-sectional studies and articles that described potential biases in this type of research. We summarize the biases that tools should be considering and assessed how well each cross-sectional study tool did this.

What did we find?

We found 67 articles reporting tools used to check for biases in cross-sectional studies, 15 articles that summarized the biases important to cross-sectional research and eight articles that were reviews of related methods or biases. We found that some of the tools were made for assessing many different types of studies, while others were developed specifically for cross-sectional studies. In the end, we found that none of the tools we identified covered all the possible biases that could influence the results of a cross-sectional study away from the 'truth'. These findings show that researchers need to improve these tools or develop new tools that include all biases that may impact cross-sectional studies.

Word count: 3,416

Figures: 1

Tables: 4

Running Title: Risk of bias in cross-sectional studies

Supplemental content: Yes

Keywords: *scoping review; risk of bias; tool; research methodology; cross-sectional study; prevalence.*

Background

Cross-sectional studies are a common epidemiological research design used to investigate the prevalence of disease, characteristics or the distribution of outcomes in populations. They may also be used to explore analytic associations between health exposures and disease outcomes but cannot provide data for causality. [1] These study designs give a 'snapshot' of a health variable (or any other factor) in a given population at one point in time, which can provide insight into the occurrence of disease, health knowledge, risk factors, patterns or behaviours. As with any research study, the robustness of the findings is contingent on minimizing factors that may compromise the validity, reliability and generalizability of the results. [2]

Bias, or systematic errors in the study design, conduct or analysis, can impact cross-sectional and prevalence studies in unique ways.[3-5] For example, selection bias is a threat to the validity of cross-sectional and prevalence studies. As such, sampling strategies must be carefully designed so as not to distort estimates of exposure-outcome relationships or prevalence rates and to maintain representativeness to a target population. Cross-sectional studies are also particularly susceptible to information bias, especially inaccuracies in the measurement of exposures or outcomes. Related to information bias, it is difficult or impossible to establish a temporal sequence for the variables under study which is a requirement for any epidemiological consideration of causation. Cross-sectional data collection inherently limits our ability to discern whether an exposure precedes an outcome or vice versa.(18) These challenges lead to cross-

sectional and survey designs being very low on any ‘hierarchy of evidence’ use for evidence-based decision-making.[6]

Despite this, cross-sectional studies offer significant advantages to researchers through efficient and low-resource data collection, and therefore, are well-utilized in many disciplines. Likewise, knowledge users often rely on cross-sectional research studies to inform planning, policy, practice or other research activities. Taking benefits and limitations into consideration, it is critical to accurately identify reliable, valid and trustworthy cross-sectional research. Structured tools or instruments have been created to assist with discernment of study quality or risk of bias (RoB) and each may be applicable to one or more epidemiologic study designs. However, there are many tools available, and it is unclear how well existing tools consider all potential biases for cross-sectional studies. This scoping review aims to identify and describe available tools to assess the RoB in cross-sectional studies and to compile the critical RoB items, domains, concepts applicable to cross-sectional studies into an item bank. A secondary objective was to determine the extent to which the retrieved RoB tools address the relevant bias items, domains and concepts identified.

Materials and Methods

The scoping review was conducted using methodology developed by Arksey and O’Malley [7] and later refined by the Joanna Briggs Institute (JBI) [8]. It is reported using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR; Supplemental Appendix, Table A) [9]. A protocol is published elsewhere.[10]

Information sources

We searched for MEDLINE, Embase and Web of Science for potentially relevant tools, methodological systematic reviews, systematic reviews of cross-sectional studies and peer-

reviewed methodological guidance specific to cross-sectional research on November 18, 2021, and updated the search on August 6, 2023. The strategies used a combination of controlled vocabulary (e.g., “Cross-Sectional Studies”, “Bias”, “Checklists”) and keywords (e.g., “cross-sectional design”, “quality rating”, “tool”). Vocabulary and syntax were adjusted across the databases. To supplement the bibliographic database search, we consulted methodological and epidemiological experts and reviewed ROB tools identified in systematic review protocols registered through The International Prospective Register of Systematic Reviews (PROSPERO). Additional details regarding the strategies appear in the Supplemental Appendix.

Eligibility criteria

This is a methodological scoping review and as such, we present the review in terms of the Population, Concept, Context (PCC) framework based on guidance from JBI.[8]

Population: Not applicable

Concept: The concepts of interest are the potential biases that are relevant to research studies employing a cross-sectional design, and whether corresponding appraisal tools address these biases sufficiently.

Context: All contexts (including setting, field and discipline) will be considered.

We sought to identify tools intended to be used to appraise RoB in cross-sectional study designs. We considered RoB appraisal to be a measure of how features of the design, conduct of a study may produce systematic error in the results of a study. As terminology for methodological quality and RoB are commonly used interchangeably and inconsistently, tools for critical appraisal and quality assessment were eligible if they contained one or more items relating to the internal validity of a study.[11] Tools could be checklists (i.e., users are asked to consider whether a study meets a predetermined set of items, usually criteria, signalling questions or probes) or scales (i.e.,

instruments structured to provide quantitative scores for each item presented for appraising research) [12](19). Tools were eligible if they were intended for assessment of: (1) observational or descriptive studies defined as cross-sectional studies, including analytic/association or prevalence/descriptive approaches; (2) multiple study designs, including cross-sectional study designs; or (3) study designs other than cross-sectional studies but were adapted or modified for application to cross-sectional studies. We excluded records published in languages other than English and tools that were not intended or modified to be applied to cross-sectional studies. Records for systematic reviews that did not cite the RoB tool used or which adapted an existing RoB tool for application to cross-sectional studies but did not make the tool available, were excluded. Methodological records were eligible if they described concepts, items, domains or tools relevant to appraisal of RoB in cross-sectional research. Methodological concepts are defined as constructs related to the validity of cross-sectional research.

Selection process

Following a pilot exercise (n=100 records), the titles and abstract for each retrieved record were screened by one reviewer (SK). A second reviewer (LGF) vetted the included and excluded records. The full text of all potentially relevant records was retrieved and assessed by one reviewer (SK). All included records were confirmed through discussion and validation with the review team. The included records from all eligible systematic reviews, PROSPERO registrations and methodological records were checked for potentially relevant RoB tools (SK, SB).

Data items and collection process

One reviewer (SK) extracted data from each identified tool, including details of the development process, structure, included items and domains, scoring or evaluation system and any reported

processes for validation. We classified tools as checklists or scales, by their self-identified use, and their intended study design application. Information was extracted using standardized Google forms (Google Inc., 2023) and then transferred to Microsoft Excel (Microsoft Corp.,2018).

Methodological records were reviewed for content applicable to the RoB in cross-sectional research designs. One reviewer (SK) extracted relevant methodological concepts, biases, items, domains or other author-identified considerations for cross-sectional research design or appraisal from each record. Content was extracted verbatim from each record and compiled by record in Microsoft Excel (Microsoft Corp.,2018). One reviewer (SK) amalgamated results into similar themes and constructs. The study team iteratively discussed results and compiled data into domains or items applicable to cross-sectional research.

Critical appraisal

As this is a scoping review, we did not conduct a critical appraisal of the included records.

Data charting and synthesis

The characteristics of the identified tools and methodological records were summarized with proportions for categorical variables, and means and standard deviations (SD), for continuous variables. Findings from the methodological records are reported descriptively. An analytic framework of RoB domains and items proposed by Wang et al. (2019) for observational studies was used to map the bias domains and items covered in each included tool [13]. Informed by the methodological concepts located in the search, the study team discussed all methodological concepts extracted. Following iterative discussions, the framework was adapted to include domains or items not currently addressed, but relevant to cross-sectional research, and to remove

those not considered relevant to the study design or not related to RoB. Tools for descriptive/prevalence or analytic/association cross-sectional studies were considered separately.

Results

Literature search

The bibliographic database search returned 5,164 records (3,527 records after deduplication) and 29 additional records were identified by content experts and review authors. A total of 27,249 records were identified using the PROSPERO database. After 1,000 records were examined prospectively, the review team agreed that an acceptable saturation of tools had been achieved given how few new unique ROB tools were identified over the last 200 records (Supplemental Appendix, Figure A). Including these 1,000 PROSPERO records, a total of 4,556 records were screened. Database records were screening using title and abstract and PROSPERO registrations using the 'Risk of bias (quality) assessment' field of each record. Of these, 348 were reviewed in full text to assess eligibility for inclusion in the scoping review. The full text of nine records could not be retrieved due to insufficient citation details provided and follow-up by email twice to corresponding authors from the original reports was unsuccessful. Following full-text review, 90 records were included[3, 13-101] (Figure 1).

Summary of included records

Of the 90 records selected for inclusion, 67 (74%) described the development of, or validation process for, appraisal tools [14-27, 29-37, 39, 41, 45-54, 56, 58, 60, 62-64, 66-72, 74, 76, 78, 81-91, 93, 97, 99-101], 15 (17%) described methodological content or theory relevant to RoB for cross-sectional studies[3, 28, 40, 42-44, 57, 59, 65, 73, 75, 77, 92, 94, 96] and 8 (9%) were records of methodological systematic reviews[13, 38, 55, 61, 79, 80, 95, 98].

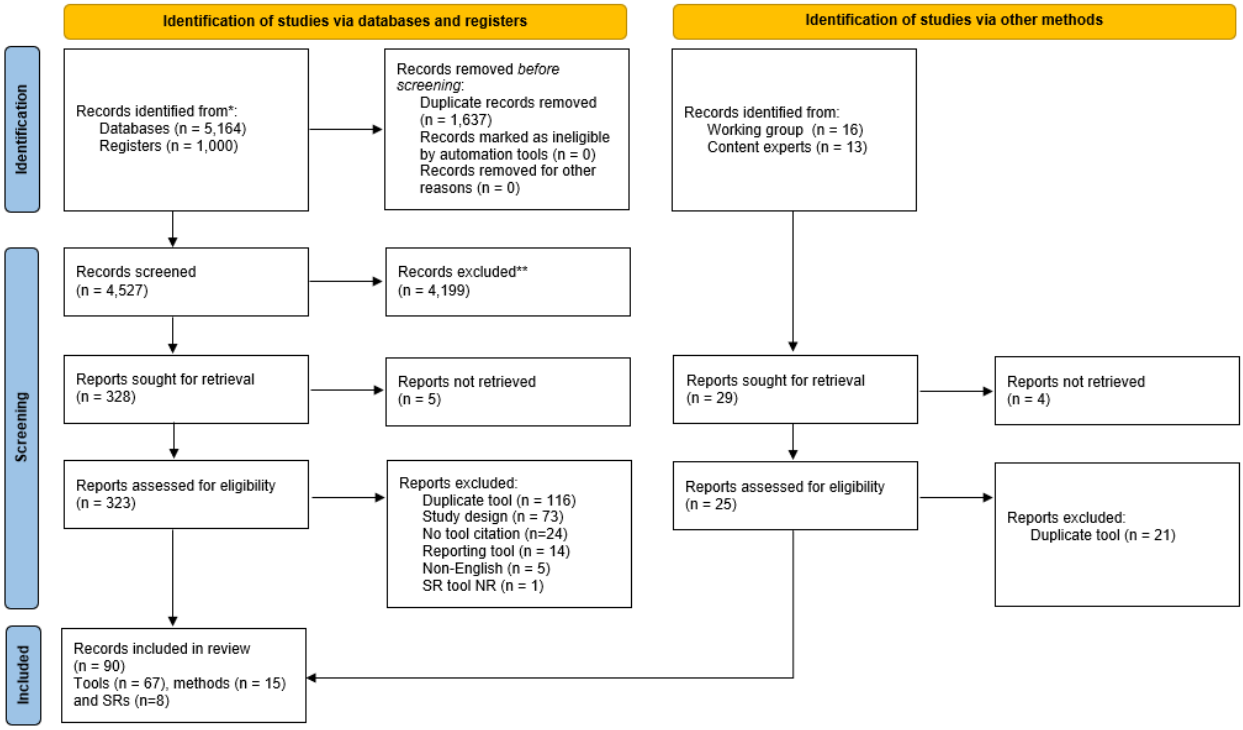


Figure 1: Flow diagram of searches of databases, registers and other sources.

Methodological concepts relevant to risk of bias in cross-sectional studies

Records used to identify methodological constructs (n=15)[3, 28, 40, 42-44, 57, 59, 65, 73, 75, 77, 92, 94, 96] were published between 1995 and 2023. The included records were generally focused on methodological descriptions of cross-sectional study designs[3, 28, 57, 59, 77, 92, 94, 96], surveys of prevalence[40, 43, 44], non-randomized research with consideration for cross-sectional studies[42, 65, 73, 75] and two records focused specifically on sources of bias in cross-sectional research[77, 96]. However, the terminology for describing cross-sectional study designs was ambiguous. Identified sources of bias in cross-sectional studies were quite broadly described, with most records highlighting the potential for selection bias related to sampling frame [1, 57, 96, 102-105], non-sampling error [96, 103, 104] and the concern that it is not possible to determine temporality (i.e., whether the outcome occurred prior to the exposure or

intervention; reverse causality) [1, 57, 96, 102-105]. Information bias related to the appropriate and valid measurement of exposure or outcomes was a commonly highlighted source of bias [3, 28, 57, 59, 77, 92, 94, 96]. In addition, several records mentioned items not related to RoB, including the handling of statistics, sample size, various reporting criteria (i.e., whether a design feature was present in the study report and/of if the details were fulsomely described), the generalizability of results or appropriateness of conclusions in context with RoB concepts.

Biases were considered separately depending on the intent of the study to provide descriptive/prevalence or analytic/association estimates[3, 57, 75, 96]. For descriptive cross-sectional studies aiming to provide estimates of prevalence, the noted threats to validity were representativeness of the sampling frame, non-sampling error from non-response or coverage and whether the outcome, disease or characteristic under study was measured in a valid and reliable way. Considerations specifically for analytic cross-sectional studies[3, 28, 57, 59, 75, 92] focused less on sampling frame, and included consideration for, and handling of, confounding variables in estimates and subgroups, appropriateness of eligibility criteria or group under study, blinding of research staff, validity and reliability of the exposure and outcome ascertainment, missingness in the data and time components related to recruitment. None of the records mentioned selective reporting, conflict of interest or funding as potential sources of bias.

Methodological systematic reviews

A total of eight systematic reviews[13, 38, 55, 61, 79, 80, 95, 98] published between 2010 and 2020 were included. The systematic reviews all endeavored to identify and describe or review use of tools for assessing quality or RoB in human research studies. The reviews focused on summarizing tools intended for prevalence studies[61, 80], observational studies[13], or all research designs[38, 55, 79, 95, 98]. Of these, two reviews focused specifically on intervention

studies, yet the tools cited were applicable to intervention and observational study designs [38, 95]. The included tools and references in each review were explored in detail to identify and extract RoB tools intended for assessment of cross-sectional studies. Not all tools reviewed were cited appropriately or reported in enough detail that the tools could be located and included in this scoping review.

Tools for RoB assessment in cross-sectional studies

From the 67 included records reporting tools, we identified 64 unique appraisal tools.[14-27, 29-37, 39, 41, 45-54, 56, 58, 60, 62-64, 66-72, 74, 76, 78, 81-91, 93, 97, 99-101] One tool was reported in four of the included records and all were combined and considered one unique tool.[14, 33-35] Table 1 provides an overview of the 64 unique RoB tools. Tools were applicable to multiple study designs, including cross-sectional studies (n=22; 34%)[14, 16, 17, 25, 29, 33-35, 39, 41, 46, 49, 50, 52, 56, 60, 72, 78, 83, 84, 89, 90], or cross-sectional designs only (n=42; 66%)[15, 18-24, 26, 27, 30-32, 36, 37, 45, 47, 48, 51, 53, 54, 58, 62-64, 66-71, 74, 76, 81, 82, 85-88, 91, 93, 97]. Twenty-two tools (34%) were identified from a non-methodological systematic review[27, 30, 36, 45, 51, 54, 56, 62, 64, 66-70, 74, 76, 84, 86-88, 91, 97]. While most tools were designed to be applied to any content area, several were developed specifically for clinical or health disciplines with the most frequent being nutrition (n=3)[17, 60, 76], pain (n=2)[47, 56] and environmental health or toxicology (n=2)[15, 93]. Other content-specific tools were for surgery[101], psychology[71], occupational therapy[52], mental health[85], medical education[25], chronic disease[82], genetics[26], psychiatry[31], public health[72] or cardio-pulmonary health[16]. Sixteen tools reported modifying or adapting an existing RoB tool for application to cross-sectional studies (n=16; 27%)[27, 30, 45, 51, 54, 62-64, 68, 69, 74, 76, 87, 91]. All modified tools were based on the Newcastle-Ottawa Scale (NOS)[106] except for Ross

et al. (2023)[76] which adapted the NUtrition QUality Evaluation Strengthening Tool (NUQUEST)[107] for cohort studies for application to cross-sectional studies in a systematic review.

Characteristics of the included tools

The characteristics of the included tools are summarized in Table 2. Tools were published between 1991 and 2023. Tools are described as being intended for critical appraisal (n=14; 22%), quality assessment (n=37; 59%) or RoB (n=11; 17%). Two additional tools did not differentiate between quality assessment or RoB in assessment or aimed to assess usefulness of research using methodological quality and trustworthiness concepts. Tools were either part of a set of tools developed to assess RoB (n=14; 22%) or standalone (n=50; 78%). Half of the tools were accompanied by some type of user guide or guidance (n=32; 50%) but only one tool had a worksheet available to assist raters with pragmatic or conceptual aspects important to the RoB assessment (n=1; <2%). Over two thirds (n=45; 70%) of the tools produced study ratings (i.e., any type of structured or unstructured judgment relating to a study's potential for risk of bias or study quality/critical appraisal)(e.g., low, unclear or high) and 30% (n=19) calculated a quantitative score.

Table 1: Overview of tools for cross-sectional studies (n=64).

Author (or Organization), Year	Abbreviation	Full tool name	Clinical focus	Study designs
Tools intended for multiple study designs, including cross-sectional studies				
Fowkes and Fulton, 1991[39]	N/A	Critical appraisal tool	Broad	RCT, C, CC, XS
Avis, 1994[29]	N/A	Critical appraisal tool	Broad	Any[a]
Law, 1998[52]	N/A	Critical Review Form for Quantitative Studies	Occupational therapy	RCT, C, CC, XS, CSt
Macfarlane, 2001[56]	N/A	Tool for XS studies reported in SR	Oro-facial pain	Cohort /XS
Margetts, 2002[60]	N/A	Checklist for reviewing nutritional epidemiological studies	Nutrition	RCT, C, CC, XS, P, E
Slim, 2003[101]	MINORS	Methodological index for non-randomized studies	Surgery	NRS, XS
Kmet, 2004[50]	N/A	Quality assessment criteria for evaluating primary research papers from a variety of fields	Broad	Any[b]
Genaidy, 2007[41]	EAI	Epidemiological Appraisal Instrument	Broad	RCT, C, CC, NRS, XS
Crowe, 2011[14, 33-35]	CCAT	Crowe Critical Appraisal tool	Broad	XS, P, RCT, NRS, C, CC, CSe, CSt, Q, SRs
Ross, 2011[99]	SAQOR	Systematic Assessment of Quality in Observational Research	Broad	C, CC, XS, Any[c,d]
Viswanathan, 2012[89, 90]	RTI	RTI Item Bank	Broad	C, CC, CSe, XS
Puzzolo, 2013[72]	LQAT	EPPI-Centre, Liverpool Quality Assessment tools	Public Health	RCT, C, CC, NRS, XS
Takahashi, 2014[84], Sedgwick, 2019[78]	NOS for XS	Newcastle-Ottawa Scale applied to cross-sectional studies	Broad	C, XS[e]
Hong, 2018[46]	MMAT -QNR	Mixed Methods Appraisal Tool for quantitative non-randomized studies	Broad	NRS[b] (including XS)
Hong, 2018[46]	MMAT-QD	Mixed Methods Appraisal Tool for quantitative descriptive studies	Broad	Any[f] (including P)
Kennedy, 2019[49]	N/A	The Evidence Project Risk of Bias tool	Broad	RCT, C, CC, NRS, PTPT, XS[b,c]
Stone, 2019[83]	MASTER	MethodologicAl STandard for Epidemiological Research (MASTER) Unified Framework for bias assessment in clinical research	Broad	Any[a]
National Heart Lung and Blood Institute (NHLBI), 2021[16]	NHLBI	NHLBI Study Quality Assessment Tool for Observational cohort and cross-sectional studies	Cardio-pulmonary	C, XS
Academy of Nutrition and Dietetics, 2022[17]	ANDEAL	Academy of Nutrition and Dietetics Evidence Analysis Library (ANDEAL): Quality Criteria Checklist for Primary Research	Nutrition	RCT, C, CC, Cse, CSt, XS, E, NRS, D, BA
Al Asmri, 2023[25]	MERSQI	Medical Education Research Study Quality Instrument	Medical education	RCT, C, CC, XS, PTPT Controlled NRS

Author (or Organization), Year	Abbreviation	Full tool name	Clinical focus	Study designs
Tools intended for cross-sectional study designs (descriptive/prevalence or analytic/association)[g]				
Boyle, 1998[31]	N/A	Guidelines for evaluating prevalence studies	Psychiatry	P
Loney, 1998[53]	N/A	Critical appraisal of the health research literature: prevalence or incidence of a health problem	Broad	P
van der Windt, 2000[86]	N/A	Tool for XS studies reported in SR	Broad	XS
Silva, 2001[82]	N/A	Tool for assessing the usefulness of prevalence studies done for surveillance purposes	Chronic/non-communicable disease surveillance	P[f]
Al-Jader, 2002[26]	N/A	Quality scoring system for epidemiological surveys of genetic disorders	Genetics	P[f]
Jackson, 2006[48]	GATE- appraise	Graphical Appraisal Tool	Broad	XS
Viswanathan, 2008[88]	AHRQ	AHRQ Design-Specific Criteria To Assess Risk of Bias in XS Studies	Broad	XS[h]
Busse, 2011[32]	CLARITY	RoB for cross-sectional surveys of attitudes and practices	Broad	XS
National Collaborating Centre on Environmental Health, 2011[93]	NCCEH-XS	NCCEH Tool for Critical Appraisal of a Cross-Sectional Study on Environmental Health	Environmental health	XS
Shamliyan, 2011[81]	MORE	Methodological Evaluation of Observational Research	Broad but developed based on topics in chronic disease or risk factors	P[f]
Yusef, 2011[97]	N/A	Tool for XS studies reported in SR	Broad[A]	XS
Hoy, 2012[47]	N/A	Tool for assessing risk of bias in prevalence studies	Broad but developed with neck and back pain focus	P[f]
Powell, 2013[70]	N/A	Tool for XS studies reported in SR	Broad	XS
DeGroot, 2014[36]	QATQS	Quality Assessment Tool for Quantitative Studies modified in SR for application to XS studies	Broad	XS
University of Oxford, Centre for Evidence-Based Medicine, 2014[58]	CEBM	CEBM Critical Appraisal Tool for Cross-sectional Studies	Broad	XS
Downes, 2016[37]	AXIS	Appraisal tool for Cross-Sectional Studies	Broad medical or veterinary	XS
Ohadike, 2016[67]	N/A	Tool for XS studies reported in SR	Broad	XS
Cardiff University, 2018[24]	SURE	Cross-sectional Studies Critical Appraisal Checklist	Broad	XS
National Toxicology Program, US Department of Health and Human Services, 2019[15]	OHAT	Office of Health Assessment and Translation (OHAT) RoB Tool	Toxicology	XS[h]

Author (or Organization), Year	Abbreviation	Full tool name	Clinical focus	Study designs
University of Oxford, Centre for Evidence-Based Medicine, 2019[23]	CAT Manager App	Critically Appraised Topic (CAT) Manager Tool for Cross-sectional Studies (Apple phone/ANDROID application)	Broad	XS
Nudelman and Otto, 2020[66]	ROBUST	Risk-of-Bias Measure for Systematic Reviews of Surveys	Broad	P
Protogerou, 2022[71]	Q-SSP	Quality of survey studies in psychology	Psychology	XS
Perez-Rios, 2022[100]	STREAMS-P	STrengthen the design and REporting of Attributable Mortality Studies using a Prevalence-based method	Attributable mortality focus	P
Cincinnati Childrens Hospital Medical Centre, 2023[20]	LEGEND-XS (INTERVENTION)	Let Evidence Guide Every New Decision (LEGEND) Tool for Cross Sectional Studies of Interventions	Broad	XS
Cincinnati Childrens Hospital Medical Centre, 2023[21]	LEGEND-XS (ETIOLOGY, RISK FACTORS)	Let Evidence Guide Every New Decision (LEGEND) Tool for Cross Sectional Studies of Etiology or Risk Factors	Broad	P
Cincinnati Childrens Hospital Medical Centre, 2023[22]	LEGEND-XS (PREVALENCE)	Let Evidence Guide Every New Decision (LEGEND) Tool for Cross Sectional Studies of Prevalence	Broad	P
Joanna Briggs Institute (JBI), 2023[18]	JBI-XS	JBI Checklist for Analytical Cross Sectional Studies	Broad	XS
Joanna Briggs Institute (JBI), 2023[19]	JBI-P	JBI Checklist for Prevalence Studies	Broad	P
Ross, 2023[76]	Modified NUQUEST	NUtrition QUality Evaluation Strengthening Tool for cohort studies modified in SR for application to XS studies	Nutrition	XS
Thomy, 2023[85]	RoB-PrevMH	A tool to assess risk of bias in studies estimating the prevalence of mental health disorders (RoB-PrevMH)	Mental Health	P
Herzog, 2013[45], Bawor, 2014[30], Alshabanat, 2015[27], Kunutsor, 2015[51], Patra, 2015[68], Perera, 2015[69], van Dijk, 2015[87], Modesti, 2016[62], Rodriguez, 2016[74], Murray, 2018[64], Moskalewicz, 2020[63], Lu, 2021[54], Wagg, 2021[91]	Modified NOS	Newcastle-Ottawa Scale adapted for cross-sectional studies	Broad[i]	XS[j]

RCT=randomized controlled trial; C=cohort study; CC=case-control study; XS=cross-sectional study; P=Prevalence study, Cse=Case series, CSt=Case study; E=Ecological study; D=diagnostic study; DS=Descriptive studies; BA=Before-after study; PTPT=Pre-test/Post-test study; NRS=non-randomized study, OBS=observational study, Q=qualitative study, SR=systematic review.

Table footnotes:

a: Must be experimental or analytic study.
b: Must be quantitative study.
c: Must be comparative.
d: Includes 'database' studies.
e: NOS was intended for cohort studies but was applied to cross-sectional studies in two systematic reviews. Each was considered a separate tool (n=2 tools).
f: Must be descriptive study.
g: Terminology for cross-sectional and prevalence studies are often used ambiguously or interchangeably.

h: Multi-study tool with subset of questions for XS studies only.
i: Some items in the tool may be SR topic specific.
j: Each was considered a separate tool (n=13 tools).

Forty-one percent of the tools grouped items or probes using categories or RoB domains (e.g., selection, sampling, exposure ascertainment etc.) as a structure (n=26; range 3 to 8 categories or domains). Tools contained, on average, 13.5 items or probes (range 3 to 43 items or probes). The number of response options available to rate studies ranged from 2 to 5 (mean 3.2 (0.8 standard deviation (SD)). Almost 40% (n=24) of the tools contained items related to reporting quality. Twenty-four tool developers (38%) made a declaration regarding their funding for the tool development but only 15 (23%) made any declaration regarding conflict of interest (COI).

Table 2: Characteristics of the risk of bias tools for cross-sectional studies (n=64).

Characteristic:	N (%)
Total included:	
Unique tools	64 (100)
Tool developed for [a]:	
Critical appraisal	14 (21.9)
Quality assessment	37 (57.8)
Risk of bias	11 (17.2)
Other	2 (3.1)[b]
Year:	
1999 or prior	5 (7.8)
2000 to 2009	10 (15.6)
2010 to 2019	33 (51.6)
2020 to 2023	18 (28.1)
Standalone or part of a tool suite:	
Standalone	50 (78.0)
Part of a suite of tools	14 (21.9)
Tool features:	
User guide/guidance provided:	
Yes	32 (50.0)
No	32 (50.0)
Worksheet available:	1 (<2%)
Assessment:	
Scored	19 (29.7)[c]
Rated	45 (70.3)[d]
Number of domains or categories	4.4 (1.7)[e,f]
Number of items/probes in tool:	13.5 (8.1)[e,g]
Number of response options:	3.2 (0.8)[e]
No response option suggested:	1 (1.6)[e]
Declarations:	
Funding:	
Declared: had funding	21 (32.8)
Declared: no funding	3 (4.7)
Not declared	40 (62.5)
Conflict of interest (COI):	
Declared – no COI	8 (12.5)
Declared – COI	7 (10.9)

Not declared	49 (76.6)

Table footnotes:

a: Author declared terminology was collected for each tool. For 'other': One made was no distinction between quality assessment or risk of bias assessment and one tool was for assessment of 'useability' which combined concepts of methodological quality and trustworthiness.
b: One reported being for 'usefulness' and one for 'validity'.
c: Quantitative score calculated based on assessment of domains and/or items, categories (I.e., a scored tool).
d: Domains and/or items used to assign a rating based on the assessment (could be overall rating and/or rating by domain, category or item, or both)(I.e., a checklist)
e: mean(SD).
f: Items are structured using domains of bias (e.g., selection, ascertainment, confounding) or another metric.
g: Based on 63 tools. One tool was not considered in this category due to the graphical format(1).

Approaches used to develop the content of the tools are reported in Table 3. Teams developing content for tools were reported to be methodologists (n=32; 50%), clinical content experts (n=16; 25%) or undefined working groups of authors with unclear expertise (n=32; 50%). No patient, caregiver, or public team members were reported to have contributed to tool content. Methods to develop the tools included a literature or systematic review of existing tools (n=19; 29.7%), expert consultation (n=10; 16%), consensus meetings (n=8; 13%) or a Delphi process (n=5; 8%). However, 60% (n=39) of the tools did not report details about the tool content or development process. Few tool developers reported pilot (n=13; 20%)[14, 26, 33-37, 41, 46, 47, 49, 50, 66, 81, 85], usability (n=7; 11%)[14, 26, 33-35, 37, 41, 49, 72, 89, 90], reliability (n=11;17%)[14, 26, 33-35, 41, 47, 49, 50, 56, 66, 81, 99, 101] or validity (n=8; 12.5%)[14, 25, 33-35, 41, 46, 66, 81, 101] testing as part of tool development. Most tools applicable to multiple study designs did not report any evaluation involving cross-sectional studies as part of their testing process. Reliability testing was based on interrater agreement at varying levels (e.g., by study, domain or item/probe). Test-retest reliability was assessed for one tool by one reviewer[101]. One study considered the experience level of the reviewers in their reliability testing[66]. Content or criterion validity was assessed using expert consultation, group consensus or Delphi process. Convergent validity was assessed using ratings from the same studies in a previous meta-analysis with RoB rated by a different tool[66].

Table 3: Development features of the risk of bias tools for cross-sectional studies (n=64).

Characteristic:	N (%)
Methods to develop tool[a]:	
Literature/systematic review	19 (29.7)
Expert consultation	10 (15.6)
Delphi Survey	5 (7.8)
Consensus meetings	8 (12.5)
Not reported/unclear	39 (60.9)
Development team:	
Content expert (Methods)	32 (50.0)
Content expert (Clinical)	16 (25.0)
Patients or Public	0 (0.0)
Not reported/unclear	32 (50.0)
Pilot tested:	
Yes	13 (20.3)
No	23 (35.9)
Not reported/unclear	28 (43.8)
Reported usability:	
Yes	7 (10.9)
No	23 (35.9)
Not reported/unclear	34 (53.1)
Reported reliability:	
Yes	11 (17.2)
No	25 (39.1)
Not reported/unclear	28 (43.8)
Reported validity:	
Yes	8 (12.5)
No	30 (46.9)
Not reported/unclear	26 (40.6)

Table footnotes:

a: Tools may have used one or more of the methods noted.

Analytic framework for coverage of RoB domains and items in cross-sectional tools

The analytic framework of RoB domains and items proposed by Wang et al. (2019) [13] for observational studies was adapted by the study team to include the domains and items most applicable and appropriate for assessment of risk of bias for cross-sectional studies (Table 4). The original framework considered eight bias domains: selection (5 items), exposure (1 item), outcome (2 items), confounding (2 items), loss to follow-up (3 items), analysis (1 item), selective reporting (1 item) and conflict of interest (COI; 1 item). The framework also had a domain for other bias (2 items) and a category that considered 'other' items not applicable to research bias (statistical power/sample size, analysis, clarity of research objectives/hypothesis, reporting of eligibility criteria and conclusions supported by results and methods) (4 items). Based on the review of methodological guidance and discussion and agreement amongst the authors, the loss

to follow-up domain was considered not applicable to this research design and removed. One item 'mechanism of non-response' was added to the selection domain and a new domain for consideration of missingness was added (including 3 items: amount of missingness, handling of missingness and mechanism for missingness). Missingness considers data coverage for study participants. Handling of missingness may include imputation to assign replacement values for some, most or all missing values, and may include consideration for invalid or inconsistent data [108]. Items considered not applicable to research bias were removed from the final framework.

The updated analytic framework was used to map the domains and items covered in each included RoB tool. Table 4 provides a summary of the frequency each item was addressed, by domain, for all tools. The RoB Items most frequently addressed in the RoB tools were validity and reliability of the exposure (53%) or outcome (65%) measurement and representativeness of the study population (59%). Most tools did not consider non-response, missingness, selective reporting or COI fully, or at all. The Supplemental Appendix (Table B) provides a detailed map of all items in the framework for each tool. Many items in the included tools did not map to any item in the framework and were considered not applicable by the study team (i.e., items related to reporting quality, statistics, precision of estimates, sample size, or clarity of objectives/hypotheses).

Table 4: Tools for cross-sectional studies reporting risk of bias framework items (framework adapted from Wang et al. (2019)[13]), n = 64 tools.

BIAS DOMAIN	BIAS CONCEPT	Addressed in tool, *n (%)*	Not addressed in tool, *n (%)*	Unclear or partially addressed in tool[a], *n (%)*
SELECTION				
	Appropriateness of eligibility criteria	13 (38.2)	48 (59.4)	3 (4.7)
	Comparability of exposure group vs. comparison group	16 (25.0)	47 (73.4)	1 (1.6)
	Non-response rate	30 (46.9)	33 (51.6)	1 (1.6)
	Non-response rate - mechanism	1 (1.6)	62 (96.9)	1 (1.6)

BIAS DOMAIN	BIAS CONCEPT	Addressed in tool, *n (%)*	Not addressed in tool, *n (%)*	Unclear or partially addressed in tool[a], *n (%)*
	Recruitment time frame	6 (9.4)	57 (89.1)	1 (1.6)
	Representativeness of sample to target population	38 (59.4)	24 (37.5)	2 (3.1)
EXPOSURE				
	Validity and reliability of exposure measurement	34 (53.1)	18 (28.1)	12 (18.8)
OUTCOME				
	Blinding of the research staff	26 (40.6)	36 (56.3)	2 (3.1)
	Validity and reliability of the outcome measurement	41 (64.1)	15 (23.4)	8 (12.5)
CONFOUNDING				
	Accounting for confounding variables	34 (53.1)	27 (42.2)	3 (4.7)
	Description of confounding variables	23[b] (35.9)	37 (57.8)	4 (6.3)
MISSINGNESS				
	Amount of missingness	7 (10.9)	51 (79.7)	6 (9.4)
	Handling of missingness	5 (7.8)	56 (87.5)	3 (4.7)
	Mechanism for missingness	0 (0.0)	63 (98.4)	1 (1.6)
SELECTIVE REPORTING				
	Selective reporting of outcomes	4 (6.3)	60 (93.8)	0 (0.0)
CONFLICT OF INTEREST				
	Conflict of interest (e.g., funding)	8 (12.5)	56 (87.5)	0 (0.0)
OTHER BIAS[c]				
	Other bias or threats to internal validity	4 (6.3)	60 (93.8)	0 (0.0)

a: Items were mapped as *unclear* if they may have covered the bias concept, but no additional documentation was available to clarify or confirm the wording or intent of the probe.
b: SR-based tools did not report items for confounding variable description as these were named and considered in the SR itself.
c: The 'other' category could result in discrepant or erroneous judgments of bias for a study(2); therefore, authors felt it was important to document how many tools used this otherwise unspecified domain.

Exploratory review of subgroups of tools developed to be applied to multiple study designs or only cross-sectional studies showed few differences when considered separately and compared (Supplemental Appendix, Table C). Multi-study tools more frequently covered most items in the selection bias domain, except for representativeness of the sample to the target population. Cross-sectional only tools more frequently addressed the validity and reliability of the outcome and exposure measurement, while multi-study tools more frequently addressed the description of confounding variables. Tools that self-identified as RoB tools (versus critical appraisal or

quality) more frequently considered non-response rate, the validity and reliability of the exposure and outcome measurement and some selection bias items.

Discussion

We conducted a comprehensive scoping review that included 15 methodological papers, 8 systematic reviews and 64 assessment tools related to the risk of bias in cross-sectional studies. The findings highlight the variety of tools that are available and have been used to consider the risk of bias in cross-sectional studies. However, variability in the content and coverage for key bias domains indicates the absence of a "gold" standard tool for cross-sectional studies. Assessing risk of bias is a key step when knowledge users move from evidence to action, for all types of evidence. The volume of tools applicable to cross-sectional studies retrieved in this scoping review confirms a need for these instruments. Our findings may explain why existing tools have not been fully accepted and identifies gaps that should be addressed to encourage uptake and use [55].

Notably, our results suggest that the methodology to develop and evaluate risk of bias tools can be improved and better reported[110]. Consistent with broader reviews of similar intent, our assessment showed that most tools for appraisal of cross-sectional studies lack rigour or empirical evidence of their measurement properties[13, 111]. Without this, we are unable to conclude that we are assessing risk of bias in a meaningful or useful way[112]. While many tools were developed to in theory to address issues of bias, or built on previous tools with extensive use cases, most were not formally assessed for reliability or validity. More attention to the empirical evaluation of the validity, reliability and utility of any proposed tool by appraisers and end-users is needed prior to dissemination [111]. Criterion validity and inter-rater reliability were most often assessed as part of the tool development process. However, the majority do not

formally consider any psychometric properties of the tool, usability or the feasibility of application. This is especially evident in tools developed specifically for, and reported within, systematic reviews. Beyond basic applicability to the cross-sectional study design and the topic under review, very few of these tools were formally evaluated. This makes these tools fit-for-purpose yet potentially limits their more generalized use. This also is related to research and resource waste, with key contributors being avoidable research design limitations and inadequate reporting or dissemination [110, 113]. With this in mind, individuals using these existing RoB tools may assume they have been demonstrated to be reliable and valid and are relevant to cross-sectional studies. Our rationale for focusing part of the scoping review on describing the particular biases that may impact cross-sectional research and then examining each tool for their coverage of each bias was to ensure these tools are indeed appropriate for their intended use, and to understand where future research in this area can avoid research waste and support future tool development.

Consultation with methodological experts and our review of methodological guidance specific to risk of bias in cross-sectional studies proved to be useful to the study team. This also facilitated our adaptation of the RoB domain framework[13] to highlight issues of greatest concern for cross-sectional studies and to differentiate between issues that may be more of a concern for studies with an aim to produce prevalence estimates versus analytic/association approximations. The framework adapted for use in this scoping review[13] provides the domains that risk of bias tools should assess in cross-sectional studies: selection, exposure, outcome, confounding, missingness, selective reporting, and COI. These domains consider empirical and theoretical evidence about the aspects of observational study designs that can introduce bias into outcomes[13, 55], and were adapted following a thorough review of risk of bias specific to cross-sectional studies.

Our aim was to provide an overview of coverage across these domains for each RoB tool applicable to cross-sectional studies. Readers can prioritize domains and items within domains that apply to their needs or content area and select a ROB tool, based on the framework. However, it is worth noting that there is currently no "gold" standard tool reflects the full breadth of the domains and items identified in the framework. As with tools applicable to RoB in other study designs, we found tools for cross-sectional studies contained items not addressing risk of bias, including reporting quality, applicability and methods for statistical analyses[114]. Often, it was unclear or ambiguous whether an item was addressing a RoB issue or reporting quality due to the phrasing or scoring/rating systems used. This was prominent in tools published before the year 2000 and there has since been some improvement. Inclusion of items not related to bias in a tool may distort the measurement properties and impact ratings. Taking the by-domain and by-item coverage assessments into account, selection of a tool for evaluating potential for risk of bias in cross-sectional studies should prioritize coverage for bias domains and items.

We found that no one tool covered all key bias domains and items for cross-sectional studies. In addition, our findings show that more guidance may be required to help users with their study assessments when using the available tools. Although tools in more recent, published records were described well and their scoring or ratings systems explained, many tools often had brief or no guidance for how to interpret the items appropriately. For tools applicable to more than one study design, the available guidance was often generally applicable to observational study designs but not specific to enable appropriate assessment of cross-sectional studies.

There are several limitations to our scoping review that are worth noting. The inherent limitation of scoping reviews is the focus on identifying the breadth of information on the topic at hand rather than depth[115]. As such, additional consideration for synthesis of the topics presented may be warranted, but we feel we have adequately covered the scope as intended. Although we

conducted a comprehensive search, methodological records are poorly indexed in bibliographic databases and difficult to find in the grey literature[116]. There were 24 records for potential eligible records that could not be retrieved as the existing methodological systematic reviews did not fulsomely cite the tools or provided only an organization name which was not enough to locate the record. We limited the included records to those published in English and so our findings are generalizable to risk of bias tools written in English. Although no formal Delphi or consensus approach was used, the study working group met weekly over the course of the study to iteratively discuss and validate review methods, findings and interpretation.

We anticipate that our findings will be of interest to individuals or organizations conducting systematic reviews of cross-sectional studies, researchers who conduct surveys and descriptive research studies, journal editors and agencies or organizations who may use cross-sectional research data to inform policy or decision-making. We plan to use the results of this scoping review to improve content and education around risk of bias in cross-sectional studies and to inform the design and reporting of future cross-sectional studies.

Conclusions

This study provides a comprehensive summary of the unique, and more general, methodological considerations for RoB in cross-sectional study designs and maps the inclusion of each consideration to existing RoB tools. We identified RoB tools designed for broad applicability across various study designs as well as those specifically tailored for cross-sectional studies. However, none of the identified tools comprehensively address all potential bias domains pertinent to cross-sectional studies (i.e., selection, exposure, outcome, confounding, missingness, selective reporting, and conflict of interest). Findings may be used to inform the selection and continued use of existing RoB tools and indicate a need for ongoing improvement. Tool

developers may use results to consider development of context-specific or more precise RoB tools for population health estimates or variable associations.

References

[1] Reeves BC, Wells GA, Waddington H. Quasi-experimental study designs series-paper 5: a checklist for classifying studies evaluating the effects on health interventions-a taxonomy without labels. Journal of Clinical Epidemiology. 2017;89:30-42.

[2] Cumpston M, Li T, Page MJ, Chandler J, Welch VA, Higgins JP, et al. Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Database Syst Rev. 2019;10:ED000142.

[3] Kesmodel US. Cross-sectional studies–what are they good for? Acta obstetricia et gynecologica Scandinavica. 2018;97:388-93.

[4] Pandis N. Cross-sectional studies. American Journal of Orthodontics and Dentofacial Orthopedics. 2014;146:127-9.

[5] Wang X, Cheng Z. Cross-sectional studies: strengths, weaknesses, and recommendations. Chest. 2020;158:S65-S71.

[6] McNair P, Lewis G. Levels of evidence in medicine. Int J Sports Phys Ther. 2012;7:474-81.

[7] Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International journal of social research methodology. 2005;8:19-32.

[8] Jordan Z, Lockwood C, Aromataris E, Pilla B, Porritt K, Klugar M, et al. JBI series paper 1: Introducing JBI and the JBI Model of EHBC. J Clin Epidemiol. 2022;150:191-5.

[9] Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. 2018;169:467-73.

[10] Kelly SE, Benkhedda K, Brooks SPJ, MacFarlane AJ, Greene-Finestone LS, Skidmore B, et al. Risk of bias in cross-sectional studies: Protocol for a scoping review of concepts and tools. MethodsX. 2024;12:102610.

[11] Stone JC, Barker TH, Aromataris E, Ritskes-Hoitinga M, Sears K, Klugar M, et al. From critical appraisal to risk of bias assessment: clarifying the terminology for study evaluation in JBI systematic reviews. JBI Evid Synth. 2023;21:472-7.

[12] Barker TH, Stone JC, Sears K, Klugar M, Leonardi-Bee J, Tufanaru C, et al. Revising the JBI quantitative critical appraisal tools to improve their applicability: an overview of methods and the development process. JBI Evidence Synthesis. 2023;21:478-93.

[13] Wang Z, Taylor K, Allman-Farinelli M, Armstrong B, Askie L, Ghersi D, et al. A systematic review: Tools for assessing methodological quality of human observational studies. 2019.

[14] Crowe Critical Appraisal Tool (v1.4). 2015.

[15] Office of Health Assessment and Translation - Risk of Bias Rating Tool for Human and Animal Studies. National Toxicology Program, US Department of Health and Human Services; 2015.

[16] Study Quality Assessment Tools: Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies. National Heart, Lung and Blood Institute, National Institutes of Health; 2021.

[17] Evidence Worksheet and Quality Criteria Checklist - Primary Research. Online: Academy of Nutrition and Dietetics Evidence Analysis Library; 2022.

[18] Personal Communication from Timothy Hugh Barker. JBI CRITICAL APPRAISAL TOOL: JBI CHECKLIST FOR ANALYTICAL CROSS-SECTIONAL STUDIES (2023 Update). University of Adelaide2023.

[19] Personal communication from Timothy Hugh Barker. JBI CRITICAL APPRAISAL TOOL: JBI CHECKLIST FOR PREVALENCE STUDIES (2023 Update). University of Adelaide2023.

[20] Evidence Evaluation Tools & Resources: LEGEND (Let Evidence Guide Every New Decision) for Cross-sectional studies of interventions. Cinncinati, OH, USA James M Anderson Center for Health Systems Excellence, Cincinnati Children's Hospital Medical Center; 2023.

[21] Evidence Evaluation Tools & Resources: LEGEND (Let Evidence Guide Every New Decision) for Cross-sectional studies of etiology or risk factors. Cinncinati, OH, USA James M Anderson Center for Health Systems Excellence, Cincinnati Children's Hospital Medical Center; 2023.

[22] Evidence Evaluation Tools & Resources: LEGEND (Let Evidence Guide Every New Decision) for Cross-sectional studies of prevalence. Cinncinati, OH, USA James M Anderson Center for Health Systems Excellence, Cincinnati Children's Hospital Medical Center; 2023.

[23] CAT Manager App for iPhone or Android. version not reported. ed. Application Store (Free): Center for Evidence Based Management. Oxford, UK.; 2023.

[24] (SURE) SUfRE. Questions to assist with the critical appraisal of cross-sectional studies. online2018.

[25] Al Asmri M, Haque MS, Parle J. A Modified Medical Education Research Study Quality Instrument (MMERSQI) developed by Delphi consensus. BMC Med Educ. 2023;23:63.

[26] Al-Jader LN, Newcombe RG, Hayes S, Murray A, Layzell J, Harper PS. Developing a quality scoring system for epidemiological surveys of genetic disorders. Clin Genet. 2002;62:230-4.

[27] Alshabanat A, Zafari Z, Albanyan O, Dairi M, FitzGerald J. Asthma and COPD overlap syndrome (ACOS): a systematic review and meta analysis. PloS one. 2015;10:e0136065.

[28] Asiamah N, Mends-Brew E, Boison BKT. A spotlight on cross-sectional research: addressing the issues of confounding and adjustment. International Journal of Healthcare Management. 2021;14:183-96.

[29] Avis M. Reading research critically. II. An introduction to appraisal: assessing the evidence. Journal of clinical nursing. 1994;3:271-7.

[30] Bawor M, Dennis BB, Anglin R, Steiner M, Thabane L, Samaan Z. Sex differences in outcomes of methadone maintenance treatment for opioid addiction: a systematic review protocol. Systematic reviews. 2014;3:1-7.

[31] Boyle M. Guidelines for evaluating prevalence studies. BMJ Mental Health. 1998;1:37-9.

[32] Busse JobotCGaMUaEP. Methods Commentary: Risk of Bias in Cross-Sectional Surveys of Attitudes and Practices. Online2011.

[33] Crowe M, Sheppard L. A general critical appraisal tool: an evaluation of construct validity. Int J Nurs Stud. 2011;48:1505-16.

[34] Crowe M, Sheppard L, Campbell A. Comparison of the effects of using the Crowe Critical Appraisal Tool versus informal appraisal in assessing health research: a randomised trial. Int J Evid Based Healthc. 2011;9:444-9.

[35] Crowe M, Sheppard L, Campbell A. Reliability analysis for a proposed critical appraisal tool demonstrated value for diverse research designs. J Clin Epidemiol. 2012;65:375-83.

[36] de Groot R, van den Hurk K, Schoonmade LJ, de Kort WL, Brug J, Lakerveld J. Urban-rural differences in the association between blood lipids and characteristics of the built environment: a systematic review and meta-analysis. BMJ global health. 2019;4:e001017.

[37] Downes MJ, Brennan ML, Williams HC, Dean RS. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). BMJ Open. 2016;6:e011458.

[38] Farrah K, Young K, Tunis MC, Zhao L. Risk of bias tools in systematic reviews of health interventions: an analysis of PROSPERO-registered protocols. Syst Rev. 2019;8:280.

[39] Fowkes F, Fulton P. Critical appraisal of published research: introductory guidelines. BMJ: British Medical Journal. 1991;302:1136.

[40] Gauthier B. Assessing Survey Research A Principled Approach. Citeseer; 2003.

[41] Genaidy A, Lemasters G, Lockey J, Succop P, Deddens J, Sobeih T, et al. An epidemiological appraisal instrument–a tool for evaluation of epidemiological studies. Ergonomics. 2007;50:920-60.

[42] Goldberg RJ, McManus DD, Allison J. Greater knowledge and appreciation of commonly-used research study designs. Am J Med. 2013;126:169 e1-8.

[43] Habibzadeh F, Habibzadeh P, Yadollahie M. The apparent prevalence, the true prevalence. Biochem Med (Zagreb). 2022;32:020101.

[44] Harder T. Some notes on critical appraisal of prevalence studies: comment on:" the development of a critical appraisal tool for use in systematic reviews addressing questions of prevalence". International journal of health policy and management. 2014;3:289.

[45] Herzog R, Álvarez-Pasquin M, Díaz C, Del Barrio JL, Estrada JM, Gil Á. Are healthcare workers' intentions to vaccinate related to their knowledge, beliefs and attitudes? A systematic review. BMC public health. 2013;13:1-17.

[46] Hong QN, Gonzalez-Reyes A, Pluye P. Improving the usefulness of a tool for appraising the quality of qualitative, quantitative and mixed methods studies, the Mixed Methods Appraisal Tool (MMAT). J Eval Clin Pract. 2018;24:459-67.

[47] Hoy D, Brooks P, Woolf A, Blyth F, March L, Bain C, et al. Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. J Clin Epidemiol. 2012;65:934-9.

[48] Jackson R, Ameratunga S, Broad J, Connor J, Lethaby A, Robb G, et al. The GATE frame: critical appraisal with pictures. BMJ Evidence-Based Medicine. 2006;11:35-8.

[49] Kennedy CE, Fonner VA, Armstrong KA, Denison JA, Yeh PT, O'Reilly KR, et al. The Evidence Project risk of bias tool: assessing study rigor for both randomized and non-randomized intervention studies. Systematic reviews. 2019;8:1-10.

[50] Kmet LM, Cook LS, Lee RC. Standard quality assessment criteria for evaluating primary research papers from a variety of fields. 2004.

[51] Kunutsor SK, Apekey TA, Laukkanen JA. Association of serum total osteocalcin with type 2 diabetes and intermediate metabolic phenotypes: systematic review and meta-analysis of observational evidence. European journal of epidemiology. 2015;30:599-614.

[52] Law M, Stewart D, Pollock N, Letts L, Bosch J, Westmorland M. Guidelines for Critical Review Form—Quantitative Studies 1998. McMaster University Occupational Therapy Evidence-based Practice Research Group. 2003.

[53] Loney PL, Chambers LW, Bennett KJ, Roberts JG, Stratford PW. Critical appraisal of the health research literature: prevalence or incidence of a health problem. Chronic Dis Can. 1998;19:170-6.

[54] Lu G, Chen Q, Li Y, Liu Y, Zhang Y, Huang Y, et al. Status of antibiotic residues and detection techniques used in Chinese milk: A systematic review based on cross-sectional surveillance data. Food Res Int. 2021;147:110450.

[55] Ma LL, Wang YY, Yang ZH, Huang D, Weng H, Zeng XT. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? Mil Med Res. 2020;7:7.

[56] Macfarlane TV, Glenny AM, Worthington HV. Systematic review of population-based epidemiological studies of oro-facial pain. J Dent. 2001;29:451-67.

[57] Maier C, Thatcher JB, Grover V, Dwivedi YK. Cross-sectional research: A critical perspective, use cases, and recommendations for IS research. Elsevier; 2023. p. 102625.

[58] Management CfEB. Critical appraisal checklist for cross-sectional study. Centre for Evidence-Based Management; 2014.

[59] Mancuso CA, Charlson ME. Does recollection error threaten the validity of cross-sectional studies of effectiveness? Medical Care. 1995:AS77-AS88.

[60] Margetts B, Vorster H, Venter C. Evidence-based nutrition-review of nutritional epidemiological studies. South African Journal of Clinical Nutrition. 2002.

[61] Migliavaca CB, Stein C, Colpani V, Munn Z, Falavigna M, Prevalence Estimates Reviews - Systematic Review Methodology G. Quality assessment of prevalence studies: a systematic review. J Clin Epidemiol. 2020;127:59-68.

[62] Modesti PA, Reboldi G, Cappuccio FP, Agyemang C, Remuzzi G, Rapi S, et al. Panethnic Differences in Blood Pressure in Europe: A Systematic Review and Meta-Analysis. PLoS One. 2016;11:e0147601.

[63] Moskalewicz A, Oremus M. No clear choice between Newcastle-Ottawa Scale and Appraisal Tool for Cross-Sectional Studies to assess methodological quality in cross-sectional studies of health-related quality of life and breast cancer. J Clin Epidemiol. 2020;120:94-103.

[64] Murray C, Marshall M, Rathod T, Bowen CJ, Menz HB, Roddy E. Population prevalence and distribution of ankle pain and symptomatic radiographic ankle osteoarthritis in community dwelling older adults: A systematic review and cross-sectional study. PLoS One. 2018;13:e0193662.

[65] Myers EF, Parrott JS, Splett P, Chung M, Handu D. Using risk of bias domains to identify opportunities for improvement in food- and nutrition-related research: An evaluation of research type and design, year of publication, and source of funding. PLoS One. 2018;13:e0197425.

[66] Nudelman G, Otto K. The Development of a New Generic Risk-of-Bias Measure for Systematic Reviews of Surveys. Methodology. 2020;16:278-98.

[67] Ohadike CO, Cheikh-Ismail L, Ohuma EO, Giuliani F, Bishop D, Kac G, et al. Systematic review of the methodological quality of studies aimed at creating gestational weight gain charts. Advances in Nutrition. 2016;7:313-22.

[68] Patra J, Bhatia M, Suraweera W, Morris SK, Patra C, Gupta PC, et al. Exposure to second-hand smoke and the risk of tuberculosis in children and adults: a systematic review and meta-analysis of 18 observational studies. PLoS Med. 2015;12:e1001835; discussion e.

[69] Perera S, Eisen R, Bawor M, Dennis B, de Souza R, Thabane L, et al. Association between body mass index and suicidal behaviors: a systematic review protocol. Syst Rev. 2015;4:52.

[70] Powell DJ, Liossi C, Moss-Morris R, Schlotz W. Unstimulated cortisol secretory activity in everyday life and its relationship with fatigue and chronic fatigue syndrome: a systematic review and subset meta-analysis. Psychoneuroendocrinology. 2013;38:2405-22.

[71] Protogerou C, Hagger M. A checklist to assess the quality of survey studies in psychology. Methods in Psychology. 2020;20:e100031.

[72] Puzzolo E, Stanistreet D, Pope D, Bruce N, Rehfuess E. Factors influencing the large-scale uptake by households of cleaner and more efficient household energy technologies. 2013.

[73] Reeves BC, Wells GA, Waddington H. Quasi-experimental study designs series—paper 5: a checklist for classifying studies evaluating the effects on health interventions—a taxonomy without labels. Journal of Clinical Epidemiology. 2017;89:30-42.

[74] Rodriguez AJ, Nunes Vdos S, Mastronardi CA, Neeman T, Paz-Filho GJ. Association between circulating adipocytokine concentrations and microvascular complications in patients with type 2 diabetes mellitus: A systematic review and meta-analysis of controlled cross-sectional studies. J Diabetes Complications. 2016;30:357-67.

[75] Rojanaworarit C. Misleading epidemiological and statistical evidence in the presence of Simpson's paradox: an illustrative study using simulated scenarios of observational study designs. Journal of Medicine and Life. 2020;13:37.

[76] Ross AB, Shertukde SP, Staffier KL, Chung M, Jacques PF, McKeown NM. The relationship between whole-grain intake and measures of cognitive decline, mood, and anxiety–a systematic review. Advances in Nutrition. 2023.

[77] Sedgwick P. Bias in observational study designs: cross sectional studies. BMJ. 2015;350:h1286.

[78] Sedgwick R, Epstein S, Dutta R, Ougrin D. Social media, internet use and suicide attempts in adolescents. Curr Opin Psychiatry. 2019;32:534-41.

[79] Seehra J, Pandis N, Koletsi D, Fleming PS. Use of quality assessment tools in systematic reviews was varied and inconsistent. J Clin Epidemiol. 2016;69:179-84 e5.

[80] Shamliyan T, Kane RL, Dickinson S. A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. J Clin Epidemiol. 2010;63:1061-70.

[81] Shamliyan TA, Kane RL, Ansari MT, Raman G, Berkman ND, Grant M, et al. Development quality criteria to evaluate nontherapeutic studies of incidence, prevalence, or risk factors of chronic diseases: pilot study of new checklists. J Clin Epidemiol. 2011;64:637-57.

[82] Silva LC, Ordúñez P, Rodríguez MP, Robles S. A tool for assessing the usefulness of prevalence studies done for surveillance purposes: the example of hypertension. Revista Panamericana de Salud Pública. 2001;10:152-60.

[83] Stone JC, Glass K, Clark J, Munn Z, Tugwell P, Doi SAR. A unified framework for bias assessment in clinical research. Int J Evid Based Healthc. 2019;17:106-20.

[84] Takahashi N, Hashizume M. A systematic review of the influence of occupational organophosphate pesticides exposure on neurological impairment. BMJ open. 2014;4:e004798.

[85] Tonia T, Buitrago-Garcia D, Peter N, Mesa-Vieira C, Li T, Furukawa TA, et al. A tool to assess risk of bias in studies estimating the prevalence of mental health disorders (RoB-PrevMH). medRxiv. 2023:2023.02. 01.23285335.

[86] van der Windt DA, Thomas E, Pope DP, de Winter AF, Macfarlane GJ, Bouter LM, et al. Occupational risk factors for shoulder pain: a systematic review. Occup Environ Med. 2000;57:433-42.

[87] van Dijk GM, Maneva M, Colpani V, Dhana K, Muka T, Jaspers L, et al. The association between vasomotor symptoms and metabolic health in peri- and postmenopausal women: a systematic review. Maturitas. 2015;80:140-7.

[88] Viswanathan M, Ansari MT, Berkman ND, Chang S, Hartling L, McPheeters M, et al. Assessing the Risk of Bias of Individual Studies in Systematic Reviews of Health Care Interventions. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville (MD)2008.

[89] Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. J Clin Epidemiol. 2012;65:163-78.

[90] Viswanathan M, Berkman ND, Dryden DM, Hartling L. Assessing Risk of Bias and Confounding in Observational Studies of Interventions or Exposures: Further Development of the RTI Item Bank. Rockville (MD)2013.

[91] Wagg E, Blyth FM, Cumming RG, Khalatbari-Soltani S. Socioeconomic position and healthy ageing: A systematic review of cross-sectional and longitudinal studies. Ageing Res Rev. 2021;69:101365.

[92] Wang X, Cheng Z. Cross-Sectional Studies: Strengths, Weaknesses, and Recommendations. Chest. 2020;158:S65-S71.

[93] Ward H. A Primer for Evaluating the Quality of Studies on Environmental Health: Critical Appraisal of Cross-Sectional Studies. National Collaborating Centre for Environmental Health; 2011.

[94] Ximenes RA, Araujo TV. [Internal validity in cross-sectional studies: comments based on an investigation on the association between socioeconomic factors and schistosomiasis]. Cad Saude Publica. 1995;11:118-27.

[95] Yang JC, Yang ZR, Yu SQ, Zhan SY, Sun F. [Introduction on 'assessing the risk of bias of individual studies' in systematic review of health-care intervention programs revised by the Agency for Healthcare Research and Quality]. Zhonghua Liu Xing Bing Xue Za Zhi. 2019;40:106-11.

[96] Yu IT, Tse SL. Workshop 6--sources of bias in cross-sectional studies; summary on sources of bias for different study designs. Hong Kong Med J. 2012;18:226-7.

[97] Yusuf E, Kortekaas MC, Watt I, Huizinga TW, Kloppenburg M. Do knee abnormalities visualised on MRI explain knee pain in knee osteoarthritis? A systematic review. Annals of the rheumatic diseases. 2011;70:60-7.

[98] Zeng X, Zhang Y, Kwong JS, Zhang C, Li S, Sun F, et al. The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. J Evid Based Med. 2015;8:2-10.

[99] Ross LE, Grigoriadis S, Mamisashvili L, Koren G, Steiner M, Dennis CL, et al. Quality assessment of observational studies in psychiatry: an example from perinatal psychiatric research. Int J Methods Psychiatr Res. 2011;20:224-34.

[100] Perez-Rios M, Rey-Brandariz J, Galan I, Fernandez E, Montes A, Santiago-Perez MI, et al. Methodological guidelines for the estimation of attributable mortality using a prevalence-based method: the STREAMS-P tool. J Clin Epidemiol. 2022;147:101-10.

[101] Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J. Methodological index for non-randomized studies (minors): development and validation of a new instrument. ANZ J Surg. 2003;73:712-6.

[102] Kesmodel US. Cross-sectional studies – what are they good for? Acta Obstetricia et Gynecologica Scandinavica. 2018;97:388-93.

[103] Sedgwick P. Bias in observational study designs: cross sectional studies. BMJ (Online). 2015;350:h1286-h.

[104] Wang X, Cheng Z. Cross-Sectional Studies: Strengths, Weaknesses, and Recommendations. Chest. 2020;158:S65-S71.

[105] Ximenes RAdA, Araújo TVBd. Validade interna em estudos de corte transversal: reflexões a partir de uma investigação sobre esquistossomose mansônica e condições socioeconômicas. Cadernos de Saúde Pública. 1995;11.

[106] Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2000.

[107] Kelly SE, Greene-Finestone LS, Yetley EA, Benkhedda K, Brooks SPJ, Wells GA, et al. NUQUEST-NUtrition QUality Evaluation Strengthening Tools: development of tools for the evaluation of risk of bias in nutrition studies. Am J Clin Nutr. 2022;115:256-71.

[108] Statistics Canada, Statistics: Power from Data!, September 2021. Available at: https://www150.statcan.gc.ca/n1/edu/power-pouvoir/toc-tdm/5214718-eng.htm (Last accessed 02 January 2024).

[109] Babic A, Pijuk A, Brázdilová L, Georgieva Y, Raposo Pereira MA, Poklepovic Pericic T, et al. The judgement of biases included in the category "other bias" in Cochrane systematic reviews of interventions: a systematic survey. BMC Medical Research Methodology. 2019;19:77.

[110] Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet. 2009;374:86-9.

[111] Crowe M, Sheppard L. A review of critical appraisal tools show they lack rigor: Alternative tool structure is proposed. J Clin Epidemiol. 2011;64:79-89.

[112] Stewart GL, Barrick MR. The folly of using research lacking rigor as a call to action. Academy of Management Learning & Education. 2012;11:278-84.

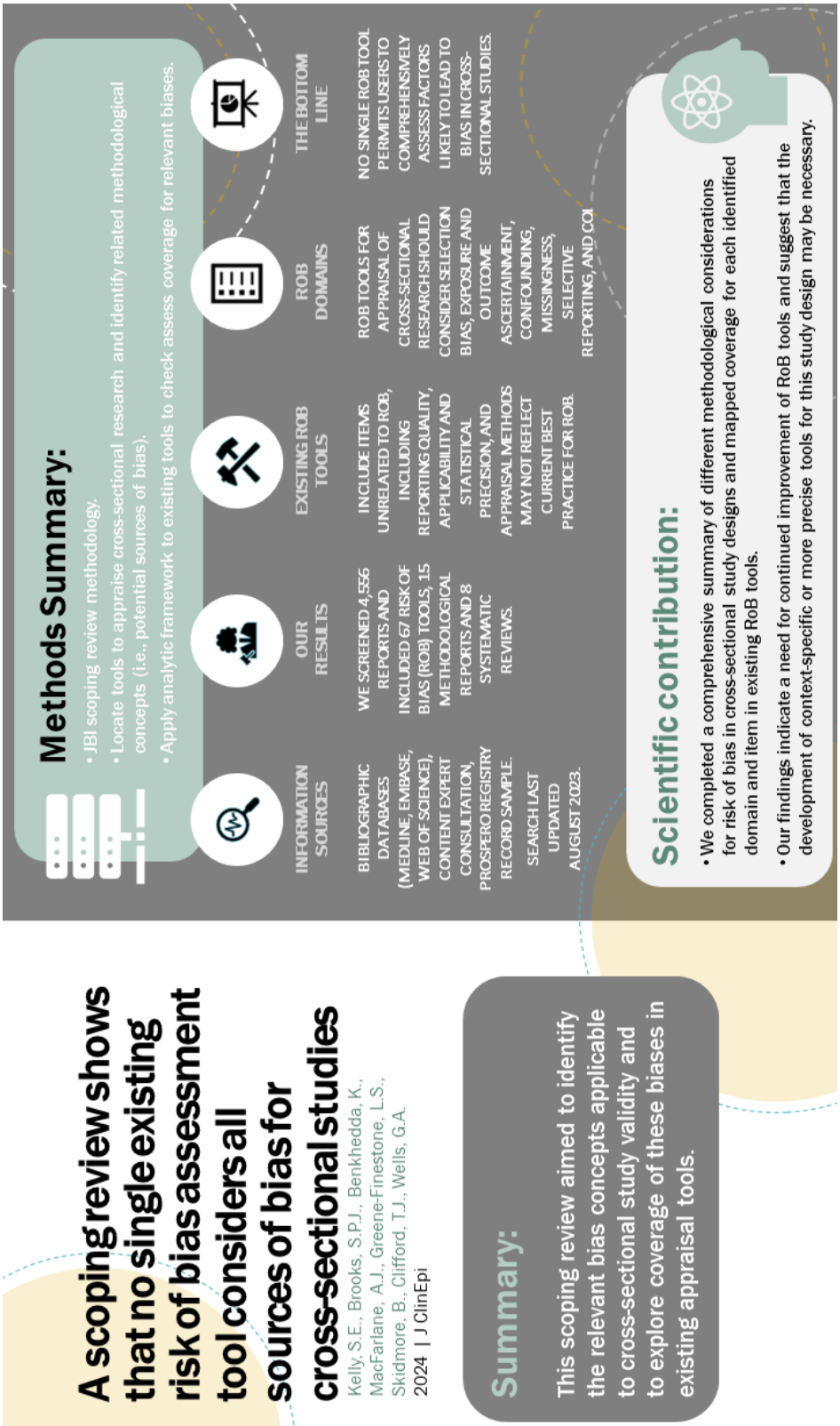
[113] Glasziou P, Chalmers I. Research waste is still a scandal—an essay by Paul Glasziou and Iain Chalmers. BMJ. 2018;363:k4645.

[114] Wang Y, Ghadimi M, Wang Q, Hou L, Zeraatkar D, Iqbal A, et al. Instruments assessing risk of bias of randomized trials frequently included items that are not addressing risk of bias issues. Journal of Clinical Epidemiology. 2022;152:218-25.

[115] Campbell F, Tricco AC, Munn Z, Pollock D, Saran A, Sutton A, et al. Mapping reviews, scoping reviews, and evidence and gap maps (EGMs): the same but different- the "Big Picture" review family. Syst Rev. 2023;12:45.

[116] Hirt J, Ewald H, Briel M, Schandelmaier S. Searching a methods topic: practical challenges and implications for search design. Journal of Clinical Epidemiology. 2024;166.

Graphical Abstract



Supplemental Appendix

PRISMA-ScR Reporting Checklist

Table A: Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
TITLE			
Title	1	Identify the report as a scoping review.	1
ABSTRACT			
Structured summary	2	Provide a structured summary that includes (as applicable): background, objectives, eligibility criteria, sources of evidence, charting methods, results, and conclusions that relate to the review questions and objectives.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known. Explain why the review questions/objectives lend themselves to a scoping review approach.	3-4
Objectives	4	Provide an explicit statement of the questions and objectives being addressed with reference to their key elements (e.g., population or participants, concepts, and context) or other relevant key elements used to conceptualize the review questions and/or objectives.	3-4
METHODS			
Protocol and registration	5	Indicate whether a review protocol exists; state if and where it can be accessed (e.g., a Web address); and if available, provide registration information, including the registration number.	3
Eligibility criteria	6	Specify characteristics of the sources of evidence used as eligibility criteria (e.g., years considered, language, and publication status), and provide a rationale.	4
Information sources*	7	Describe all information sources in the search (e.g., databases with dates of coverage and contact with authors to identify additional sources), as well as the date the most recent search was executed.	4-5
Search	8	Present the full electronic search strategy for at least 1 database, including any limits used, such that it could be repeated.	4-5
Selection of sources of evidence†	9	State the process for selecting sources of evidence (i.e., screening and eligibility) included in the scoping review.	5
Data charting process‡	10	Describe the methods of charting data from the included sources of evidence (e.g., calibrated forms or forms that have been tested by the team before their use, and whether data charting was done independently or in duplicate) and any processes for obtaining and confirming data from investigators.	5
Data items	11	List and define all variables for which data were sought and any assumptions and simplifications made.	5
Critical appraisal of individual sources of evidence§	12	If done, provide a rationale for conducting a critical appraisal of included sources of evidence; describe the methods used and how this information was used in any data synthesis (if appropriate).	N/A
Synthesis of results	13	Describe the methods of handling and summarizing the data that were charted.	5
RESULTS			
Selection of sources of evidence	14	Give numbers of sources of evidence screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally using a flow diagram.	6, Fig. 1
Characteristics of sources of evidence	15	For each source of evidence, present characteristics for which data were charted and provide the citations.	Table 1
Critical appraisal within sources of evidence	16	If done, present data on critical appraisal of included sources of evidence (see item 12).	N/A
Results of individual sources of evidence	17	For each included source of evidence, present the relevant data that were charted that relate to the review questions and objectives.	6-11
Synthesis of results	18	Summarize and/or present the charting results as they relate to the review questions and objectives.	6-14
DISCUSSION			
Summary of evidence	19	Summarize the main results (including an overview of concepts, themes, and types of evidence available), link to the review questions and objectives, and consider the relevance to key groups.	15
Limitations	20	Discuss the limitations of the scoping review process.	15-16

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
Conclusions	21	Provide a general interpretation of the results with respect to the review questions and objectives, as well as potential implications and/or next steps.	16
FUNDING			
Funding	22	Describe sources of funding for the included sources of evidence, as well as sources of funding for the scoping review. Describe the role of the funders of the scoping review.	No funding.

JBI = Joanna Briggs Institute; PRISMA-ScR = Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews.

* Where *sources of evidence* (see second footnote) are compiled from, such as bibliographic databases, social media platforms, and Web sites.

† A more inclusive/heterogeneous term used to account for the different types of evidence or data sources (e.g., quantitative and/or qualitative research, expert opinion, and policy documents) that may be eligible in a scoping review as opposed to only studies. This is not to be confused with *information sources* (see first footnote).

‡ The frameworks by Arksey and O'Malley (6) and Levac and colleagues (7) and the JBI guidance (4, 5) refer to the process of data extraction in a scoping review as data charting.

§ The process of systematically examining research evidence to assess its validity, results, and relevance before using it to inform a decision. This term is used for items 12 and 19 instead of "risk of bias" (which is more applicable to systematic reviews of interventions) to include and acknowledge the various sources of evidence that may be used in a scoping review (e.g., quantitative and/or qualitative research, expert opinion, and policy document).

From: Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMAScR): Checklist and Explanation. Ann Intern Med. 2018;169:467–473. doi: 10.7326/M18-0850.

Information sources

A search strategy was designed to capture potentially relevant tools, methodological systematic reviews, systematic reviews of cross-sectional studies and peer-reviewed methodological guidance specific to cross-sectional studies. An experienced medical information specialist (BS) developed and tested the search strategies through an iterative process in consultation with the review team. We searched Ovid MEDLINE® ALL and Embase Classic+Embase on the Ovid platform, followed by a search of Web of Science (core collection). All searches were performed on November 18, 2021, and updated on August 6, 2023. The strategies used a combination of controlled vocabulary (e.g., “Cross-Sectional Studies”, “Bias”, “Checklists”) and keywords (e.g., “cross-sectional design”, “quality rating”, “tool”). Vocabulary and syntax were adjusted across the databases. Results were downloaded and deduplicated using EndNote version 9.3.3 (Clarivate Analytics) and uploaded to Covidence systematic review software (Veritas Health Innovation, Melbourne, Australia. Available at www.covidence.org). Specific details regarding the strategies appear in the Supplemental Appendix.

Search of PROSPERO records

To supplement the bibliographic database search, we reviewed ROB tools identified in systematic review protocols registered through The International Prospective Register of Systematic Reviews (PROSPERO). One reviewer (SB) used the search function to review registrations that considered cross-sectional studies (January 22, 2022). The '*Risk of bias (quality) assessment*' field in each record was used to identify potentially eligible RoB tools. Starting with the most recent registration, records were reviewed using the principle of diminishing returns, whereby a pragmatic decision to halt review of PROSPERO records was made based on agreement from the review team that further screening was unlikely to result in the identification of additional unique RoB tools. A second independent reviewer (SK) deduplicated results against the other identified tools.

Search strategy

2021 Nov 18
Updated 2023 Aug 6

MEDLINE

Database: Ovid MEDLINE(R) ALL <1946 to November 17, 2021>
Search Strategy:
---
1 Cross-Sectional Studies/ (398684)
2 ((cross-section* or crosssection*) adj2 (study or studies or design?)).ti,kw,kf. (48153)
3 1 or 2 [CROSS-SECTIONAL STUDIES] (412208)
4 Cross-Sectional Studies/st [standards] (41)
5 Research Design/st [standards] (12411)
6 Evidence-Based Medicine/mt, st [methods,standards] (9222)
7 or/4-6 [RESEARCH DESIGN STANDARDS] (21297)
8 3 and 7 [CROSS-SECTIONAL STUDIES - RESEARCH DESIGN STANDARDS] (320)
9 exp bias/ (72351)
10 bias$2.ti,kw,kf. (38345)
11 ((assess* or apprais* or estimat* or grade? or grading) adj3 (bias* or quality)).tw,kw,kf. (118427)
12 ((susceptib* or risk?) adj3 bias*).tw,kw,kf. (33354)
13 ((rate or rated or rates or rating) adj2 quality).tw,kw,kf. (7834)
14 Research Design/ (116838)
15 (research adj3 (appropriate* or design? or method* or protocol? or strateg* or valid*)).tw,kw,kf. (132941)
16 (methodological* adj3 (appropriate* or consider* or decision? or design? or issue? or quality or valid*)).tw,kw,kf. (33428)
17 ((design? or report* or study or studies) adj3 (appropriate* or quality or valid*)).tw,kw,kf. (160543)
18 "Reproducibility of Results"/ (431267)
19 reproducib*.ti,kw,kf. (18259)
20 (reliabilit* or reliabl*).ti,kw,kf. (60000)
21 (replicable or replicat*).ti,kw,kf. (58248)
22 valid*.ti,kw,kf. (142161)
23 confound*.ti,kw,kf. (5461)
24 Data Accuracy/ (3315)
25 ((calculat* or data or measur* or statistic*) adj3 (accurat* or accurac* or appropriate* or objective)).tw,kw,kf. (141906)

26 ((calculat* or data or measur* or statistic*) adj3 (correct* or exact* or precis*)).tw,kw,kf. (50759)
27 or/9-26 [QUALITY] (1285786)
28 3 and 27 [CSS - QUALITY] (38988)
29 Checklists/ (7539)
30 (checklist* or check list*).ti,kw,kf. (8020)
31 (tool or tools).ti,kw,kf. (100645)
32 (instrument or instruments).ti,kw,kf. (37759)
33 (framework? or frame work?).ti,kw,kf. (58639)
34 ((assess* or apprais* or grade? or grading or quality) adj3 (criteri* or guide or guideline? or guidance? or standard?)).ti,kw,kf. (7910)
35 or/29-34 [TOOLS/FRAMEWORKS] (215219)
36 28 and 35 [CSS - QUALITY ASPECTS - TOOLS] (1844)
37 8 or 36 [CSS - QUALITY ASPECTS - TOOLS/STANDARDS] (2136)
38 limit 37 to yr="2010-current" (1713)

Data items and collection

One reviewer (SK) extracted general information for each identified tool *(first author or organization, year, tool name, scope, development and testing*) and details related to the items, domains and assessment of RoB (*characteristics, response options, weighting or calculations, availability of guidance, declarations or funding*). We classified tools as checklists or scales, by their self-identified use (*RoB, quality assessment, critical appraisal*), and their intended study design application (*only cross-sectional, multi-study including cross-sectional, adapted for cross-sectional*). Information was extracted using standardized Google forms (Google Inc., 2023) and then transferred to Microsoft Excel (Microsoft Corp.,2018).

References

1. Jackson R, Ameratunga S, Broad J, Connor J, Lethaby A, Robb G, Wells S, Glasziou P, Heneghan C. *The GATE frame: critical appraisal with pictures*. BMJ Evidence-Based Medicine. 2006;11(2):35-38.

2. Babic A, Pijuk A, Brázdilová L, Georgieva Y, Raposo Pereira MA, Poklepovic Pericic T, Puljak L. *The judgement of biases included in the category "other bias" in Cochrane systematic reviews of interventions: a systematic survey*. BMC Medical Research Methodology. 2019 2019/04/11;19(1):77. doi:10.1186/s12874-019-0718-8.

Additional Tables

Table B: Risk of bias framework item coverage by tool

Author, Year	Tool Acronym	Item:																
		1						2	3		4		5			6	7	8
		1.1	1.2	1.3	1.4	1.5	1.6	2.1	3.1	3.2	4.1	4.2	5.1	5.2	5.3	6.1	7.1	8.1
Crowe, 2011	CCAT	X	-	X	-	-	X	X	P	X	X	X	-	X	-	-	-	-
National Heart Lung and Blood Institute, 2021	NHLBI	X	-	X	-	X	-	X	X	X	X	X	-	-	-	-	-	-
Joanna Briggs Institute, 2023	JBI-XS	X	-	X	-	-	X	X	-	X	-	-	X	-	-	-	-	-
Joanna Briggs Institute, 2023	JBI-P	X	-	-	-	-	-	X	-	X	P	X	-	P	-	-	-	-
Slim, 2003	MINORS	-	-	-	-	X	-	-	X	-	X	-	-	-	-	-	-	-
Downes, 2016	AXIS	-	-	X	-	-	X	-	-	X	-	-	-	-	-	-	X	-
Viswanathan, 2008	AHRQ	X	-	-	-	-	-	X	X	X	X	X	-	X	P	X	-	-
Viswanathan, 2012	RTI	X	X	-	-	-	-	X	X	X	X	P	-	-	-	X	-	-
Boyle, 1998	N/A	-	-	-	-	-	X	P	-	P	-	-	-	-	-	-	-	-
Al-Jader, 2002	N/A	-	-	-	-	-	-	-	-	-	-	-	P	-	-	-	-	-
Hoy, 2012	N/A	-	-	X	-	-	X	X	-	X	-	-	-	-	-	-	-	-
Loney, 1998	N/A	X	-	-	-	X	-	-	X	X	-	-	-	-	-	-	-	-
Shamliyan, 2011	MORE	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Silva, 2001	N/A	-	-	-	-	-	X	X	-	X	P	P	-	-	-	-	-	-
Avis, 1994	N/A	-	X	X	X	-	X	P	P	P	-	X	X	-	-	-	-	-
Fowkes and Fulton, 1991	N/A	X	X	X	-	X	X	X	X	X	X	X	X	-	-	-	-	-
Kmet, 2004	N/A	-	-	-	-	-	X	X	-	X	X	-	-	-	-	-	-	-
Law, 1998	N/A	-	-	-	-	-	-	P	-	X	-	-	-	-	-	-	-	X
Margetts, 2002	N/A	-	-	X	-	-	-	X	-	X	-	-	-	-	-	-	-	-
Cardiff University, 2018	SURE	X	-	-	-	-	P	X	-	X	P	P	-	P	-	-	X	-
Academy of Nutrition and Dietetics, 2022	ANDEAL	X	X	X	-	P	X	P	X	X	X	X	-	-	-	-	X	-
Al Asmri, 2023	MERSQI	-	-	X	-	-	-	P	-	P	-	-	-	-	-	-	-	-
Jackson, 2006	GATE- appraise	-	-	P	-	-	P	-	X	-	-	-	P	-	-	-	-	X
Macfarlane 2001	N/A	-	-	X	-	-	X	-	-	X	-	-	-	-	-	-	-	-
National Toxicology Program, USDHSS, 2019	OHAT	X	X	-	-	X	-	X	X	X	X	-	P	-	-	X	-	X
Ohadike, 2016	N/A	-	-	-	-	X	X	-	-	-	-	-	-	-	-	-	-	-
van der Windt, 2000	N/A	-	X	X	P	-	-	X	X	-	X	-	-	-	-	-	-	-
Powell, 2013	N/A	-	-	-	-	-	-	X	-	-	X	SR	-	X	-	-	-	-

Author, Year	Tool Acronym	Item:																
		1						2	3		4		5			6	7	8
		1.1	1.2	1.3	1.4	1.5	1.6	2.1	3.1	3.2	4.1	4.2	5.1	5.2	5.3	6.1	7.1	8.1
Yusef, 2011	N/A	-	-	X	-	-	-	P	X	X	X	P	-	-	-	-	-	-
Puzzolo, 2013	LQAT	-	-	-	-	-	X	-	-	P	X	-	-	-	-	-	-	-
Busse, 2011	CLARITY	-	-	X	-	-	X	P	-	X	-	-	X	-	-	-	-	-
Thomy, 2023	RoB-PrevMH	-	-	X	-	-	X	X	-	X	-	-	-	-	-	-	-	-
National Collaborating Centre on Environmental Health, 2011	NCCEH-XS	-	X	-	-	-	X	X	-	X	-	-	-	-	-	-	X	X
Kennedy, 2019	N/A	-	X	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-
DeGroot, 2014	QATQS	-	-	X	-	-	X	-	X	-	X	SR	-	-	-	-	-	-
Stone, 2019	MASTER	-	X	X	-	-	-	X	X	X	X	X	X	X	-	-	X	-
Genaidy, 2007	EAI	-	-	X	-	-	-	-	-	-	-	X	-	-	-	X	-	-
Nudelman and Otto, 2020	ROBUST	-	-	X	-	-	X	P	-	P	-	-	-	P	-	-	-	-
Ross, 2023	modified NUQUEST	-	X	X	-	-	-	X	X	X	X	X	-	-	-	-	-	-
Ross, 2011	SAQOR	-	-	-	-	-	X	X	-	P	X	-	-	-	-	-	-	-
Protogerou, 2022	Q-SSP	-	P	-	-	-	-	P	-	P	-	-	P	-	-	-	-	-
Perez-Rios, 2022	STREAMS-P	-	-	X	-	-	X	P	-	-	-	-	-	-	-	-	-	-
Hong, 2018	MMAT -QNR	-	-	-	-	-	X	X	-	X	X	-	-	-	-	-	-	-
Hong, 2018	MMAT-QD	X	-	X	-	-	X	P	-	P	-	-	-	-	-	-	-	-
University of Oxford, Centre for Evidence-Based Medicine, 2019	CAT Manager App	-	-	-	-	-	-	X	-	X	-	-	-	-	-	-	-	-
University of Oxford, Centre for Evidence-Based Medicine, 2014	CEBM	-	-	X	-	-	X	-	-	-	X	-	-	-	-	-	-	-
Modesti, 2016	Modified NOS	-	-	X	-	-	X	X	X	X	X	-	-	-	-	-	-	-
Alshabanat, 2015	Modified NOS	-	-	-	-	-	X	X	X	X	X	-	-	-	-	-	-	-
Bawor, 2014	Modified NOS	-	X	-	-	-	X	-	-	X	X	X	X	-	-	-	-	-
Herzog, 2013	Modified NOS	-	-	X	-	-	X	X	X	-	X	-	-	-	-	-	-	-
Kunutsor, 2015	Modified NOS	-	-	X	-	-	X	X	X	X	X	SR	-	-	-	-	-	-
Lu, 2021	Modified NOS	-	-	-	-	-	X	X	X	X	X	-	-	-	-	-	-	-
Moskalewicz, 2020	Modified NOS	-	-	X	-	-	-	X	X	-	X	SR	-	-	-	-	-	-
Murray, 2018	Modified NOS	P	X	-	-	-	X	X	X	X	X	X	-	-	-	-	-	-
Patra, 2015	Modified NOS	-	-	X	-	-	X	X	X	X	X	SR	-	-	-	-	-	-
Perera, 2015	Modified NOS	X	X	-	-	-	X	-	X	X	X	X	X	X	-	-	-	-
Rodriguez, 2016	Modified NOS	-	-	-	-	-	X	-	-	-	X	-	-	-	-	-	-	-
Takahashi, 2014	NOS for XS	P	X	-	-	-	X	X	X	X	X	X	-	-	-	-	-	-
van Dijk, 2015	Modified NOS	-	-	X	-	-	X	X	X	X	X	SR	-	-	-	-	-	-
Wagg, 2021	Modified NOS	-	-	X	-	-	X	X	-	X	X	-	-	-	-	-	-	-

Author, Year	Tool Acronym	Item:																
		1						2	3		4		5			6	7	8
		1.1	1.2	1.3	1.4	1.5	1.6	2.1	3.1	3.2	4.1	4.2	5.1	5.2	5.3	6.1	7.1	8.1
Sedgwick, 2019	NOS for XS	P	X	-	-	-	X	X	X	X	X	X	-	-	-	-	-	-
Cincinnati Childrens Hospital Medical Centre, 2023	LEGEND-XS (INTERVENTION)	-	-	-	-	-	-	-	-	X	-	X	P	-	-	-	X	-
Cincinnati Childrens Hospital Medical Centre, 2023	LEGEND-XS (ETIOLOGY RISK FACTORS)	-	-	-	-	-	-	-	-	X	-	X	P	-	-	-	X	-
Cincinnati Childrens Hospital Medical Centre, 2023	LEGEND-XS (PREVALENCE)	-	X	-	-	-	-	P	-	X	-	-	-	-	-	-	X	-

Table footnotes: Identifying domains and items:

1: Selection domain
  1.1: Item = Appropriateness of eligibility criteria
  1.2: Item = Comparability of exposure group vs. Comparison group
  1.3: Item = Non-response rate
  1.4: Item = Mechanism for non-response rate
  1.5: Item = Recruitment time frame
  1.6: Item = Representativeness of sample to target population
2: Exposure domain
  2.1: Item = Validity and reliability of exposure measurement
3: Outcome domain
  3.1: Item = Blinding of the research staff
  3.2: Item = Validity and reliability of the outcome measurement
4: Confounding domain
  4.1: Item = Accounting for confounding
  4.2: Item = Description of confounding variables
5: Missingness domain
  5.1: Item = Amount of missingness
  5.2: Item = Handling of missingness
  5.3: Item = Mechanism for missingness
6: Selective reporting domain
  6.1: Item = Selective reporting of outcomes
7: Conflict of interest domain
  7.1: Item = Conflict of interest (e.g., funding)
8: Other bias domain
  8.1: Item = Any other concerns regarding the internal validity of the study

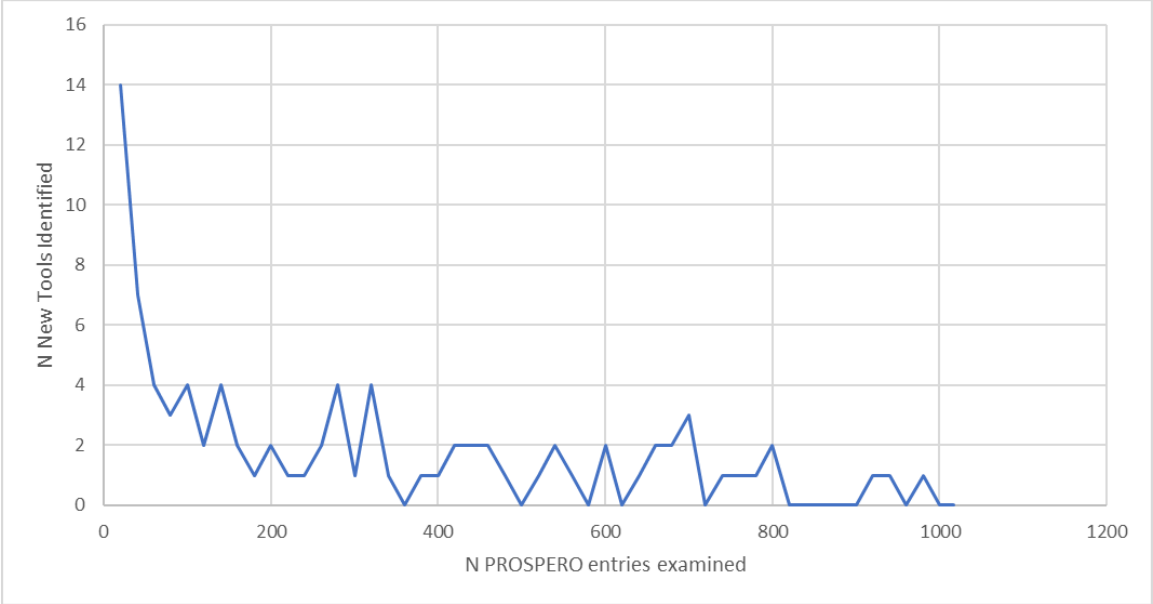
Table C: Risk of bias framework items addressed by subgroups of
a) multi-study tools (n=22)
b) cross-sectional specific tools (n=42), or
c) tools focused on RoB (n=13) (versus quality/critical appraisal) (framework adapted from Wang et al. (2019)(1)).

BIAS DOMAIN:	BIAS CONCEPT:	Addressed by multi-study tools, *n (%)*	Addressed by XS only tools, *n (%)*	RoB Tools, *n (%)*
SELECTION				
		N=22	N=42	N=13
	Appropriateness of eligibility criteria	6 (27.3)	7 (16.7)	6 (46.2)
	Comparability of exposure group vs. comparison group	8 (36.4)	8 (19.0)	6 (46.2)
	Non-response rate	11 (50.0)	19 (45.2)	8 (61.5)
	Non-response rate - mechanism	1 (4.5)	0 (0.0)	0 (0.0)
	Recruitment time frame	3 (13.6)	3 (7.1)	1 (7.8)
	Representativeness of sample to target population	13 (59.1)	26 (61.9)	7 (53.9)
EXPOSURE				
	Validity and reliability of exposure measurement	11 (50.0)	23 (54.8)	9 (69.2)
OUTCOME				
	Blinding of the research staff	8 (36.4)	18 (42.9)	6 (46.2)
	Validity and reliability of the outcome measurement	13 (59.1)	28 (66.7)	11 (84.6)
CONFOUNDING				
	Accounting for confounding variables	13 (59.1)	22 (52.4)	6 (46.2)
	Description of confounding variables	9 (40.9)	8 (19.0)	5 (38.5)
MISSINGNESS				
	Amount of missingness	3 (13.6)	4 (9.5)	3 (23.1)
	Handling of missingness	2 (9.1)	3 (7.1)	2 (15.4)
	Mechanism for missingness	0 (0.0)	0 (0.0)	0 (0.0)
SELECTIVE REPORTING				
	Selective reporting of outcomes	2 (9.1)	2 (4.8)	3 (23.1)
CONFLICT OF INTEREST				
	Conflict of interest (e.g., funding)	2 (9.1)	6 (14.3)	2 (15.4)
OTHER BIAS				
	Other bias or threats to internal validity	1 (4.5)	3 (7.1)	1 (7.8)

1. Wang Z, Taylor K, Allman-Farinelli M, Armstrong B, Askie L, Ghersi D, McKenzie J, Norris SL, Page MJ, Rooney A. A systematic review: Tools for assessing methodological quality of human observational studies. 2019.

Additional Figures

Figure A: Tools compiled through PROSPERO record checking



1. Wang Z, Taylor K, Allman-Farinelli M, Armstrong B, Askie L, Ghersi D, McKenzie J, Norris SL, Page MJ, Rooney A. *A systematic review: Tools for assessing methodological quality of human observational studies*. 2019.

Chapter 5
Development and validation of nutrition-specific risk of bias tools for cross-sectional studies

5.1. Chapter Overview

This chapter outlines the rationale and scientific approach to be used for the development, testing and establishment of validity and reliability for two nutrition-specific risk of bias (RoB) tools applicable to cross-sectional research designs. This research is ongoing[4] with planned completion in 2025. One tool will focus on descriptive cross-sectional studies (i.e., those studies aiming to provide estimates of prevalence of disease, traits, knowledge or behaviour) and the second on analytic cross-sectional studies (i.e., those studies aiming to assess associations between different parameters under study). The intent is to include the tools, and those developed in Chapter 3 for randomized controlled trials, cohort studies and case-control studies, and assemble a suite of RoB tools in the NUtrition QUality Evaluation Strengthening Tools (NUQUEST) with integrated consideration for dietary exposure assessment.

Chapter 5, presented as a rationale and research strategy, addresses both thesis research questions:

How can nutrition-specific risk of bias assessment tools be developed to address the unique methodological challenges posed by studies in nutritional epidemiology?

[4] This research, while led by Shannon Kelly, is part of an ongoing collaborative effort within a multidisciplinary working group. As such, the project's timeline is influenced by collective processes and the availability of resources, which broadly extends beyond the student's direct control.

To what extent are the newly developed nutrition-specific risk of bias tools valid and reliable when applied across different study designs in nutritional research, and do any additional components developed alongside the nutrition-specific risk of bias tools enhance reliability or usability?

5.2. Objective

The goal of development is to create an evidence-based, valid and reliable nutrition-specific rating tools which can be applied to assess RoB in cross-sectional research studies for either prevalence/descriptive or analytic/association purposes.

5.2.1. Research Questions

In order to meet this objective, this portion of the thesis will address the following research questions:

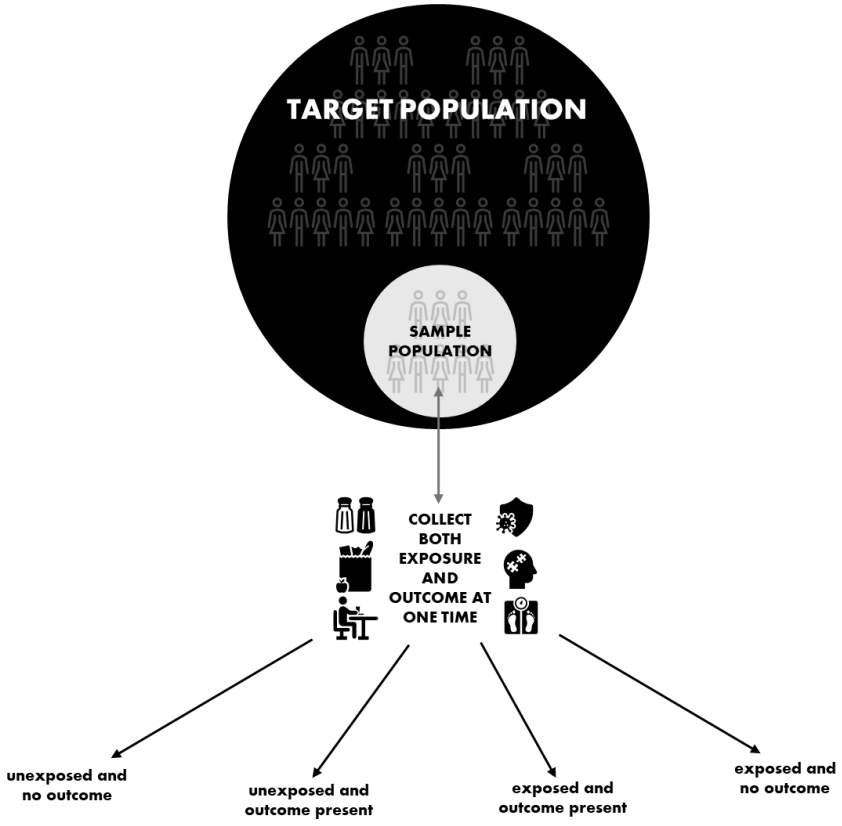
> *Question 1: Do the NUQUEST checklists for cross-sectional studies provide a valid approach to assess the potential for risk of bias?*
>
> *Question 2: Do the NUQUEST checklists for cross-sectional studies provide a reliable approach to assess the potential for risk of bias?*

5.3. Background and Rationale

Robust, well-designed research studies are required to determine causal relations between dietary exposures and human health outcomes. As in other disciplines, randomized controlled trials (RCTs) are engrained in the field of nutrition as the gold-standard (1). However, randomized studies are not always feasible or appropriate and thus, observational and correlational studies play an important role in establishing policies and recommendations in the absence of RCTs (2). Additionally, while RCTs offer many advantages, their limited external validity often leads to

the desire for “real world” studies to enhance generalizability (3). Cross-sectional studies are a well-established research design in nutritional epidemiology used to concurrently consider dietary exposures and human disease (Figure 5-1). They play an important role in describing the nature or extent of human disease by measuring prevalence of disease or outcomes and provide information about factors that may be associated with that prevalence, such as nutritional status (4). While noted limitations curtail the use of these studies to inform causal assessments, they do provide valuable evidence to inform design of subsequent research studies and inform planning of health services and public health measures and response.

Figure 5-1: Cross-sectional study design: Schematic applied to nutrition



As with any study design, some cross-sectional studies of dietary exposure and human health outcomes are planned and executed more robustly than others. Formal assessments of RoB in

research studies are conducted so that interpretation of research findings, whether from a single study or across a body of evidence, is derived relative to the perceived strengths and limitations of the research (5). Therefore, when considering research to inform policy or arrive at recommendations, some evidence may carry more weight than others, and a robust risk of bias tool can aid in those deliberations. When considering instruments available to assess the potential RoB in epidemiological research reports, a number of factors must be considered especially when multiple tools are available for the same use. Selection of an instrument for use involves consideration of the aim of the instrument, the study design(s) for which it is intended, its evaluated validity and reliability and the feasibility of use (6). Additionally, users may consider whether a suitable tool includes 'noise' from items or domains not related to RoB (7) or whether the tool falls in line with current best knowledge or practice.

A related scoping review of existing RoB tools applicable to cross-sectional research found that existing bias assessment tools for cross-sectional studies do not consistently or fulsomely consider RoB aspects relevant to cross-sectional study designs (8). Further, many were not assessed to be valid or reliable and no single tool comprehensively considered all aspects of bias relevant to cross-sectional research. A variety of tools (n=64) were located, yet very few were empirically evaluated for validity and reliability. Although criterion reliability and inter-rater agreement testing were more often assessed to demonstrate consistency in application, very few were empirically evaluated for validity using cross-sectional studies. Further, the reported use cases were narrow in scope, consisted of small number of studies or relied on a small number of raters with limited or undocumented testing.

There are several well-developed RoB tools available for assessment of trials and observational studies, including a set of nutrition-specific checklists for randomized controlled trials, cohort

studies and case-control studies (9), yet there is no well-tested or robust RoB tool for cross-sectional studies that adequately covers all relevant RoB domains and that gives adequate consideration to dietary exposure assessments. As dietary exposure assessments are a critical issue in evaluating human nutrition studies, the objective of this research is to develop, test, and establish psychometric properties for a new, valid and reliable tool intended for assessment of the RoB for cross-sectional epidemiological studies in the field of nutrition. The resulting tool and any accompanying guidance will additionally help to inform knowledge of bias unique to nutrition studies and impart skills required to comprehensively and consistently assess cross-sectional nutrition research studies.

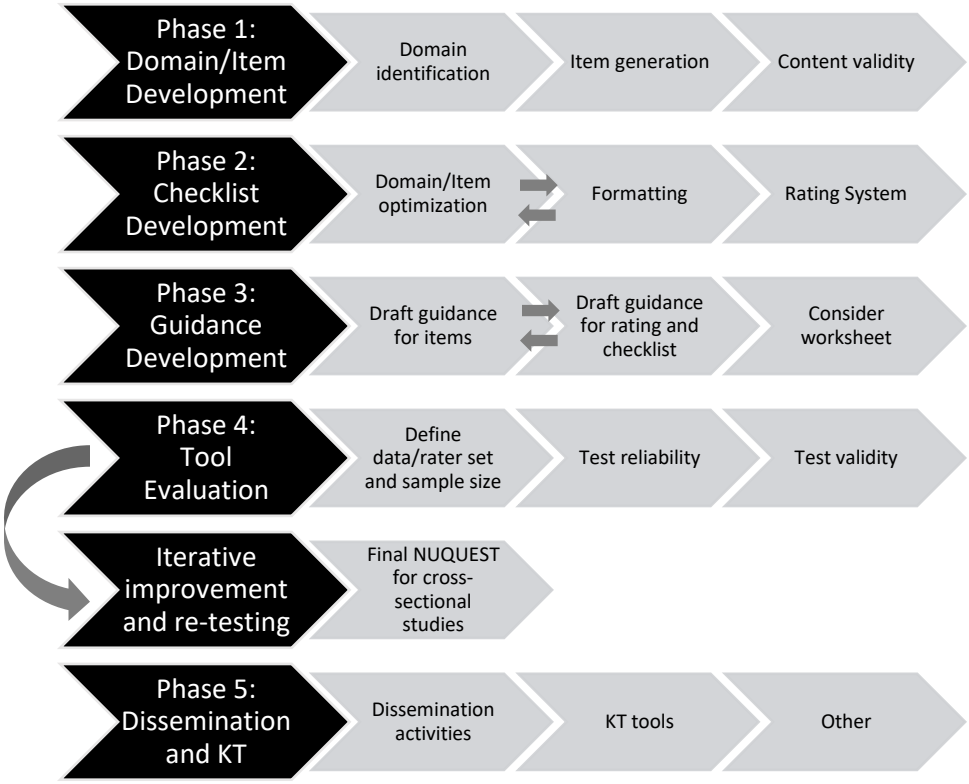
5.4. Methods

A multidisciplinary working group of six US and Canadian experts in the areas of nutrition science and epidemiology who were previously involved in the development of NUtrition QUality Evaluation Strengthening Tools (NUQUEST), a suite of RoB checklists for nutrition studies (9). For this work, the term "nutrition studies" would broadly apply to studies in which the focus is on the intake of a food substance (nutrient or other food component), food, dietary pattern or biomarker(s) of nutritional intake/status. The working group's overall plan was to develop a nutrition-specific RoB tool that captured the generic RoB concepts applicable to cross-sectional research while adding specificity for assessing nutrition issues in nutrition studies. The working group defined the goals, focus, and general approach for the project.

The development of the nutrition-specific RoB tools for cross-sectional studies followed the broad steps used previously for NUQUEST checklists for RCTs, cohorts and case-control studies with consideration included for new tool development (instead of adaptation of an existing tool) (9). The existing checklists provided a valuable reference for the structure of the sections, RoB

items and the response and rating system. While no existing tool is available for adaptation, the development goal was to ensure alignment in both look and feel, and application aspects, of the existing NUQUEST checklists. A 5-phase approach was used: 1) item development; 2) checklist development; 3) guidance development; 4) checklist and guidance evaluation; and 5) dissemination and knowledge translation (Figure 5-2). An iterative approach was used to refine and re-test the tool and guidance in phases prior to agreement on a final NUQUEST checklists for cross-sectional studies. The anticipated outcome is a tool (one each for prevalence/ descriptive and analytic/association research) that can be used to assess the potential RoB for cross-sectional nutrition studies accompanied by comprehensive user guidance. Similar to previous NUQUEST checklists, development of a worksheet to supplement the guidance, enhance consistency and structure the assessment of key variables was also developed.

Figure 5-3: Phased approach for tool development.



KT=knowledge translation.

Adapted from Boateng et al. (2018) (6).

5.4.1. Phase 1: Domain and Item Development

The objective of the item development phase was to come up with a set of initial sections (i.e., domains) and questions (i.e., items) that would constitute an eventual draft NUQUEST checklist for cross-sectional studies. In lay terms, this involves deciding on what questions to ask that measure potential bias in cross-sectional studies, identifying nutrition-specific items, eliminating any items not required and assessing how to conceptually group items into separate domains for evaluation. Despite use of the term 'checklist', NUQUEST is completed using a domain-based assessment, where assessments are made separately for each domain in the tool. Each domain contains one or more items. Individual steps within the phase include 1) domain and item identification; 2) domain and item generation; 3) nutrition-specific item identification; 4)

nutrition-specific item generation; and 5) consideration for the content validity of the identified domains and items.

Identify Generic Domains and Items

To formulate the generic bias domains and items the working group felt were essential to measure in cross-sectional studies, a comprehensive scoping review was conducted. The protocol, methods and detailed results for the scoping review are reported elsewhere (10, 11). Briefly, the scoping review endeavored to identify reports published of any related methodological concepts or tools used to consider RoB in health research reported to be cross-sectional including consideration for both descriptive/ prevalence or analytic/association research.

Development of Generic Domains and Items

Results from the scoping review were discussed and used to generate the dimensions of each potential domain and then the subsequent appropriate items that fit into the domains. An adapted analytic framework (Table 3-1:) (12) which resulted from the scoping review was used to serve as a starting point for developing a nutrition-specific RoB tool for cross-sectional studies.

Identify Nutrition-specific Items

Preliminary nutrition-specific appraisal issues were compiled from all available sources. The working group evaluated the appraisal issues and used each to develop appropriate items to add to a nutrition-specific evaluation domain. As previously applied in the development of the

Table 3-1:Analytic framework

Bias Domain	Bias Concept
Selection	Appropriateness of eligibility criteria

	Comparability of exposure group vs. comparison group
	Non-response rate
	Non-response rate - mechanism
	Recruitment time frame
	Representativeness of sample to target population
Exposure	Validity and reliability of exposure measurement
Outcome	Blinding of the research staff
	Validity and reliability of the outcome measurement
Confounding	Accounting for confounding variables
	Description of confounding variables
Missingness	Amount of missingness
	Handling of missingness
	Mechanism for missingness
Selective Reporting	Selective reporting of outcomes
Conflict of Interest	Conflict of interest (e.g., funding)

original NUQUEST checklists, nutrition-specific issues were defined as *"those related to nutrition exposure measurement and the interpretation of nutrition data"* and 'dietary exposure assessment' was broadly considered to "*encompass multiple components of exposure assessment including accuracy and completeness of exposure assessment methodologies, assurances of adherence to interventions and useful context information for interpreting dietary exposure data*" (9). These definitions were broadly considered, regardless of their specific applicability to cross-sectional studies, to maintain consistency across NUQUEST checklists. The group reviewed nutrition-specific tools and any associated guidance or use-case experience identified in the scoping review to identify nutrition-specific appraisal issues that they considered to not be fulsomely addressed or missing from existing RoB tools (11, 13).

The experience among the working group members in evaluating nutrition studies in a variety of policy and practice settings also proved informative to identify of items. We consulted external

content experts who identified practical issues with existing tools related to nutrition-specific study evaluation.

Generate Nutrition-specific items

The identification of nutrition-specific items was informed by items inadequately addressed or missing in a comprehensive scoping review of existing RoB tools. The working group, consisting of experienced nutrition scientists and research methodologists, then reviewed the items and provided practical nutrition-specific feedback, context and item suggestions. To refine and validate our list, we engaged external knowledge users and content experts in nutrition research and policy for an independent evaluation, who offered critical feedback and item additions or deletions. Additionally, we assessed the applicability of nutrition-specific questions from previous NUQUEST checklists by evaluating their use, relevance and effectiveness in detecting nutrition-specific bias. This process combined evidence review, expert input and applied use evaluation and separately considered the use of descriptive/prevalence or analytic/association applications. This enabled development of a robust set of nutrition-specific items tailored to address unique sources of bias in cross-sectional nutrition studies, increasing confidence that NUQUEST for cross-sectional studies is a RoB tool that is both evidence-based and grounded in practical expertise.

Pre-testing content validity

To assess the content validity of both generic and nutrition-specific risk of bias (RoB) items, we implemented a structured evaluation process combining expert judgment and empirical evaluation. To assess whether the generic and nutrition-specific items adequately measure the domains, the nutrition-specific domain and items were incorporated into the analytic framework output from the scoping review. The working group drafted brief text to accompany each item to

explain the rationale for inclusion in the NUQUEST checklists for cross-sectional studies and discussed until a set of final domains and items were agreed upon. The group circulated the domain and item pool to ten content experts in both methodology and nutrition science who were asked to consider the content relevance, representativeness and value of the assembled domains and items with consideration for their own area of application and broader application across the field of nutritional epidemiology. Experts were asked to consider the appropriateness, accuracy and interpretability of the items and provide feedback to accept, reject or modify. The working group considered the feedback and discussed until agreement was achieved on all items.

A broader group of target users in a range of policy, scientific and regulatory settings internationally was then consulted to evaluate each item in the context of their representative individual experience. Interviews with six end-users were conducted to evaluate face validity of the domains and generic and nutrition-specific items, to subjectively confirm that each item measures the constructs appropriately for the domain given the assessment aim. We developed a survey and experts were asked to rate each item on a Likert scale (e.g., 1 to 5, where 1 is not at all appropriate/relevant and 5 is extremely appropriate/relevant) based on their initial judgment of whether the generic and nutrition-specific items appear to measure constructs of bias as intended. In addition, experts were encouraged to provide open-ended feedback on each item. They were directed to consider the relevance, comprehensiveness, and appropriateness of the items in the context of their professional experience. The qualitative feedback was analyzed to identify common thought patterns, issues and suggestions for the new NUQUEST checklists. This helped the working group better understand the specific feedback of the expert respondents.

Based on the feedback and working group discussions, revisions were made to address any concerns about relevance, clarity, and appropriateness, ensuring that the items accurately

reflected a broad potential set of use or application contexts. Items identified as being least relevant to cross-sectional and/or nutrition studies were discussed by the working group and agreement was achieved prior to item or domain exclusion.

5.4.2. Phase 2: Checklist Development

In the tool development phase, a draft NUQUEST checklist for cross-sectional studies was drafted for both descriptive/prevalence and analytic/association uses. The assembled domain and item pool for generic RoB and nutrition-specific appraisal issues. During this phase the NUQUEST format, response options and the rating system were confirmed or adapted as necessary. A decision tool was also developed to help users select the appropriate checklist for descriptive/prevalence and analytic/association research.

Formatting and rating system

The draft checklists were structured using the NUQUEST format and response options for each item, domain and the overall rating for the study. The current rating system for NUQUEST includes up to 5 item response options (*yes/probably yes/probably no/no/not applicable*) and the "*Not applicable*" response option was only incorporated for items where the working group deemed a response would not be pertinent. Individuals applying NUQUEST to assess cross-sectional research translate their judgment for each item and domain rating into the provided response options. The available domain rating options are: *good* (+), *neutral* (0) and *poor* (-). A rating of *good* indicates a completed assessment of low RoB, *neutral* indicates moderate RoB and *poor* indicates high RoB. The overall ratings for each section are carried forward to the Overall Study Rating section and used when considering the final assessment of the study (response options: poor, neutral, or good). Guidance for determining overall ratings, including the consideration of the relative importance of specific items within the context of the study

design, was edited to reflect the updated format and response options. The formatted version of the checklist with the rating system and all included domains and items was considered the test-ready draft of NUQUEST.

5.4.3. Phase 3: Guidance Development

The working group drafted *de novo* item-specific guidance for all items in the NUQUEST cross-sectional study checklist. Draft guidance was reviewed and revised iteratively and pertinent nutrition-specific examples relevant to the study design and potential variation in use or context were included. Two individuals (non-working group members) with context expertise in nutrition and epidemiology reviewed the draft guidance, including the generic and nutrition-specific items and guidance. Guidance and related nutrition examples were refined for clarity and included in the test-ready draft of the NUQUEST checklist for cross-sectional studies.

5.4.4. Phase 4: Tool Evaluation

Sample Size Estimation

A psychometrically sound tool measuring all relevant latent constructs should be generalizable, meaning that the items the tool is measuring should be transferable to all nutrition studies using a cross-sectional design (14). The working group assembled a comprehensive and representative convenience sample of 30 analytic and 30 descriptive studies for testing of the NUQUEST checklists for cross-sectional studies using a systematic and robust approach.

Eligible studies were identified using published systematic reviews that included more than one cross-sectional study and were focused on human nutrition-related topics in humans (e.g., dietary intake, nutritional status, dietary patterns, nutrition interventions). Systematic reviews were published after 01 January 2018, reported the citations for any included cross-sectional studies, assessed the RoB for the studies included and reported the results for the study-level assessments

by item/domain. Reviews reporting only a summary or overall RoB rating or score for the assessed cross-sectional studies were not considered.

Eligible studies were descriptive or analytic cross-sectional studies with a focus on nutrition-related topics in humans. Studies presented as abstracts, in-progress studies, protocols for studies, methodological cross-sectional studies assessing new instruments and duplicate publications were excluded. Eligible studies were classified using their assessed cross-sectional approach (descriptive, analytic), study quality/RoB (high, moderate or low RoB) and broad content area (dietary intake, nutritional status, dietary patterns, nutrition interventions).

Data set

A random sample of 120 cross-sectional studies was selected from the comprehensive pool identified. The sample size was estimated based on a minimum of 5 studies applied for each item anticipated to be included in the tools (A range 10 to 12 items is anticipated to be included in each tool, and there are two tools). The studies were stratified by assessed methodological quality or RoB (good, moderate poor) and by cross-sectional design (descriptive and analytic studies) and the sample broadly included a range of nutrition-specific use cases (dietary intake, nutritional status, dietary patterns, nutrition interventions) and publication dates (older and newer). This sampling approach ensures a generalizable and representative distribution of study quality and content, thereby providing a robust foundation for assessing the RoB tools.

Raters

A set of raters were selected to apply the NUQUEST checklists for cross-sectional studies to the selected set of studies, in pairs. Each tool was applied to 60 studies. The raters were selected to have varied backgrounds and expertise, encompassing methodology, epidemiology and nutrition. Raters were also selected based on geographic location, ensuring a comprehensive evaluation of

the tools based on different use cases, perspectives and contextual understandings. This diversity in raters' experience and location is intended to add to the robustness and generalizability of the tool assessments.

Tests of Reliability

Reliability was evaluated to show the ability of the NUQUEST checklists for cross-sectional studies to measure the items and domains consistently (6, 15, 16). Reliability was demonstrated using inter-rater and intra-rater reliability.

a) *Inter-rater reliability:* To evaluate inter-rater reliability, two independent reviewers applied NUQUEST for cross-sectional studies to the selected studies (n=60 analytic, n=60 descriptive). The level of agreement between reviewers was assessed using Cohen's kappa (κ) for each item and domain rating in the tool and for the overall rating. A κ value of 0.60-0.79 was considered moderate, 0.80-0.89 as strong, and ≥0.90 as almost perfect agreement. A threshold for the coefficient alpha of .70 was used for acceptable reliability of the new checklist (17).

b) *Intra-rater reliability:* To determine intra-rater reliability, one reviewer reassessed 24 of the previously assessed studies (20%) after a two-week interval. The intra-rater agreement was repeated using the same criteria. (16).

Tests of Validity

Validity was evaluated to show the ability of the NUQUEST checklists for cross-sectional studies to measure the attributes of the nutrition-specific risk of bias constructs specific to descriptive/prevalence or analytic/association studies (15).

a) *Face Validity:* To assess face validity, six end users/reviewers were asked to provide a subjective assessment as to whether NUQUEST measures the construct of interest and to review NUQUEST for grammar, syntax, organization, appropriateness and whether it appears to flow logically. This provides valuable insight into how potential raters may interpret and respond to the items proposed.

b) *Content Validity:* Content validity was established by including all relevant generic RoB items from a comprehensive analytic framework and those specifically related to dietary exposure (16) beyond the identified generic RoB items. Feedback on initial content validity testing was used to inform the final checklist content and this was vetted by the same ten content experts in both methodology and nutrition science.

c) *Concurrent Validity:* The concurrent validity was established by comparing ratings from NUQUEST to ratings from the original systematic review (using two independent reviewers). Agreement for generic and nutrition-specific items, sections and overall was compared, and differences in agreement were investigated to ascertain whether the nutrition-specific section was a key contributor to the difference.

As there is no accepted 'gold standard' nutrition-specific assessment instrument, it was not possible to assess criterion validity (i.e., how well NUQUEST is in alignment with a gold standard comparator).

5.5. Plans for Continuation of Research Activities

Given the limitations of existing RoB tools identified in our scoping review, we are considering publishing the RoB tools, once tested, as a valid and reliable option for both nutrition-specific (NUQUEST) and generic RoB applications to cross-sectional research in any subject area

(QUEST). Plans for continuation research are expected to broadly follow the schedule outlined in Table 5-2. This research, while led by Shannon Kelly, is part of an ongoing collaborative effort within a multidisciplinary working group. As such, the project's timeline is influenced by collective processes and the availability of resources, which broadly extends beyond the student's direct control. Following completion of the tools in 2025, they will be disseminated in a peer-reviewed manuscript.

Table 5-2: Schedule for continuation of research

Phase	Task	Start	End	Milestone
Phase 1: Pre-testing Content Validity	Expert Review	Complete		Finalize initial item pool with revisions from expert input.
	End-User Interviews and Survey	In-progress	Jan-25	Integrate face validity feedback from end-users.
	Final Revisions for Content Validity	Feb-25	Feb-25	Finalized item set for NUQUEST checklists for cross-sectional studies.
	Phase 1 Completion	--	Mar-25	Final domains and items agreed upon following expert and end-user feedback.
Phase 2: Checklist Development	Drafting Checklist	Apr-25	Apr-25	Initial checklist draft with domain, items, and rating structure completed.
	Format and Rating System Confirmation	Apr-25	Apr-25	Checklist format and rating system refined.
	Development of Decision Tool	In-progress	Apr-25	Decision tool finalized to guide checklist selection for study types.
	Test-ready Draft	Apr-25	Jun-25	NUQUEST draft prepared for pilot testing.
	Phase 2 Completion	--	Jun-25	Draft and finalize the NUQUEST checklist with response and rating options confirmed.
Phase 3: Guidance Development	Drafting Guidance	Apr-25	Jul-25	Initial draft guidance for items and domains reviewed.
	Iterative Revisions with Expert Feedback	Jul-25	Jul-25	Finalize item-specific guidance with contextual examples.
	Finalize Test-ready Checklist	Jul-25	Jul-25	Checklist and guidance document ready for tool evaluation.

Phase	Task	Start	End	Milestone
	Phase 3 Completion		Jul-25	Final guidance document with nutrition-specific examples completed.
Phase 4: Tool Evaluation	Sample Size Estimation and Selection	Jul-25	Aug-25	Representative studies selected and classified for evaluation.
	Testing (Reliability and Validity)	Sept-25	Nov-25	Reliability (inter- and intra-rater) and validity assessments completed.
	Final Adjustments	Nov-25	Dec-25	Checklist revisions and final tool preparation completed based on evaluation results.
	Phase 4 Completion	--	Dec-25	Tool validation and reliability assessments completed for finalization.
Dissemination	Manuscript prepared for peer-reviewed publication	Apr-25	Dec-25	Manuscript prepared and ready to submit following completed NUQUEST tools.

5.6. References

1. Lichtenstein AH, Petersen K, Barger K, Hansen KE, Anderson CAM, Baer DJ, Lampe JW, Rasmussen H, Matthan NR. *Perspective: Design and Conduct of Human Nutrition Randomized Controlled Trials*. Adv Nutr. 2021 Feb 1;12(1):4-20. Epub 2020/11/18. doi:10.1093/advances/nmaa109. Cited in: Pubmed; PMID 33200182.

2. Lichtenstein AH, Yetley EA, Lau J. *Application of Systematic Review Methodology to the Field of Nutrition*. The Journal of Nutrition. 2008;138(12):2297-2306. doi:10.3945/jn.108.097154.

3. Deaton A, Cartwright N. *Understanding and misunderstanding randomized controlled trials*. Soc Sci Med. 2018 Aug;210:2-21. eng. Epub 20171225. doi:10.1016/j.socscimed.2017.12.005. Cited in: Pubmed; PMID 29331519.

4. Freudenheim JL. *Study design and hypothesis testing: issues in the evaluation of evidence from research in nutritional epidemiology*. Am J Clin Nutr. 1999 Jun;69(6):1315S-1321S. doi:10.1093/ajcn/69.6.1315S. Cited in: Pubmed; PMID 10359232.

5. Cumpston M, Li T, Page MJ, Chandler J, Welch VA, Higgins JP, Thomas J. *Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane Database Syst Rev. 2019 Oct 3;10:ED000142. Epub 2019/10/24. doi:10.1002/14651858.ED000142. Cited in: Pubmed; PMID 31643080.

6. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quinonez HR, Young SL. *Best Practices for Developing and Validating Scales for Health, Social, and Behavioral Research: A Primer*. Front Public Health. 2018;6:149. Epub 20180611. doi:10.3389/fpubh.2018.00149. Cited in: Pubmed; PMID 29942800.

7. Wang Y, Ghadimi M, Wang Q, Hou L, Zeraatkar D, Iqbal A, Ho C, Yao L, Hu M, Ye Z, Couban R, Armijo-Olivo S, Bassler D, Briel M, Gluud LL, Glasziou P, Jackson R, Keitz SA, Letelier LM, Ravaud P, Schulz KF, Siemieniuk RAC, Brignardello-Petersen R, Guyatt GH. *Instruments assessing risk of bias of randomized trials frequently included items that are not addressing risk of bias issues*. Journal of Clinical Epidemiology. 2022;152:218-225. doi:10.1016/j.jclinepi.2022.10.018.

8. Kelly SE, Brooks SPJ, Benkhedda K, MacFarlane AJ, Greene-Finestone LS, Skidmore B, Clifford TJ, Wells GA. *A scoping review shows that no single existing risk of bias assessment tool considers all sources of bias for cross-sectional studies*. Journal of Clinical Epidemiology. 2024;172. doi:10.1016/j.jclinepi.2024.111408.

9. Kelly SE, Greene-Finestone LS, Yetley EA, Benkhedda K, Brooks SPJ, Wells GA, MacFarlane AJ. *NUQUEST-NUtrition QUality Evaluation Strengthening Tools: development of tools for the evaluation of risk of bias in nutrition studies*. Am J Clin Nutr. 2022 Jan 11;115(1):256-271. doi:10.1093/ajcn/nqab335. Cited in: Pubmed; PMID 34605544.

10. (PRE-PRINT) Kelly SE, Brooks SP, Benkhedda K, MacFarlane AJ, Greene-Finestone LS, Skidmore B, Clifford TJ, Wells G. *A Scoping Review Shows that No Single Existing Risk of Bias*

Assessment Tool Considers All Sources of Bias for Cross-Sectional Studies. Available at SSRN 4731722. 2024.

11. Kelly SE, Benkhedda K, Brooks SPJ, MacFarlane AJ, Greene-Finestone LS, Skidmore B, Clifford TJ, Wells GA. *Risk of bias in cross-sectional studies: Protocol for a scoping review of concepts and tools*. MethodsX. 2024 2024/06/01/;12:102610. doi:https://doi.org/10.1016/j.mex.2024.102610.

12. Wang Z, Taylor K, Allman-Farinelli M, Armstrong B, Askie L, Ghersi D, McKenzie J, Norris SL, Page MJ, Rooney A. *A systematic review: Tools for assessing methodological quality of human observational studies*. 2019.

13. Kelly SE, Brooks SPJ, Benkhedda K, MacFarlane AJ, Greene-Finestone LS, Skidmore B, Clifford TJ, Wells GA. *A scoping review shows that no single existing risk of bias assessment tool considers all sources of bias for cross-sectional studies*. J Clin Epidemiol. 2024 Aug;172:111408. Epub 20240604. doi:10.1016/j.jclinepi.2024.111408. Cited in: Pubmed; PMID 38844117.

14. El-Den S, Schneider C, Mirzaei A, Carter S. *How to measure a latent construct: Psychometric principles for the development and validation of measurement instruments*. International Journal of Pharmacy Practice. 2020;28(4):326-336. doi:https://doi.org/10.1111/ijpp.12600.

15. DeVon HA, Block ME, Moyle-Wright P, Ernst DM, Hayden SJ, Lazzara DJ, Savoy SM, Kostas-Polston E. *A Psychometric Toolbox for Testing Validity and Reliability*. Journal of nursing scholarship. 2007;39(2):155-164. doi:10.1111/j.1547-5069.2007.00161.x.

16. Cronbach LJ, Meehl PE. *Construct validity in psychological tests*. Psychological bulletin. 1955;52(4):281-302. doi:10.1037/h0040957.

17. DeVellis RF, Thorpe CT. Scale development : theory and applications. Fifth edition. ed. Thousand Oaks, California: SAGE Publications, Inc.; 2022. ISBN: 9781544379340.

Chapter 6
Discussion

6.1. Chapter Overview

This thesis explored risk of bias (RoB) concepts and their application to nutrition research. Specifically, the NUtrition QUality Evaluation Strengthening Tools (NUQUEST) were developed and tested for randomized controlled trials, cohort studies and case-control studies (**Chapter 3**), a scoping review was used to explore RoB concepts and tools relevant to cross-sectional research (**Chapter 4**) and the development of NUQUEST for cross-sectional studies was planned and initiated (**Chapter 5**). Three manuscripts were produced as part of this thesis and all three are published in peer-reviewed journals. Chapter 5 presents a rationale and strategy for research in-progress for the two planned NUQUEST checklists for cross-sectional studies (one for descriptive/prevalence studies and a second for analytic/association studies) at the time of this thesis submission.

Chapter 6 presents a discussion for the thesis as a whole. The individual manuscripts offer a discussion specific to each individual study. More specifically, this final chapter is intended to:

- summarize the rationale and objective for the thesis
- provide a discussion of findings
- consider the implications of the thesis for practice and policy
- describe key considerations for user selection of RoB tools
- highlight the novelty of the thesis

- describe the strengths and weaknesses
- provide suggestions for future research in this area

6.2. Summary of overall rationale for the thesis

The objective of this thesis was to develop a set of reliable and valid RoB tools specifically tailored for nutrition research with the aim to enhance the rigor and reliability of RoB assessment by addressing unique methodological challenges inherent to the nutrition field.

To meet this aim, two research questions were addressed:

Research Question 1: How can nutrition-specific risk of bias assessment tools be developed to address the unique methodological challenges posed by randomized controlled trials (RCTs), cohort studies, case-control studies, and cross-sectional studies in nutritional epidemiology?

Research Question 2: To what extent are the newly developed nutrition-specific risk of bias tools valid and reliable when applied across different study designs in nutritional research, and do any additional components developed alongside the nutrition-specific risk of bias tools enhance reliability or usability?

6.3. Discussion of findings

The products that are included in this thesis are comprised of one protocol manuscript, two results manuscripts and a work-in-progress presented in three related chapters (Chapters 3 to 5). A brief summary of the key findings from each chapter in relation to the two thesis research questions is presented in Table 6-1.

.

Table 6-1: Summary of findings from the thesis

Research question	Chapter	Findings
How can nutrition-specific risk of bias assessment tools be developed to address the unique methodological challenges posed by randomized controlled trials, cohort studies, case-control studies, and cross-sectional studies in nutritional epidemiology?	Chapter 3	• Dietary exposure assessments are a critical issue in evaluating human nutrition studies, yet nutrition-specific criteria are not consistently included in existing bias assessment tools. • NUQUEST is grounded in the scientific concepts reported in a broadly used generic assessment instrument, while improving the usability for application to nutrition studies. • NUQUEST provides nutrition research users with robust, user-friendly tools to appraise studies. • NUQUEST for RCTs, cohort studies, and case-control studies are intended for the evaluation of individual nutrition studies or studies included in systematic reviews. • These tools may additionally support researchers in the improved design, conduct, and reporting of future nutrition research studies through their focus on nutrition-specific appraisal items. • The availability of these new tools will foster consistency among future reviews of nutrition studies while enabling future systematic refinement, as warranted, by experience and as science evolves. • These tools were developed for study designs commonly used in the field of nutrition and are applicable to single research studies or may be integrated into the processes for an evidence synthesis. Application of NUQUEST contributes to an improved understanding of the evidence base for nutrition or dietary recommendations and health policy decisions.
	Chapter 4	• Assessing cross-sectional studies involves unique risk of bias (RoB) considerations. • This thesis identified RoB tools tailored for cross-sectional studies and those designed for broad applicability across various study designs including cross-sectional studies • None of the 64 identified tools comprehensively address the potential biases pertinent to cross-sectional studies. • The findings indicate a need for continued improvement of RoB tools and suggest that the development of context-specific or more precise tools for this study design may be necessary.

Research question	Chapter	Findings
	Chapter 5	• There are unique RoB considerations for cross-sectional research and based on further consideration of the findings in Chapter 4, potential nutrition-specific aspects which are not comprehensively, or consistently, considered by existing tools. • Two additional NUQUEST checklists are in development – each considering different applications of cross-sectional studies (descriptive/prevalence and analytic/association) with a nutrition focus. • The methods for the design and testing of these *de novo* NUQUEST checklists are described along with the planned testing. • This research is in progress with results expected in 2025. This research is part of an ongoing collaborative effort within a multidisciplinary working group. As such, the project's timeline is influenced by collective processes and the availability of resources, which broadly extends beyond the direct control of the student for this thesis.
To what extent are the newly developed nutrition-specific risk of bias tools valid and reliable when applied across different study designs in nutritional research, and do any additional components developed alongside the nutrition-specific risk of bias tools enhance reliability or usability?	Chapter 3	• A set of reliable and valid nutrition-specific RoB tools to assess nutrition studies of commonly used study designs was developed. • Using a convenience sample of 45 studies, the overall interrater reliability for NUQUEST across all study designs (n = 45 studies) was substantial and moderate to almost perfect for the individual instruments for RCTs, cohort studies, and case-control studies. • NUQUEST integrates nutrition-specific criteria with generic risk of bias criteria from assessment tools with demonstrated reliability and validity. • When compared to published ratings, NUQUEST ratings for evaluated studies demonstrated high concurrent validity (perfect or near-perfect agreement). Where there was disagreement, the nutrition-specific component was a contributing factor in discerning exposure methodological issues.
	Chapter 4	• There are some valid and reliable tools considering RoB in cross-sectional studies, however, none comprehensively address all potential biases pertinent to cross-sectional studies.
	Chapter 5	• A set of reliable and valid nutrition-specific RoB tools to assess nutrition studies with a cross-sectional study design is being developed.

6.4. Disciplinary Implications

The development of NUQUEST as a set of reliable and valid nutrition-specific RoB tools has various potential implications for policy and practice. The availability of NUQUEST contributes to improved quality of nutrition research, facilitates evidence-based practice, influences health policy and guidelines, promotes accountability for research practice and reporting and advances the science of nutrition. NUQUEST permits a nuanced, nutrition-specific assessment of research study in a variety of research designs commonly used in the discipline. By having a comprehensive understanding of both the generic and nutrition-specific sources of bias, researchers, policymakers and practitioners can ensure that the conclusions formed on the evidence reflect this. The availability of a reliable, tailored tool focused on both the study design and nutrition-specific biases will optimistically reduce the amount of ad-hoc modifications to existing RoB tools and provide a trustworthy tool for application. The detailed guidance with examples and the worksheet accompanying NUQUEST facilitate the application and assist raters with considering exposures, outcomes and confounders in advance, which contributes to improved inter-rater reliability and ease of application. Having assessed the validity of NUQUEST, one can be confident that the ratings evaluated are based on appropriate and accurate assessments of relevant biases for nutrition studies.

By providing a comprehensive and consistent approach for users to consider the potential for bias in a research study, NUQUEST enables researchers to more effectively translate findings to the public which, in turn, will foster improvements in public trust pertaining to the legitimacy of dietary guidelines and health claims. Guideline and policy makers rely on nutrition research to inform their dietary recommendations related to human health, which, in turn, inform practice and individual- and population-level behaviours related to diet and nutrient intake. When nutrition research consumers can differentiate between different strengths and weaknesses of a

study of body of evidence, and there is accurate consideration for dietary intake tied to human health, the quality and the stability of the resulting recommendations and guidelines is improved. This impacts all citizens as practitioners and regulators uptake these guidelines as they are used to influence facets encountered daily such as food labels, public health campaigns and the design of food assistance or nutritional supplement programs as examples. The ability to consistently identify common sources of bias in existing nutrition studies indirectly influences the planning and execution of future research studies. The items and sections in NUQUEST can be used to develop methodologically robust nutrition studies as they will inform both the key aspects of study design and conduct, and the nutrition-specific aspects which may influence the findings. In addition, by using NUQUEST to identify risk of bias, researchers can identify more fulsomely where knowledge gaps exist and use this to prioritize new studies in areas where evidence is needed. NUQUEST also contributes to enhancing the public consumption of nutrition research. It is hoped that this contributes to more critical consideration of or reflection on the dietary claims reported in the media and by industry.

6.5. Choice of Risk of Bias Tools

Results from the rapid review done to inform NUQUEST for RCTs, cohort studies and case-control studies (Chapter 3) and the scoping review used to inform the ongoing development of the NUQUEST checklists for cross-sectional studies (Chapter 4) reveal the range and quantity of RoB instruments available across study designs. NUQUEST offers a nutrition-specific, valid and reliable approach to RoB, which fills in a specific gap identified in the field. While these are highly useful and discipline-specific tools, it is important to recognize that a wide range of RoB tools exist, and individuals may find value in selecting or using other tools that better suit their specific research needs, preferences, and the availability of resources (1). Cochrane recommends the use of comprehensive and rigorous assessment tools for RoB assessments (2, 3) Ultimately,

RoB is a subjective science highly reliable on the user applying the tool (1) and studies have shown that any one RoB tool, however methodologically rigorous, may not generate ratings which are comparable for the same primary research study (4-6). This may have consequences on meta-analyses and related sensitivity analyses (4), however, results in the literature are conflicting and this may be context-specific (5). This in turn, could influence how findings are interpreted and put into practice or policy. Importantly, users must know that all RoB tools will have strengths and weaknesses and it is essential to use this knowledge to inform the choice of a RoB tool that will align with the specific objectives, use case, resources and methodological requirements at hand (1). Users should consider a number of factors to ensure the selected RoB tool aligns with their own research needs and context (Table 6-2).

Although RoB assessment is a fundamental step for any high-quality SR, users may apply RoB tools outside of this context, for example to a single research study. In this case, individuals should consider the significance of any decision or relevant policy at stake. For example, when a single research study is available to inform a very impactful decision at-hand, the detailed assessment of the direction and magnitude of any potential bias may be extremely informative to the understanding of the study results. As such, RoB assessors may value the comprehensiveness and rigour of a tool focused on RoB which will give a detailed account of any identified biases impacting study outcomes of interest regardless of the resources required to complete the RoB assessment. Conversely, in resource or time-limited settings, where a systematic assessment of RoB for a higher volume of research is required, a more feasible and practical approach to RoB may be required. In resource-limited settings, RoB tools that require a resource-intensive approach may not be valued as highly despite their depth and precision, but evidence has shown that rigour in a tool may not always lead to reliable (7-9) or valid results for RoB (10). A study

applying the ROBINS-E tool, developed to accompany ROB2 and ROBINS-I and intended for human observational studies of exposures, identified practical concerns with the tool (10). As it was built on the foundations of the other RoB tools, some items in ROBINS-E are derived from questions pertaining to interventions, which are not feasible or appropriate to apply to observational studies. Authors identified deficiencies in the content which results in inadequate assessment of bias related to confounding, practical concerns that application was time consuming and difficult to apply across multiple study de3signs and noted that the tool fails to distinguish between studies with a single RoB in one domain and those with many RoBs across the domains assessed (10). One study comparing three tools (Newcastle-Ottawa Scale (NOS), Risk Of Bias In Nonrandomized Studies of Interventions (ROBINS-I), Quality of Cohort Studies (C-Qoh)) found that NOS was more conservative than other tools (RoB ratings are more likely to be lower RoB than higher) (5). Differences documented in ratings were at the high and low ends of the RoB spectrum for all three tools and users considered the lack of a domain-based approach in NOS a weakness

Feasibility and usability are practical concerns, as tools can vary in complexity, requiring different levels of expertise and time commitment. Researchers must also assess whether a tool reflects current best practices based on evidence-based approaches to RoB assessment. Date of publication (i.e., older versus newer tools), however, should not automatically be assumed to be a good indicator of coverage for key biases as demonstrated in the results in the scoping review (11, 12). Organizational and methodological preferences and values also play a key role in selection of a tool, and sometimes, choice is not permitted. For example, As described in the Cochrane Handbook, the Methodological Expectations of Cochrane Intervention Reviews (MECIR) manual stipulates that the RoB 2.0 tool must be used for assessment of RCTs in Cochrane SRs (2, 13). The Methodological Expectations of Campbell Collaboration Intervention

Reviews (MECCIR) recommends use of the same tool for RCTs but does not make this mandatory (14). Instead, they value use of a tool that includes coverage for domains such as selection bias, confounding and attrition that are appropriate for the study designs that are included. Other organizations with standardized methodological expectations for knowledge syntheses also have more permissible options for RoB. For example, the JBI Manual for Evidence Synthesis offers guidance for specifying the critical appraisal process and transparent descriptions of methods, but does not require the use of the organizations' JBI critical appraisal tools (15). In the scoping review, results showed that many tools contained elements unrelated to RoB, such as reporting quality or generalizability and applicability (11, 16). This may introduce 'noise' into the assessment process and detract from bias-specific assessments for the studies under consideration. Researchers must also assess whether the tool reflects current best practices based on evidence-based approaches to RoB assessment. Date of publication (i.e., older versus newer tools), however, should not automatically be assumed to be a good indicator of coverage for key biases as demonstrated in the results in the scoping review (11, 12).

When considering RoB tools, reliability and validity are vital. Users should value tools which have been assessed as valid and reliable, regardless of their research focus. Users in turn, may further value tools shown to be valid and reliable in their specific research field, such as NUQUEST (16). Choice of tool may additionally consider whether a tool focuses distinctly on RoB, or serves a broader critical appraisal purpose. Other tool characteristics may factor in similarly, such as the tool's structure (i.e., whether it was designed as a scale, checklist, or domain-based rating tool), applicability and specificity across study designs, and how well it captures relevant biases and provides interpretable outputs for their use, such as numerical scores or categorical ratings. When considering results from an RoB assessment, the tool's output format and ease of interpretation is critical, especially if it employs numeric scores or arbitrary

thresholds, and these should be supported by evidence. For researchers using evidence grading systems like GRADE, RoB tools which were designed to integrate seamlessly into these systems (i.e., they offer coverage for relevant RoB concepts and report RoB ratings by outcome) may offer a more straightforward path from RoB assessment to synthesis and certainty around the findings.

Additionally, a higher value may be placed on selection of instruments which are part of a suite of tools applicable across various study designs (as compared to a standalone RoB tool) which, in the context of a SR, can introduce consistency in RoB assessments and their interpretation. Comprehensive guidance accompanying the tool can serve to enhance user confidence in ratings by offering detailed instructions, examples, rules of thumb or advice for varying scenarios in application, fostering standardization across studies and reviewers. Lastly, the availability of accompanying implements, such as worksheets, rating decision algorithms or digital tools, can streamline the rating process and improve usability. Those with less experience with RoB assessments or raters new to a specific tool may value these implements, however, they are arguably useful for all raters and should be reviewed in full regardless of experience-level or expertise.

Table 6-2: Considerations for selecting a RoB tool for primary research studies

Category	Consideration	Implications
Purpose	Are you assessing one standalone study (or more) or are you assessing studies as part of a systematic review?	When a single study assessment may be influential to guidelines, policy, are a decision at hand, a tool may be selected that prioritizes comprehensive coverage of relevant biases with precision in the estimates for the magnitude and direction of bias. However, when a tool is applied in the context of a systematic review, resource or time constraints may necessitate a more feasible and practical to risk of bias assessment with a tool that is fit for purpose.

Category	Consideration	Implications
Organizational or methodological requirements	What tool is recommended or required by the organization for which a rating or systematic review is being conducted?	Organizations like Cochrane or JBI may require or recommend risk of bias tools to meet their stated methodological requirements. Organizations or governments may also have preferences for risk of bias tools based on legal or policy applications at hand.
Focus	Is the tool intended for a general critical appraisal, evaluation of methodological quality, or risk of bias assessment?	Although quality and risk of bias concepts are often used interchangeably, users wishing to focus on risk of bias may seek out a tool with this emphasis and with coverage of appropriate domains and items.
Structure	Is it a scale, a checklist, or a domain-based rating tool?	Users may wish to consider what approach, or structure, best suits their needs and is most appropriate for the risk of bias assessment. Users may want to consider the current state-of-the-art (e.g., which suggests a numeric score for a rated study is not informative) Additionally, they may consider if specific potential for biases are identifiable within the tool if relevant, and if an overall rating for the study is presented.
Tool applicable to which study design(s)	Is the risk of bias tool applicable to a single primary study design (e.g., RCT), one or more primary study designs (RCTs, NRSIs, cohort studies and case-control studies), or any study design (including systematic reviews)?	Users may consider if a tool is applicable to a specific study design or whether it may be applicable to more than one study design. Included items (or biases) addressed may not all be applicable to the study design under consideration or the tool may not be tailored enough to catch the associated nuances of any one study design. However, it may be practical and desirable to be able to apply one tool to a variety of study designs.
Interpretation or output	Is the tool structured to produce a score, count, scale, other a different rating approach?	A user may consider if they can feasibly interpret the results of the risk of bias tool for their purposes, and whether a simple count of items or a numeric score is appropriate compared to another rating approach. Some tools may use a hybrid approach combining numeric scores and another ratings approach. Any applicable thresholds or cut-offs used to delineate bias or quality should be checked to see if they are supported by evidence or arbitrary.
Specificity to risk of bias	Does the tool include items which are not relevant to risk of bias?	Inclusion of items not related to risk of bias may detract from true biases and introduce noise into an assessment. This includes consideration for reporting quality, statistical precision, generalizability or applicability. These may be of interest and relevant to the user but are not related to the risk of bias.

Category	Consideration	Implications
		In addition, items evaluating the presence or absence of study features may inform reporting guidelines, but not necessarily risk of bias judgements.
Reliability and validity	Has it been assessed for reliability? Has it been assessed for validity? Are the findings published?	Although many tools are proposed to assess the risk of bias in any study design, not all will have assessed or published the reliability or validity of the tools. In addition, users may consider giving priority to selection of a tool with documented reliability and validity. Users may also what to consider whether the properties of a tool were assessed within a particular use case, field or research context and whether they feel the results of that assessment may be applicable to their own research context and use case.
Feasibility/Useability	How complex is the tool to apply? How much time is needed to apply?	Some tools may be comprehensive for risk of bias, but complex to apply and ascertain related judgments. Use of a comprehensive tool may be justified or desirable, however, it may not be feasible to complete the ratings in a timely fashion (for example within an SR with many included studies). Also, some tools require training or more extensive rater experience to apply, regardless of any accompanying guidance. Users should consider their capacity and knowledge to implement, and the time needed to complete an assessment (or many!).
Comprehensiveness	Does the tool focus on a comprehensive range of potential risk of bias considerations, or a select subset of biases?	Users may select a tool based on their own needs and the depth of evaluation needed. Some tools may be simplistic and provide assessments which are fit-for-purpose. Others are very comprehensive in their consideration for risk of bias but users may not require precision in terms of the magnitude or direction of bias to inform their needs.
State-of-the-art	Is the tool considered "outdated", missing key items, or applying consideration for bias concepts which are more or less relevant based on our current knowledge?	Science in the area of risk of bias continues to advance, and new approaches or focuses on specific elements of bias may change over time. Critical consideration for the current state-of-the-art should be considered even though there are opposing views and no gold standard. For example, risk of bias may differ according to the outcome, and so users may consider if a tool offers outcome-specific

Category	Consideration	Implications
		assessment, or if subjective/stakeholder-reported or objective outcomes may be assessed separately[5].
Suite or single tool	Is the tool stand-alone or one of many tools in a suite of study designs?	User may consider their needs for risk of bias assessment and whether a suite of tools applicable across study designs may facilitate a more standardized, interpretable or consistent rating.
Available guidance	Is there user guidance accompanying the tool to inform assessments? How comprehensive and user-friendly is the guidance?	Available tools may or may not have accompanying guidance to facilitate use and application of the tool and guidance may be brief or comprehensive. Guidance may help with the interpretation of items, assist with scoring or rating assessment and serve to support standardization of ratings across reviewers or across studies. In some cases, guidance will provide examples to inform a particular use case or context or they may dictate rules of thumb that may be applied or more firm decision thresholds or cutoffs.
Research context	Are the tools broadly applicable or were they developed for a specific field or research context?	Users should assess a tool's ability to evaluate biases specific to their field. Users may consider if a tool was developed to capture biases that may may be broadly relevant to a particular study design (i.e., generic assessment of bias) or whether a tool was developed specific to a particular field or research context. Regardless of whether a tool is generic or developed for a particular field or context, users should not assume all risk of bias concepts are considered.
Accompanying implements to facilitate ratings	Are there any implements available to assist with use of the tool?	Risk of bias tools may or may not have accompanying implements such as a worksheet, visual algorithm, figures or tables structured to support ratings. This may support users in many ways as they apply a tool as they consider potential biases for their study, field, or the decisions that may be applied to ratings. This may also encourage more consistent evaluations across raters, regardless of experience or knowledge. In addition, some tools may be available in alternative formats which facilitate use, for example as webpages, online forms, forms within systematic review management software, spreadsheet templates, or apps that can be used on personal devices. Users may want to consider whether these implements are desirable or

[5] ROB tools not designed to be outcome-specific can in theory be applied this way, as is the case with NUQUEST.

Category	Consideration	Implications
		helpful for their use when selecting a risk of bias tool.
Intent to use evidence grading system	Will the output from the risk of bias assessment be used to inform evidence grading frameworks or guideline processes?	Certain risk of bias tools have been developed with formal linkages to evidence-based grading or guideline systems in mind. For example, the Scottish Intercollegiate Guidelines Network (SIGN) tools are used to inform guideline development processes by the group. Many other Professional or Clinical Guideline development groups may have similar tools or suggested tools to inform their processes. Similarly, other tools may directly be informative to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework. Although ratings from any tool are useable for these same uses, risk of bias ratings from select tools may be more instantly transferrable into these platforms. Also, the guidance for applying GRADE, for example, may provide examples for risk of bias from a particular tool or features of a tool. Users should consider whether a transferrable rating may be desirable in their use case and research context.

Users may find several published studies or reviews that assist in their decisions and in formally evaluating some of the available RoB tools. Such evidence or information will not be available for all tools, and research examining the use and metrics of various RoB tools may be more obtainable for tools that are more recent or more widely used. It is crucial to acknowledge that every RoB tool, even the NUQUEST checklists developed here, come with inherent strengths and limitations. Therefore, selecting the most appropriate tool requires careful consideration of how these characteristics align with the specific objectives, context, and methodological demands of the research. Users must evaluate these factors to ensure that the chosen RoB tool enables a robust and meaningful RoB assessment. To facilitate future consideration of NUQUEST, Table 6-3 summarizes the characteristics, strengths and limitations of NUQUEST.

Table 6-3: Considerations for NUQUEST

Category	NUQUEST Consideration
Purpose	NUQUEST can be applied to a standalone study(ies) or as part of RoB assessment to inform a systematic review process.
Ties to organizations or methods groups	NUQUEST was developed independent from any specific organizational or methodological standards or funding.
Focus	NUQUEST is intended to assess potential for risk of bias. Resulting from the use of a bolt-on approach to include nutrition-specific biases, the NUQUEST checklists do contain optional content which may be informative to assessments of applicability.
Structure	NUQUEST is rating tool organized by four sections (Selection, comparability, ascertainment, nutrition-specific) and a section for the overall rating for the study. The exact title of each section varies by study design. Ratings are completed for each item within a section, overall for the section, and then for the overall rating of the study. The number of items varies by the study design and section. Nutrition-specific items are separated from generic biases.
Tool applicable to which study design(s)	NUQUEST checklists each cover a specific study design. NUQUEST is available for RCTs, cohort studies, case-control studies and soon, cross-sectional studies. All relevant biases specific to these study designs are considered (RCTs, cohort studies, case-control studies and soon, cross-sectional studies). NUQUEST for RCTs may be applied to any trial design (e.g., cluster or crossover RCTs) but guidance focuses on parallel RCTs and may not specifically address the nuances for other RCT designs. Any type of cohort study may be assessed with NUQUEST, including those collecting prospective or retrospective data, or a hybrid of both. NUQUEST can be applied to case-control studies and nested case-control studies. The planned NUQUEST checklists for cross-sectional studies will be applicable to studies reporting descriptive/prevalence or analytic/association relationships.
Interpretation or output	The NUQUEST rating system for items in each of four sections include a range of response options (*yes/probably yes/probably no/no*) and a "not applicable" option which is only included for items where a response would not be applicable. The overall rating options for each section are: *good* (+; low RoB); *neutral* (0; moderate RoB); and *poor* (−; high RoB). The overall ratings for each section are carried forward to an overall study rating section and used when considering the final assessment of the study (response options: *poor*, *neutral, or good*). The use of NUQUEST has several advantages for supporting SR development and nutrition guidance decisions, including clarity of the key research question, flexibility and transparency in developing the PICO and worksheet, and transparency of the strengths and weaknesses of the relied-upon evidence. NUQUEST, with its emphasis on assessment of exposure data, can support future nutrition guidance initiatives.

Category	NUQUEST Consideration
Specificity to risk of bias	Reporting quality, statistical precision, generalizability or applicability are often relevant to the user, but unrelated to RoB.
Reliability and validity	NUQUEST for RCTs, cohort studies and case-control studies subsume the existing properties of the SIGN instruments, which are valid, reliable, and have extensive use experience (17). Overall interrater reliability for NUQUEST across all study designs (n = 45 studies) was substantial (0.73; 95% CI: 0.52, 0.93) and moderate to almost perfect for the individual instruments for RCTs (0.59; 95% CI: 0.19, 1.00), cohort studies (0.57; 95% CI: 0.15, 1.00), and case-control studies (1.00; 95% CI: 1.00, 1.00). As there is no current gold-standard approach for the assessment of RoB in nutrition studies, the working group could not assess the criterion validity of NUQUEST. Instead, the research here represents the initial assessment of content, construct, and concurrent validity. During testing, NUQUEST performed well in comparison to other generic RoB tools and, where there was disagreement, the nutrition-specific items influenced the overall rating. The reliability and validity of the NUQUEST for cross-sectional studies will be reported once completed.
Feasibility/Useability	Realizing that some existing assessment instruments are challenging or resource-intensive to use, NUQUEST is user-friendly (regardless of experience or expertise), offers an intuitive rating system, well- organized and supported by detailed nutrition-specific guidance. During testing, raters felt that NUQUEST was easy to apply and feasible to complete within a reasonable time by raters with general knowledge of epidemiology or nutrition science (average: 16.5 min/study, range: 10–33 min). NUQUEST for cohort studies generally took more time for raters to complete than the others, but time to completion improved with each completed study assessment.
Comprehensiveness	NUQUEST is grounded in the scientific concepts of bias reported considered broadly in other instruments, while improving the usability for application to nutrition studies. NUQUEST focuses on the extent to which nutrition studies were designed, conducted, analyzed, and reported to the highest possible methodological standards, but is not intended to quantify the magnitude of bias that may be attributable to methodological flaws.
State-of-the-art	Current RoB science widely recommends that RoB be outcome specific. While NUQUEST does not directly facilitate outcome-specific RoB assessment, it can be used to generate outcome specific RoB assessments (as can most other tools) which is facilitated by the accompanying NUQUEST worksheet.

Category	NUQUEST Consideration
Suite or single tool	NUQUEST is a suite of RoB tools applicable to RCTs, cohort studies and case control studies. Ongoing work will expand the suite of tools to include two tools for cross-sectional studies.
Available guidance	Comprehensive, user guidance specific to each study design was adapted and developed to accompany NUQUEST. All generic RoB item guidance is supplemented with nutrition-related context where relevant. Guidance pertaining to the assessment of exposure and intake is detailed and includes consideration for accuracy and validity in a nutrition context. Topics covered include baseline assessments, adherence to intervention, self-report, supplement and food product composition, dietary patterns or exposures, biomarkers and appropriate documentation, and the appropriateness of dietary assessment methodologies for different research purposes. The guidance includes examples from nutrition studies to illustrate specific issues and facilitate tool application and decision-making. Even with the guidance and assessment criteria defined (See accompanying implements below), user judgment will be required.
Research context	NUQUEST is intended to be applied to nutrition studies. While 'generic' into sections focus on selection, comparability, and ascertainment, they are supplemented with nutrition-specific RoB items to recognize the nature and impact of the challenges when considering nutrition exposures.
Accompanying implements to facilitate ratings	A worksheet is offered (with an accompanying worked example) to elucidate topic-specific issues important for study assessment and to support standardized ratings across multiple reviewers of varying experience-levels. As is common practice, we recommend training on a set of studies to test topic-specific guidance and use of the worksheet, and discussing consensus on items among raters. Initial NUQUEST testing found that the inclusion of a worksheet improved interrater reliability and increased discussion among raters to achieve consensus.
Intent to use evidence grading system	NUQUEST RoB ratings may be used to inform the RoB portion of an evidence grading framework (e.g., GRADE) or transferred into guideline processes as needed. However, it was not specifically designed to fit any specific evidence grading framework.
Outcome-specificity	While NUQUEST does not directly facilitate outcome-specific RoB assessment, it can be used to generate outcome specific RoB assessments (as can most other tools) which is facilitated by the accompanying NUQUEST worksheet.

6.6. Novelty of the thesis

This thesis presents the first suite of nutrition-specific RoB tools, NUQUEST, which permit comprehensive assessment of potential biases in primary research studies focused on

relationships between diet and human health. Each chapter in this thesis contributes uniquely to our knowledge in this area, as demonstrated in the overview provided in Table 6-4. This program of research makes an important contribution to the literature around nutrition-specific risk of bias and will be useful and informative for both practice and policy in Canada and globally. As demonstrated through the early use and application of NUQUEST (Section 6.5), NUQUEST fills in an important gap for users looking to assess RoB in RCTs, cohort studies and case-control studies. In addition, early adaptations of NUQUEST for application to cross-sectional studies provide evidence to support the need nutrition-specific tools for these studies. As shown in Appendix 3 (Table 7-1), NUQUEST has been disseminated internationally and has been used as part of key training and capacity-building initiatives.

Table 6-4: Overview of key contributions of this thesis

Chapter	Area of Contribution	Unique Contribution(s)
Chapter 3	Nutrition-related challenges that contribute to uncertainty in nutrition studies	• When this research was conceived in 2019, there was little methodological guidance specific to biases which uniquely impact nutrition research available. • As part of the evidence base required to develop NUQUEST, issues related to nutrition exposure measurement and the interpretation of nutrition data were considered based on published literature describing the experiences of scientists who had either conducted or used nutrition-based SRs (including those on the NUQUEST working group), recommendations of expert groups and articles that had identified nutrition-specific appraisal issues. • Nutrition-specific appraisal issues were compiled from all available sources, vetted within the working group and then sorted by their applicability to randomized controlled trials, cohort studies and case-control studies. • The rigorous evaluation and transparency afforded by the application of NUQUEST to nutrition studies can highlight critical areas that necessitate further advancement of the science or the development of innovative statistical methodologies to address the limitations inherent in dietary exposure data.

Chapter	Area of Contribution	Unique Contribution(s)
	Nutrition-specific risk of bias tools for RCTs, cohort studies and case-control studies	• When this research was conceived in 2019, nutrition-specific criteria were not consistently included in existing bias assessment tools and • NUQUEST for RCTs, cohort studies and case-control studies developed in this thesis is the first suite of comprehensive risk of bias assessment tools which integrate nutrition-specific concepts with accepted general concepts of study design and conduct. • NUQUEST provides a standardized, robust approach for considering risk of bias in nutrition research. • Given that the field of nutrition overlaps with epidemiology, public health, basic sciences and medicine, this context-specific approach for assessing risk of bias provides a model for fields facing similar challenges with unique risk of bias considerations. A 'bolt-on' approach may provide a standardized and comprehensive approach for evaluating potential context-specific biases using existing reliable and valid generic tools.
Chapter 4	Risk of bias of specific to cross-sectional research	• Based on a scoping search conducted prior to initiation of this work in 2021, methodological literature relevant to cross-sectional research was focused on bias in surveys and did not comprehensively consider potential biases specific to analytic designs. • This scoping review fills in an important methodological gap in the literature and provides comprehensive summary of biases specific to cross-sectional research, for both descriptive and analytic approaches. • The framework adapted provides a structure to identify potential biases in cross-sectional studies, which may inform the methodological rigour of future cross-sectional research. As such, this thesis identified that existing risk of bias tools fail to capture all potentially relevant biases. • Findings are directly informing the development of two comprehensive and nutrition-specific risk of bias tools for cross-sectional studies which will be more precise for evaluation of both descriptive and analytic research.
Chapter 5	Nutrition-specific risk of bias tools for descriptive and analytic cross-sectional studies	• At the time this research was conceived, no existing tools comprehensively considered risk of bias relevant to either descriptive/prevalence or analytic/association cross-sectional research, regardless of field. • Both NUQUEST checklists in development will be the first to comprehensively consider biases unique to these research

Chapter	Area of Contribution	Unique Contribution(s)
		designs and provide a structure to identify and evaluate nutrition-specific risk of bias. • While cross-sectional research may be thought of as a lower value form of evidence, it is widely used in nutrition. NUQUEST will provide a standardized, detailed approach for accurate assessment of cross-sectional nutrition research studies and enables discrimination between studies at varying risk of bias. • As NUQUEST will provide a tool for more nuanced interpretation of studies, this adds to the capacity of knowledge users and decision-makers to appropriately weight the value of study findings in future syntheses, statistical analyses and policy-making.

RCT = randomized controlled trial; SR = systematic review.

6.7. Strengths and limitations of the thesis

The individual manuscripts presented in Chapters 3 through 5 each include a discussion of the strengths and limitations of importance to the individual research projects. As such, the strengths and limitations presented below apply broadly across the thesis.

6.7.1. Strengths

In this thesis, a suite of reliable and valid nutrition-specific RoB tools were developed, guided by a multidisciplinary working group of US and Canadian experts in the areas of nutrition science, biostatistics and epidemiology (16). This thesis addresses study designs that are relevant and widely used in the field of nutrition, namely RCTs, cohort studies, case control studies and cross-sectional studies. The tools developed (and still under development) in this thesis (NUQUEST) have robust consideration for RoB concepts applicable across fields and an added unique set of nutrition-specific items which enable consistent and transparent evaluation of potential RoB related to dietary exposure assessment commonly encountered in human nutrition studies. NUQUEST permits users to consider differentially weight the items assessed for each section and for the overall rating where it would be pertinent or informative to the research question

being assessed. This facilitates ratings which are further nuanced to the specific topic under study, and provides the capacity for raters to stress potential bias for items which may be particularly influential in a study. NUQUEST is available online through an open-access publication. All checklists, guidance and the worksheet are provided in the supplemental materials.

NUQUEST was developed to be user-friendly, and as such, development included an improved rating system, structure and detailed nutrition-specific guidance. The guidance includes prompts to consider both general and topic-specific criteria when rating studies. Similar to other tools (18, 19), the NUQUEST rating system eliminates issues raised in previous RoB tools, such as "fence sitting" by raters reluctant to commit to a definitive rating based on unclear or not reported details within a study (18). The use of response options such as *probably yes* and *probably no*, facilitates assessment of risk of bias in cases where information to inform the assessment may not be reported, and guides users towards making a judgment based on something likely to have, or have not, occurred in the study. NUQUEST is accompanied by a worksheet to facilitate ratings, and an example of a completed worksheet was also developed to illustrate worksheet use in the context of a specific research question. The worksheet was created to support users through common challenges in assessing the RoB in nutrition studies, particularly for observational studies, in which study authors frequently neglect to adequately describe the procedure used to establish the accuracy and appropriateness of their dietary exposure assessment, The NUQUEST worksheet facilitates consideration of the accuracy, reliability, and appropriateness of dietary intake in selected studies, as well as the appropriateness of the dietary methodology for meeting research objectives.

Following the publication of the seminal NUQUEST article in the American Journal of Clinical Nutrition, NUQUEST received two Canadian awards through Health Canada, including a Foods Directorate Award for Excellence in Science (2022) and the Health Products and Foods Branch Assistant Deputy Minister Award for Excellence in Science (2022). Later in 2022, NUQUEST was also nominated for a Departmental Award for Excellence in Science by Health Canada.

6.7.2. Limitations

Despite the robust and comprehensive nature of this work, there are important limitations that should be noted. While comprehensive and structured to encourage consistency of ratings, each item and section in NUQUEST is rated based on the subjective assessment of the individual applying the tool, which could lead to variability in assessments. NUQUEST, as with other RoB tools, rely heavily on transparent and comprehensive research reporting. Although the response options in NUQUEST encourage raters to assign a rating even when there is absence of evidence for important RoB items, the accuracy of the overall RoB rating may be impacted by studies with poor reporting quality. Users may consider this within the range of other tool characteristics presented in Section 6.6 to inform their selection of an RoB tool that suits their individual needs and resources. As NUQUEST was built on the existing properties of the Scottish Intercollegiate Guidelines Network (SIGN) appraisal tools, we maintained the wording for the generic RoB items in each checklist. There may be instances where changes to wording would have enhanced specificity to individual biases. The *Applicability* section of the SIGN tools was also maintained for continuity despite the fact that it does not relate to RoB, although assessment of this section is optional. The nutrition-specific tools developed in this thesis are designed to be user-friendly and feasible to complete, however users with general knowledge in nutrition science and epidemiology should additionally consider the training and time required for preparation for applying the tools, especially for cohort studies which broadly took more time to complete than

the others. While these tools facilitate nutrition-specific judgements, these are subjective assessments of individual biases which facilitate an overall consideration for the collective effect of risk of bias in a study. NUQUEST were designed and tested to comprehensively consider biases relevant to nutrition science, and particularly dietary intake assessment, however, testing completed in Chapter 3 and proposed in Chapter 4 represents preliminary assessment of the tools in a set of studies which may not fulsomely reflect the range of applications across nutritional research contexts.

6.8. Directions for future research

Based on the results of this thesis and the limitations identified in Section 6.7, several directions for future research are proposed.

Since publication in the American Journal of Clinical Nutrition in 2022 (Appendix 2), NUQUEST has been successfully applied in several systematic reviews, with initial applications demonstrating its value in evaluating biases specific to nutrition studies (20-27). Notably, some SR authors have already published adaptations of NUQUEST for cohort studies for application to cross-sectional studies, which underscores the need for the tools in development (Chapter 5). The applications of NUQUEST to rate studies within the context of nutrition systematic reviews highlights the tool's relevance for addressing methodological challenges in this field.

Of note, the SRs citing NUQUEST have frequently adapted or modified it prior to application, including both the items assessed and the rating system. To illustrate adaptations of NUQUEST for cross-sectional studies, a systematic review by Ross et al (2023) modified NUQUEST for cohort studies for this purpose (26). Authors modified NUQUEST by prospectively answering ‘No’ for any items they considered to be not applicable for cross-sectional research, or seven

items in total[6]. An SR by Cara et al. (2024) also adapted the NUQUEST checklist for cohort studies to apply to included cross sectional studies (21). The review authors modified the NUQUEST to exclude items irrelevant to the cross-sectional design and expanded on items related to dietary assessment tools by adding to items in the *Nutrition-specific* domain. Specifically, item 2.1 (*The frequency and quantity of the exposure under study are accurately and reliably measured*) was split into items: 2.1a (*The tool used to assess the exposure level was specific to the country or population included in this study (e.g., a national nutrient database)*) and 2.1b (*The tool used to assess the exposure level was appropriate for the specific exposure of interest in the study (e.g., the study used a database with data on added sugar)*). Additionally, SRs are published which have changed the NUQUEST rating system. Ross et al (2023) (26) changed the NUQUEST rating system by converting it to a quantitative scoring system for all study designs assessed, including randomized controlled trials, cohort studies and case-control studies. Cara et al. (2024) modified NUQUEST ratings by downgrading all overall ROB ratings one level (from "Good" to "Neutral") (21). Accordingly, even when a study received 'good' ratings for all domains, the overall judgement for the study was 'neutral'. A later SR of RCTs by Cara et al. utilized a different rating system altogether. Review authors assigned quantitative points to each item assessed in NUQUEST (22). A scoring algorithm was then applied to define the overall risk of bias rating for each study assessed.

[6] **Selection of Cohorts:** The likelihood that some eligible subjects might have the outcome at the time of enrollment is assessed and considered in the analysis. **Comparability of Cohorts:** The percentage of individuals or clusters recruited into each arm of the study that dropped out before the study was completed is reasonable / Comparison is made between full participants and those lost to follow up, by exposure / Exposure level or prognostic factor is assessed more than once. **Ascertainment of Outcomes:** Where blinding was not possible, there is some recognition that knowledge of exposure could have influenced the assessment of outcome**. Nutrition-Specific:** The baseline exposures of all groups have been maintained over the course of the study / The interval between the exposure and outcome is of sufficient duration to observe an effect, if there is one.

That some SRs have changed the rating system for NUQUEST as part of their application may provide evidence for related challenges that persist in the field which warrants further investigation. It is unclear why these modifications were completed, if users found it difficult to apply consistently using the rating system developed, or whether these simply reflect a trend of poor methodological approaches to SRs in the nutrition field more broadly. The modifications to NUQUEST lack consistency and limit any assessment of NUQUESTs utility or value across reviews and it is important to monitor use to observe whether these trends will persist. If trends persist, this could indicate that improvements to the description of the NUQUEST checklists or additional clarity or flexibility in the related guidance should be considered in the future. Regardless, while it is encouraging to see uptake of NUQUEST in the field, modifications to NUQUEST outside of the proposed structure and rating system undermine standardization and perpetuate existing quality issues with SRs of nutrient-health relationships. Yet, the SRs cited do provide justification for the need for a NUQUEST checklist applicable to cross-sectional studies which would standardize ratings and comprehensively consider all relevant biases. Further evidence is needed for NUQUEST application to all study designs, especially RCTs and case-control studies using the tool and rating system as designed.

Future research should ideally expand the suite of available NUQUEST checklists to other study designs, including controlled before-after studies, interrupted time series studies (28) and ecological studies (29) among others. Ecological studies are valuable for nutrition researchers because they can identify population-level risk factors that are relatively uniformly distributed but vary significantly over time or among different populations. (30, 31). NUQUEST for ecologic studies would give researchers a nutrition-specific approach to considering biases that are common in the design, and in nutrition contexts, more precisely (such as confounding by group, effect measure modification by group, and non-differential exposure misclassification)

(32). Interrupted time-series studies are often used in nutrition to consider individual and population-level changes following large-scale intervention, public health or policy change (e.g., changes in labelling of foods, new recommendations on dietary intake, or restrictions on advertising) or even an interruption such as the global COVID-19 pandemic (33-35). In such studies, variations in single or controlled group designs and knowledge of relevant concurrent history (for which alternative interventions, interruptions or policy changes may contribute to changes in results) commonly threaten validity and limit uptake (36). Similarly, within the current suite of NUQUEST checklists, future research should consider rectifying limitations identified, such as facilitating outcome-specific RoB more concretely. Future improvements to the checklists may expand the information section to note which outcome is being assessed and the guidance or worksheet may be updated to address outcome specific concerns or examples.

Artificial intelligence (AI) models and machine learning technologies are increasingly being employed in a number of policy and practice settings, including nutrition (37-45). Yang et al. (2020) were early adopters of AI in their application of a semi-automated approach to retrieve and analyze the reporting completeness of nutrition research articles according to the Strengthening the Reporting of Observational Studies in Nutritional Epidemiology (STROBE-NUT) Statement (37, 46, 47). Intuitively these same principles could be applied to facilitate AI-assisted risk of bias assessments using NUQUEST. AI tools such as ChatGPT have been suggested to completely or semi-automate RoB assessment (48), but at the time this thesis was completed, there is only one known application using the RoB2 tool (49). When ChatGPT RoB assessments were compared to assessments from an existing set of Cochrane systematic reviews (n=157 unique RCTs in n=34 systematic reviews), reliability was only ‘slight’ to ‘fair’ when compared to human reviewers. Authors concluded that ChatGPT is not currently able to reliably assess risk of bias of RCTs and advised against using the platform for RoB assessments. A

Canadian assessment of machine versus human RoB assessment using the 'RobotReviewer' tool which is based on the 2011 version of the Cochrane RoB tool for RCTs. Results showed that while time and burden on reviewers to assess each study decreased, authors were concerned over a lack of evidence for the validation of the tool itself (50, 51). Assessed inter-rater agreement varied from no agreement (related to blinding of participants) to good (related to random sequence generation). A large proportion of the text gathered by text-mining to support the RobotReviewer ratings was considered to be irrelevant. Study authors concluded that applications such as RobotReviewer may support RoB ratings but can never replace human assessment (50). Future research should consider the limitations of past approaches, the viability of innovative AI approaches for RoB and these should be evaluated empirically while encouraging a culture of high-quality, high-value and sustainable methodological research (37, 48).

As emerging technologies such as machine learning models and AI-supported applications are increasingly contributing or influencing nutrition science in other ways. More and more, digital tools are being used for data collection, analysis and dissemination (39). Digital databases have made it easier to access nutritional profiles for thousands of foods and software is available to connect and estimate dietary profiles with increased accuracy. Wearable technologies, including fitness trackers, digital watches, clothing or jewelry and individual glucose monitors are informing the advancement of methods behind more personalized nutrition, providing tailored dietary recommendations based on machine learning algorithms (39, 43, 52). This patient-generated data is also being used to inform research (43) and accordingly, this must be considered when RoB is assessed. In the future, research should consider if NUQUEST requires additional items or guidance to consider how intake and outcome data generated through AI methods may introduce bias into a study. While these tools show promise, the data generated

may essentially be from a “black box”, and it may not be intuitive to individuals outside of the development team how these data are processed, connected or summarized, which makes consideration for their RoB potentially difficult. Individuals applying RoB tools to research studies informed by AI may not have information they need to fulsomely consider the related biases and they may not have the expertise to understand how digital tools and AI may influence dietary intake methods or study findings.

Findings from this thesis demonstrate that selection of an appropriate RoB tool goes beyond a “one size fits all” approach and there is a role for context-specific assessment of biases in all research. In academic terms, it is important to emphasize that the selection of RoB tools should be guided by the specific methodological strengths, limitations, and contextual appropriateness of each tool, or suite of tools, relative to the study design and research question. While the prominence or influence of a research group or organization may add to the credibility of a RoB tool, this is not a proxy for formal assessment of the reliability, validity and feasibility. Further research efforts should focus on creation of tools or frameworks to permit systematic and structured identification of tools applicable to specific study designs and contexts, with consideration for their various features, strengths and weaknesses to facilitate choice. Similar databases for reporting quality (EQUATOR Network), outcome measurement (COSMIN, OMERACT, PROQOLID) are already widely in use. Some work in this area showed promise for RoB tools but was not maintained beyond the funded project work (53). This may serve to combat the ongoing issue of users making ad hoc modifications to existing RoB tools which persist in evidence syntheses. To supplement this, future qualitative or mixed methods research should focus on understanding what leads to author decisions to modify tools over applying existing tools. An improved understanding and documenting where true gaps exist to inform future tool improvements or development. Further, it may provide evidence for future ‘bolt-on’

contextual-specific RoB considerations to extend existing RoB tools with demonstrated reliability, validity and feasibility. This may offer a better 'fit' for detecting study-specific biases or addressing nuances particular to a research field. Emphasizing methodological and context appropriateness will ensure that RoB assessment is in alignment with best practices, promotes methodological rigour, and enhances the uptake of evidence in fields with complex or context-sensitive biases, such as nutrition science.

6.9. Conclusions

This thesis demonstrates the importance of considering and assessing nutrition-specific biases in nutrition studies to inform health policy decisions and recommendations related to diet and health. The development of NUQUEST, a set of nutrition-specific RoB tools designed for RCTs, cohort studies, case-control studies (and cross-sectional studies), advances knowledge in this area for the field of nutrition science and addresses gaps and issues in consistency of consideration for these biases in existing tools. As currently available RoB tools lack sensitivity to consider dietary exposure assessments, NUQUEST integrates nutrition-specific assessment criteria into validated generic assessment tools. Developed and assessed through an iterative, eight-step development process, NUQUEST provides a structured means of capturing nutrition-specific sources of bias which provides a more accurate and complete summary of the strengths and limitations of the research for evidence consumers, including providers, recommendation and guideline developers and policymakers.

Initial testing of NUQUEST, completed using a purposive sample of RCTs, cohort studies and case-control studies, demonstrated high inter-rater reliability and usability in real-world research contexts, which was improved with the addition of a NUQUEST worksheet and experience in application. This reflects the capacity of NUQUEST to enable consistent assessments across

study designs and users, who are supported by detailed nutrition-specific guidance unique to each NUQUEST checklist. The NUQUEST checklists developed as part of this thesis additionally show high concurrent validity with published ratings for the same set of studies extracted from systematic reviews. Where disagreements were documented, initial testing of NUQUEST indicated the nutrition-specific component was a contributing factor in discerning methodological issues related to dietary exposure. This underscores the utility of incorporating nutrition-specific items into RoB tools which can be applied to nutrition studies, as they are helpful in discerning nuances relevant to exposure methodologies that may influence study findings.

Initial development of the NUQUEST checklists precipitated further consideration for research designs often used in the field of nutrition, namely, cross-sectional studies. As no valid and reliable tools applicable to cross-sectional research were available to apply a similar 'bolt on' approach to an existing tool, additional research was considered. This thesis comprehensively adds to our methodological knowledge of the specific biases applicable to cross-sectional studies as identified in the related scoping review. In addition, the unique considerations for both descriptive and analytic cross-sectional research is addressed, as these types of evidence may be considered differently in policy and practice settings. As with the initial NUQUEST suite of tools, the evidence base described in this thesis is being used to develop two new NUQUEST checklists and their development is ongoing. Early use of NUQUEST in citing systematic reviews provides ample justification for the development and tailoring of nutrition-specific risk of bias tools for cross-sectional studies.

NUQUEST provides a transparent, reliable and valid approach for evaluating RoB issues related to dietary exposure assessment commonly encountered in human nutrition studies. This plays an

important role in supporting policy and practice leaders who require a thorough understanding of human health and how various nutrients and dietary patterns effect health status and quality of life. NUQUEST is useful for users who require RoB assessment tools for individual nutrition studies or for application to multiple studies in a systematic review. Additionally, the nutrition-specific appraisal items in NUQUEST may support evidence developers to enhance the design, conduct, and reporting of future nutrition research studies. In turn, it is hoped that use of NUQUEST will foster consistency across future systematic reviews that synthesize nutrition research.

6.10. References

1. De Cassai A, Boscolo A, Zarantonello F, Pettenuzzo T, Sella N, Geraldini F, Munari M, Navalesi P. *Enhancing study quality assessment: an in-depth review of risk of bias tools for meta-analysis-a comprehensive guide for anesthesiologists*. J Anesth Analg Crit Care. 2023 Nov 6;3(1):44. eng. Epub 20231106. doi:10.1186/s44158-023-00129-z. Cited in: Pubmed; PMID 37932825.

2. Cumpston M, Li T, Page MJ, Chandler J, Welch VA, Higgins JP, Thomas J. *Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane Database Syst Rev. 2019 Oct 3;10:ED000142. Epub 2019/10/24. doi:10.1002/14651858.ED000142. Cited in: Pubmed; PMID 31643080.

3. Higgins JPT, Savović J, Page MJ, Elbers RG, Sterne JAC. *Assessing risk of bias in a randomized trial*. Cochrane handbook for systematic reviews of interventions. 2019:205-228.

4. Hartling L, Bond K, Vandermeer B, Seida J, Dryden DM, Rowe BH. *Applying the risk of bias tool in a systematic review of combination long-acting beta-agonists and inhaled corticosteroids for persistent asthma*. PloS one. 2011;6(2):e17242.

5. Losilla J-M, Oliveras I, Marin-Garcia JA, Vives J. *Three risk of bias tools lead to opposite conclusions in observational research synthesis*. Journal of clinical epidemiology. 2018;101:61-72.

6. Eick SM, Goin DE, Chartres N, Lam J, Woodruff TJ. *Assessing risk of bias in human environmental epidemiology studies using three tools: different conclusions from different tools*. Systematic reviews. 2020;9:1-13.

7. Crocker TF, Lam N, Jordao M, Brundle C, Prescott M, Forster A, Ensor J, Gladman J, Clegg A. *Risk-of-bias assessment using Cochrane's revised tool for randomized trials (RoB 2) was useful but challenging and resource-intensive: observations from a systematic review*. J Clin Epidemiol. 2023 Sep;161:39-45. Epub 20230624. doi:10.1016/j.jclinepi.2023.06.015. Cited in: Pubmed; PMID 37364620.

8. Minozzi S, Cinquini M, Gianola S, Castellini G, Gerardi C, Banzi R. *Risk of bias in nonrandomized studies of interventions showed low inter-rater reliability and challenges in its application*. J Clin Epidemiol. 2019 Aug;112:28-35. Epub 2019/04/15. doi:10.1016/j.jclinepi.2019.04.001. Cited in: Pubmed; PMID 30981833.

9. Minozzi S, Cinquini M, Gianola S, Gonzalez-Lorenzo M, Banzi R. *The revised Cochrane risk of bias tool for randomized trials (RoB 2) showed low interrater reliability and challenges in its application*. J Clin Epidemiol. 2020 Oct;126:37-44. Epub 20200618. doi:10.1016/j.jclinepi.2020.06.015. Cited in: Pubmed; PMID 32562833.

10. Bero L, Chartres N, Diong J, Fabbri A, Ghersi D, Lam J, Lau A, McDonald S, Mintzes B, Sutton P. *The risk of bias in observational studies of exposures (ROBINS-E) tool: concerns arising from application to observational studies of exposures*. Systematic reviews. 2018;7:1-11.

11. Kelly SE, Benkhedda K, Brooks SPJ, MacFarlane AJ, Greene-Finestone LS, Skidmore B, Clifford TJ, Wells GA. *Risk of bias in cross-sectional studies: Protocol for a scoping review of concepts and tools*. MethodsX. 2024 Jun;12:102610. Epub 20240212. doi:10.1016/j.mex.2024.102610. Cited in: Pubmed; PMID 38371462.

12. Kelly SE, Brooks SPJ, Benkhedda K, MacFarlane AJ, Greene-Finestone LS, Skidmore B, Clifford TJ, Wells GA. *A scoping review shows that no single existing risk of bias assessment tool considers all sources of bias for cross-sectional studies*. J Clin Epidemiol. 2024 Aug;172:111408. Epub 20240604. doi:10.1016/j.jclinepi.2024.111408. Cited in: Pubmed; PMID 38844117.

13. Higgins JPT LT, Thomas J, Flemyng E, Churchill R. *Methodological Expectations of Cochrane Intervention Reviews*. London: Cochrane.

14. Aloe AM, Dewidar O, Hennessy EA, Pigott T, Stewart G, Welch V, Wilson DB, Campbell MWG. *Campbell Standards: Modernizing Campbell's Methodologic Expectations for Campbell Collaboration Intervention Reviews (MECCIR)*. Campbell Systematic Reviews. 2024;20(4):e1445.

15. Aromataris E MZ, (Editors). JBI Manual for Evidence Synthesis [internet]. Adelaide:: JBI; 2020 June 20, 2022. Available from: https://synthesismanual.jbi.global.

16. Kelly SE, Greene-Finestone LS, Yetley EA, Benkhedda K, Brooks SPJ, Wells GA, MacFarlane AJ. *NUQUEST-NUtrition QUality Evaluation Strengthening Tools: development of tools for the evaluation of risk of bias in nutrition studies*. Am J Clin Nutr. 2022 Jan 11;115(1):256-271. doi:10.1093/ajcn/nqab335. Cited in: Pubmed; PMID 34605544.

17. Bai A, Shukla VK, Bak G, G W. Quality Assessment Tools Project Report. Ottawa, Ontario, Canada; 2012. Canadian Agency for Drugs and Technologies in Health

18. Sterne JAC, Savovic J, Page MJ, Elbers RG, Blencowe NS, Boutron I, Cates CJ, Cheng HY, Corbett MS, Eldridge SM, Emberson JR, Hernan MA, Hopewell S, Hrobjartsson A, Junqueira DR, Juni P, Kirkham JJ, Lasserson T, Li T, McAleenan A, Reeves BC, Shepperd S, Shrier I, Stewart LA, Tilling K, White IR, Whiting PF, Higgins JPT. *RoB 2: a revised tool for assessing risk of bias in randomised trials*. BMJ. 2019 Aug 28;366:l4898. Epub 20190828. doi:10.1136/bmj.l4898. Cited in: Pubmed; PMID 31462531.

19. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, Henry D, Altman DG, Ansari MT, Boutron I, Carpenter JR, Chan AW, Churchill R, Deeks JJ, Hrobjartsson A, Kirkham J, Juni P, Loke YK, Pigott TD, Ramsay CR, Regidor D, Rothstein HR, Sandhu L, Santaguida PL, Schunemann HJ, Shea B, Shrier I, Tugwell P, Turner L, Valentine JC, Waddington H, Waters E, Wells GA, Whiting PF, Higgins JP. *ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions*. BMJ. 2016 Oct 12;355:i4919. Epub 20161012. doi:10.1136/bmj.i4919. Cited in: Pubmed; PMID 27733354.

20. Alhaj OA, Ahmad L, Alenezi AF, Abodoleh G, Ghazzawi H, Trabelsi K, Bragazzi NL, Mehta BM, Faye B, Jahrami HA. *Does total caloric count of camel milk differ by species, country, season and year of publication: A systematic review, meta-analysis and meta-regression*. International Journal of Dairy Technology. 2024;77(1):15-34.

21. Cara KC, Fan Z, Chiu Y-H, Jiang X, Alhmly HF, Chung M. *Associations between Intake of Dietary Sugars and Diet Quality: A Systematic Review of Recent Literature*. Nutrients. 2024;16(11):1549.

22. Cara KC, Taylor SF, Alhmly HF, Wallace TC. *The effects of vitamin D intake and status on symptom severity and quality-of-life in adults with irritable bowel syndrome (IBS): a systematic review and meta-analysis*. Critical Reviews in Food Science and Nutrition. 2024:1-14. doi:10.1080/10408398.2024.2400603.

23. El Sherbiny S, Squillacioti G, Colombi N, Ghelli F, Lenta E, Dalla Costa C, Bono R. *The Effect of Dietary Patterns and Nutrient Intake on Oxidative Stress Levels in Pregnant Women: A Systematic Review*. Antioxidants. 2023;12(7):1427.

24. Feng Y, Zhao Y, Liu J, Huang Z, Yang X, Qin P, Chen C, Luo X, Li Y, Wu Y. *Consumption of dairy products and the risk of overweight or obesity, hypertension, and type 2 diabetes mellitus: a dose–response meta-analysis and systematic review of cohort studies*. Advances in Nutrition. 2022;13(6):2165-2179.

25. Ma W, Zhang Y, Pan L, Wang S, Xie K, Deng S, Wang R, Guo C, Qin P, Wu X. *Association of egg consumption with risk of all-cause and cardiovascular disease mortality: A systematic review and dose–response meta-analysis of observational studies*. The Journal of nutrition. 2022;152(10):2227-2237.

26. Ross AB, Shertukde SP, Staffier KL, Chung M, Jacques PF, McKeown NM. *The relationship between whole-grain intake and measures of cognitive decline, mood, and anxiety–a systematic review*. Advances in Nutrition. 2023.

27. Villoz F, Filippini T, Ortega N, Kopp-Heim D, Voortman T, Blum MR, Del Giovane C, Vinceti M, Rodondi N, Chocano-Bedoya PO. *Dairy intake and risk of cognitive decline and dementia: A systematic review and dose-response meta-analysis of prospective studies*. Advances in Nutrition. 2023:100160.

28. Reeves BC, Wells GA, Waddington H. *Quasi-experimental study designs series—paper 5: a checklist for classifying studies evaluating the effects on health interventions—a taxonomy without labels*. Journal of Clinical Epidemiology. 2017 2017/09/01/;89:30-42. doi:https://doi.org/10.1016/j.jclinepi.2017.02.016.

29. Hiller JE, McMichael AJ. Ecological studies. In: Margetts BM, Nelson M, editors. Design Concepts in Nutritional Epidemiology. Oxford University Press; 1997. Oxford Academic; p. 0.

30. Tominaga S, Kuroishi T. *An ecological study on diet/nutrition and cancer in Japan*. Int J Cancer. 1997;Suppl 10:2-6. eng. doi:10.1002/(sici)1097-0215(1997)10+<2::aid-ijc2>3.0.co;2-c. Cited in: Pubmed; PMID 9209011.

31. Galmés S, Serra F, Palou A. Current State of Evidence: Influence of Nutritional and Nutrigenetic Factors on Immunity in the COVID-19 Pandemic Framework. Nutrients. 2020;12(9). doi: 10.3390/nu12092738

32. Webster TF. *Bias magnification in ecologic studies: a methodological investigation*. Environ Health. 2007 Jul 5;6:17. eng. Epub 20070705. doi:10.1186/1476-069x-6-17. Cited in: Pubmed; PMID 17615079.

33. Burkart S, Parker H, Weaver RG, Beets MW, Jones A, Adams EL, Chaput JP, Armstrong B. *Impact of the COVID-19 pandemic on elementary schoolers' physical activity, sleep, screen time and diet: A quasi-experimental interrupted time series study*. Pediatric obesity. 2022;17(1):e12846.

34. Rogers NT, Cummins S, Jones CP, Mytton O, Rayner M, Rutter H, White M, Adams J. *Estimated changes in free sugar consumption one year after the UK soft drinks industry levy came into force: controlled interrupted time series analysis of the National Diet and Nutrition Survey (2011–2019)*. J Epidemiol Community Health. 2024;78(9):578-584.

35. Yau A, Berger N, Law C, Cornelsen L, Greener R, Adams J, Boyland EJ, Burgoine T, de Vocht F, Egan M. *Changes in household food and drink purchases following restrictions on the advertisement of high fat, salt, and sugar products across the Transport for London network: A controlled interrupted time series analysis*. PLoS medicine. 2022;19(2):e1003915.

36. Linden A. *Challenges to validity in single-group interrupted time series analysis*. Journal of evaluation in clinical practice. 2017;23(2):413-418.

37. Yang C, Hawwash D, De Baets B, Bouwman J, Lachat C. *Perspective: Towards Automated Tracking of Content and Evidence Appraisal of Nutrition Research*. Adv Nutr. 2020 Sep 1;11(5):1079-1088. doi:10.1093/advances/nmaa057. Cited in: Pubmed; PMID 32504536.

38. Thomas DM, Kleinberg S, Brown AW, Crow M, Bastian ND, Reisweber N, Lasater R, Kendall T, Shafto P, Blaine R, Smith S, Ruiz D, Morrell C, Clark N. *Machine learning modeling practices to support the principles of AI and ethics in nutrition research*. Nutr Diabetes. 2022 Dec 2;12(1):48. Epub 20221202. doi:10.1038/s41387-022-00226-y. Cited in: Pubmed; PMID 36456550.

39. Arslan S. *Decoding dietary myths: The role of ChatGPT in modern nutrition*. Clin Nutr ESPEN. 2024 Apr;60:285-288. Epub 20240223. doi:10.1016/j.clnesp.2024.02.022. Cited in: Pubmed; PMID 38479923.

40. Chen D, Lu C, Bai H, Xia K, Zheng M. *Integrating AI with medical industry chain data: enhancing clinical nutrition research through semantic knowledge graphs*. Front Digit Health. 2024;6:1439113. Epub 20241003. doi:10.3389/fdgth.2024.1439113. Cited in: Pubmed; PMID 39421754.

41. Daungsupawong H, Wiwanitkit V. *ChatGPT in modern nutrition: Comment*. Clin Nutr ESPEN. 2024 Jun;61:451. Epub 20240327. doi:10.1016/j.clnesp.2024.03.025. Cited in: Pubmed; PMID 38777469.

42. Hieronimus B, Hammann S, Podszun MC. *Can the AI tools ChatGPT and Bard generate energy, macro- and micro-nutrient sufficient meal plans for different dietary patterns?* Nutr Res. 2024 Aug;128:105-114. Epub 20240714. doi:10.1016/j.nutres.2024.07.002. Cited in: Pubmed; PMID 39102765.

43. Lee HA, Liu CY, Hsu CY. *Precision Nutrition Management in Continuous Care: Leveraging AI for User-Reported Dietary Data Analysis*. Stud Health Technol Inform. 2024 Jul 24;315:256-261. doi:10.3233/SHTI240148. Cited in: Pubmed; PMID 39049264.

44. Papastratis I, Konstantinidis D, Daras P, Dimitropoulos K. *AI nutrition recommendation using a deep generative model and ChatGPT*. Sci Rep. 2024 Jun 25;14(1):14620. Epub 20240625. doi:10.1038/s41598-024-65438-x. Cited in: Pubmed; PMID 38918477.

45. Thomas DM, Knight R, Gilbert JA, Cornelis MC, Gantz MG, Burdekin K, Cummiskey K, Sumner SCJ, Pathmasiri W, Sazonov E, Gabriel KP, Dooley EE, Green MA, Pfluger A, Kleinberg S. *Transforming Big Data into AI-ready data for nutrition and obesity research*. Obesity (Silver Spring). 2024 May;32(5):857-870. Epub 20240301. doi:10.1002/oby.23989. Cited in: Pubmed; PMID 38426232.

46. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, Initiative S. *The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies*. J Clin Epidemiol. 2008 Apr;61(4):344-9. doi:10.1016/j.jclinepi.2007.11.008. Cited in: Pubmed; PMID 18313558.

47. Lachat C, Hawwash D, Ocke MC, Berg C, Forsum E, Hornell A, Larsson C, Sonestedt E, Wirfalt E, Akesson A, Kolsteren P, Byrnes G, De Keyzer W, Van Camp J, Cade JE, Slimani N, Cevallos M, Egger M, Huybrechts I. *Strengthening the Reporting of Observational Studies in Epidemiology-Nutritional Epidemiology (STROBE-nut): An Extension of the STROBE Statement*. PLoS Med. 2016 Jun;13(6):e1002036. Epub 2016/06/09. doi:10.1371/journal.pmed.1002036. Cited in: Pubmed; PMID 27270749.

48. Trevino-Juarez AS. *Assessing Risk of Bias Using ChatGPT-4 and Cochrane ROB2 Tool*. Med Sci Educ. 2024 Jun;34(3):691-694. Epub 20240405. doi:10.1007/s40670-024-02034-8. Cited in: Pubmed; PMID 38887420.

49. Pitre T, Jassal T, Talukdar JR, Shahab M, Ling M, Zeraatkar D. *ChatGPT for assessing risk of bias of randomized trials using the RoB 2.0 tool: A methods study*. medRxiv. 2024:2023.11.19.23298727. doi:10.1101/2023.11.19.23298727.

50. Armijo-Olivo S, Craig R, Campbell S. *Comparing machine and human reviewers to evaluate the risk of bias in randomized controlled trials*. Research Synthesis Methods. 2020;11(3):484-493.

51. Marshall IJ, Kuiper J, Wallace BC. *RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials*. Journal of the American Medical Informatics Association. 2015;23(1):193-201. doi:10.1093/jamia/ocv044.

52. Berciano S, Figueiredo J, Brisbois TD, Alford S, Koecher K, Eckhouse S, Ciati R, Kussmann M, Ordovas JM, Stebbins K, Blumberg JB. *Precision nutrition: Maintaining scientific integrity while realizing market potential*. Front Nutr. 2022;9:979665. eng. Epub 20220902. doi:10.3389/fnut.2022.979665. Cited in: Pubmed; PMID 36118748.

53. Risk of Bias Tools. [internet]: National Toxicology Program of the United States Department of Health and Human Services; 2020 June 13, 2022. Available from:

https://ntp.niehs.nih.gov/whatwestudy/assessments/noncancer/tools/index1.html?utm_source=direct&utm_medium=prod&utm_campaign=ntpgolinks&utm_term=ohat_tools.

Chapter 7
Thesis Appendices

APPENDIX 1: Supplements to Chapter 2

Permission to reproduce figure included in Chapter 2

BMJ PUBLISHING GROUP LTD. LICENSE
TERMS AND CONDITIONS
Oct 11, 2024

This Agreement between Shannon E. Kelly ("You") and BMJ Publishing Group Ltd. ("BMJ Publishing Group Ltd.") consists of your license details and the terms and conditions provided by BMJ Publishing Group Ltd. and Copyright Clearance Center.

License Number	5886191252876
License date	Oct 11, 2024
Licensed Content Publisher	BMJ Publishing Group Ltd.
Licensed Content Publication	BMJ Evidence-Based Medicine
Licensed Content Title	New evidence pyramid
Licensed Content Author	M Hassan Murad,Noor Asi,Mouaz Alsawas,Fares Alahdab
Licensed Content Date	Aug 1, 2016
Licensed Content Volume	21
Licensed Content Issue	4
Type of Use	Dissertation/Thesis
Requestor type	Individual
Format	Print and electronic
Portion	Figure/table/extract
Number of figure/table/extracts	1
Description of figure/table/extracts	Figure 1 to be reproduced with full attribution in the background section of a PHD thesis in Epidemiology.
Will you be translating?	No
Circulation/distribution	15
Order reference number	N/A
Title of new work	DEVELOPMENT OF TOOLS FOR THE EVALUATION OF RISK OF BIAS IN NUTRITION STUDIES
Institution name	University of Ottawa
Expected presentation date	Oct 2024
Order reference number	N/A
Portions	Figure 1 to be reproduced with full attribution in the background section of a PHD thesis in Epidemiology.
The Requesting Person / Organization to Appear on the License	Shannon E. Kelly Ms. Shannon Kelly
Requestor Location	 Attn: Ms. Shannon Kelly
Publisher Tax ID	GB674738491
Billing Type	Invoice

Billing Address	Ms. Shannon Kelly
	Attn: Ms. Shannon Kelly
Total	0.00 CAD

APPENDIX 2: Supplements to Chapter 3

Summary of Dissemination Activities

The development and publication of the NUQUEST checklists for randomized controlled trials, cohort studies and case-control studies provided a fundamental advancement in the field of nutrition research. Nearing the end of the NUQUEST development, and after the open-access publication was made available online, extensive dissemination and capacity-building activities were initiated to promote awareness, understanding, and application of NUQUEST across a variety of stakeholders.

Table 7-1 summarizes these dissemination activities and provides an overview of their format, duration, and target audience and their reach. Activities included information seminars, webinars, and comprehensive workshops ranging from one- to three-day sessions. These activities were held for interested individuals at regional, national and international levels, engaging a broad array of stakeholders, including trainees, researchers, practitioners, policy-makers and academics both within Canada and globally. Collectively, these initiatives facilitated dialogue on the importance of robust risk of bias assessment in nutrition research and supported the integration of the NUQUEST checklists into existing research practices. A summary of related citations is provided below Table 7-1 and a sample presentation is included as an example of the dissemination activities lead by the student.

Table 7-1. Summary of Dissemination and Knowledge Translation Activities

Timing	Activity	Format	Host/Venue	Audience	Purpose	Progress
October 2024	Seminar (30 minutes)	Hybrid	Cochrane India	Knowledge synthesis group	Dissemination of completed tool, training nutrition-specific risk of bias concepts	Completed
July 2024	Seminar (90 minutes)	Hybrid	Tufts University, Boston, MA, USA	Graduate students in nutrition science program	Dissemination of completed tool, training risk of bias concepts	Completed
February 2023	Poster and presentation (60 minutes)	Hybrid	Health Canada Science Forum, Ottawa, ON.	National Regulators, Scientists, Trainees, Policy Analysts	Dissemination of completed tool	Completed
December 2022	Presentation (30 minutes)	Hybrid	School Assembly, School of Epidemiology and Public Health, University of Ottawa. Ottawa, ON.	Local Professors/ Adjunct Professors, Scientists, Administrative staff, Students	Dissemination of completed tool and plans for additional research	Completed
November 2022	Presentation (30 minutes)	Hybrid	Health Canada Science Seminar – Health Products and Food Branch. Ottawa, ON.	National Regulators, Scientists, Trainees, Policy Analysts	Dissemination of completed tool, highlight linkages to regulatory and scientific activities	Completed

Timing	Activity	Format	Host/Venue	Audience	Purpose	Progress
October and November 2022	Virtual workshop series (Three half-day workshops over three weeks)	Online, moderated with facilitators for breakout sessions	Educational workshop series. Canadian Nutrition Society.	National/ International	Dissemination of completed tool, training risk of bias concepts, use and application of completed tool for different epidemiological study designs.	Completed
July 2022	Webinar (1 hour)	Online presentation, moderated question and answer session	WHO Cochrane Cornell Summer Institute for Systematic Reviews in Nutrition for Global Policy Making. Department of Nutritional Sciences. Cornell University. Ithica, New York **See presentation in Appendix 4 for example**	International	Dissemination of completed tool	Completed
	Workshop (Half day)	Online, moderated with facilitators		International	Application of completed tool to RCTs	Completed
	Workshop (Half day)	Online, moderated with facilitators		International	Application of completed tool to cohort studies	Completed
June 2022	Webinar (1 hour)	Online presentation, moderated question and answer session	Education Webinar Series, Canadian Nutrition Society.	National Trainees, Academics, Professionals, Scientists		Completed

Timing	Activity	Format	Host/Venue	Audience	Purpose	Progress
December 2021	Presentation| (60 minutes)	Online Presentation	Nordic Nutrition Recommendations group. Oslo, Norway.	International	Dissemination and feedback on draft tool	Completed
March 2021	Presentation (60 minutes)	Online Presentation	European Food Safety Authority, International Liaison Group on Nutrient Reference Values Methodologies.	International	Dissemination and feedback on draft tool	Completed
May 2019	Workshop (Half day)	In-person, Pre-Conference Workshop	Canadian Nutrition Society Conference	National Trainees, Academics, Professionals	Dissemination of draft tool	Completed
May 2018	Scientific Poster	In-person Scientific Presentation	Canadian Nutrition Society Conference, Halifax, NS	National Trainees, Academics, Professionals	Dissemination of draft tool	Completed

NUQUEST Related Citations for Dissemination Activities

As of 10 October 2024

a) K. Benkhedda, **S.E. Kelly**, A. MacFarlane, L.S. Greene-Finestone, S. Brooks, G.A. Wells, and E.A. Yetley. *Development and validation of NUQUEST—Nutrition Quality Evaluation Strengthening Tools for the evaluation of risk of bias in nutrition studies.* Health Canada Science Forum. February 13-17, 2023. Ottawa, ON.

b) K. Benkhedda, A. MacFarlane, **S.E. Kelly**, L.S. Greene-Finestone, S. Brooks, G.A. Wells, and E.A. Yetley. *Development and validation of NUQUEST—Nutrition Quality Evaluation Strengthening Tools for the evaluation of risk of bias in nutrition studies*. Health Canada Science Seminar – Health Products and Food Branch. November 2022. Ottawa, ON.

c) **S.E. Kelly**, A.J. MacFarlane, L.S. Greene-Finestone, G.A. Wells. *Introduction to epidemiologic study designs and risk of bias concepts*. October 2022 (Virtual Workshop series – Part 1). Canadian Nutrition Society.

d) **S.E. Kelly**, L.S. Greene-Finestone, E.A. Benkhedda, G.A. Wells. *Introduction to NUQUEST*. November 2022 (Virtual Workshop series – Part 2). Canadian Nutrition Society.

e) **S.E. Kelly**, A.J. MacFarlane, L.S. Greene-Finestone, E.A. Benkhedda, S.P.J. Brooks, G.A. Wells. *Application of NUQUEST*. November 2022 (Virtual Workshop series – Part 3). Canadian Nutrition Society.

f) **S.E. Kelly**, A.J. MacFarlane. *Risk of bias tools for nutrition studies*. WHO Cochrane Cornell Summer Institute for Systematic Reviews in Nutrition for Global Policy Making. July 2022. Department of Nutritional Sciences. Cornell University. Ithica, New York. USA.

g) **S.E. Kelly**, A.J. MacFarlane, G.A. Wells. *Randomized Controlled Trial Workshop*. WHO Cochrane Cornell Summer Institute for Systematic Reviews in Nutrition for Global Policy Making. July 2022. Department of Nutritional Sciences. Cornell University. Ithica, New York. USA.

h) **S.E. Kelly**, S.P.J. Brooks, G.A. Wells. *Cohort Study Workshop*. WHO Cochrane Cornell Summer Institute for Systematic Reviews in Nutrition for Global Policy Making. July 2022. Department of Nutritional Sciences. Cornell University. Ithica, New York. USA.

i) **S.E. Kelly**, A.J. MacFarlane on behalf of the NUQUEST Working Group, NUQUEST—NUtrition QUality Evaluation Strengthening Tools: Risk of bias tools for nutrition studies. Canadian Nutrition Society. June 20, 2022, (Virtual webinar). https://cns-scn.ca/education-portal/view/nuquest-a-risk-of-bias-tool-for-nutrition-studies

j) A.J. MacFarlane, **S.E. Kelly**, L.S. Greene-Finestone, E.A. Yetley, K. Benkhedda, S.P.J. Brooks, G.A. Wells. *NUQUEST—NUtrition QUality Evaluation Strengthening Tools:*

development of tools for the evaluation of risk of bias in nutrition studies. Nordic Nutrition Recommendations group. December 8, 2021. Oslo, Norway (Online).

k) A.J. MacFarlane**, S.E. Kelly**, L.S. Greene-Finestone, E.A. Yetley, K. Benkhedda, S.P.J. Brooks, G.A. Wells. *NUQUEST—NUtrition QUality Evaluation Strengthening Tools: development of tools for the evaluation of risk of bias in nutrition studies.* International Liaison Group on Nutrient Reference Values Methodologies. European Food Safety Authority. March 22, 2021. (Online).

l) L. Greene-Finestone, S.E. Kelly, G.A. Wells. *Assessing the methodological quality of nutrition studies: Checklists for cohort, case-control and controlled trials.* Canadian Nutrition Society Annual Conference. Niagara Falls, Ontario, Canada. May 2, 2019. (Workshop).

m) L.S. Greene-Finestone, **S.E. Kelly**, A. MacFarlane, K. Benkhedda, S. Brooks, G.A. Wells, and E.A. Yetley. *Developing a nutrition-specific set of tools to assess the quality of nutrition studies Developing a nutrition-specific set of tools to assess the quality of nutrition studies.* Canadian Nutrition Society Conference, Halifax, May 5-6 2018 (Poster).

Articles Citing NUQUEST

As of 18 October 2024, the manuscript has been cited 14 times according in Google Scholar or the Scopus database. These citations include nine systematic reviews (1-9), one review (10), one commentary (11), one primary research study (12) and one foreign-language record of unknown design or premise (13). one scoping review (14)(Note this is a self-citation from the scoping review in Chapter 4).

1. Alhaj OA, Ahmad L, Alenezi AF, Abodoleh G, Ghazzawi H, Trabelsi K, Bragazzi NL, Mehta BM, Faye B, Jahrami HA. *Does total caloric count of camel milk differ by species, country, season and year of publication: A systematic review, meta-analysis and meta-regression*. International Journal of Dairy Technology. 2024;77(1):15-34.

2. El Sherbiny S, Squillacioti G, Colombi N, Ghelli F, Lenta E, Dalla Costa C, Bono R. *The Effect of Dietary Patterns and Nutrient Intake on Oxidative Stress Levels in Pregnant Women: A Systematic Review*. Antioxidants. 2023;12(7):1427.

3. Feng Y, Zhao Y, Liu J, Huang Z, Yang X, Qin P, Chen C, Luo X, Li Y, Wu Y. *Consumption of dairy products and the risk of overweight or obesity, hypertension, and type 2 diabetes mellitus: a dose–response meta-analysis and systematic review of cohort studies*. Advances in Nutrition. 2022;13(6):2165-2179.

4. Ma W, Zhang Y, Pan L, Wang S, Xie K, Deng S, Wang R, Guo C, Qin P, Wu X. *Association of egg consumption with risk of all-cause and cardiovascular disease mortality: A systematic review and dose–response meta-analysis of observational studies*. The Journal of nutrition. 2022;152(10):2227-2237.

5. Ross AB, Shertukde SP, Staffier KL, Chung M, Jacques PF, McKeown NM. *The relationship between whole-grain intake and measures of cognitive decline, mood, and anxiety–a systematic review*. Advances in Nutrition. 2023.

6. Villoz F, Filippini T, Ortega N, Kopp-Heim D, Voortman T, Blum MR, Del Giovane C, Vinceti M, Rodondi N, Chocano-Bedoya PO. *Dairy intake and risk of cognitive decline and dementia: A systematic review and dose-response meta-analysis of prospective studies*. Advances in Nutrition. 2023:100160.

7. Wingrove K, Lawrence MA, Machado P, Stephens LD, McNaughton SA. *Using the Hierarchies of Evidence Applied to Lifestyle Medicine (HEALM) Approach to Assess the Strength of Evidence on Associations between Dietary Patterns and All-Cause Mortality*. Nutrients. 2022;14(20):4340.

8. Cara KC, Fan Z, Chiu Y-H, Jiang X, Alhmly HF, Chung M. *Associations between Intake of Dietary Sugars and Diet Quality: A Systematic Review of Recent Literature*. Nutrients. 2024;16(11):1549.

9. Cara KC, Taylor SF, Alhmly HF, Wallace TC. *The effects of vitamin D intake and status on symptom severity and quality-of-life in adults with irritable bowel syndrome (IBS): a systematic review and meta-analysis*. Critical Reviews in Food Science and Nutrition. 2024:1-14. doi:10.1080/10408398.2024.2400603.

10. Kris-Etherton PM, Petersen KS, LaMarche B, Karmally W, Guyton JR, Champagne C, Lichtenstein AH, Bray GA, Sacks F, Maki KC. *The Role of Nutrition-Related Clinical Trials in Informing Dietary Recommendations for Health and Treatment of Diseases*. Journal of Clinical Lipidology. 2024.

11. Weaver C, Ahles S, Murphy KJ, Shyam S, Cade J, Plat J, Schwingshackl L, Roche HM, Lachat C, Minihane A-M. *Perspective: peer evaluation of recommendations for CONSORT guidelines for randomized controlled trials in nutrition*. Advances in Nutrition. 2023:100154.

12. Ali AAH. *Evaluation of Macro-micronutrient Intake in Overweight People in Grodno, Belarus*. Journal of Biomedicine and Biochemistry. 2022;1(3):70-79.

13. Bitzer M. *NUQUEST–ein Instrument zur Bewertung des Biasrisikos in Ernährungsstudien*. Aktuel Ernahrungsmed. 2022;47:158-166.

14. Kelly SE, Brooks SP, Benkhedda K, MacFarlane AJ, Greene-Finestone LS, Skidmore B, Clifford TJ, Wells GA. *A scoping review shows that no single existing risk of bias assessment tool considers all sources of bias for cross-sectional studies*. Journal of Clinical Epidemiology. 2024:111408.

Permission to re-print the published journal article

Email correspondence from Elsevier Permissions Helpdesk (received 30 January, 2024)

Dear Shannon Kelly

We hereby grant you permission to reprint the material below at no charge in your thesis subject to the following conditions:

RE: NUQUEST—NUtrition QUality Evaluation Strengthening Tools: development of tools for the evaluation of risk of bias in nutrition studies, The American Journal of Clinical Nutrition, Volume 115, Issue 1, 2022, Pages 256-271, Kelly et al.

1. If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgment to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies.

2. Suitable acknowledgment to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows: "*This article was published in Publication title, Vol number, Author(s), Title of article, Page Nos, Copyright Elsevier (or appropriate Society name) (Year).*"

3. Your thesis may be submitted to your institution in either print or electronic form.

4. Reproduction of this material is confined to the purpose for which permission is hereby given.

5. This permission is granted for non-exclusive world English rights only. For other languages please reapply separately for each one required. Permission excludes use in an electronic form other than submission. Should you have a specific electronic project in mind please reapply for permission.

6. As long as the article is embedded in your thesis, you can post/share your thesis in the University repository.

7. Should your thesis be published commercially, please reapply for permission.

8. Posting of the full article/ chapter online is not permitted. You may post an abstract with a link to the Elsevier website www.elsevier.com, or to the article on ScienceDirect if it is available on that platform.

Kind regards,
Roopa Lingayath

Senior Copyrights Specialist
ELSEVIER | HCM - Health Content Management

Re-print of the published journal article

NUQUEST—NUtrition QUality Evaluation Strengthening Tools: development of tools for the evaluation of risk of bias in nutrition studies

Shannon E Kelly,[1] *Linda S Greene-Finestone,*[2] *Elizabeth A Yetley,*[3] *Karima Benkhedda,*[4] *Stephen PJ Brooks,*[4,5] *George A Wells,*[1] *and Amanda J MacFarlane*[4,5]

[1]Cardiovascular Research Methods Centre, University of Ottawa Heart Institute, Ottawa, Ontario, Canada; [2]Applied Research Division, Public Health Agency of Canada, Ottawa , Ontario, Canada; [3]US NIH (retired), Bethesda, MD, USA; [4]Bureau of Nutritional Sciences, Health Canada, Ottawa, Ontario, Canada; and [5]Department of Biology, Carleton University, Ottawa, Ontario, Canada

ABSTRACT

Background: Dietary exposure assessments are a critical issue in evaluating human nutrition studies; however, nutrition-specific criteria are not consistently included in existing bias assessment tools.

Objectives: Our objective was to develop a set of risk of bias (RoB) tools that integrated nutrition-specific criteria into validated generic assessment tools to address RoB issues, including those specific to dietary exposure assessment.

Methods: The Nutrition QUality Evaluation Strengthening Tools (NUQUEST) development and validation process included 8 steps. The first steps identified *1*) a development strategy; *2*) generic assessment tools with demonstrated validity; and *3*) nutrition-specific appraisal issues. This was followed by *4*) generation of nutrition-specific items and *5*) development of guidance to aid users of NUQUEST. The final steps used established ratings of selected studies and feedback from independent raters to *6*) assess reliability and validity; *7*) assess formatting and usability; and *8*) finalize NUQUEST.

Results: NUQUEST is based on the Scottish Intercollegiate Guidelines Network checklists for randomized controlled trials, cohort studies, and case-control studies. Using a purposive sample of 45 studies representing the 3 study designs, interrater reliability was high (Cohen's κ: 0.73; 95% CI: 0.52, 0.93) across all tools and at least moderate for individual tools (range: 0.57–1.00). The use of a worksheet improved usability and consistency of overall interrater agreement across all study designs (40% without worksheet, 80%–100% with worksheet). When compared to published ratings, NUQUEST ratings for evaluated studies demonstrated high concurrent validity (93% perfect or near-perfect agreement). Where there was disagreement, the nutrition-specific component was a contributing factor in discerning exposure methodological issues.

Conclusions: NUQUEST integrates nutrition-specific criteria with generic criteria from assessment tools with demonstrated reliability and validity. NUQUEST represents a consistent and transparent approach for evaluating RoB issues related to dietary exposure assessment commonly encountered in human nutrition studies. *Am J Clin Nutr* 2022;115:256–271.

Keywords: quality assessment instrument, nutrition, randomized controlled trial, cohort study, case-control study, risk of bias tool

Introduction

Scientific consensus committees tasked with evaluating diet and health–disease relations for policy, clinical, and educational purposes are increasingly applying evidence-based methodologies to the evaluation of nutrition studies (1). However, a universal observation is that although the concepts and methods of evidence-based medicine are applicable to nutrition topics, diet-related challenges in nutrition studies require consideration (1–7). These include the difficulty in obtaining accurate and comprehensive exposure estimates (i.e., intake or status estimates), the potential for confounding and interactive effects of food

Supported by Health Canada.

AJM is an Associate Editor for *The American Journal of Clinical Nutrition* and played no role in the Journal's evaluation of the manuscript.

Supplemental Files 1–7 are available from the "Supplementary data" link in the online posting of the article and from the same link in the online table of contents at https://academic.oup.com/ajcn/.

GAW and AJM contributed equally to this work as senior authors.

Address correspondence to AJM (e-mail: amanda.macfarlane@hc-sc.gc.ca).

Abbreviations used: AHRQ, Agency for Healthcare Research and Quality; NUQUEST, Nutrition QUality Evaluation Strengthening Tools; PICO, Population, Intervention (or exposure), Comparison, and Outcome; QAI, quality assessment instrument; RCT, randomized controlled trial; RoB, risk of bias; SIGN, Scottish Intercollegiate Guidelines Network; SR, systematic review.

Received March 7, 2021. Accepted for publication September 29, 2021.

First published online November 23, 2021; doi: https://doi.org/10.1093/ajcn/nqab335.

substances, and the impact of participant baseline nutritional status on the relation between exposure to a particular food substance and health outcome. If not properly accounted for, these challenges can lead to bias, misleading interpretation, and erroneous conclusions. Thus, the development of an assessment approach that integrates nutrition-specific concepts with accepted general concepts of study design and conduct is of paramount importance to users of nutrition evidence (8–10).

Our objective was to develop a set of reliable and valid tools—Nutrition QUality Evaluation Strengthening Tools (NUQUEST)—to aid in evaluating the types of study designs commonly used for nutrition studies while retaining generic assessment components from existing tools. We found inconsistencies among publications regarding the appropriate terminology as to whether our goal was best described by the term "quality" or "risk of bias" (RoB) (11). These terms, often applied interchangeably, describe conditions related to study design and conduct associated with the validity of study results (12). Quality assessment instruments (QAIs) assess the quality of a study from conception to interpretation, whereas RoB tools assess the accuracy of estimates of benefit and risk. A QAI, an older term than RoB, includes the assessment of precision and generalizability, as well as RoB. Existing QAIs and RoB tools (assessment instruments) are usually generic and focus on the evaluation of general study design and conduct issues that are universally important. Examples of generic tools for assessing RoB previously used in the nutrition literature include Cochrane RoB for randomized controlled trials (RCTs) and ROBINS-I for observational studies (10, 13). However, they do not address specific nutrition-related challenges that contribute to uncertainty in nutrition studies, but, on occasion, are adapted to address them. For example, the Agency for Healthcare Research and Quality (AHRQ) added several nutrient-specific questions to the Cochrane RoB tool for RCTs and the Newcastle-Ottawa Scale for observational studies to address uncertainties associated with dietary assessment measures (10, 14–16). Although modifications can improve the sensitivity of these assessment instruments for nutrition studies, inconsistent adaptation and application limit the validity and comparability of review outcomes.

Here we present the development and validation of NUQUEST, a suite of RoB tools for evaluating nutrition RCTs, cohort studies, and case-control studies. They combine RoB components from an existing assessment instrument with nutrition-specific criteria, detailed guidance, and worksheets that address RoB issues related to dietary exposure assessment. We intended NUQUEST to be helpful to both nonnutrition scientists (e.g., research methodologists, epidemiologists) who may be unfamiliar with nutritional issues, and nutrition scientists who may lack in-depth expertise in human study design and conduct. NUQUEST complements nutrition reporting guidelines, guidelines for conducting nutrition systematic reviews (SRs), and guidelines for grading nutrition evidence, and it can be used in conjunction with these tools when appropriate.

Methods

The coauthors constituted a multidisciplinary working group of 7 US and Canadian experts in the areas of nutrition science and epidemiology with the goal to develop a set of nutrition-specific instruments to be based on validated generic instruments for the assessment of RoB in nutrition research studies. For this work, the term "nutrition studies" would broadly apply to studies in which the focus is on the intake of a food substance (nutrient or other food component), food, dietary pattern, or biomarkers of nutritional intake/status. The development of the nutrition-specific RoB tools for RCTs, cohort studies, and case-control studies followed an 8-step approach: *1*) identify a development strategy; *2*) identify existing and appropriate assessment instruments for adaptation; *3*) identify nutrition-specific appraisal issues; *4*) generate nutrition-specific items; *5*) adapt and/or develop guidance for use; *6*) assess reliability and validity; *7*) assess formatting and usability; and *8*) finalize NUQUEST (**Figure 1**).

Identify a development strategy

The group's overall strategy was to develop a nutrition-specific RoB tool that captured the generic RoB concepts addressed in existing instruments while adding specificity for assessing nutrition issues in nutrition studies. The working group defined the goals, focus, and general approach for the project.

Identify existing and appropriate assessment instruments for adaptation

A rapid, systematic process was used to identify instruments from systematic or comparative review articles that assessed existing QAIs and RoB tools that could serve as a starting point for developing nutrition-specific RoB tools. The rationale for starting with existing instruments was to leverage their demonstrated properties (i.e., validity) and experience of use in the field. We performed a literature search in December 2018 of PubMed, MEDLINE, and Google Scholar using a combination of keywords and Medical Subject Headings syntax relevant to critical appraisal, quality assessment, or RoB to locate potentially relevant articles (**Supplemental File 1**). We included reviews that formally evaluated the appraisal domains and properties of different types of study designs (e.g., RCTs or observational studies). A working group member screened titles and abstracts and then the full text of potentially relevant articles. Included reviews were vetted by the working group until agreement on inclusion status was achieved. In addition, we searched bibliographic databases to identify instruments published after the search date of the most recent review. Individual generic instruments were identified from recommendations made by included reviews, and generic instruments were assessed by the working group using a process that considered the value and feasibility of adding nutrition-specific items in the context of the existing RoB items.

Identify nutrition-specific appraisal issues

In order to develop appropriate nutrition-specific items to add to the generic instruments, the working group identified nutrition-specific appraisal issues. Nutrition-specific issues were defined as those related to nutrition exposure measurement and the interpretation of nutrition data. We considered "dietary exposure assessment" broadly to encompass multiple components

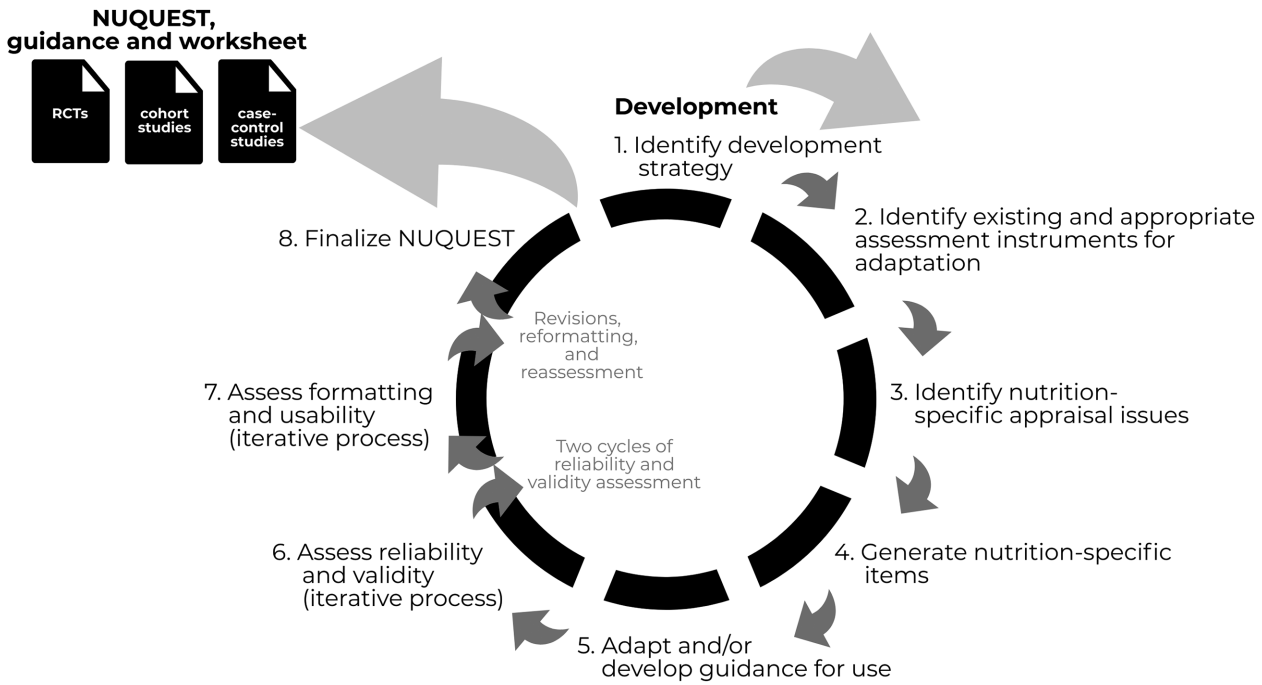


FIGURE 1 Approach for the development of NUQUEST, guidance and worksheet. NUQUEST, NUtrition QUality Evaluation Strengthening Tools; RCT, randomized controlled trial.

of exposure assessment including accuracy and completeness of exposure assessment methodologies, assurances of adherence to interventions, and useful context information for interpreting dietary exposure data. We drew on published literature describing the experiences of scientists who had either conducted or used nutrition-based SRs (7–10), recommendations of expert groups (4, 5, 17–21), articles that had identified nutrition-specific appraisal issues (2, 3), and considered the experiences among the working group members in evaluating nutrition studies. Preliminary nutrition-specific appraisal issues were compiled from all available sources and then vetted within the working group.

Generate nutrition-specific items

The working group drafted a set of nutrition-specific RoB items for each study type (RCTs, cohort studies, case-control studies) based on the identified nutrition-specific appraisal issues with the intention of "bolting" them onto the selected generic instruments as an additional checklist section.

Adapt and/or develop guidance for use

The working group reviewed the existing item-specific guidance for the generic RoB items and edited or revised the text as needed to include context and examples relevant to nutrition studies and the specific study design. We developed detailed guidance de novo for each nutrition-specific item using a structured, iterative process. Two additional experts (non–working group members) in clinical and population nutrition reviewed NUQUEST, including the generic and nutrition-specific items and guidance. The proposed guidance was refined and finalized using an iterative approach.

Assess reliability and validity

To assess the impact of the nutrition-specific items on overall study rating, we used a purposive sample of 45 nutrition-based studies representing 3 study designs (**Supplemental File 2**). Studies were selected based on 2 criteria. First, the study had to have been included in an SR that had examined the association between a nutrition exposure and a health outcome in humans and assigned an overall study rating based on an assessment instrument. Second, the studies as a whole had to cover a diverse range of study ratings, as determined in the original SR, which allowed for comparisons between the original SR overall rating and the NUQUEST overall rating across a range of ratings. The sample consisted of 15 RCTs, 15 cohort studies, and 15 case-control studies on a range of topics. Reliability testing was performed in 2 rounds; the first round consisted of 10 studies for each study design and the second consisted of 5 additional studies for each design. Two raters (not members of the working group), with knowledge of research methodology and critical appraisal or nutrition science, independently applied NUQUEST to the selected studies for each set of studies/study designs.

Reliability.

We assessed reliability using Cohen's κ statistic for interrater agreement, comparing the overall rating of "poor" with "neutral" or "good." A κ value was determined for each checklist and for all checklists combined (22). The working group regarded κ values $\leq$0 as less than chance agreement; 0.01–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, almost perfect agreement (23).

In the first round of reliability testing, the working group assessed the proportion of agreement between raters across 10 studies for each of the 3 types of study design. Interrater reliability was assessed using the proportion of agreement between raters (i.e., raters' assessments of good, neutral, or poor were exactly the same) and the proportion of near agreement (i.e., raters' assessments of good, neutral, or poor either were exactly the same or differed by only 1 category of good, neutral, or poor). After assessing the first 10 studies for each checklist, raters met with the working group to discuss sources of disagreement and strategies for improving agreement. The working group made modifications, as needed, before raters completed the remaining 5 studies for each checklist. Consensus for the 45 sets of NUQUEST ratings was reached by discussion between the 2 raters with adjudication by a third rater.

Validity.

Because there is no accepted "gold-standard" nutrition-specific assessment instrument, it was not possible to assess criterion validity (i.e., how well NUQUEST is in alignment with a gold-standard comparator). A degree of validity was established by identifying a generic tool on which to build NUQUEST that had a long history of application and has performed well in identifying studies of different RoB to the satisfaction of authors and users (Step 2). In addition, content and concurrent validity of NUQUEST was assessed. Content validity was established by including issues specifically related to dietary exposure and other key aspects needed when evaluating nutrition studies (Steps 3 and 4), which provided specific content that was in alignment with the important issues related to nutrition studies, over and above the general methods items of the selected generic tool on which NUQUEST was based. Concurrent validity, how a new tool compares to an existing tool, was assessed by comparing the NUQUEST overall rating for each study (poor, neutral, and good) to the overall rating reported in the original SR from which the study was selected. The ratings from the original SR assessment were recoded into categories of poor, neutral, and good to allow for comparison to NUQUEST.

Assess formatting and usability

Verbal and written feedback was sought from the raters pertaining to the usability of NUQUEST, clarity of the appraisal items and corresponding guidance, and their overall experience applying the tools. They also tracked the time it took to complete each NUQUEST assessment, and documented instances where they perceived scoring to be challenging or guidance unclear.

Finalize NUQUEST

Taking into consideration the final feedback from raters, the wording of the NUQUEST guidance was edited for clarity and made consistent across the 3 study design–specific checklists.

Results

Identify a development strategy

The working group identified their initial goals for the development of NUQUEST: *1*) to integrate accepted generic criteria for assessing study design and conduct with nutrition-specific methodological issues relating primarily to nutritional intake and exposure; and *2*) to have the resultant NUQUEST be useful both to epidemiological experts who may have limited knowledge of nutritional methodologies and to nutritional scientists who may have limited knowledge of human study design and conduct issues. To achieve these goals, the group decided to adapt an existing generic assessment instrument to maintain the scientific concepts and qualities associated with the items included in the instrument, with some limited modifications to enhance the usability for nutrition contexts. In addition, the group decided that "bolt-on" nutrition-specific items could be added to the generic assessment instruments when those concepts were not addressed by items in the generic instrument. The group agreed that NUQUEST should initially focus on common study designs used in nutrition science.

Identify existing and appropriate assessment instruments for adaptation

The rapid review (24, 25) located 6 SRs (26–31) comparing various assessment instruments after a review of 367 titles and abstracts and 14 full-text articles. Six potentially relevant records (5, 9, 32–35) identified through additional top-up searching were also considered. One unpublished environmental scan was identified (36, 37), but eventually excluded because it identified, but did not evaluate, a number of available assessment instruments.

The working group used the most recent, comprehensive, and rigorously conducted SR by Bai et al. (26) to select an assessment instrument for adaptation. The Bai et al. (26) SR is an agency-sponsored report completed by the Canadian Agency for Drugs and Technologies in Health that updated the findings of a comprehensive 2002 AHRQ report (38) that identified and evaluated assessment instruments. Briefly, a total of 267 unique assessment instruments (57 for SRs, 94 for RCTs, 99 for observational studies, 17 for multiple designs, and 60 evidence grading systems) were included after a comprehensive multidatabase search of published and unpublished sources and applying a rigorous approach to comparisons and evaluation. Each assessment instrument was assessed independently by $\geq$2 review authors and external content experts based on its adequacy in addressing 5 study-level domains important for generic assessment instruments: comparability of subjects, exposure ascertainment/intervention, outcome measure/ascertainment, statistical analysis, and funding. Bai et al. noted that a variety of instruments for appraising studies exist; yet, no gold standard has been identified. They recommended 3 generic assessment instruments produced by the Scottish Intercollegiate Guidelines Network (SIGN) (39, 40) (1 checklist each for RCTs, cohort studies, and case-control studies) based on internal validity and the potential for feasible, consistent, and systematic application of the checklists during evidence appraisal.

There are many assessment instruments but only a few that have been developed to assess with consistency a range of study designs including RCTs and observational studies (cohort studies and case-control studies), such as those developed by the SIGN (39, 40) and the Joanna Briggs Institute (41–43). Other advantages of the SIGN instruments are that they are simple to use, include key criteria for RoB and clear guidance for their application and interpretation, and have been used extensively. Also, SIGN instruments are not outcome-specific so they can be used for a wide range of health outcomes. The working group concluded that the advantages of the SIGN approach supported its use as the basis for the concurrent development of nutrition-specific RoB tools across 3 commonly used nutrition study designs (i.e., RCTs, cohort studies, and case-control studies) for which a suite of SIGN tools were available. Because a SIGN checklist for cross-sectional studies was not available, a de novo process will be required for the future development of NUQUEST for cross-sectional studies.

Identify nutrition-specific appraisal issues

Based on findings from a broad literature review (1–5, 17, 19, 34, 44–47) and building on the experiential knowledge of the working group members in performing assessments of nutrition studies, the group identified 5 essential nutrition-specific issues to consider when evaluating studies. These issues are explained in detail in **Table 1** and broadly relate to *1*) accuracy of exposure estimates; *2*) baseline exposure; *3*) exposure throughout the study; *4*) consideration of the dietary context; and *5*) the duration of exposure relative to outcome.

The first issue identified was the challenge involved in obtaining accurate estimates of exposure and status—an issue that has received considerable attention (1, 4, 5, 17, 44, 45, 47). Several authoritative reports have offered guidance on how to deal with these issues in reporting and evaluating study RoB, and in analyzing study results (4, 5, 17, 19, 34). The second issue—the need for accurate baseline exposure or status measurements—has been recognized as an important issue for cohort studies (34) but is also important for all nutrition study designs (2, 3, 46). The third issue—the need to assess exposure/status throughout the study—affects all study designs, and is especially relevant to the issue of adherence to study protocols in intervention studies (46). The fourth issue—the need to consider dietary context when evaluating the effect of specific food substances or dietary patterns on health outcomes—is important because it can affect the exposure–outcome relation. Diets are complex mixtures containing multiple biologically active components that affect food substance bioavailability and metabolic utilization, and any food substance of interest may be correlated with other food substances in the diet (1, 3, 4). Energy intakes are important for understanding the full effect of an addition or deletion of a food, or substitution of one food for another, for identifying potential errors in dietary intake reporting (e.g., under- or overreporting of dietary intake) (44, 45, 48). The various dietary and biological complexities and their effects on food substance and outcome relations of interest may vary systematically between subpopulation groups. The last nutrition issue—the duration of exposure relative to outcome—is important because the duration of exposure must be sufficient to reasonably expect the full effect of the exposure on development of the outcome from both an effectiveness and a safety perspective (46, 49–51).

Generate nutrition-specific items

Reflecting on the identified nutrition-specific issues from Step 3, the working group developed nutrition-specific items to address the issues when generic items of the existing assessment instrument addressed them either insufficiently or not at all. In the case of the former, the nutrition-specific items were designed to build on concepts addressed by generic RoB items, thereby emphasizing specific important details related to nutrition exposure. The items were focused on measurement of exposure, assessment of baseline exposure, exposure assessment throughout the study, the dietary context of the exposure and potential for confounding, and the duration of exposure relative to the outcome. They were "bolted-on" to the SIGN checklists, as appropriate for each study design. An iterative process among the working group and raters was used to refine the wording of the nutrition-specific items. The final nutrition-specific items vary in number and wording according to the study design and consider methodological rigor, complexity, and user-friendliness (**Table 2**). We added 5 nutrition-specific items to the RCT checklist and 4 to the cohort checklist. For the case-control checklist, we retained the original SIGN item for exposure assessment and added 2 additional nutrition-specific items.

Adapt and/or develop guidance for use

Key nutrition concepts applied to both generic and nutrition-specific items. As such, we added nutrition-specific guidance to the existing guidance for the generic SIGN items, including contextual examples pertinent to nutrition studies (**Table 3** presents examples of nutrition-specific guidance for generic items included in NUQUEST for cohorts). We also developed de novo guidance for each nutrition-specific item (**Table 4** presents examples of guidance for nutrition-specific items included in NUQUEST for cohorts). Although exposure was assessed by generic items, we emphasized the importance of accurate exposure assessment in the context of nutrition studies (NUQUEST for RCTs, cohort studies, and case-control studies with complete guidance are in **Supplemental Files 3**, **4**, and **5**, respectively). In NUQUEST for case-control studies, the generic item on exposure assessment received greatly enhanced nutrition-specific guidance, similar to that of NUQUEST for RCTs and cohort studies.

Guidance pertaining to the assessment of exposure and intake was expanded with each round of validation to include discussion of accuracy and validity in a nutrition context. Topics covered include baseline assessments, adherence to intervention, self-report, supplement and food product composition, dietary patterns or exposures, biomarkers and appropriate documentation, and the appropriateness of dietary assessment methodologies for different research purposes. Guidance was edited to include examples from nutrition studies to illustrate specific issues and facilitate tool application and decision-making.

Assess reliability and validity

Raters completed the assessment of 45 nutrition studies (i.e., 15 RCTs, 15 cohort studies, and 15 case-control studies)

TABLE 1 Nutrition-specific issues[1]

Nutrition-specific issue	Rationale for consideration of the issue in the NUQUEST checklists
Accuracy of exposure estimate	This issue refers to the ability to assess nutritional exposures accurately and with minimum error: • The 2 main issues are measurement error and validation of the assessment methodology. • Measurement error is the difference between the measured value and the true value. It has 2 forms: random error and systematic bias. • Random error (day-to-day or within-person variation) is related to an individual's measured intake on a specific day of testing vs. an individual's long-term average intake ("usual" intake) based on multiple administrations of the same instrument: ○ These data are imprecise but not biased. ○ With repeated measures on a subsample of a population, statistical modeling can be used to adjust for these effects (i.e., to estimate "usual" intakes). ○ Failure to deal with random error results in overestimates of tail probabilities, attenuated relations, and loss of power. • Systematic error (bias) results in measurements that depart from the "true intake" in the same direction: ○ This can result from under- or overreporting or misreporting intakes and may also be related to individual characteristics (e.g., BMI). Some biomarker methodologies can also exhibit bias. ○ This type of error cannot be reduced or eliminated by taking repeated measurements. ○ Bias can result in loss of power to detect diet–health relations or in errors in establishing accurate intake–response curves. ○ The magnitude of bias varies by the type of methodology used, among different nutrients, and with participant characteristics. • Validation studies are conducted to determine how accurately exposure instruments measure true exposures: ○ One type, rarely used, assesses validation by collecting reference measures by direct observation or feeding studies for a time period that is exactly consistent with when each exposure measurement is taken. ○ A second type collects reference measures such as recovery biomarkers or less biased assessment instruments for a time period that is not exactly the same as the measured exposure assessment. Bias is the difference between the average reported intake and average true intake at the group level. ○ Correlation coefficients between measured and true usual intakes are related to the loss of power to detect exposure–outcome relations. ○ Unbiased reference measures are limited, but include some recovery biomarkers and accurately collected data from feeding studies or direct observation. ○ Validation of one method using a comparison with a second method that has a known bias or imprecision (i.e., a weak or flawed reference instrument) can propagate the bias. A common example of this is when one FFQ with a known bias is used as the comparator for another new FFQ. • Sources of error and lack of validation for exposure assessments can result in misleading results for assessing diet–health relations, intake–response curves, and population prevalence estimates. • Differences in research objectives can affect the appropriateness of different dietary assessment approaches. • Failure to consider measurement error and validation can lead to erroneous and misleading study conclusions.
Baseline exposure	This issue refers to the need for consideration of baseline exposure in assessing exposure–outcome relations: • Allows for correct description/identification of study groups. • Allows for comparison of nutritional and dietary similarities and differences between study groups. • Facilitates the evaluation of generalizability of study populations to user target populations. • Facilitates the identification of potential nutrition-based confounders and nutrient–nutrient interactions that can influence the relations between the exposure of interest and the outcome, and which therefore should be accounted for in the analysis. • Observational studies: allows for appropriate assignment of participants to study groups. • Intervention trials: baseline exposures can inform on total exposure (from both diet/supplements and intervention, for example), allowing for the determination of accurate intake–response curves.
Exposure throughout study	This issue refers to the need to document changes to exposure throughout the study that could affect exposure–outcome relations: • Identifies exposure changes during the study that can occur owing to self-initiated changes in participant dietary patterns and supplement use or compositional changes in marketed food products used by study participants that occur after baseline assessment. These changes can affect exposure–outcome relations in unpredictable ways and lead to erroneous conclusions. Accounting for these changes in analyses is necessary to prevent misleading and erroneous results. • Intervention studies: adherence to intervention products should be monitored because it can affect the actual exposure and thus the exposure–outcome relation.

(*Continued*)

TABLE 1 (*Continued*)

Nutrition-specific issue	Rationale for consideration of the issue in the NUQUEST checklists
Dietary context	This issue refers to the need to consider the complexity of the dietary exposure to determine its impact, if any, on the exposure–outcome relation: • Food substances may correlate with other food substances in the diet, thereby confounding the attribution of an exposure–outcome relation to a single specific food substance. • Bioavailability differences among food substance sources (e.g., naturally occurring or synthetic) and/or the varying effects of different dietary patterns on the bioavailability of the food substance of interest can affect the exposure–outcome relation. • Direct or indirect interactions between the food substance and other dietary components can alter its effect on the outcome. • Dietary context provides information on total intakes or exposure of the food substance (e.g., background diet plus supplements, bioavailable vs. less bioavailable formats), thus more accurately capturing the effect of the food substance on the outcome. • Dietary context can provide information on energy intakes. This information is important in adjusting for potential biases in nutrient exposures related to energy intakes, when evaluating equivalencies of food or dietary pattern substitutions, or in identifying the potential for under- or overreporting of food substances of interest. • Dietary context provides information to calculate ranges or distributions of the food substance, thus enhancing the ability to create intake–response curves. • Dosing or eating conditions (e.g., single bolus vs. multiple exposures per day; with meals vs. between meals; chemical forms of the supplement) can affect the intake–response relation. • Intervention studies: the ability to "blind" a study may be affected when foods or dietary patterns are the intervention; in studies where there is an addition or deletion of a food, or a food substitution is made, energy equivalency between comparison groups must be considered.
Duration of exposure relative to outcome	This issue refers to the need for the exposure to the food substance to be of sufficient duration to see the full effect on the outcome: • The time needed to show a significant and nontransient change in the outcome may vary by food substance and outcome assessed. • Baseline exposure of study participants can affect the length of time of exposure needed to see an effect of the food substance on the outcome. • The dose of the exposure may affect the duration required to observe a change in the outcome. For example, low doses consumed for a longer duration can have the same effect on the outcome as high doses consumed for a shorter time. • Similarly, the format of the exposure may affect the duration required to observe a change in the outcome. For example, a food substance in supplement form may be more bioavailable than as part of a dietary intervention. • Outcomes based on a surrogate marker might reasonably be expected to occur in a shorter time period than the occurrence of a disease outcome.

[1]For the purpose of this work, in the context of a nutrition study, "exposure" refers to intake or nutritional status of the food substance or dietary pattern (e.g., folate, Mediterranean diet) of interest. The food substance definition in the table can be found in the checklists. We considered "dietary exposure" broadly to encompass multiple components of exposure assessment including accuracy and completeness of exposure assessment methodologies, assurances of adherence to interventions, and useful context information for interpreting dietary exposure data of interest. The National Cancer Institute's Dietary Assessment Primer provides comprehensive and detailed information on how to assess and use exposure assessment data (17, 48). NUQUEST, Nutrition QUality Evaluation Strengthening Tools.

selected from SRs in the nutrition literature in 2 separate rounds (10 of each study design in round 1 without a worksheet; 5 in a second round with a worksheet; **Supplemental File 6**). The working group developed the NUQUEST worksheet based on rater feedback after the first round of testing. This worksheet is intended to provide, as needed, a framework for raters to a priori learn how to assess each item as it relates to the research question; the Population, Intervention (or exposure), Comparison, and Outcome (PICO); and methods of assessing intake/nutritional status, outcomes, and confounders. Each of the NUQUEST sections, based on items grouped according to selection, comparability, ascertainment of outcomes (or exposure), and the overall assessment, was rated independently by each rater as "good," "neutral," or "poor."

Reliability.

Overall interrater reliability for NUQUEST across all study designs (n = 45 studies) was substantial (0.73; 95% CI: 0.52, 0.93) and moderate to almost perfect for the individual instruments for RCTs (0.59; 95% CI: 0.19, 1.00; n = 15 studies), cohort studies (0.57; 95% CI: 0.15, 1.00; n = 15 studies), and case-control studies (1.00; 95% CI: 1.00, 1.00; n = 15 studies). κ Values were not robust when considering only the first round of reliability testing owing to the limited sample size (n = 30 studies) and the corresponding number of nominal categories, therefore only the combined κ values are presented (n = 45 studies).

The proportion of agreement of the 2 raters was generally satisfactory for the individual sections and NUQUEST overall rating (**Table 5**). The proportion of "near agreement" (i.e., the raters' assessments of good, neutral, or poor were either exactly

TABLE 2 Nutrition-specific items in Nutrition QUality Evaluation Strengthening Tools (NUQUEST)

Study design	Nutrition-specific items
Randomized controlled trial	• The frequency and quantity of the intervention exposure under study are accurately and reliably measured. • The relevant baseline exposure is measured and taken into account. • The baseline exposures of all groups have been maintained over the course of the study. • Adherence to the intervention and control intakes is monitored in a reliable and accurate manner. • The interval between the intervention and outcome is of sufficient duration to observe an effect, if there is one.
Cohort study	• The frequency and quantity of the exposure under study are accurately and reliably measured. • The relevant exposure at baseline is measured and taken into account. • The baseline exposure differences between the groups have been maintained over the course of the study. • The interval between the exposure and outcome is of sufficient duration to observe an effect, if there is one.
Case-control study[1]	• The exposure under study is accurately and reliably measured. • The baseline exposure differences between the groups have been maintained over the course of the study. • The interval between the exposure and outcome is of sufficient duration to observe an effect, if there is one.

[1] All items except for the exposure assessment in NUQUEST for case-control studies are found in the nutrition-specific section of NUQUEST. An exception was made for NUQUEST for case-control studies because the generic assessment items for this study design included an exposure ascertainment item. Nutrition-specific guidance was developed for this generic item to highlight the specific issues in assessing nutrition exposures, and to highlight the fact that retrospective case-control study designs are particularly vulnerable to recall bias.

the same or differed by only 1 category of good, neutral, or poor) on the selection, comparability, ascertainment, and nutrition sections of NUQUEST, as well as the overall rating, ranged from 80% to 100% for RCTs, from 80% to 100% for cohort studies, and from 87% to 100% for case-control studies. The proportion for "agreement" (i.e., raters' assessments of good, neutral, or poor were exactly the same) was lower. In particular, for the overall rating, agreement was only 40% based on the first 10 studies assessed for each of the 3 study designs. For the assessment of the remaining 5 studies for each study design, the agreement improved to 100% for RCTs, 80% for cohort studies, and 100% for case-control studies when raters used the completed worksheets.

Rater feedback indicated that it would have been helpful to meet to review their consensus on NUQUEST items after assessment of the first 2–3 studies before completing the

TABLE 3 Examples of nutrition-specific guidance developed for generic items of Nutrition QUality Evaluation Strengthening Tools (NUQUEST) for cohort studies

Concept	Item	Sample guidance
Selection of participants/creation of study groups	1.1 The groups being studied are selected from source populations that are comparable in all respects other than the exposure under investigation.	…Groups that are formed from 2 different source populations or are selected from the same source population using different approaches are at higher risk of selection bias. For example, comparing a group of women of childbearing age from an urban center who took folic acid during pregnancy with those from a rural area who did not would be problematic…
	1.2 The study indicates how many of the people asked to take part did so in each of the groups being studied.	…Groups selected from different source populations, or from the same source population using different approaches, may have differential participation rates and should be evaluated more closely. For example, one would be less concerned about differential participation rates in a cohort of women of childbearing age who were selected from the same urban area and among whom groups were formed based on whether or not they took folic acid during their pregnancy. In contrast, if populations from rural and urban areas are compared, we would have more concern that participation rates might differ because, for instance, access to the study centers may be more difficult for the rural participants.
Comparability of groups	1.8 The main potential confounders are identified and/or taken into account in the design and analysis.	…There may be fundamental differences between the groups that may affect nutrient intake or exposure and these should be considered and accounted for in a study (e.g., health-related behavior or personal/lifestyle variables). These factors may be known or unknown. For example, in a cohort of women of childbearing age where folic acid intake (compared with no intake) is being studied, we may want to consider potential confounders that may be related to the nutrient intake such as education, physical activity, and race or ethnicity. In addition, and particularly when we have different source populations (e.g., different geographies), a variety of known and unknown differences between the groups may confound the nutrient intake or exposure, including but not limited to population structure, culture, genetic diversity, and health or personal behaviors.

TABLE 4 Examples of nutrition-specific items and guidance developed for the nutrition-specific items of Nutrition QUality Evaluation Strengthening Tools (NUQUEST) for cohort studies[1]

Concept	Item	Sample guidance
Exposure of interest/intervention intake	2.1 The frequency and quantity of the exposure under study are accurately and reliably measured.	…*The study should report the method used to assess exposures or status and the validity of such methods.* Validity refers to the degree to which the exposure assessment methodology accurately measures the aspect of exposure it was intended to measure. Exposure assessment methods are validated in different ways. Comparing an exposure method to a "gold-standard" or reference method thought to capture all the food and supplement sources containing the exposure of interest provides the best validation approach. Comparing a self-reported method to another self-reported method requires caution because the 2 methods may be affected by the same types of errors and biases. An exposure assessment method could be modified to reflect the population under study if the food/supplement supply differs for the new study population. Validation of a new questionnaire against a previous questionnaire targeting the exposure under study allows for comparability among similarly validated studies… Currently available assessment methods are often subject to random errors and systematic biases that can result in misleading and erroneous conclusions. Optimal methodologies as well as the nature and severity of potential errors and biases vary among food substances and, therefore, must be considered on a case-by-case basis. Methods used in the study under evaluation need to be evaluated for their potential for RoB that includes both general and topic-specific criteria. Topic-specific RoB criteria are issues related specifically to the exposure of interest. For example, issues related to assessing sodium intake will be different from those of other nutrients or total energy… *Food substance intakes from self-reported diets* (e.g., FFQ, food record, or 24-h recall). Methods have different strengths and weaknesses (known errors and biases)… The frequency of evaluation of the exposure will often depend on the food substance, the dosage conditions, and the outcome. *Food substance intakes from supplements or specific food products* (e.g., bars, liquid drinks, fortified foods). The study should report product composition details and the methods used to confirm the composition of the product (i.e., assay procedure), and how adherence to the use of these products was monitored (e.g., supplement count, assessment of changes in nutritional status). *For dietary pattern or food intakes*, the study should include ranges or distributions of any prespecified food substance or status exposures of interest. The study should also include exposure measurement characteristics and errors, of which some can be controlled for statistically (e.g., "usual" intakes). This can be accomplished by intake methodologies that represent averages over a specified time period or by statistically adjusting for within-person variabilities in 24-h recalls. Some intake methodologies cannot adjust for usual intakes (e.g., if they are unknown or unmeasured; or subject to a systematic bias such as underreporting) and this can introduce bias. *Biomarkers of intake and exposure* are usually metabolites recovered from the blood or urine that objectively and accurately assess intake over a period of time. However, biomarker assays can vary in accuracy and reliability. If a biomarker has been accurately measured and appropriately qualified for its intended purpose, it is more reliable than exposure assessed by dietary intake estimation methods. Note that use of nonstandardized methodologies could affect status measures, even when the method itself includes standards and controls. For example, 25-hydroxyvitamin D values differ depending on the assay platform and the laboratory performing the measure. A measure from a laboratory that has been certified for the method will be more reliable than one from an uncertified laboratory and standardized procedures are more reliable than nonstandardized procedures. *Other important considerations for assessing exposure include the composition, the form, any potential interactions with other exposure components, supplement brand name, and food types or recipes. The importance and weight of these issues (and other relevant characteristics of the exposure) must be determined a priori by the reviewer.* Knowledge about the exposure form (e.g., naturally occurring vs. synthetic food substance) and intake conditions (e.g., taken with or without meals; bolus vs. multiple daily exposures; preparation practices for meal components) can be important, especially when bioavailability and/or bioactivity are issues. Ideally, product composition would be confirmed by analysis because food/supplement label values and/or composition databases do not always accurately reflect actual composition. *If the exposure is not accurately measured or adequately characterized, confidence in the degree of exposure is reduced. A value on a product label could be acceptable but might not accurately reflect composition and should be interpreted with caution.* Energy intakes are another consideration. They are important for understanding the full effect of an addition or deletion of a food, or substitution of one food for another, and for identifying potential errors in dietary intake reporting (e.g., under- or overreporting of dietary intake).

(*Continued*)

TABLE 4 (*Continued*)

Concept	Item	Sample guidance
		Rating: Documentation of exposure should be sufficient to determine its validity. For example, intake/nutritional status based on a comparison to an accurate external method would have the highest rating. Self-reported intake or use of a biomarker assay procedure with accuracy or bias concerns that have been described/evaluated would have higher uncertainty. Self-reported intake with no description of accuracy or bias would have the lowest rating. In addition, because of large day-to-day variabilities in intakes by a given individual, higher rating should be given to results expressed as "usual" or long-term intakes. If adherence to the exposure is assessed by a supplement count then there is greater confidence in the measurement than if the assessment is achieved by recall. The tolerance for uncertainty will vary depending on the purpose of your evaluation and will be reflected in the study rating. The rationale for the rating system used, including a description of tolerance for uncertainty, should be defined a priori.
Baseline nutrient intake or exposure	2.2 The relevant exposure at baseline is measured and taken into account.	Accurate baseline intake/status data are necessary to ensure appropriate group assignments; to determine whether changes in intake occurred during the study period (which could confound study results); and to compare results across studies in different populations. If supplement use is the exposure, baseline intakes can be used to ascertain the nutritional similarity of exposed and unexposed groups at baseline and estimate total intake. If expressed as distributions, baseline intakes of study groups can aid in the interpretation of results. For example, if a wide range of baseline statuses is present among groups of a supplement study, the overall effect of the exposure may be null; this could happen if nutrient-depleted participants experience benefit but replete participants experience no effect or even harm…
Intake adherence and maintenance	2.3 The baseline exposure differences between the groups have been maintained over the course of the study.	…All participants are exposed to diets that include food substances, foods, or dietary patterns of interest. Changes in intake over the duration of the study for any or all of the groups could make it more difficult to detect differences in outcome due to the effect of the baseline exposures. Has the potential for this been monitored? If a change in intake has occurred, does the study have appropriate methods to deal with it?…
Intake/outcome interval	2.4 The interval between the exposure and outcome is of sufficient duration to observe an effect, if there is one.	In order to determine whether a "true" relation exists, the duration of the interval between the exposure and the outcome must be sufficient to allow the outcome to occur. False positives can occur in studies of inadequate duration. For example, transient decreases in LDL cholesterol may be observed in short-term lipid studies but may not persist in long-term studies. Confounding factors such as seasonal changes can transiently affect status or intake. Even study enrolment can transiently affect intake or status in the short term…

[1]RoB, risk of bias.

assessment for all of the studies. They felt that this would have increased their agreement for items for which ratings were inconsistent. The multidisciplinary backgrounds of the raters underscored the need to understand the nature and sources of discrepancies between their ratings, which were generally dependent on the expertise of each individual rater (study appraisal methodology or nutrition).

Validity.

As noted, the SIGN instruments on which NUQUEST was built demonstrate high content validity (26) by virtue of their original development process and extensive history of use. When the working group compared the overall study ratings from NUQUEST (based on consensus from 2 independent raters) to the overall study ratings from the original SR (**Figure 2**), there was agreement for 21 of the 45 studies (46.7%; 10 RCTs, 9 cohort studies, 2 case-control studies) and at least near agreement for 42 studies (93.3%; 14 RCTs, 15 cohort studies, 13 case-control studies). When the overall ratings differed between NUQUEST and the original SR, the rating was lower 79% of the time (19 of 24 studies) and higher 21% of the time (5 of 24 studies). For the 3 studies for which there was not at least near agreement (study numbers 1, 32, and 39; see Figure 2), the nutrition bolt-on section affected the NUQUEST overall study rating negatively and led to greater disagreement between the NUQUEST overall rating and SR overall rating. In summary, the overall study ratings generally demonstrated agreement between NUQUEST and the original SR ratings and, in instances of major disagreement, the nutrition-specific section was a key contributor to the difference.

Assess formatting and usability

Based on feedback from raters and consensus discussions among the working group members, we modified the format and organization of NUQUEST to improve the logical structure and usability while maintaining the scientific concepts outlined in the original SIGN instruments (Supplemental Files 1–3). We edited the original SIGN items and guidance slightly to improve clarity and, where appropriate, to highlight their relevance to nutrition studies. NUQUEST was reorganized into 5 sections that varied in title depending on the study design: *1*) *Selection*—Selection of participants (RCT)/Selection of cohorts (cohort)/Creation of study groups (case-control); *2*) *Comparability*—Comparability of study groups (RCT, case-control)/Comparability of cohorts (cohort); *3*) *Ascertainment*—Ascertainment of outcomes (RCT, cohort)/Exposure ascertainment (case-control); *4*) *Nutrition*—Nutrition-specific issues (RCT, cohort, and case-control); and *5*)

TABLE 5 Rater agreement on individual sections and overall rating using Nutrition QUality Evaluation Strengthening Tools (NUQUEST)[1]

			NUQUEST assessment in agreement, %				
Study design	Studies, *n*	Degree of agreement	Selection[2]	Comparability[3]	Ascertainment[4]	Nutrition[5]	Overall[6]
Randomized controlled trial	15	Near agreement	87	80	93	100	87
		Agreement	40	47	93	80	60
	10	Agreement without worksheet	40	40	80	100	40
	5	Agreement with worksheet	40	60	80	80	100
Cohort	15	Near agreement	100	80	80	93	93
		Agreement	67	47	33	87	53
	10	Agreement without worksheet	60	30	30	80	40
	5	Agreement with worksheet	80	80	40	100	80
Case-control	15	Near agreement	93	87	93	100	100
		Agreement	60	47	40	60	60
	10	Agreement without worksheet	70	20	40	40	40
	5	Agreement with worksheet	40	100	40	100	100

[1]Agreement/near agreement was determined section by section and overall. Percentage in agreement was calculated for each section/overall using the number of studies with rater agreement or near agreement out of the total number of studies assessed. Near agreement: raters' assessments of good, neutral, or poor were either exactly the same or within 1 category of good, neutral, or poor. Agreement: raters' assessments of good, neutral, or poor were exactly the same. RCT, randomized controlled trial.

[2]Selection = selection of participants (RCT)/selection of cohorts (cohort)/creation of study groups (case-control).

[3]Comparability = comparability of study groups (RCT, case-control)/comparability of cohorts (cohort).

[4]Ascertainment = ascertainment of outcomes (RCT, cohort)/exposure ascertainment (case-control).

[5]Nutrition = nutrition-specific (RCT, cohort, and case-control).

[6]Overall = NUQUEST overall rating.

Overall study rating. A table in the guidance section of each checklist describes the individual sources of bias that NUQUEST is designed to address (i.e., performance, attrition, detection, selection, misclassification, and dietary exposure assessment biases). The revised structure and improved guidance brought focus to the generic concepts of selection, comparability, and ascertainment and to the de novo developed bolt-on nutrition concepts, and facilitated the determination of the overall rating for a specific concept (e.g., selection), which in turn assisted the decision-making required for the overall study rating.

The original SIGN rating scheme for each item included the following response options: yes/no/can't say/does not apply. In the first round of validation (30 studies), the raters frequently gave "fence-sitting" responses where they used the response option of "can't say"; however, the working group determined that this rating was not always justified. In response, we modified the rating system for individual items to include a broader range of response options (yes/probably yes/probably no/no) and removed the "can't say" option. The renamed response option "not applicable" was included only for items where a response would not be pertinent.

The overall rating options for each section are: good (+), where almost all criteria are met, there is little or no concern, and low RoB; neutral (0), where most criteria are met but there are some flaws with an associated concern, and there is a moderate RoB; and poor (−), where either most or all criteria are not met, there are significant flaws, and there is a high RoB. The overall ratings for each section are carried forward to the Overall Study Rating section and used when considering the final assessment of the study (response options: poor, neutral, or good). Guidance for determining overall ratings, including the consideration of the relative importance of specific items within the context of the study design, was edited to reflect the updated format and response options.

Based on qualitative feedback, the raters felt that NUQUEST was easy to apply and feasible to complete within a reasonable time by raters with general knowledge of epidemiology or nutrition science (average: 16.5 min/study, range: 10–33 min). NUQUEST for cohort studies generally took more time for raters to complete than the others, but time to completion improved with each completed study assessment. To aid the future use of NUQUEST, we identified points to consider when developing a review (**Box 1**) and when training raters to use the worksheet (**Box 2**).

The raters indicated that they felt more comfortable rating items that were within their area of expertise than rating items that were not. Specifically, raters with expertise in epidemiology/study design methodology felt confident to assign stricter ratings to the generic items, whereas raters with expertise in nutrition felt confident to assign stricter ratings for the nutrition-specific items. In addition, raters were more likely to use the "no information" option for items outside of their expertise. Often other information in the article could have permitted the rater to assess whether something likely occurred or not. With the addition of a completed worksheet in the second validation round (completed example in **Supplemental File 7**), the raters indicated that they were more confident in applying a rating for items for which they had less expertise and overall more likely to be stricter when applying ratings to all items. This resulted in a tendency to downgrade individual items and overall study ratings. The addition of the worksheet also decreased the use of the "no information" rating; however, raters still tended to use this option in areas for which they had less expertise. As such, the working group decided to remove this option in the final checklists to avoid uninformative answers. Users are now forced to make a judgment on each item, even in cases where details are reported poorly, while taking into account stated accounts of study design, conduct, and analysis and items that the rater judges are likely to have (or have not) occurred in the study.

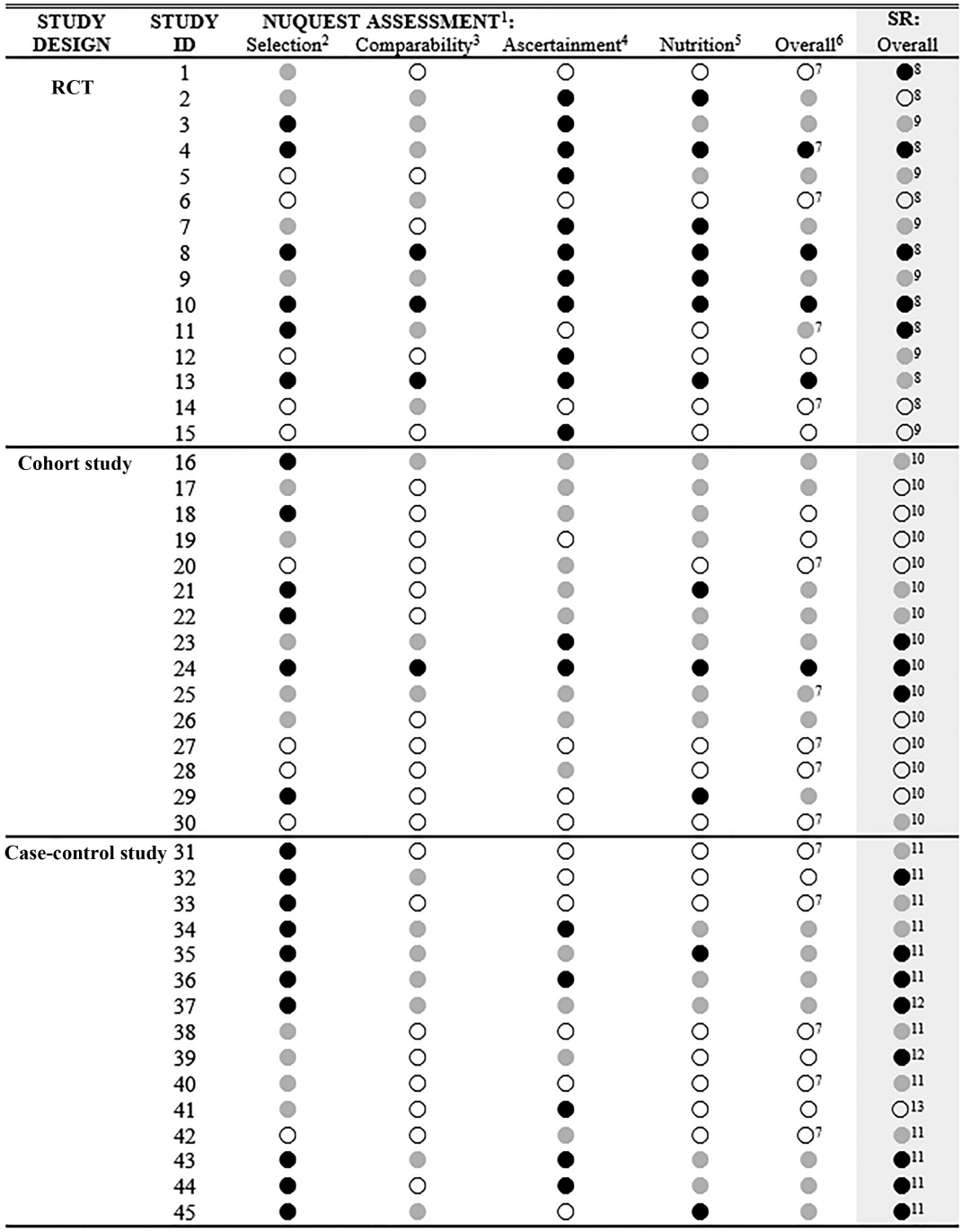


FIGURE 2 Nutrition QUality Evaluation Strengthening Tools (NUQUEST) ratings for individual sections and overall and the overall rating from the original SRs. [1]NUQUEST assessments are by section and overall by study: ● *good*, where almost all criteria are met, little or no concern, and low RoB; ● *neutral*, where most criteria are met, there are some flaws, and moderate RoB; and ○ *poor*, where either most or all criteria are not met, there are significant flaws, and high RoB. [2]Selection = selection of participants (RCT)/selection of cohorts (cohort)/creation of study groups (case-control). [3]Comparability = comparability of study groups (RCT, case-control)/comparability of cohorts (cohort). [4]Ascertainment = ascertainment of outcomes (RCT, cohort)/exposure ascertainment (case-control). [5]Nutrition = nutrition-specific (RCT, cohort, case-control). [6]Overall = overall NUQUEST rating. [7]Assessed using the NUQUEST worksheet. [8]Evaluated by the Cochrane RoB Tool modified for nutrition studies. [9]Evaluated by the Cochrane RoB Tool. [10]Evaluated by the NOS modified for nutrition cohort studies. [11]Evaluated by the NOS. [12]Evaluated by the JBI critical appraisal tool for case-control studies. [13]Evaluated by the JBI critical appraisal tool for cross-sectional studies. Because the original SRs used a variety of tools to assess RoB, the original scores were recoded to "poor, neutral, and good" to allow for comparison to NUQUEST scores. Recoding for the Cochrane RoB Tool or the NOS modified for nutrition cohort studies was as follows: low RoB, it was assigned good ●; moderate RoB, it was assigned neutral ●; and high RoB, it was assigned poor ○. Recoding for the NOS was as follows: 1–3, it was assigned ○; 4–6, it was assigned ●; and 7–9, it was assigned ●. Recoding for the JBI cross-sectional tool was as follows: 1–4, it was assigned ○; and 5–8, it was assigned ●. Recoding for the JBI case-control tool was as follows: 1–3, it was assigned ○; 4–7, it was assigned ●; and 8–10, it was assigned ●. Where the authors of the original SR did not make an overall RoB judgment, section-specific judgments and related text within the report were examined to assign an overall RoB assessment by 2 members of the working group. ID, identification; JBI, Joanna Briggs Institute; NOS, Newcastle-Ottawa Scale; RCT, randomized controlled trial; RoB, risk of bias; SR, systematic review.

BOX 1
If Initiating a Review, Points to Consider When Developing the Worksheet[1]

- Is there a clearly stated key research question for the review?
  - In cases where the study's purpose differs from the key research question, does the study design provide information that can still be used to evaluate the key research question?
- Has a PICO been prepared before the identification of articles to assist in study selection/exclusion decisions?
  - Are the selection/exclusion criteria sufficiently broad to include a range of acceptable study methodologies that can subsequently be subjected to assessment with application of the NUQUEST criteria?
- Have the rating criteria to assist reviewers in using NUQUEST been prepared before initiation of the reviews?
  - Does the worksheet include criteria for evaluating a range of exposure methodologies that are specific for the food substance and for the analysis and expression of results?
  - Does the worksheet differentiate between key confounders and other potential confounders of lesser importance?

[1]NUQUEST, NUtrition QUality Evaluation Strengthening Tools; PICO, population, intervention/exposure, comparator, outcome.

Finalize NUQUEST

NUQUEST for RCTs, cohort studies, and case-control studies, and guidance for each instrument are available in the Supplementary data, where we have also included an example of a completed worksheet (Supplemental File 7). The wording of the guidance is critical for the application of NUQUEST and refinement is expected to continue with future use.

Discussion

With an emphasis on using the best available evidence to inform public health and nutrition initiatives (52), several evidence appraisal instruments have been developed; however, current instruments do not adequately address methodological challenges inherent to nutrition studies. These challenges contribute to uncertainty and lower our ability to establish nutrition–health outcome associations with confidence. Here we present a series of developed and tested RoB tools that retain generic study appraisal components but with the addition of nutrition-specific appraisal items to address challenges common to nutrition studies. NUQUEST presents an integrated approach for combining generic and nutrition-specific issues in a single tool for assessing the RoB of individual nutrition studies. Other complementary tools include guidelines for designing and conducting nutrition studies (46), guidelines for reporting nutrition studies (2, 34, 35), procedures for conducting nutrition SRs and meta-analyses (53), and guidelines for grading the evidence (32, 54–57). The rigorous evidentiary evaluation and transparency resulting from the use of RoB tools for nutritional epidemiology can identify problematic areas for which more research or creative statistical approaches for dealing with limitations of dietary exposure data are needed (17).

BOX 2
For the Initiators of the Review, Points to Consider When Preparing Raters to Use the Worksheet in Study Assessments[1]

- Do raters understand the worksheet's purpose for the review, the PICO, and the rating criteria?
- If possible, do raters represent multidisciplinary expertise (e.g., epidemiology, nutrition)?
- If there is >1 rater, are differences adjudicated?
- Do raters recognize the types of situations that may require judgment?

For example:

- If detailed study design and methods information is primarily available in companion articles, have these articles been accessed?
- Where double blinding may not be possible (e.g., 1 of the groups receives an educational program or a whole food substitution), have the study investigators taken steps to reduce the potential biases (e.g., detection bias)?
- If randomization failed to some degree, have the study investigators taken steps to reduce potential biases by treating important variables as covariates in analyses?
- If comparability of groups is based on a broad range of potential confounders or limited to commonly used demographic factors, how well have study investigators dealt with the key critical confounders?
- If intention-to-treat analysis was not specifically mentioned, was there other information (e.g., analysis descriptions) to suggest that it was or was not the approach used?
- When evaluating nutrient exposures, have the complexities and nuances of exposure assessments for different nutrients, different purposes, different study designs, and different statistical analyses been taken into account?
- If nutrient exposure is an intermediate factor (e.g., when education is the intervention), is it still possible to evaluate the exposure–outcome relation of interest?
- Has the rater maintained focus on the evaluation of the scientific quality of reviewed studies and not been distracted by the potentially trivial nature of a study topic? Trivial or not, does the study provide information that can be used to evaluate the key research question?

[1]In many cases, ≥2 raters will perform the assessment. However, we recognize that in some cases, a single rater will conduct the review. In this latter case, adaptation of the "Points to Consider" may be appropriate. PICO, population, intervention/exposure, comparator, outcome.

We built on the existing properties of the SIGN instruments, which are valid, reliable, and have extensive use experience (26). The SIGN instruments have a long history of application and have performed well identifying studies of different quality over the years to the satisfaction of authors and users. The SIGN instruments provided a complementary suite of tools for evaluating 3 different study designs commonly used in nutrition studies on which a nutrition-specific instrument could be developed. We enhanced the clarity of the SIGN items by reorganizing them into sections focused on selection, comparability, and ascertainment, and supplemented the items by adding nutrition-related guidance. Nutrition-specific RoB items were added to recognize the nature and impact of the challenges when considering nutrition exposures. We offer suggestions in Boxes 1 and 2 and in the guidance and a worksheet template to elucidate topic-specific issues important for study assessment (see Supplemental File 6). We provide an example of a completed worksheet (Supplemental File 7) to illustrate its use in the context of a specific research question. We also include reminders in the guidance to consider both general and topic-specific criteria when rating studies. NUQUEST focuses on the extent to which nutrition studies were designed, conducted, analyzed, and reported to the highest possible methodological standards, but is not intended to quantify the magnitude of bias that may be attributable to methodological flaws.

Our nutrition-specific RoB tools can be used for assessing single studies or multiple studies included in an SR, the latter being similar to the adaptation of ROBINS-I for evaluation of nonrandomized studies of exposures (58). Realizing that some existing assessment instruments are challenging to use (59, 60), we aimed to make NUQUEST user-friendly for both epidemiology and nutrition users via an improved rating system, improved organization, and detailed nutrition-specific guidance. Our multidisciplinary raters felt that NUQUEST was feasible: assessments could be completed in a reasonable amount of time, time for completion improved with repeated use, and the inclusion of a completed worksheet facilitated study evaluation. Importantly, we wanted NUQUEST to be educational for users and to promote improvement in nutrition study design, conduct, and reporting.

Although SRs and meta-analyses are important for informing evidence-based nutrition policy, some nutrition scientists have expressed concern that there is potential for misuse if they are poorly conducted (61). Indeed, a recent SR and meta-analysis revealed that the vast majority of a large sample of nutrition SRs had serious methodological issues, including no RoB assessment of included studies and a lack of formal evaluation of the certainty of evidence (62). Given that most nutrition evidence comes from observational studies where exposure is based on self-reported intakes with systematic biases and random errors that can produce inaccurate and misleading results (4), it is imperative that nutrition-focused SRs are rigorous and that RoB is assessed in a consistent and transparent fashion. Our results showed that the effect of nutrition-specific criteria on overall ratings varied, with some ratings increasing and others decreasing when comparing NUQUEST to other instruments. The use of NUQUEST has several advantages for supporting SR development and nutrition guidance decisions, including clarity of the key research question, flexibility and transparency in developing the PICO and worksheet, and transparency of the strengths and weaknesses of the relied-upon evidence. In addition, policy and guidance decisions are not solely dependent on the strength of the evidence but also consider other factors (e.g., public health impact, economics) (52). We believe that NUQUEST, with its emphasis on assessment of exposure data, can support future nutrition guidance initiatives.

A common challenge in assessing the RoB in nutrition studies, particularly for observational studies, is the failure of study authors to adequately describe the procedure used to establish the accuracy and appropriateness of their dietary exposure assessment. Researchers completing the NUQUEST worksheet will be responsible for establishing rating criteria for assessing the accuracy, reliability, and appropriateness of dietary intake in selected studies, as well as the appropriateness of the dietary methodology for meeting research objectives. This need can be partially addressed if journal editors require that authors follow reporting guidelines for dietary studies (2, 34, 35). Resources exist to aid in these decisions, such as the National Cancer Institute's Dietary Assessment Primer (17), which provides a comprehensive description of the available dietary assessment instruments, the key concerns in dietary assessment (i.e., measurement error and validation), and recommendations for choosing a dietary assessment approach for different research objectives. In addition, the evolving use of statistical procedures to improve the ability to estimate "true" intakes in nutritional epidemiology research (47) can also provide useful information for establishing NUQUEST worksheet criteria. The failure to document validation procedures or the use of validation procedures generally considered to increase the potential for RoB (e.g., repeatability properties of an exposure methodology) would likely result in lower nutrition-specific ratings; documentation of validation against an external "gold standard" (e.g., use of a recovery biomarker or controlled feeding trials), particularly when also combined with statistical adjustments for other dietary components and personal characteristics, would warrant higher ratings.

Although NUQUEST strengthens the ability to assess nutrition studies, these instruments have limitations. As is common practice, we recommend training on a set of studies to test topic-specific guidance and use of the worksheet, and discussing consensus on items among raters. The working group did not address issues of outcome specificity. However, in practice, any tool can be made outcome-specific, as can be done with the use of specific criteria in the outcome section of the worksheet, keeping in mind potential biases related to the ascertainment of each outcome of interest (63). Ultimately, even with the guidance and assessment criteria defined, user judgment will be required. We found that the inclusion of a worksheet improved interrater reliability and increased discussion among raters to achieve consensus. These limitations underscore the need for an interdisciplinary team, including experts with knowledge of study rating systems, study design, and nutrition. The effective application of NUQUEST will require the team to develop a priori rating criteria and estimate their relative impact on study RoB. Because there is no current gold-standard approach for the assessment of RoB in nutrition studies, the working group could not assess the criterion validity of NUQUEST. Instead, the exercise here represents an initial assessment of content, construct, and concurrent validity. Our results suggest NUQUEST performed well in comparison to other RoB tools

and that, where there was disagreement, the nutrition-specific items influenced the overall rating. Finally, the absence of a SIGN checklist for cross-sectional studies precluded the development of a NUQUEST for cross-sectional studies, at least for the moment. We recognize the need for such a tool given the plethora of cross-sectional nutrition studies. As such, NUQUEST for cross-sectional studies is being developed de novo and will be reported separately.

In conclusion, we developed a set of validated nutrition-specific RoB tools to assess nutrition studies of commonly used study designs. NUQUEST is grounded in the scientific concepts reported in a broadly used generic assessment instrument, while improving the usability for application to nutrition studies. These nutrition-specific RoB tools for RCTs, cohort studies, and case-control studies are intended for the evaluation of individual nutrition studies or studies included in SRs. In addition, NUQUEST will help researchers improve the design, conduct, and reporting of future nutrition research studies through its focus on nutrition-specific appraisal items. Finally, the availability of these new tools will foster consistency among future reviews of nutrition studies while enabling future systematic refinement, as warranted, by experience and as science evolves.

We gratefully acknowledge the SIGN for their permission to adapt the SIGN checklists in the development of the NUQUEST checklists. We thank Joan Peterson, Dr. Zemin Bai, Amy Johnston, Dr. Shu-Ching Hsieh, Dr. Matthew Parrot, Alexandre Martel, Dr. Jesse Elliott, and Cynthia Boudrault for their contributions as raters for the validation of the QAIs. We greatly appreciate Drs. Stephanie Atkinson, Alice Lichtenstein, and Maria Makrides for their time and insight on the nutrition-specific items and guidance, and/or individual study RoB.

The authors' responsibilities were as follows—AJM, EAY, GAW, and LSG-F: conceptualized the project; SEK: drafted the manuscript; AJM and GAW: had primary responsibility for the final content; and all authors: developed the research plan, conducted the research, analyzed the data, and contributed to the final manuscript. All authors read and approved the final manuscript. The authors report no conflicts of interest.

Data Availability

Data described in the article are publicly and freely available without restriction (available from the "Supplementary data" link in the online posting of the article and from the same link in the online table of contents at https://academic.oup.com/ajcn/).

References

1. Yetley EA, MacFarlane AJ, Greene-Finestone LS, Garza C, Ard JD, Atkinson SA, Bier DM, Carriquiry AL, Harlan WR, Hattis D, et al. Options for basing Dietary Reference Intakes (DRIs) on chronic disease endpoints: report from a joint US-/Canadian-sponsored working group. Am J Clin Nutr 2017;105(1):249S–85S.
2. Chung M, Balk EM, Ip S, Raman G, Yu WW, Trikalinos TA, Lichtenstein AH, Yetley EA, Lau J. Reporting of systematic reviews of micronutrients and health: a critical appraisal. Am J Clin Nutr 2009;89(4):1099–113.
3. Lichtenstein AH, Yetley EA, Lau J. Application of systematic review methodology to the field of nutrition. J Nutr 2008;138(12):2297–306.
4. National Academies of Sciences, Engineering, and Medicine. Guiding principles for developing Dietary Reference Intakes based on chronic disease. Washington (DC): The National Academies Press; 2017.
5. National Academies of Sciences, Engineering, and Medicine. Dietary Reference Intakes for sodium and potassium. Washington (DC): The National Academies Press; 2019.
6. Trepanowski JF, Ioannidis JPA. Perspective: limiting dependence on nonrandomized studies and improving randomized trials in human nutrition research: why and how. Adv Nutr 2018;9(4):367–77.
7. Zeraatkar D, Johnston BC, Guyatt G. Evidence collection and evaluation for the development of dietary guidelines and public policy on nutrition. Annu Rev Nutr 2019;39:227–47.
8. Suh JR, Herbig AK, Stover PJ. New perspectives on folate catabolism. Annu Rev Nutr 2001;21:255–82.
9. Academy of Nutrition and Dietetics. Evidence analysis manual: steps in the Academy Evidence Analysis Process. Chicago (IL): Academy of Nutrition and Dietetics; 2016.
10. Newberry SJ, Chung M, Anderson CAM, Chen C, Fu Z, Tang A, Zhao N, Booth M, Marks J, Hollands S, et al. Sodium and potassium intake: effects on chronic disease outcomes and risks. Comparative Effectiveness Review, No. 206. Rockville (MD): Agency for Healthcare Research and Quality (US); 2018.
11. Guyatt G, Busse J. Methods commentary: risk of bias in randomized trials 1. [Internet]. Ottawa (Canada): Evidence Partners; 2021. [Accessed 2021 Oct 8]. Available from: https://www.evidencepartners.com/resources/methodological-resources/risk-of-bias-commentary.
12. Hartling L, Ospina M, Liang Y, Dryden DM, Hooton N, Krebs Seida J, Klassen TP. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. BMJ 2009;339(7728):b4012.
13. Handel MN, Rohde JF, Jacobsen R, Nielsen SM, Christensen R, Alexander DD, Frederiksen P, Heitmann BL. Processed meat intake and incidence of colorectal cancer: a systematic review and meta-analysis of prospective observational studies. Eur J Clin Nutr 2020;74(8):1132–48.
14. Balk EM, Adams GP, Langberg V, Halladay C, Chung M, Lin L, Robertson S, Yip A, Steele D, Smith BT, et al. Omega-3 fatty acids and cardiovascular disease: an updated systematic review. Evid Rep Technol Assess (Full Rep) 2016;(223):1–1252.
15. Newberry SJ, Chung M, Shekelle PG, Booth MS, Liu JL, Maher AR, Motala A, Cui M, Perry T, Shanman R, et al. Vitamin D and calcium: a systematic review of health outcomes (update). Evid Rep Technol Assess (Full Rep) 2014;(217):1–929.
16. Newberry SJ, Chung M, Booth M, Maglione MA, Tang AM, O'Hanlon CE, Wang DD, Okunogbe A, Huang C, Motala A, et al. Omega-3 fatty acids and maternal and child health: an updated systematic review. Evid Rep Technol Assess (Full Rep) 2016;(224):1–826.
17. Thompson FE, Kirkpatrick SI, Subar AF, Reedy J, Schap TE, Wilson MM, Krebs-Smith SM. The National Cancer Institute's Dietary Assessment Primer: a resource for diet research. J Acad Nutr Diet 2015;115(12):1986–95.
18. Del Valle HB, Yaktine AL, Taylor CL, Ross AC. Dietary Reference Intakes for calcium and vitamin D. Washington (DC): National Academies Press; 2011.
19. European Food Safety Authority (EFSA). Outcome of public consultations on the Scientific Opinions of the EFSA Panel on Nutrition, Novel Foods and Food Allergens (NDA) on Dietary Reference Values for sodium and chloride. EFSA Support Publ 2019;16(9):1679E.
20. National Heart, Lung, and Blood Institute (NHLBI). Lifestyle interventions to reduce cardiovascular risk: systematic evidence review from the Lifestyle Work Group, 2013. Bethesda (MD): NHLBI; 2015.
21. World Health Organization (WHO). Evidence-informed guideline development for World Health Organization nutrition-related normative work: continuous quality improvement for efficiency and impact. Geneva (Switzerland): WHO; 2018.
22. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960;20(1):37–46.
23. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull 1968;70(4):213–20.
24. Garritty C, Gartlehner G, Nussbaumer-Streit B, King VJ, Hamel C, Kamel C, Affengruber L, Stevens A. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. J Clin Epidemiol 2021;130:13–22.
25. Tricco AC, Antony J, Zarin W, Strifler L, Ghassemi M, Ivory J, Perrier L, Hutton B, Moher D, Straus SE. A scoping review of rapid review methods. BMC Med 2015;13(1):224.
26. Bai A, Shukla VK, Bak G, Wells G. Quality Assessment Tools project report. Ottawa (Canada): Canadian Agency for Drugs and Technologies in Health; 2012.

27. Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar VS, Grimmer KA. A systematic review of the content of critical appraisal tools. BMC Med Res Method 2004;4(1):22.
28. Lohr KN. Rating the strength of scientific evidence: relevance for quality improvement programs. Int J Qual Health Care 2004;16(1):9–18.
29. Sanderson S, Tatt ID, Higgins J. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. Int J Epidemiol 2007;36(3):666–76.
30. Whiting P, Davies PA, Savović S, Caldwell DM, Churchill R. Evidence to inform the development of ROBIS, a new tool to assess the risk of bias in systematic reviews. Bristol (UK): School of Social and Community Medicine, University of Bristol; 2013.
31. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, Petticrew M, Altman DG. Evaluating non-randomised intervention studies. Health Technol Assess 2003;7(27):iii–x, 1-173.
32. Schwingshackl L, Knuppel S, Schwedhelm C, Hoffmann G, Missbach B, Stelmach-Mardas M, Dietrich S, Eichelmann F, Kontopantelis E, Iqbal K, et al. Perspective: NutriGrade: a scoring system to assess and judge the meta-evidence of randomized controlled trials and cohort studies in nutrition research. Adv Nutr 2016;7(6):994–1004.
33. Yang C, Pinart M, Kolsteren P, Van Camp J, De Cock N, Nimptsch K, Pischon T, Laird E, Perozzi G, Canali R, et al. Perspective: essential study quality descriptors for data from nutritional epidemiologic research. Adv Nutr 2017;8(5):639–51.
34. Hörnell A, Berg C, Forsum E, Larsson C, Sonestedt E, Åkesson A, Lachat C, Hawwash D, Kolsteren P, Byrnes G, et al. Perspective: an extension of the STROBE statement for observational studies in nutritional epidemiology (STROBE-nut): explanation and elaboration. Adv Nutr 2017;8(5):652–78.
35. Lachat C, Hawwash D, Ocké MC, Berg C, Forsum E, Hörnell A, Larsson C, Sonestedt E, Wirfält E, Åkesson A, et al. Strengthening the Reporting of Observational Studies in Epidemiology—Nutritional Epidemiology (STROBE-nut): an extension of the STROBE statement. PLoS Med 2016;13(6):e1002036.
36. Farrah K, Young K, Tunis MC, Zhao L. Risk of bias tools in systematic reviews of health interventions: an analysis of PROSPERO-registered protocols. Syst Rev 2019;8(1):280.
37. Charlton P, Doucet S, Azar R, Nagel DA, Boulos L, Luke A, Mears K, Kelly KJ, Montelpare WJ. The use of the environmental scan in health services delivery research: a scoping review protocol. BMJ Open 2019;9(9):e029805.
38. West S, King V, Carey TS, Lohr KN, McKoy N, Sutton SF, Lux L. Systems to rate the strength of scientific evidence. Evid Rep Technol Assess (Summ) 2002;(47):1–11.
39. Scottish Intercollegiate Guidelines Network (SIGN). SIGN 50 methodology. Critical appraisal: checklists. Checklist 2: randomised controlled trials, methodology. Checklist 3: cohort studies, methodology. Checklist 4: case-control studies. Edinburgh (United Kingdom): SIGN; 2014 [Internet]. Available from: https://www.sign.ac.uk/what-we-do/methodology/checklists/.
40. Scottish Intercollegiate Guidelines Network (SIGN). SIGN 50 methodology. Critical appraisal: notes. Edinburgh (United Kingdom): SIGN; 2014 [Internet]. Available from: https://www.sign.ac.uk/what-we-do/methodology/checklists/.
41. Joanna Briggs Institute (JBI). Critical appraisal checklist for randomized controlled trials. Adelaide (Australia): JBI; 2020.
42. Joanna Briggs Institute (JBI). Critical appraisal checklist for cohort studies. Adelaide (Australia): JBI; 2020.
43. Joanna Briggs Institute (JBI). Critical appraisal checklist for case control studies. Adelaide (Australia): JBI; 2020.
44. Freedman LS, Commins JM, Moler JE, Arab L, Baer DJ, Kipnis V, Midthune D, Moshfegh AJ, Neuhouser ML, Prentice RL, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. Am J Epidemiol 2014;180(2):172–88.
45. Freedman LS, Commins JM, Moler JE, Willett W, Tinker LF, Subar AF, Spiegelman D, Rhodes D, Potischman N, Neuhouser ML, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for potassium and sodium intake. Am J Epidemiol 2015;181(7):473–87.
46. Lichtenstein AH, Petersen K, Barger K, Hansen KE, Anderson CAM, Baer DJ, Lampe JW, Rasmussen H, Matthan NR. Perspective: design and conduct of human nutrition randomized controlled trials. Adv Nutr 2021;12(1):4–20.
47. Zheng C, Beresford SA, Van Horn L, Tinker LF, Thomson CA, Neuhouser ML, Di C, Manson JE, Mossavar-Rahmani Y, Seguin R, et al. Simultaneous association of total energy consumption and activity-related energy expenditure with risks of cardiovascular disease, cancer, and diabetes among postmenopausal women. Am J Epidemiol 2014;180(5):526–35.
48. NIH National Cancer Institute (NCI). Dietary assessment primer. [2021 Internet]. Bethesda (MD): NCI. [Accessed 2021 Oct 28]. Available from: https://dietassessmentprimer.cancer.gov/.
49. Houghton LA, Gray AR, Rose MC, Miller JC, Hurthouse NA, Gregory JF 3rd. Long-term effect of low-dose folic acid intake: potential effect of mandatory fortification on the prevention of neural tube defects. Am J Clin Nutr 2011;94(1):136–41.
50. Nguyen P, Tam C, O'Connor DL, Kapur B, Koren G. Steady state folate concentrations achieved with 5 compared with 1.1 mg folic acid supplementation among women of childbearing age. Am J Clin Nutr 2009;89(3):844–52.
51. Tighe P, Ward M, McNulty H, Finnegan O, Dunne A, Strain J, Molloy AM, Duffy M, Pentieva K, Scott JM. A dose-finding trial of the effect of long-term folic acid intervention: implications for food fortification policy. Am J Clin Nutr 2011;93(1):11–8.
52. Johnston BC, Seivenpiper JL, Vernooij RWM, de Souza RJ, Jenkins DJA, Zeraatkar D, Bier DM, Guyatt GH. The philosophy of evidence-based principles and practice in nutrition. Mayo Clin Proc Innov Qual Outcomes 2019;3(2):189–99.
53. Cumpston M, Li T, Page MJ, Chandler J, Welch VA, Higgins JPT, Thomas J. Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Database Syst Rev 2019;10:ED000142.
54. Johnston BC, Alonso-Coello P, Bala MM, Zeraatkar D, Rabassa M, Valli C, Marshall C, El Dib R, Vernooij RWM, Vandvik PO, et al. Methods for trustworthy nutritional recommendations NutriRECS (Nutritional Recommendations and accessible Evidence summaries Composed of Systematic reviews): a protocol. BMC Med Res Method 2018;18(1):162.
55. EFSA Scientific Committee, Hardy A, Benford D, Halldorsson T, Jeger MJ, Knutsen HK, More S, Naegeli H, Noteborn H, Ockleford C, et al. Guidance on the use of the weight of evidence approach in scientific assessments. EFSA J 2017;15(8):e04971.
56. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336(7650):924–6.
57. Katz DL, Karlsen MC, Chung M, Shams-White MM, Green LW, Fielding J, Saito A, Willett W. Hierarchies of evidence applied to lifestyle medicine (HEALM): introduction of a strength-of-evidence approach based on a methodological systematic review. BMC Med Res Method 2019;19(1):178.
58. Morgan RL, Thayer KA, Santesso N, Holloway AC, Blain R, Eftim SE, Goldstone AE, Ross P, Ansari M, Akl EA, et al. A risk of bias instrument for non-randomized studies of exposures: a users' guide to its application in the context of GRADE. Environ Int 2019;122:168–84.
59. Minozzi S, Cinquini M, Gianola S, Castellini G, Gerardi C, Banzi R. Risk of bias in nonrandomized studies of interventions showed low inter-rater reliability and challenges in its application. J Clin Epidemiol 2019;112:28–35.
60. Thomson H, Craig P, Hilton-Boon M, Campbell M, Katikireddi SV. Applying the ROBINS-I tool to natural experiments: an example from public health. Syst Rev 2018;7(1):15.
61. Tobias DK, Wittenbecher C, Hu FB. Grading nutrition evidence: where to go from here? Am J Clin Nutr 2021;113(6):1385–7.
62. Zeraatkar D, Bhasin A, Morassut RE, Churchill I, Gupta A, Lawson DO, Miroshnychenko A, Sirotich E, Aryal K, Mikhail D, et al. Characteristics and quality of systematic reviews and meta-analyses of observational nutritional epidemiology: a cross-sectional study. Am J Clin Nutr 2021;113(6):1578–92.
63. Viswanathan M, Patnode CD, Berkman ND, Bass EB, Chang S, Hartling L, Murad MH, Treadwell JR, Kane RL. Recommendations for assessing the risk of bias in systematic reviews of health-care interventions. J Clin Epidemiol 2018;97:26–34.

APPENDIX 3: Supplements to Chapter 4

Permissions to re-print the published journal articles

Received via email on 29 October, 2024

Dear Shannon Kelly,

Thank you for contacting us.

RE: Risk of bias in cross-sectional studies: Protocol for a scoping review of concepts and tools, MethodsX, Volume 12, June 2024, 102610.

Proposed use: Thesis.

As an Elsevier journal author, you retain the right to Include the article in a thesis or dissertation (provided that this is not to be published commercially) whether in full or in part, subject to proper acknowledgment; see https://www.elsevier.com/about/policies/copyright for more information. As this is a retained right, no written permission from Elsevier is necessary.\

As outlined in our permissions licenses, this extends to the posting to your university's digital repository of the thesis provided that if you include the published journal article (PJA) version, it is embedded in your thesis only and not separately downloadable.

Thank you.

Kind regards,

Thomas Rexson Yesudoss
Copyrights Specialist
ELSEVIER | HCM - Health Content Management

Visit Elsevier Permissions

Received via email on 04 November, 2024

Dear Shannon Kelly,

Thank you for contacting us.

RE: Kelly, Shannon E. et al. Journal of Clinical Epidemiology, Volume 172, 111408.

Proposed use: Thesis.

As an Elsevier journal author, you retain the right to Include the article in a thesis or dissertation (provided that this is not to be published commercially) whether in full or in part, subject to proper acknowledgment; see https://www.elsevier.com/about/policies/copyright for more information. As this is a retained right, no written permission from Elsevier is necessary.

As outlined in our permissions licenses, this extends to the posting to your university's digital repository of the thesis provided that if you include the published journal article (PJA) version, it is embedded in your thesis only and not separately downloadable.

Thank you.

Kind regards,

Thomas Rexson Yesudoss
Copyrights Specialist
ELSEVIER | HCM - Health Content Management

Re-print of the published journal article (Protocol for a scoping review)

MethodsX 12 (2024) 102610

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/methodsx

Risk of bias in cross-sectional studies: Protocol for a scoping review of concepts and tools

Shannon E. Kelly[a,b,*], Karima Benkhedda[c], Stephen P.J. Brooks[c,d], Amanda J. MacFarlane[d,e], Linda S. Greene-Finestone[f], Becky Skidmore[g], Tammy J. Clifford[a], George A. Wells[a,b]

[a] School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada
[b] Cardiovascular Research Methods Centre, University of Ottawa Heart Institute, Ottawa, Ontario, Canada
[c] Bureau of Nutritional Sciences, Health Canada, Ottawa, Ontario, Canada
[d] Department of Biology, Carleton University, Ottawa, Ontario, Canada
[e] Agrilife Evidence Center at Texas A&M University, College Station, Fort Worth, Texas, USA
[f] Applied Research Division, Public Health Agency of Canada, Ottawa, Ontario, Canada
[g] Independent Information Scientist, Ottawa, Ontario, Canada

ARTICLE INFO

Method name:
Scoping review protocol

Keywords:
Risk of bias
Cross-sectional
Scoping review
Tool
Methodological quality
Critical appraisal

ABSTRACT

Cross-sectional studies are commonly used to study human health and disease, but are especially susceptible to bias. This scoping review aims to identify and describe available tools to assess the risk of bias (RoB) in cross-sectional studies and to compile the key bias concepts relevant to cross-sectional studies into an item bank. Using the JBI scoping review methodology, the strategy to locate relevant RoB concepts and tools is a combination of database searches, prospective review of PROSPERO registry records; and consultation with knowledge users and content experts. English language records will be included if they describe tools, checklists, or instruments which describe or permit assessment of RoB for cross-sectional studies. Systematic reviews will be included if they consider eligible RoB tools or use RoB tools for RoB of cross-sectional studies. All records will be independently screened, selected, and extracted by one researcher and checked by a second. An analytic framework will be used to structure the extraction of data. Results for the scoping review are pending. Results from this scoping review will be used to inform future selection of RoB tools and to consider whether development of a new RoB tool for cross-sectional studies is needed.

* Corresponding author at: School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada.
E-mail address: skell102@uottawa.ca (S.E. Kelly).

https://doi.org/10.1016/j.mex.2024.102610
Received 9 January 2024; Accepted 9 February 2024

Specifications table

Subject area:	Research Methods
More specific subject area:	Epidemiology, knowledge synthesis methods, study designs, risk of bias
Name of your protocol:	Risk of bias in cross-sectional studies: protocol for a scoping review of concepts and tools
Reagents/tools:	Not applicable
Experimental design:	Not applicable. This protocol describes a scoping review.
Trial registration:	Not applicable. This is not a trial. This scoping review is registered with the Open Science Framework (https://osf.io/y7vst/)
Ethics:	This work did not involve human subjects, animal experiments or data collected from social media platforms.
Value of the Protocol:	Scientific contribution: • Comprehensive summary of different methodological considerations for risk of bias in cross-sectional study designs. • Maps coverage for key bias considerations in existing risk of bias tools. • Identified gaps may be used to select tools for use and/or to inform future risk of bias tool development.

Description of protocol

A cross-sectional study is a type of observational research study that is used to examine a group of individuals at a specific point in time in order to understand their characteristics and behaviors. The objective of this scoping review is to identify tools used to evaluate the risk of bias (RoB) in cross-sectional studies of humans; and to identify, describe, and to compile the key items, domains or concepts relevant to RoB in cross-sectional studies into an item bank.

Methods

The proposed scoping review will be conducted in accordance with the JBI methodology for scoping reviews and reported in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) [1,2].

Review questions

1. What tools are used to assess the risk of bias in cross-sectional studies?
2. What risk of bias items, domains or concepts are included in tools intended for cross-sectional study designs?
3. How do risk of bias concepts vary when tools are intended to be applied to analytic studies of association or prevalence?

Inclusion criteria

The Cochrane definition will be used to consider RoB as a measure of how features of the design, conduct or analysis of a study may cause systematic error in the results of a study; That is, whether the prevalence or associated effects for an outcome in a cross-sectional study may be over- or under-estimated. Since tools may be described inconsistently using language on methodological quality, quality tools will be eligible if they are judged by reviewers to contain one or more items relating to internal validity of the study.

This review will consider records for inclusion if they meet the following criteria:

- Permit the assessment of RoB in human observational studies defined as cross-sectional studies, including analytic or prevalent/survey approaches;
- Describe or permit global assessment of RoB for multiple study designs and originally developed for application to cross-sectional study designs; or,
- Describe the application of a tool not originally designed for application to cross-sectional study designs which has been adapted or modified for application to cross-sectional study designs for the assessment of RoB; and,
- Published or available in English.

“Tools” may be described as checklists, scored instruments, ‘question/item-based‘ or ’domain-based’ according to groupings or categories of bias. Systematic reviews of RoB tools will also be included if they endeavoured to locate and/or include relevant tools/items/concepts for cross-sectional studies.

Records identified will be excluded if they 1) report tools intended for other study designs and are not intended for evaluation of cross-sectional study designs; or, 2) report single-study or multi-study tools modified for application to cross-sectional studies in a systematic review and the modified tool is not reported or cited.

Types of sources

This scoping review will consider records of any study design for inclusion. As there is a limited number of RoB tools published in peer-reviewed literature, four approaches will be used to search for and locate eligible RoB tools relevant to cross-sectional studies:

1. **Bibliographic database search:** This approach will be used to identify published RoB tools and relevant systematic reviews of RoB tools. For the search of bibliographic databases, an experienced medical information specialist developed and tested search strategies through an iterative process in consultation with the review team.

2. **Prospective scan of PROSPERO registrations:** This approach will be used to identify RoB tools identified in 1000 prospectively registered systematic review protocols in the PROSPERO database. In order to identify currently used tools, all records will be searched for proposals including the term "cross-sectional" and systematic reviews indicating the proposed inclusion of cross-sectional studies in the analysis will be retained. Applicable RoB tools will be identified from the details in the "Risk of bias (quality) assessment" section of the registration.
3. **Outreach to key knowledge users:** This approach will be used to gather information regarding RoB tools being applied in practice for cross-sectional studies. A group of users in the field of nutrition will be selected as they commonly use or assess cross-sectional research. The email-based outreach will briefly ask participants to reply if they are willing to share a citation or the name of a relevant RoB tool they use or are aware of.
4. **Experiential knowledge of content experts:** This approach will supplement the other approaches by collecting the experiential knowledge of key content experts to identify additional tools not represented they are aware of.

Bibliographic database search strategy

An experienced medical information specialist developed and tested the search strategies through an iterative process in consultation with the review team. A full search strategy for OvidMEDLINE® ALL, including Epub Ahead of Print, In-Process & Other Non-Indexed Citations, and Embase Classic+Embase will be searched on the Ovid platform, followed by a search of Web of Science (see Appendix I). The strategies utilize a combination of controlled vocabulary (e.g.,"Cross-Sectional Studies", "Bias","Checklists") and keywords (e.g.,"cross-sectional design", "quality rating","tool"). The search strategy will be adapted for each included information source. The search strategy will aim to locate both published and unpublished tools. Results will be limited to the publication years 2011 to the present. Results will be downloaded and deduplicated using EndNote version 20.4.1 (Clarivate Analytics, PA, USA) and uploaded to Covidence systematic review software for screening (Veritas Health Innovation, Melbourne, Australia. Available at www.covidence.org). The database search will be updated prior to completion of the scoping review.

Evidence selection

Following the search, all identified database records will be collated and uploaded into Covidence systematic review software (Veritas Health Innovation, Melbourne, Australia. Available at www.covidence.org) and EndNote and duplicates removed. Following a pilot test, titles and abstracts will then be screened by two independent reviewers for assessment against the inclusion criteria for the review. Potentially relevant records will be retrieved in full and tracked through Covidence. The full text of selected citations will be assessed in detail against the inclusion criteria by one reviewer and included records will be confirmed by a second independent reviewer.

The included study lists of all eligible systematic reviews will be screened to identify potentially relevant RoB tools. All records identified through systematic reviews, the environmental scan, expert working group and PROSPERO registry search will be retrieved in full, assessed in detail against the inclusion criteria by one reviewer and confirmed by a second independent reviewer. All non-database records will be tracked using an MS Excel spreadsheet (Microsoft Corp. (2018).

Following a detailed assessment, a super-set of unique RoB tools relevant to cross-sectional studies will be created once de-duplication of records and tools across the approaches is complete.

Reasons for exclusion of full-text records that do not meet the inclusion criteria will be recorded and reported in the scoping review. Any disagreements that arise between the reviewers at each stage of the selection process will be resolved through discussion or with a third reviewer. The results of the search will be reported in full in the final scoping review and presented in a PRISMA flow diagram [3].

Data extraction

Data will be extracted from tools included in the scoping review by one reviewer and checked by a second reviewer. The data extracted will focus on descriptive information, characteristics of the tool, items and domains used to evaluate risk of bias and any psychometric properties reported, including:

- citation, tool name;
- reported study design(s) intended for use/application (single or multiple study designs);
- tool development and testing: reported methods used to develop the tool, reported reliability, validity or usability testing, including whether the tool was developed or tested in a particular clinical area;
- characteristics of the tool including but not limited to: use of domains/items/rating probes, format for completing the item or overall rating, whether rating summaries of quality or bias are used (e.g. poor/good/moderate) or if scoring is calculated;
- discrimination between concepts of quality, risk of bias, and reporting quality;
- how and why an existing tool was modified for application to cross-sectional studies; and,
- declared organization supporting the tool development, any sponsors or reported funding.

Data extraction will be standardized using a framework reported for RoB tools by the National Toxicology Program of the United States Department of Health and Human Services [4]. The extraction framework focuses on key bias domains (*selection, exposure, outcome, confounding, loss-to-follow-up, analysis, selective reporting, conflict of interest and other*), the topic(s) within each domain, the questions and available response options. The extraction framework will be adapted if necessary for considerations specific to cross-

sectional studies during a pilot to assess the feasibility of the process for extracting data from each included record. Any modifications will be detailed in the full scoping review. Any disagreements that arise between the reviewers will be resolved through discussion or with a third reviewer. All extracted information will be compiled into detailed evidence tables in Microsoft Excel.

Data analysis and presentation

Results will be presented in tables and figures (if applicable) and described in detail. A narrative summary will accompany the charted results and will describe how the results relate to the scoping review objective and questions. The final evidence product will be an item bank of risk of bias concepts relevant to cross-sectional studies.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Supplementary material *and/or* additional information

Supplementary material

Search strategy

MEDLINE
Database: Ovid MEDLINE(R) ALL ⟨1946 to November 17, 2021⟩

1 Cross-Sectional Studies/ (398,684)
2 ((cross-section* or crosssection*) adj2 (study or studies or design?)).ti,kw,kf. (48,153)
3 1 or 2 [CROSS-SECTIONAL STUDIES] (412,208)
4 Cross-Sectional Studies/st [standards] (41)
5 Research Design/st [standards] (12,411)
6 Evidence-Based Medicine/mt, st [methods,standards] (9222)
7 or/4–6 [RESEARCH DESIGN STANDARDS] (21,297)
8 3 and 7 [CROSS-SECTIONAL STUDIES - RESEARCH DESIGN STANDARDS] (320)
9 exp bias/ (72,351)
10 bias$2.ti,kw,kf. (38,345)
11 ((assess* or apprais* or estimat* or grade? or grading) adj3 (bias* or quality)).tw,kw,kf. (118,427)
12 ((susceptib* or risk?) adj3 bias*).tw,kw,kf. (33,354)
13 ((rate or rated or rates or rating) adj2 quality).tw,kw,kf. (7834)
14 Research Design/ (116,838)
15 (research adj3 (appropriate* or design? or method* or protocol? or strateg* or valid*)).tw,kw,kf. (132,941)
16 (methodological* adj3 (appropriate* or consider* or decision? or design? or issue? or quality or valid*)).tw,kw,kf. (33,428)
17 ((design? or report* or study or studies) adj3 (appropriate* or quality or valid*)).tw,kw,kf. (160,543)
18 "Reproducibility of Results"/ (431,267)
19 reproducib*.ti,kw,kf. (18,259)
20 (reliabilit* or reliabl*).ti,kw,kf. (60,000)
21 (replicable or replicat*).ti,kw,kf. (58,248)
22 Valid*.ti,kw,kf. (142,161)
23 Confound*.ti,kw,kf. (5461)
24 Data Accuracy/ (3315)
25 ((calculat* or data or measur* or statistic*) adj3 (accurat* or accurac* or appropriate* or objective)).tw,kw,kf. (141,906)
26 ((calculat* or data or measur* or statistic*) adj3 (correct* or exact* or precis*)).tw,kw,kf. (50,759)
27 or/9–26 [QUALITY] (1,285,786)
28 3 and 27 [CSS - QUALITY] (38,988)
29 Checklists/ (7539)
30 (checklist* or check list*).ti,kw,kf. (8020)
31 (tool or tools).ti,kw,kf. (100,645)
32 (instrument or instruments).ti,kw,kf. (37,759)
33 (framework? or frame work?).ti,kw,kf. (58,639)
34 ((assess* or apprais* or grade? or grading or quality) adj3 (criteri* or guide or guideline? or guidance? or standard?)).ti,kw,kf. (7910)
35 or/29–34 [TOOLS/FRAMEWORKS] (215,219)
36 28 and 35 [CSS - QUALITY ASPECTS - TOOLS] (1844)
37 8 or 36 [CSS - QUALITY ASPECTS - TOOLS/STANDARDS] (2136)
38 limit 37 to yr="2010-current" (1713)

Additional information

In a cross-sectional study, a sample of individuals is selected from a larger population and data is collected from each individual in the sample. This data can then be analyzed to draw conclusions about the population as a whole (5). Researchers may use cross-sectional studies for population-based surveys, to estimate the prevalence of outcomes and/or intervention/exposures or to measure association between the intervention/exposure and outcome by calculating an odds ratio [5,6].

The strengths of this design are that studies can be conducted relatively more resource efficiently when compared to other study designs, they can provide information that can be used to plan subsequent research studies (hypothesis generating), as well as provide data to support planning, monitoring and evaluation of health services or public health programs [6]. They often employ survey questionnaires which can be quick and simple approaches to collect prevalence data [7]. At the population level, the resulting prevalence estimates can be informative and generalizable. However, there are several limitations associated with cross-sectional studies. Causal relationships cannot be derived from a one-time measurement of intervention/exposure and outcome and there are no temporal elements within the design to strengthen estimates or sufficiently determine trends [6,7]. It can be difficult to plan and conduct a high-quality cross-sectional study as this design is prone to real or potential bias related to difficulties obtaining representative samples, inaccuracies in collected information, unclear temporal relationships and the possibility of reverse causality. Like other observational study designs, a major limitation is the inability to account for unobservable confounding when associations of interest are identified [6,8–10].

When trying to answer research questions, knowledge users may use or synthesize evidence from cross-sectional studies. Including a risk-of-bias (RoB) assessment is crucial to understanding relationships among interventions, exposures and the parameters of interest, and interpreting study findings. Although several tools exist for assessing the RoB in cross-sectional studies, there is no consensus on a 'gold standard' tool.

Recent reviews of RoB tools have shown that existing tools may apply to one or more study design, but vary in their approach to scoring or rating, the included RoB concepts, and their response options, and variably include additional items related to generalizability, precision or reporting [11,12]. Previous research concluded that many RoB tools are developed explicitly for experimental research and do not adequately address key biases common in observational research [13,14]. Many tools developed broadly for multiple observational study designs are not specifically applicable to cross-sectional studies or fail to sufficiently assess important aspects of bias common in cross-sectional studies [14–16]. While multi-study RoB tools cover a breadth of methodological bias concepts, using design-specific RoB tools can help ensure that all relevant sources of bias are considered and evaluated appropriately. These design-specific tools may be more applicable, user-friendly and easier to interpret where there is no control group present [15].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Shannon E. Kelly: Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing, Visualization. **Karima Benkhedda:** Conceptualization, Methodology, Writing – review & editing. **Stephen P.J. Brooks:** Conceptualization, Methodology, Writing – review & editing. **Amanda J. MacFarlane:** Conceptualization, Methodology, Writing – review & editing. **Linda S. Greene-Finestone:** Conceptualization, Methodology, Writing – review & editing. **Becky Skidmore:** Methodology, Validation, Writing – original draft, Writing – review & editing. **Tammy J. Clifford:** Supervision, Methodology, Writing – review & editing. **George A. Wells:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Data availability

No data was used for the research described in the article.

Acknowledgments

The review team would like to acknowledge Beth Yetley for her early involvement and for help with conceptualization of the work.

References

[1] J. McGowan, S. Straus, D. Moher, E.V. Langlois, K.K. O'Brien, T. Horsley, et al., Reporting scoping reviews-PRISMA ScR extension, J. Clin. Epidemiol. 123 (2020) 177–179.
[2] E. Aromataris, Z. Munn (Editors). JBI Manual for Evidence Synthesis [internet] JBI; 2020 [Available from: https://synthesismanual.jbi.global.
[3] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, et al., The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, Int. J. Surg. 88 (2021) 105906.
[4] Risk of Bias Tools [internet]: National Toxicology Program of the United States Department of Health and Human Services; 2020 [Available from: https://ntp.niehs.nih.gov/whatwestudy/assessments/noncancer/tools/index1.html?utm_source=direct&utm_medium=prod&utm_campaign=ntpgolinks&utm_term=ohat_tools.

[5] B.C. Reeves, G.A. Wells, H. Waddington, Quasi-experimental study designs series—Paper 5: a checklist for classifying studies evaluating the effects on health interventions—A taxonomy without labels, J. Clin. Epidemiol. 89 (2017) 30–42.
[6] M.S. Setia, Methodology series module 3: cross-sectional studies, Indian J. Dermatol. 61 (3) (2016) 261–264.
[7] W.A Van der Stede, A manipulationist view of causality in cross-sectional survey research, Account. Organ. Soc. 39 (7) (2014) 567–574.
[8] I. Yu, S. Tse, Workshop 6–sources of bias in cross-sectional studies; summary on sources of bias for different study designs, Hong Kong Med. J. 18 (3) (2012) 226–227 = Xianggang yi xue za zhi.
[9] P. Sedgwick, Bias in observational study designs: cross sectional studies, BMJ Br. Med. J. (2015) 350 (Online).
[10] P. Sedgwick, Cross sectional studies: advantages and disadvantages, BMJ Br. Med. J. (2014) 348 (Online).
[11] M.J. Downes, M.L. Brennan, H.C. Williams, R.S. Dean, Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS), BMJ Open 6 (12) (2016) e011458.
[12] Z. Wang, K. Taylor, M. Allman-Farinelli, B. Armstrong, L. Askie, D. Ghersi, J. McKenzie, S.L. Norris, M. Page, A. Rooney, T. Woodruff, L. Bero. A systematic review: tools for assessing methodological quality of human observational studies. 2019 [Internet]. Available from: osf.io/preprints/metaarxiv/pnqmy.
[13] C. Mallen, G. Peat, P. Croft, Quality assessment of observational studies is not commonplace in systematic reviews, J. Clin. Epidemiol. 59 (8) (2006) 765–769.
[14] T. Shamliyan, R.L. Kane, S Dickinson, A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases, J. Clin. Epidemiol. 63 (10) (2010) 1061–1070.
[15] M.J. Glasgow, R. Edlin, J.E. Harding, Comparison of risk-of-bias assessment approaches for selection of studies reporting prevalence for economic analyses, BMJ Open 10 (9) (2020) e037324.
[16] A. Moskalewicz, M. Oremus, No clear choice between newcastle-ottawa scale and appraisal tool for cross-sectional studies to assess methodological quality in cross-sectional studies of health-related quality of life and breast cancer, J. Clin. Epidemiol. 120 (2020) 94–103.

Re-print of the published journal article (Scoping review)

Journal of Clinical Epidemiology 172 (2024) 111408

Journal of Clinical Epidemiology

ORIGINAL RESEARCH

A scoping review shows that no single existing risk of bias assessment tool considers all sources of bias for cross-sectional studies

Shannon E. Kelly[a,b,*], Stephen P.J. Brooks[c], Karima Benkhedda[c], Amanda J. MacFarlane[d,e], Linda S. Greene-Finestone[f], Becky Skidmore[g], Tammy J. Clifford[a], George A. Wells[a,b]

[a]*School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada*
[b]*Cardiovascular Research Methods Centre, University of Ottawa Heart Institute, Ottawa, Ontario, Canada*
[c]*Bureau of Nutritional Sciences, Health Canada, Ottawa, Ontario, Canada*
[d]*Department of Biology, Carleton University, Ottawa, Ontario, Canada*
[e]*Texas A&M Agriculture, Food, and Nutrition Evidence Center, Fort Worth, Texas, USA*
[f]*Applied Research Division, Public Health Agency of Canada, Ottawa, Ontario, Canada*
[g]*Independent contractor, Ottawa, Ontario, Canada*

Accepted 28 May 2024; Published online 4 June 2024

Abstract

Objectives: Different tools to assess the potential risk of bias (RoB) for cross-sectional studies have been developed, but it is unclear whether all pertinent bias concepts are addressed. We aimed to identify RoB concepts applicable to cross-sectional research validity and to explore coverage for each in existing appraisal tools.

Study Design and Setting: This scoping review followed the Joanna Briggs Institute methodology. We included records of any study design describing or reporting methods, concepts or tools used to consider RoB in health research reported to be descriptive/prevalence survey or analytic/association (cross-sectional) study designs. Synthesis included quantitative and qualitative analysis.

Results: Of the 4556 records screened, 90 were selected for inclusion; 67 (74%) described the development of, or validation process for, appraisal tools, 15 (17%) described methodological content or theory relevant to RoB for cross-sectional studies and 8 (9%) records of methodological systematic reviews. Review of methodological reports identified important RoB concepts for both descriptive/prevalence and analytic/association studies. Tools identified (n = 64 unique tools) were either intended to appraise quality or assess RoB in multiple study designs including cross-sectional studies (n = 21; 33%) or cross-sectional designs alone (n = 43; 67%). Several existing tools were modified (n = 17; 27%) for application to cross-sectional studies. The RoB items most frequently addressed in the RoB tools were validity and reliability of the exposure (53%) or outcome (65%) measurement and representativeness of the study population (59%). Most tools did not consider nonresponse or missingness appropriately or at all.

Conclusion: Assessing cross-sectional studies involve unique RoB considerations. We identified RoB tools designed for broad applicability across various study designs as well as those specifically tailored for cross-sectional studies. However, none of the identified tools comprehensively address all potential biases pertinent to cross-sectional studies. Our findings indicate a need for continued improvement of RoB tools and suggest that the development of context-specific or more precise tools for this study design may be necessary.

Keywords: Scoping review; Risk of bias; Tool; Research methodology; Cross-sectional study; Prevalence

* Corresponding author. School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Suite 315 - 600 Peter Morand Cres, Ottawa, Ontario K1G 5Z3, Canada.
E-mail address: skell102@uottawa.ca (S.E. Kelly).

https://doi.org/10.1016/j.jclinepi.2024.111408

Plain Language Summary

What is the problem?

Cross-sectional research looks at a group of people at one specific point in time to understand how common certain traits or behaviors are and to find connections between different factors, and often, health or disease. These studies provide a snapshot at one time point but do not show changes over time. Results of these studies are often used to provide initial insights that can inform important health-care—related decisions by patients, policy makers, researchers and health providers.

We rely on these studies to be well conducted and trustworthy, but there may be problems with how these studies are designed and conducted that may influence the results. These problems are referred to as biases. There are tools designed to help researchers identify possible biases in cross-sectional studies which may influence the results, but it's not clear if these tools cover all the important issues, some of which may be unique to these types of studies.

What did we do?

We searched for and identified tools that could be used to evaluate possibles biases in cross-sectional studies and articles that described potential biases in this type of research. We summarize the biases that tools should be considering and assessed how well each cross-sectional study tool did this.

What did we find?

We found 67 articles reporting tools used to check for biases in cross-sectional studies, 15 articles that summarized the biases important to cross-sectional research and eight articles that were reviews of related methods or biases. We found that some of the tools were made for assessing many different types of studies, while others were developed specifically for cross-sectional studies. In the end, we found that none of the tools we identified covered all the possible biases that could influence the results of a cross-sectional study away from the 'truth'. These findings show that researchers need to improve these tools or develop new tools that include all biases that may impact cross-sectional studies.

1. Background

Cross-sectional studies are a common epidemiological research design used to investigate the prevalence of disease, characteristics or the distribution of outcomes in populations. They may also be used to explore analytic associations between health exposures and disease outcomes but cannot provide data for causality [1]. These study designs give a 'snapshot' of a health variable (or any other factor) in a given population at one point in time, which can provide insight into the occurrence of disease, health knowledge, risk factors, patterns or behaviors. As with any research study, the robustness of the findings is contingent on minimizing factors that may compromise the validity, reliability and generalizability of the results [2].

Bias, or systematic errors in the study design, conduct or analysis, can impact cross-sectional and prevalence studies in unique ways [3—5]. for example, selection bias is a threat to the validity of cross-sectional and prevalence studies. As such, sampling strategies must be carefully designed so as not to distort estimates of exposure-outcome relationships or prevalence rates and to maintain representativeness to a target population. Cross-sectional studies are also particularly susceptible to information bias, especially inaccuracies in the measurement of exposures or outcomes. Related to information bias, it is difficult or impossible to establish a temporal sequence for the variables under study which is a requirement for any epidemiological consideration of causation. Cross-sectional data collection inherently limits our ability to discern whether an exposure precedes an outcome or vice versa [5]. These challenges lead to cross-sectional and survey designs being very low on any 'hierarchy of evidence' use for evidence-based decision-making [6].

Despite this, cross-sectional studies offer significant advantages to researchers through efficient and low-resource data collection, and therefore, are well-utilized in many disciplines. Likewise, knowledge users often rely on cross-sectional research studies to inform planning, policy, practice or other research activities. Taking benefits and limitations into consideration, it is critical to accurately identify reliable, valid and trustworthy cross-sectional research. Structured tools or instruments have been created to assist with discernment of study quality or risk of bias (RoB)

What is new?

Key findings

- This is a comprehensive scoping review describing 64 tools, instruments or checklists and 23 methodological articles or systematic reviews applicable to risk of bias assessment for cross-sectional studies.
- Based on an analytic framework of key bias domains and concepts, no single tool considers all potential sources of bias that may impact cross-sectional research and very few are assessed for validity and reliability.

What this adds to what was known?

- Existing tools for appraising cross-sectional research may be intended for critical appraisal, assessment of methodological quality or risk of bias; Regardless, no single tool permits users to comprehensively judge factors likely to lead to bias.
- Tools include items unrelated to risk of bias, including reporting quality, applicability and statistical precision, and may not reflect current best practices, such as avoiding quantitative scales, focusing appraisal on internal validity and having coverage for biases relevant to the study design.

What is the implication and what should change now?

- To improve assessment of research quality, tool developers should report utility, validity and reliability metrics (or a protocol to assess) and build on the strengths and limitations of existing tools.
- Users should consider whether a tool is valid and reliable, follows current best practice and assesses factors most likely to lead to bias in cross-sectional research approaches (Prevalence/survey or analytic/association).

and each may be applicable to one or more epidemiologic study designs. However, there are many tools available, and it is unclear how well existing tools consider all potential biases for cross-sectional studies. This scoping review aims to identify and describe available tools to assess the RoB in cross-sectional studies and to compile the critical RoB items, domains, concepts applicable to cross-sectional studies into an item bank. A secondary objective was to determine the extent to which the retrieved RoB tools address the relevant bias items, domains and concepts identified.

2. Materials and methods

The scoping review was conducted using methodology developed by Arksey and O'Malley [7] and later refined by the Joanna Briggs Institute (JBI) [8]. It is reported using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR; Supplemental Appendix, Table A) [9]. A protocol is published elsewhere [10].

2.1. Information sources

We searched for MEDLINE, Embase, and Web of Science for potentially relevant tools, methodological systematic reviews, systematic reviews of cross-sectional studies and peer-reviewed methodological guidance specific to cross-sectional research on November 18, 2021, and updated the search on August 6, 2023. The strategies used a combination of controlled vocabulary (eg, "Cross-Sectional Studies", "Bias", "Checklists") and keywords (eg, "cross-sectional design", "quality rating", "tool"). Vocabulary and syntax were adjusted across the databases. To supplement the bibliographic database search, we consulted methodological and epidemiological experts and reviewed ROB tools identified in systematic review protocols registered through The International Prospective Register of Systematic Reviews (PROSPERO). Additional details regarding the strategies appear in the Supplemental Appendix.

2.2. Eligibility criteria

This is a methodological scoping review and as such, we present the review in terms of the Population, Concept, Context (PCC) framework based on guidance from JBI [8].

Population: Not applicable Concept: The concepts of interest are the potential biases that are relevant to research studies employing a cross-sectional design, and whether corresponding appraisal tools address these biases sufficiently. Context: All contexts (including setting, field and discipline) will be considered.

We sought to identify tools intended to be used to appraise RoB in cross-sectional study designs. We considered RoB appraisal to be a measure of how features of the design, conduct of a study may produce systematic error in the results of a study. As terminology for methodological quality and RoB are commonly used interchangeably and inconsistently, tools for critical appraisal and quality assessment were eligible if they contained one or more items relating to the internal validity of a study [11]. Tools could be checklists (ie, users are asked to consider whether a study meets a predetermined set of items, usually criteria, signaling questions or probes) or scales (ie, instruments structured to provide quantitative scores for each item presented for appraising research) [12]. Tools were eligible if they were intended for assessment of: (1) observational or descriptive studies defined

as cross-sectional studies, including analytic/association or prevalence/descriptive approaches; (2) multiple study designs, including cross-sectional study designs; or (3) study designs other than cross-sectional studies but were adapted or modified for application to cross-sectional studies. We excluded records published in languages other than English and tools that were not intended or modified to be applied to cross-sectional studies. Records for systematic reviews that did not cite the RoB tool used or which adapted an existing RoB tool for application to cross-sectional studies but did not make the tool available, were excluded. Methodological records were eligible if they described concepts, items, domains or tools relevant to appraisal of RoB in cross-sectional research. Methodological concepts are defined as constructs related to the validity of cross-sectional research.

2.3. Selection process

Following a pilot exercise ($n = 100$ records), the titles and abstract for each retrieved record were screened by one reviewer (SK). A second reviewer (LGF) vetted the included and excluded records. The full text of all potentially relevant records was retrieved and assessed by one reviewer (SK). All included records were confirmed through discussion and validation with the review team. The included records from all eligible systematic reviews, PROSPERO registrations and methodological records were checked for potentially relevant RoB tools (SK, SB).

2.4. Data items and collection process

One reviewer (SK) extracted data from each identified tool, including details of the development process, structure, included items and domains, scoring or evaluation system and any reported processes for validation. We classified tools as checklists or scales, by their self-identified use, and their intended study design application. Information was extracted using standardized Google forms (Google Inc, 2023) and then transferred to Microsoft Excel (Microsoft Corp., 2018).

Methodological records were reviewed for content applicable to the RoB in cross-sectional research designs. One reviewer (SK) extracted relevant methodological concepts, biases, items, domains or other author-identified considerations for cross-sectional research design or appraisal from each record. Content was extracted verbatim from each record and compiled by record in Microsoft Excel (Microsoft Corp., 2018). One reviewer (SK) amalgamated results into similar themes and constructs. The study team iteratively discussed results and compiled data into domains or items applicable to cross-sectional research.

2.5. Critical appraisal

As this is a scoping review, we did not conduct a critical appraisal of the included records.

2.6. Data charting and synthesis

The characteristics of the identified tools and methodological records were summarized with proportions for categorical variables, and means and standard deviations (SDs), for continuous variables. Findings from the methodological records are reported descriptively. An analytic framework of RoB domains and items proposed by Wang et al (2019) for observational studies was used to map the bias domains and items covered in each included tool [13]. Informed by the methodological concepts located in the search, the study team discussed all methodological concepts extracted. Following iterative discussions, the framework was adapted to include domains or items not currently addressed, but relevant to cross-sectional research, and to remove those not considered relevant to the study design or not related to RoB. Tools for descriptive/prevalence or analytic/association cross-sectional studies were considered separately.

3. Results

3.1. Literature search

The bibliographic database search returned 5164 records (3527 records after deduplication) and 29 additional records were identified by content experts and review authors. A total of 27,249 records were identified using the PROSPERO database. After 1000 records were examined prospectively, the review team agreed that an acceptable saturation of tools had been achieved given how few new unique ROB tools were identified over the last 200 records (Supplemental Appendix, Figure A). Including these 1000 PROSPERO records, a total of 4556 records were screened. Database records were screening using title and abstract and PROSPERO registrations using the 'Risk of bias (quality) assessment' field of each record. Of these, 348 were reviewed in full text to assess eligibility for inclusion in the scoping review. The full text of nine records could not be retrieved due to insufficient citation details provided and follow-up by email twice to corresponding authors from the original reports was unsuccessful. Following full-text review, 90 records were included [3,13−50],[1,51−79],[5,80−99] (Fig).

3.2. Summary of included records

Of the 90 records selected for inclusion, 67 (74%) described the development of, or validation process for, appraisal tools [14−27,29−37,39,41,45−54],[56,58,60,62−64,66−73,75,77,80−91,95,97−99], 15 (17%) described methodological content or theory relevant to RoB for cross-sectional studies [1,3,5,28,40,42−44,57,59,65,74,76,92,94] and 8 (9%) were records of methodological systematic reviews [13,38,55,61,78,79,93,96].

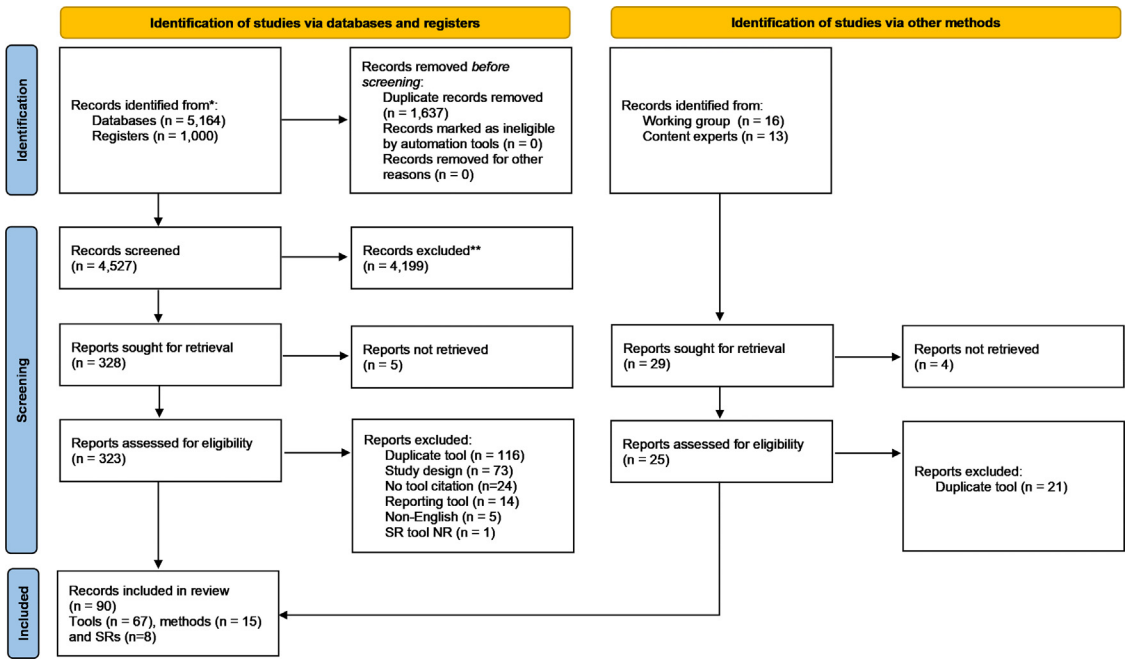


Figure. Flow diagram of searches of databases, registers and other sources. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.3. Methodological concepts relevant to risk of bias in cross-sectional studies

Records used to identify methodological constructs (n = 15) [1,3,5,28,40,42–44,57,59,65,74,76,92,94] were published between 1995 and 2023. The included records were generally focused on methodological descriptions of cross-sectional study designs [3,5,28,57,59,76,92,94], surveys of prevalence [40,43,44], nonrandomized research with consideration for cross-sectional studies [1,42,65,74] and two records focused specifically on sources of bias in cross-sectional research [76,94]. However, the terminology for describing cross-sectional study designs was ambiguous. Identified sources of bias in cross-sectional studies were quite broadly described, with most records highlighting the potential for selection bias related to sampling frame [1,3,5,57,76,94,100], nonsampling error [5,76,94] and the concern that it is not possible to determine temporality (ie, whether the outcome occurred prior to the exposure or intervention; reverse causality) [1,3,5,57,76,94,100]. Information bias related to the appropriate and valid measurement of exposure or outcomes was a commonly highlighted source of bias [3,5,28,57,59,76,92,94]. In addition, several records mentioned items not related to RoB, including the handling of statistics, sample size, various reporting criteria (ie, whether a design feature was present in the study report and/of if the details were fulsomely described), the generalizability of results or appropriateness of conclusions in context with RoB concepts.

Biases were considered separately depending on the intent of the study to provide descriptive/prevalence or analytic/association estimates [3,57,74,94]. For descriptive cross-sectional studies aiming to provide estimates of prevalence, the noted threats to validity were representativeness of the sampling frame, nonsampling error from nonresponse or coverage and whether the outcome, disease or characteristic under study was measured in a valid and reliable way. Considerations specifically for analytic cross-sectional studies [3,5,28,57,59,74] focused less on sampling frame, and included consideration for, and handling of, confounding variables in estimates and subgroups, appropriateness of eligibility criteria or group under study, blinding of research staff, validity and reliability of the exposure and outcome ascertainment, missingness in the data and time components related to recruitment. None of the records mentioned selective reporting, conflict of interest or funding as potential sources of bias.

3.4. Methodological systematic reviews

A total of eight systematic reviews [13,38,55,61,78,79,93,96] published between 2010 and 2020 were included. The systematic reviews all endeavored to identify and describe or review use of tools for assessing quality or RoB in human research studies. The reviews focused on summarizing tools intended for prevalence studies [61,79], observational studies [13], or all research designs [38,55,78,93,96]. Of these, two reviews focused specifically on intervention studies, yet the tools cited were applicable to intervention and observational study designs [38,93]. The included tools and references in each review were explored in detail to identify and extract RoB tools intended for assessment of cross-sectional studies. Not all tools

Table 1. Overview of tools for cross-sectional studies (n = 64)

Author (or organization), year	Abbreviation	Full tool name	Clinical focus	Study designs
Tools intended for multiple study designs, including cross-sectional studies				
Fowkes and Fulton, 1991 [39]	N/A	Critical appraisal tool	Broad	RCT, C, CC, XS
Avis, 1994 [29]	N/A	Critical appraisal tool	Broad	Any[a]
Law, 1998 [52]	N/A	Critical Review Form for Quantitative Studies	Occupational therapy	RCT, C, CC, XS, CSt
Macfarlane, 2001 [56]	N/A	Tool for XS studies reported in SR	Oro-facial pain	Cohort/XS
Margetts, 2002 [60]	N/A	Checklist for reviewing nutritional epidemiological studies	Nutrition	RCT, C, CC, XS, P, E
Slim, 2003 [99]	MINORS	Methodological index for nonrandomized studies	Surgery	NRS, XS
Kmet, 2004 [50]	N/A	Quality assessment criteria for evaluating primary research papers from a variety of fields	Broad	Any[b]
Genaidy, 2007 [41]	EAI	Epidemiological Appraisal Instrument	Broad	RCT, C, CC, NRS, XS
Crowe, 2011 [14,33–35]	CCAT	Crowe Critical Appraisal tool	Broad	XS, P, RCT, NRS, C, CC, CSe, CSt, Q, SRs
Ross, 2011 [97]	SAQOR	Systematic Assessment of Quality in Observational Research	Broad	C, CC, XS, Any[c,d]
Viswanathan, 2012 [88,89]	RTI	RTI Item Bank	Broad	C, CC, CSe, XS
Puzzolo, 2013 [72]	LQAT	EPPI-Centre, Liverpool Quality Assessment tools	Public Health	RCT, C, CC, NRS, XS
Takahashi, 2014 [83], Sedgwick, 2019 [77]	NOS for XS	Newcastle-Ottawa Scale applied to cross-sectional studies	Broad	C, XS[e]
Hong, 2018 [46]	MMAT -QNR	Mixed Methods Appraisal Tool for quantitative nonrandomized studies	Broad	NRS[b] (including XS)
Hong, 2018 [46]	MMAT-QD	Mixed Methods Appraisal Tool for quantitative descriptive studies	Broad	Any[f] (including P)
Kennedy, 2019 [49]	N/A	The Evidence Project Risk of Bias tool	Broad	RCT, C, CC, NRS, PTPT, XS[b,c]
Stone, 2019 [82]	MASTER	MethodologicAl STandard for Epidemiological Research (MASTER) Unified Framework for bias assessment in clinical research	Broad	Any[a]
National Heart Lung and Blood Institute (NHLBI), 2021 [16]	NHLBI	NHLBI Study Quality Assessment Tool for Observational cohort and cross-sectional studies	Cardio-pulmonary	C, XS
Academy of Nutrition and Dietetics, 2022 [17]	ANDEAL	Academy of Nutrition and Dietetics Evidence Analysis Library (ANDEAL): Quality Criteria Checklist for Primary Research	Nutrition	RCT, C, CC, Cse, CSt, XS, E, NRS, D, BA
Al Asmri, 2023 [25]	MERSQI	Medical Education Research Study Quality Instrument	Medical education	RCT, C, CC, XS, PTPT Controlled NRS
Tools intended for cross-sectional study designs (descriptive/prevalence or analytic/association)[g]				
Boyle, 1998 [31]	N/A	Guidelines for evaluating prevalence studies	Psychiatry	P
Loney, 1998 [53]	N/A	Critical appraisal of the health research literature: prevalence or incidence of a health problem	Broad	P
van der Windt, 2000 [85]	N/A	Tool for XS studies reported in SR	Broad	XS

(*Continued*)

Table 1. Continued

Author (or organization), year	Abbreviation	Full tool name	Clinical focus	Study designs
Silva, 2001 [81]	N/A	Tool for assessing the usefulness of prevalence studies done for surveillance purposes	Chronic/ noncommunicable disease surveillance	P[f]
Al-Jader, 2002 [26]	N/A	Quality scoring system for epidemiological surveys of genetic disorders	Genetics	P[f]
Jackson, 2006 [48]	GATE- appraise	Graphical Appraisal Tool	Broad	XS
Viswanathan, 2008 [87]	AHRQ	AHRQ Design-Specific Criteria To Assess Risk of Bias in XS Studies	Broad	XS[h]
Busse, 2011 [32]	CLARITY	RoB for cross-sectional surveys of attitudes and practices	Broad	XS
National Collaborating Center on Environmental Health, 2011 [91]	NCCEH-XS	NCCEH Tool for Critical Appraisal of a Cross-Sectional Study on Environmental Health	Environmental health	XS
Shamliyan, 2011 [80]	MORE	Methodological Evaluation of Observational Research	Broad but developed based on topics in chronic disease or risk factors	P[f]
Yusef, 2011 [95]	N/A	Tool for XS studies reported in SR	Broad[a]	XS
Hoy, 2012 [47]	N/A	Tool for assessing risk of bias in prevalence studies	Broad but developed with neck and back pain focus	P[f]
Powell, 2013 [70]	N/A	Tool for XS studies reported in SR	Broad	XS
DeGroot, 2014 [36]	QATQS	Quality Assessment Tool for Quantitative Studies modified in SR for application to XS studies	Broad	XS
University of Oxford, Center for Evidence-Based Medicine, 2014 [58]	CEBM	CEBM Critical Appraisal Tool for Cross-sectional Studies	Broad	XS
Downes, 2016 [37]	AXIS	Appraisal tool for Cross-Sectional Studies	Broad medical or veterinary	XS
Ohadike, 2016 [67]	N/A	Tool for XS studies reported in SR	Broad	XS
Cardiff University, 2018 [24]	SURE	Cross-sectional Studies Critical Appraisal Checklist	Broad	XS
National Toxicology Program, US Department of Health and Human Services, 2019 [15]	OHAT	Office of Health Assessment and Translation (OHAT) RoB Tool	Toxicology	XS[h]
University of Oxford, Center for Evidence-Based Medicine, 2019 [23]	CAT Manager App	Critically Appraised Topic (CAT) Manager Tool for Cross-sectional Studies (Apple phone/ANDROID application)	Broad	XS
Nudelman and Otto, 2020 [66]	ROBUST	Risk-of-Bias Measure for Systematic Reviews of Surveys	Broad	P
Protogerou, 2022 [71]	Q-SSP	Quality of survey studies in psychology	Psychology	XS
Perez-Rios, 2022 [98]	STREAMS-P	STrengthen the design and REporting of Attributable Mortality Studies using a Prevalence-based method	Attributable mortality focus	P
Cincinnati Children's Hospital Medical Center, 2023 [20]	LEGEND-XS (INTERVENTION)	Let Evidence Guide Every New Decision (LEGEND) Tool for Cross Sectional Studies of Interventions	Broad	XS
Cincinnati Children's Hospital Medical Center, 2023 [21]	LEGEND-XS (ETIOLOGY, RISK FACTORS)	Let Evidence Guide Every New Decision (LEGEND) Tool for Cross Sectional Studies of Etiology or Risk Factors	Broad	P

(*Continued*)

Table 1. Continued

Author (or organization), year	Abbreviation	Full tool name	Clinical focus	Study designs
Cincinnati Children's Hospital Medical Center, 2023 [22]	LEGEND-XS (PREVALENCE)	Let Evidence Guide Every New Decision (LEGEND) Tool for Cross Sectional Studies of Prevalence	Broad	P
Joanna Briggs Institute (JBI), 2023 [18]	JBI-XS	JBI Checklist for Analytical Cross Sectional Studies	Broad	XS
Joanna Briggs Institute (JBI), 2023 [19]	JBI-P	JBI Checklist for Prevalence Studies	Broad	P
Ross, 2023 [75]	Modified NUQUEST	NUtrition QUality Evaluation Strengthening Tool for cohort studies modified in SR for application to XS studies	Nutrition	XS
Thomy, 2023 [84]	RoB-PrevMH	A tool to assess risk of bias in studies estimating the prevalence of mental health disorders (RoB-PrevMH)	Mental Health	P
Herzog, 2013 [45], Bawor, 2014 [30], Alshabanat, 2015 [27], Kunutsor, 2015 [51], Patra, 2015 [68], Perera, 2015 [69], van Dijk, 2015 [86],Modesti, 2016 [62], Rodriguez, 2016 [73], Murray, 2018 [64] Moskalewicz, 2020 [63], Lu, 2021 [54], Wagg, 2021 [90]	Modified NOS	Newcastle-Ottawa Scale adapted for cross-sectional studies	Broad[i]	XS[j]

BA, Before-after study; C, cohort study; CC, case-control study; Cse, Case series; CSt, Case study; D, diagnostic study; DS, Descriptive studies; E, Ecological study; NOS, Newcastle-Ottawa Scale; NRS, nonrandomized study; NUQUEST, NUtrition QUality Evaluation Strengthening Tool; OBS, observational study; P, Prevalence study; PTPT, Pretest/Posttest study; Q, qualitative study; RCT, randomized controlled trial; SR, systematic review; XS, cross-sectional study.

[a] Must be experimental or analytic study.
[b] Must be quantitative study.
[c] Must be comparative.
[d] Includes 'database' studies.
[e] NOS was intended for cohort studies but was applied to cross-sectional studies in two systematic reviews. Each was considered a separate tool (n = 2 tools).
[f] Must be descriptive study.
[g] Terminology for cross-sectional and prevalence stu0000-0002-2356-2300dies are often used ambiguously or interchangeably.
[h] Multistudy tool with subset of questions for XS studies only.
[i] Some items in the tool may be SR topic specific.
[j] Each was considered a separate tool (n = 13 tools).

reviewed were cited appropriately or reported in enough detail that the tools could be located and included in this scoping review.

3.5. Tools for RoB assessment in cross-sectional studies

From the 67 included records reporting tools, we identified 64 unique appraisal tools [14–27,29–37,39,41,45–54,56,58],[60,62–64,66–73,75,77,80–91,95,97–99]. One tool was reported in four of the included records and all were combined and considered one unique tool [14,33–35]. Table 1 provides an overview of the 64 unique RoB tools. Tools were applicable to multiple study designs, including cross-sectional studies (n = 22; 34%) [14,16,17,25,29,33–35,39,41,46,49,50,52,56,60,72,77,82,83,88,89], or cross-sectional designs only (n = 42; 66%) [15,18–24,26,27,30–32,36,37,45,47,48,51,53,54,58,62–64,66–71,73,75,80,81,84–87,90,91,95]. Twenty-two tools (34%) were identified from a nonmethodological systematic review [27,30,36,45,51,54,56,62,64,66–70,73,75,83,85–87,90,95]. While most tools were designed to be applied to any content area, several were developed specifically for clinical or health disciplines with the most frequent being nutrition (n = 3) [17,60,75], pain (n = 2) [47,56] and environmental health or toxicology (n = 2) [15,91]. Other content-specific tools were for surgery [99], psychology [71], occupational therapy [52], mental health [84], medical education [25], chronic disease [81], genetics [26], psychiatry [31], public health [72] or cardio-pulmonary

health [16]. Sixteen tools reported modifying or adapting an existing RoB tool for application to cross-sectional studies (n = 16; 27%) [27,30,45,51,54,62−64,68,69,73,75,86,90]. All modified tools were based on the Newcastle-Ottawa Scale [101] except for Ross et al (2023) [75] which adapted the NUtrition QUality Evaluation Strengthening Tool (NUQUEST) [102] for cohort studies for application to cross-sectional studies in a systematic review.

3.6. Characteristics of the included tools

The characteristics of the included tools are summarized in Table 2. Tools were published between 1991 and 2023. Tools are described as being intended for critical appraisal (n = 14; 22%), quality assessment (n = 37; 59%) or RoB (n = 11; 17%). Two additional tools did not differentiate between quality assessment or RoB in assessment or aimed to assess usefulness of research using methodological quality and trustworthiness concepts. Tools were either part of a set of tools developed to assess RoB (n = 14; 22%) or standalone (n = 50; 78%). Half of the tools were accompanied by some type of user guide or guidance (n = 32; 50%) but only one tool had a worksheet available to assist raters with pragmatic or conceptual aspects important to the RoB assessment (n = 1; <2%). Over two-thirds (n = 45; 70%) of the tools produced study ratings (ie, any type of structured or unstructured judgment relating to a study's potential for RoB or study quality/critical appraisal) (eg, low, unclear or high) and 30% (n = 19) calculated a quantitative score.

Forty-one percent of the tools grouped items or probes using categories or RoB domains (eg, selection, sampling, exposure ascertainment etc.) as a structure (n = 26; range 3 to 8 categories or domains). Tools contained, on average, 13.5 items or probes (range 3 to 43 items or probes). The number of response options available to rate studies ranged from 2 to 5 (mean 3.2 (0.8 SD). Almost 40% (n = 24) of the tools contained items related to reporting quality. Twenty-four tool developers (38%) made a declaration regarding their funding for the tool development but only 15 (23%) made any declaration regarding conflict of interest (COI).

Approaches used to develop the content of the tools are reported in Table 3. Teams developing content for tools were reported to be methodologists (n = 32; 50%), clinical content experts (n = 16; 25%) or undefined working groups of authors with unclear expertise (n = 32; 50%). No patient, caregiver, or public team members were reported to have contributed to tool content. Methods to develop the tools included a literature or systematic review of existing tools (n = 19; 29.7%), expert consultation (n = 10; 16%), consensus meetings (n = 8; 13%) or a Delphi process (n = 5; 8%). However, 60% (n = 39) of the tools did not report details about the tool content or development process. Few tool developers reported pilot (n = 13; 20%) [14,26,33−37,41,46,47,49,50,66,80,84], usability (n = 7; 11%) [14,26,33−35,37,41,49,72,88,89], reliability (n = 11; 17%) [14,26,33−35,41,47,49,50,56,66,80,97,99] or validity (n = 8; 12.5%) [14,25,33−35,41,46,66,80,99] testing as part of tool development. Most tools applicable to multiple study designs did not report any evaluation involving cross-sectional studies as part of their testing process. Reliability testing was based on interrater agreement at varying levels (eg, by study, domain or item/probe). Test-retest reliability was assessed for one tool by one reviewer [99]. One study considered the experience level of the reviewers in their reliability testing [66]. Content or criterion validity was assessed using expert consultation, group consensus or Delphi process. Convergent validity was assessed using ratings from the same studies in a previous meta-analysis with RoB rated by a different tool [66].

3.7. Analytic framework for coverage of RoB domains and items in cross-sectional tools

The analytic framework of RoB domains and items proposed by Wang et al (2019) [13] for observational studies was adapted by the study team to include the domains and items most applicable and appropriate for assessment of RoB for cross-sectional studies (Table 4). The original framework considered eight bias domains: *selection* (5 items), *exposure* (1 item), *outcome* (2 items), *confounding* (2 items), *loss to follow-up* (3 items), *analysis* (1 item), *selective reporting* (1 item) and *conflict of interest* (COI; 1 item). The framework also had a domain for *other bias* (2 items) and a category that considered 'other' items not applicable to research bias (*statistical power/sample size, analysis, clarity of research objectives/hypothesis, reporting of eligibility criteria* and *conclusions supported by results and methods*) (4 items). Based on the review of methodological guidance and discussion and agreement amongst the authors, *the loss to follow-up* domain was considered not applicable to this research design and removed. One item '*mechanism of non-response*' was added to the selection domain and a new domain for consideration of *missingness* was added (including 3 items: *amount of missingness*, *handling of missingness* and *mechanism for missingness*). Missingness considers data coverage for study participants. Handling of missingness may include imputation to assign replacement values for some, most or all missing values, and may include consideration for invalid or inconsistent data [103]. Items considered not applicable to research bias were removed from the final framework.

The updated analytic framework was used to map the domains and items covered in each included RoB tool. Table 4 provides a summary of the frequency each item was addressed, by domain, for all tools. The RoB Items most frequently addressed in the RoB tools were validity and reliability of the exposure (53%) or outcome (65%) measurement and representativeness of the study population (59%). Most tools did not consider nonresponse,

Table 2. Characteristics of the risk of bias tools for cross-sectional studies ($n = 64$)

Characteristic:	N (%)
Total included:	
Unique tools	64 (100)
Tool developed for[a]:	
Critical appraisal	14 (21.9)
Quality assessment	37 (57.8)
Risk of bias	11 (17.2)
Other	2 (3.1)[b]
Year:	
1999 or prior	5 (7.8)
2000 to 2009	10 (15.6)
2010 to 2019	33 (51.6)
2020 to 2023	18 (28.1)
Standalone or part of a tool suite:	
Standalone	50 (78.0)
Part of a suite of tools	14 (21.9)
Tool features:	
User guide/guidance provided:	
Yes	32 (50.0)
No	32 (50.0)
Worksheet available:	1 (<2%)
Assessment:	
Scored	19 (29.7)[c]
Rated	45 (70.3)[d]
Number of domains or categories	4.4 (1.7)[e,f]
Number of items/probes in tool:	13.5 (8.1)[e,g]
Number of response options:	3.2 (0.8)[e]
No response option suggested:	1 (1.6)[e]
Declarations:	
Funding:	
Declared: had funding	21 (32.8)
Declared: no funding	3 (4.7)
Not declared	40 (62.5)
Conflict of interest (COI):	
Declared — no COI	8 (12.5)
Declared — COI	7 (10.9)
Not declared	49 (76.6)

[a] Author declared terminology was collected for each tool. For 'other': One made was no distinction between quality assessment or risk of bias assessment and one tool was for assessment of 'useability' which combined concepts of methodological quality and trustworthiness.

[b] One reported being for 'usefulness' and one for 'validity'.

[c] Quantitative score calculated based on assessment of domains and/or items, categories (i.e., a scored tool).

[d] Domains and/or items used to assign a rating based on the assessment (could be overall rating and/or rating by domain, category or item, or both) (ie, a checklist).

[e] mean(SD).

[f] Items are structured using domains of bias (eg, selection, ascertainment, confounding) or another metric.

[g] Based on 63 tools. One tool was not considered in this category due to the graphical format [48].

missingness, selective reporting or COI fully, or at all. The Supplemental Appendix (Table B) provides a detailed map of all items in the framework for each tool. Many items in the included tools did not map to any item in the framework and were considered not applicable by the study team (ie, items related to reporting quality, statistics, precision of estimates, sample size, or clarity of objectives/hypotheses).

Exploratory review of subgroups of tools developed to be applied to multiple study designs or only cross-sectional studies showed few differences when considered separately and compared (Supplemental Appendix, Table C). Multistudy tools more frequently covered most items in the *selection* bias domain, except for *representativeness of the sample to the target population*. Cross-sectional only tools more frequently addressed the *validity and reliability of the outcome* and *exposure* measurement, while multistudy tools more frequently addressed the *description of confounding variables*. Tools that self-identified as RoB tools (vs. critical appraisal or quality) more frequently considered *nonresponse rate*, the *validity and reliability of the exposure* and *outcome* measurement and some *selection bias* items.

4. Discussion

We conducted a comprehensive scoping review that included 15 methodological papers, 8 systematic reviews and 64 assessment tools related to the RoB in cross-sectional studies. The findings highlight the variety of tools that are available and have been used to consider the RoB in cross-sectional studies. However, variability in the content and coverage for key bias domains indicates the absence of a "gold" standard tool for cross-sectional studies. Assessing RoB is a key step when knowledge users move from evidence to action, for all types of evidence. The volume of tools applicable to cross-sectional studies retrieved in this scoping review confirms a need for these instruments. Our findings may explain why existing tools have not been fully accepted and identifies gaps that should be addressed to encourage uptake and use [55].

Notably, our results suggest that the methodology to develop and evaluate RoB tools can be improved and better reported [105]. Consistent with broader reviews of similar intent, our assessment showed that most tools for appraisal of cross-sectional studies lack rigor or empirical evidence of their measurement properties [13,106]. Without this, we are unable to conclude that we are assessing RoB in a meaningful or useful way [107]. While many tools were developed to in theory to address issues of bias, or built on previous tools with extensive use cases, most were not formally assessed for reliability or validity. More attention to the empirical evaluation of the validity, reliability and utility of any proposed tool by appraisers and end-users is

Table 3. Development features of the risk of bias tools for cross-sectional studies (n = 64)

Characteristic:	*N* (%)
Methods to develop tool[a]:	
Literature/systematic review	19 (29.7)
Expert consultation	10 (15.6)
Delphi survey	5 (7.8)
Consensus meetings	8 (12.5)
Not reported/unclear	39 (60.9)
Development team:	
Content expert (methods)	32 (50.0)
Content expert (clinical)	16 (25.0)
Patients or public	0 (0.0)
Not reported/unclear	32 (50.0)
Pilot tested:	
Yes	13 (20.3)
No	23 (35.9)
Not reported/unclear	28 (43.8)
Reported usability:	
Yes	7 (10.9)
No	23 (35.9)
Not reported/unclear	34 (53.1)
Reported reliability:	
Yes	11 (17.2)
No	25 (39.1)
Not reported/unclear	28 (43.8)
Reported validity:	
Yes	8 (12.5)
No	30 (46.9)
Not reported/unclear	26 (40.6)

[a] Tools may have used one or more of the methods noted.

needed prior to dissemination [106]. Criterion validity and interrater reliability were most often assessed as part of the tool development process. However, the majority do not formally consider any psychometric properties of the tool, usability or the feasibility of application. This is especially evident in tools developed specifically for, and reported within, systematic reviews. Beyond basic applicability to the cross-sectional study design and the topic under review, very few of these tools were formally evaluated. This makes these tools fit-for-purpose yet potentially limits their more generalized use. This also is related to research and resource waste, with key contributors being avoidable research design limitations and inadequate reporting or dissemination [105,108]. With this in mind, individuals using these existing RoB tools may assume they have been demonstrated to be reliable and valid and are relevant to cross-sectional studies. Our rationale for focusing part of the scoping review on describing the particular biases that may impact cross-sectional research and then examining each tool for their coverage of each bias was to ensure these tools are indeed appropriate for their intended use, and to understand where future research in this area can avoid research waste and support future tool development.

Consultation with methodological experts and our review of methodological guidance specific to RoB in cross-sectional studies proved to be useful to the study team. This also facilitated our adaptation of the RoB domain framework [13] to highlight issues of greatest concern for cross-sectional studies and to differentiate between issues that may be more of a concern for studies with an aim to produce prevalence estimates vs. analytic/association approximations. The framework adapted for use in this scoping review [13] provides the domains that RoB tools should assess in cross-sectional studies: selection, exposure, outcome, confounding, missingness, selective reporting, and COI. These domains consider empirical and theoretical evidence about the aspects of observational study designs that can introduce bias into outcomes [13,55], and were adapted following a thorough review of RoB specific to cross-sectional studies.

Our aim was to provide an overview of coverage across these domains for each RoB tool applicable to cross-sectional studies. Readers can prioritize domains and items within domains that apply to their needs or content area and select a ROB tool, based on the framework. However, it is worth noting that there is currently no "gold" standard tool reflects the full breadth of the domains and items identified in the framework. As with tools applicable to RoB in other study designs, we found tools for cross-sectional studies contained items not addressing RoB, including reporting quality, applicability and methods for statistical analyses [109]. Often, it was unclear or ambiguous whether an item was addressing a RoB issue or reporting quality due to the phrasing or scoring/rating systems used. This was prominent in tools published before the year 2000 and there has since been some improvement. Inclusion of items not related to bias in a tool may distort the measurement properties and impact ratings. Taking the by-domain and by-item coverage assessments into account, selection of a tool for evaluating potential for RoB in cross-sectional studies should prioritize coverage for bias domains and items.

We found that no one tool covered all key bias domains and items for cross-sectional studies. In addition, our findings show that more guidance may be required to help users with their study assessments when using the available tools. Although tools in more recent, published records were described well and their scoring or ratings systems explained, many tools often had brief or no guidance for how to interpret the items appropriately. For tools applicable to more than one study design, the available guidance was often generally applicable to observational study designs but not specific to enable appropriate assessment of cross-sectional studies.

There are several limitations to our scoping review that are worth noting. The inherent limitation of scoping

Table 4. Tools for cross-sectional studies reporting risk of bias framework items (framework adapted from Wang et al (2019) [13]), $n = 64$ tools

BIAS domain	BIAS concept	Addressed in tool, *n* (%)	Not addressed in tool, *n* (%)	Unclear or partially addressed in tool[a], *n* (%)
Selection				
	Appropriateness of eligibility criteria	13 (38.2)	48 (59.4)	3 (4.7)
	Comparability of exposure group vs comparison group	16 (25.0)	47 (73.4)	1 (1.6)
	Nonresponse rate	30 (46.9)	33 (51.6)	1 (1.6)
	Nonresponse rate - mechanism	1 (1.6)	62 (96.9)	1 (1.6)
	Recruitment time frame	6 (9.4)	57 (89.1)	1 (1.6)
	Representativeness of sample to target population	38 (59.4)	24 (37.5)	2 (3.1)
Exposure				
	Validity and reliability of exposure measurement	34 (53.1)	18 (28.1)	12 (18.8)
Outcome				
	Blinding of the research staff	26 (40.6)	36 (56.3)	2 (3.1)
	Validity and reliability of the outcome measurement	41 (64.1)	15 (23.4)	8 (12.5)
Confounding				
	Accounting for confounding variables	34 (53.1)	27 (42.2)	3 (4.7)
	Description of confounding variables	23[b] (35.9)	37 (57.8)	4 (6.3)
Missingness				
	Amount of missingness	7 (10.9)	51 (79.7)	6 (9.4)
	Handling of missingness	5 (7.8)	56 (87.5)	3 (4.7)
	Mechanism for missingness	0 (0.0)	63 (98.4)	1 (1.6)
Selective reporting				
	Selective reporting of outcomes	4 (6.3)	60 (93.8)	0 (0.0)
Conflict of interest				
	Conflict of interest (e.g., funding)	8 (12.5)	56 (87.5)	0 (0.0)
Other bias[c]				
	Other bias or threats to internal validity	4 (6.3)	60 (93.8)	0 (0.0)

[a] Items were mapped as *unclear* if they may have covered the bias concept, but no additional documentation was available to clarify or confirm the wording or intent of the probe.

[b] SR-based tools did not report items for confounding variable description as these were named and considered in the SR itself.

[c] The 'other' category could result in discrepant or erroneous judgments of bias for a study [104]; therefore, authors felt it was important to document how many tools used this otherwise unspecified domain.

reviews is the focus on identifying the breadth of information on the topic at hand rather than depth [110]. As such, additional consideration for synthesis of the topics presented may be warranted, but we feel we have adequately covered the scope as intended. Although we conducted a comprehensive search, methodological records are poorly indexed in bibliographic databases and difficult to find in the gray literature [111]. There were 24 records for potential eligible records that could not be retrieved as the existing methodological systematic reviews did not fulsomely cite the tools or provided only an organization name which was not enough to locate the record. We limited the included records to those published in English and so our findings are generalizable to RoB tools written in English. Although no formal Delphi or consensus approach was used, the study working group met weekly over the course of the study to iteratively discuss and validate review methods, findings and interpretation.

We anticipate that our findings will be of interest to individuals or organizations conducting systematic reviews of cross-sectional studies, researchers who conduct surveys and descriptive research studies, journal editors and agencies or organizations who may use cross-sectional research data to inform policy or decision-making. We plan to use the results of this scoping review to improve content and education around RoB in cross-sectional studies and to inform the design and reporting of future cross-sectional studies.

5. Conclusions

This study provides a comprehensive summary of the unique, and more general, methodological considerations for RoB in cross-sectional study designs and maps the inclusion of each consideration to existing RoB tools. We identified RoB tools designed for broad applicability across various study designs as well as those specifically tailored for cross-sectional studies. However, none of the identified tools comprehensively address all potential bias domains pertinent to cross-sectional studies (ie, selection, exposure, outcome, confounding, missingness, selective reporting, and conflict of interest). Findings may be used to inform the selection and continued use of existing RoB tools and indicate a need for ongoing improvement. Tool developers may use results to consider development of context-specific or more precise RoB tools for population health estimates or variable associations.

CRediT authorship contribution statement

Shannon E. Kelly: Writing − review & editing, Writing − original draft, Validation, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Stephen P.J. Brooks:** Writing − review & editing, Validation, Methodology, Investigation, Conceptualization. **Karima Benkhedda:** Writing − review & editing, Methodology, Conceptualization. **Amanda J. MacFarlane:** Writing − review & editing, Supervision, Methodology, Conceptualization. **Linda S. Greene-Finestone:** Writing − review & editing, Validation, Methodology, Conceptualization. **Becky Skidmore:** Writing − review & editing, Methodology, Data curation. **Tammy J. Clifford:** Writing − review & editing, Supervision, Conceptualization. **George A. Wells:** Writing − review & editing, Validation, Supervision, Methodology, Conceptualization.

Data availability

Data will be made available on request.

Declaration of competing interest

There are no competing interests for any author.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2024.111408.

References

[1] Reeves BC, Wells GA, Waddington H. Quasi-experimental study designs series-paper 5: a checklist for classifying studies evaluating the effects on health interventions-a taxonomy without labels. J Clin Epidemiol 2017;89:30−42.
[2] Cumpston M, Li T, Page MJ, Chandler J, Welch VA, Higgins JP, et al. Updated guidance for trusted systematic reviews: a new edition of the Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Database Syst Rev 2019;10:ED000142.
[3] Kesmodel US. Cross-sectional studies−what are they good for? Acta Obstet Gynecol Scand 2018;97:388−93.
[4] Pandis N. Cross-sectional studies. Am J Orthod Dentofacial Orthop 2014;146:127−9.
[5] Wang X, Cheng Z. Cross-sectional studies: strengths, weaknesses, and recommendations. Chest 2020;158:S65−71.
[6] McNair P, Lewis G. Levels of evidence in medicine. Int J Sports Phys Ther 2012;7:474−81.
[7] Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8:19−32.
[8] Jordan Z, Lockwood C, Aromataris E, Pilla B, Porritt K, Klugar M, et al. JBI series paper 1: Introducing JBI and the JBI Model of EHBC. J Clin Epidemiol 2022;150:191−5.
[9] Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169:467−73.
[10] Kelly SE, Benkhedda K, Brooks SPJ, MacFarlane AJ, Greene-Finestone LS, Skidmore B, et al. Risk of bias in cross-sectional studies: protocol for a scoping review of concepts and tools. MethodsX 2024;12:102610.
[11] Stone JC, Barker TH, Aromataris E, Ritskes-Hoitinga M, Sears K, Klugar M, et al. From critical appraisal to risk of bias assessment: clarifying the terminology for study evaluation in JBI systematic reviews. JBI Evid Synth 2023;21:472−7.
[12] Barker TH, Stone JC, Sears K, Klugar M, Leonardi-Bee J, Tufanaru C, et al. Revising the JBI quantitative critical appraisal tools to improve their applicability: an overview of methods and the development process. JBI Evid Synth 2023;21:478−93.
[13] Wang Z, Taylor K, Allman-Farinelli M, Armstrong B, Askie L, Ghersi D, et al. A systematic review: Tools for assessing methodological quality of human observational studies 2019. Available at: https://osf.io/preprints/metaarxiv/pnqmy. Accessed July 2, 2024.
[14] Crowe Critical Appraisal Tool (v1.4). Available at: https://conchra.com.au/2015/12/08/crowe-critical-appraisal-tool-v1-4/. Accessed July 2, 2024.
[15] Office of Health Assessment and Translation. Risk of Bias Rating Tool for Human and Animal Studies. Research Triangle Park, NC: National Toxicology Program, US Department of Health and Human Services; 2015.
[16] Study Quality Assessment Tools: Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies. National Heart, Lung and Blood Institute. Bethesda, MD: National Institutes of Health; 2021.
[17] Evidence Worksheet and Quality Criteria Checklist - Primary Research. Online: Academy of Nutrition and Dietetics Evidence Analysis Library 2022. Available at: https://www.andeal.org/evidence-analysis-manual. Accessed July 2, 2024.
[18] Personal Communication from Timothy Hugh Barker. JBI Critical Appraisal Tool: JBI Checklist For Analytical Cross-Sectional Studies (2023 Update). Adelaide, SA: University of Adelaide; 2023.
[19] Personal Communication from Timothy Hugh Barker. JBI Critical Appraisal Tool: JBI Checklist For Prevalence Studies (2023 Update). Adelaide, SA: University of Adelaide; 2023.
[20] Evidence Evaluation Tools & Resources: LEGEND (Let Evidence Guide Every New Decision) for Cross-sectional studies of interventions. Cinncinati, OH, USA: James M Anderson Center for Health Systems Excellence, Cincinnati Children's Hospital Medical Center; 2023.
[21] Evidence Evaluation Tools & Resources: LEGEND (Let Evidence Guide Every New Decision) for Cross-sectional studies of etiology or risk factors. Cinncinati, OH, USA: James M Anderson Center for

Health Systems Excellence, Cincinnati Children's Hospital Medical Center; 2023.
[22] Evidence Evaluation Tools & Resources: LEGEND (Let Evidence Guide Every New Decision) for Cross-sectional studies of prevalence. Cinncinati, OH, USA: James M Anderson Center for Health Systems Excellence, Cincinnati Children's Hospital Medical Center; 2023.
[23] CAT Manager App for iPhone or Android. version not reported. Oxford, UK: Application Store (Free): Center for Evidence Based Management; 2023.
[24] (SURE) SURE. Questions to assist with the critical appraisal of cross-sectional studies. online. Specialist Unit for Review Evidence; 2018.
[25] Al Asmri M, Haque MS, Parle J. A modified medical education research study quality instrument (MMERSQI) developed by Delphi consensus. BMC Med Educ 2023;23:63.
[26] Al-Jader LN, Newcombe RG, Hayes S, Murray A, Layzell J, Harper PS. Developing a quality scoring system for epidemiological surveys of genetic disorders. Clin Genet 2002;62:230−4.
[27] Alshabanat A, Zafari Z, Albanyan O, Dairi M, FitzGerald J. Asthma and COPD overlap syndrome (ACOS): a systematic review and meta analysis. PLoS One 2015;10:e0136065.
[28] Asiamah N, Mends-Brew E, Boison BKT. A spotlight on cross-sectional research: addressing the issues of confounding and adjustment. Int J Healthc Manag 2021;14:183−96.
[29] Avis M. Reading research critically. II. An introduction to appraisal: assessing the evidence. J Clin Nurs 1994;3:271−7.
[30] Bawor M, Dennis BB, Anglin R, Steiner M, Thabane L, Samaan Z. Sex differences in outcomes of methadone maintenance treatment for opioid addiction: a systematic review protocol. Syst Rev 2014;3:1−7.
[31] Boyle M. Guidelines for evaluating prevalence studies. BMJ Mental Health 1998;1:37−9.
[32] Methods Commentary: Risk of Bias in Cross-Sectional Surveys of Attitudes and Practices. Online. DistillerSR and contributed by the CLARITY Group at McMaster University; 2011.
[33] Crowe M, Sheppard L. A general critical appraisal tool: an evaluation of construct validity. Int J Nurs Stud 2011;48:1505−16.
[34] Crowe M, Sheppard L, Campbell A. Comparison of the effects of using the Crowe Critical Appraisal Tool versus informal appraisal in assessing health research: a randomised trial. Int J Evid Based Healthc 2011;9:444−9.
[35] Crowe M, Sheppard L, Campbell A. Reliability analysis for a proposed critical appraisal tool demonstrated value for diverse research designs. J Clin Epidemiol 2012;65:375−83.
[36] de Groot R, van den Hurk K, Schoonmade LJ, de Kort WL, Brug J, Lakerveld J. Urban-rural differences in the association between blood lipids and characteristics of the built environment: a systematic review and meta-analysis. BMJ Glob Health 2019;4:e001017.
[37] Downes MJ, Brennan ML, Williams HC, Dean RS. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). BMJ Open 2016;6:e011458.
[38] Farrah K, Young K, Tunis MC, Zhao L. Risk of bias tools in systematic reviews of health interventions: an analysis of PROSPERO-registered protocols. Syst Rev 2019;8:280.
[39] Fowkes F, Fulton P. Critical appraisal of published research: introductory guidelines. BMJ Br Med J (Clin Res Ed) 1991;302:1136.
[40] Gauthier B. Assessing Survey Research A Principled Approach. Hull, Québec Canada: Citeseer; Circum Network inc; 2003.
[41] Genaidy A, Lemasters G, Lockey J, Succop P, Deddens J, Sobeih T, et al. An epidemiological appraisal instrument−a tool for evaluation of epidemiological studies. Ergonomics 2007;50:920−60.
[42] Goldberg RJ, McManus DD, Allison J. Greater knowledge and appreciation of commonly-used research study designs. Am J Med 2013;126:169.e1.
[43] Habibzadeh F, Habibzadeh P, Yadollahie M. The apparent prevalence, the true prevalence. Biochem Med 2022;32:020101.
[44] Harder T. Some notes on critical appraisal of prevalence studies: comment on:" the development of a critical appraisal tool for use in systematic reviews addressing questions of prevalence". Int J Health Pol Manag 2014;3:289.
[45] Herzog R, Álvarez-Pasquin M, Díaz C, Del Barrio JL, Estrada JM, Gil Á. Are healthcare workers' intentions to vaccinate related to their knowledge, beliefs and attitudes? A systematic review. BMC Public Health 2013;13:1−17.
[46] Hong QN, Gonzalez-Reyes A, Pluye P. Improving the usefulness of a tool for appraising the quality of qualitative, quantitative and mixed methods studies, the Mixed Methods Appraisal Tool (MMAT). J Eval Clin Pract 2018;24:459−67.
[47] Hoy D, Brooks P, Woolf A, Blyth F, March L, Bain C, et al. Assessing risk of bias in prevalence studies: modification of an existing tool and evidence of interrater agreement. J Clin Epidemiol 2012;65:934−9.
[48] Jackson R, Ameratunga S, Broad J, Connor J, Lethaby A, Robb G, et al. The GATE frame: critical appraisal with pictures. BMJ Evid Based Med 2006;11:35−8.
[49] Kennedy CE, Fonner VA, Armstrong KA, Denison JA, Yeh PT, O'Reilly KR, et al. The Evidence Project risk of bias tool: assessing study rigor for both randomized and non-randomized intervention studies. Syst Rev 2019;8:1−10.
[50] Kmet LM, Cook LS, Lee RC. Standard quality assessment criteria for evaluating primary research papers from a variety of fields. Alberta Heritage Foundation for Medical Research; 2004.
[51] Kunutsor SK, Apekey TA, Laukkanen JA. Association of serum total osteocalcin with type 2 diabetes and intermediate metabolic phenotypes: systematic review and meta-analysis of observational evidence. Eur J Epidemiol 2015;30:599−614.
[52] Law M, Stewart D, Pollock N, Letts L, Bosch J, Westmorland M. Guidelines for Critical Review Form—Quantitative Studies 1998. Hamilton, Ontario, Canada: McMaster University Occupational Therapy Evidence-based Practice Research Group; 2003.
[53] Loney PL, Chambers LW, Bennett KJ, Roberts JG, Stratford PW. Critical appraisal of the health research literature: prevalence or incidence of a health problem. Chronic Dis Can 1998;19:170−6.
[54] Lu G, Chen Q, Li Y, Liu Y, Zhang Y, Huang Y, et al. Status of antibiotic residues and detection techniques used in Chinese milk: a systematic review based on cross-sectional surveillance data. Food Res Int 2021;147:110450.
[55] Ma LL, Wang YY, Yang ZH, Huang D, Weng H, Zeng XT. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? Mil Med Res 2020;7:7.
[56] Macfarlane TV, Glenny AM, Worthington HV. Systematic review of population-based epidemiological studies of oro-facial pain. J Dent 2001;29:451−67.
[57] Maier C, Thatcher JB, Grover V, Dwivedi YK. Cross-sectional research: A critical perspective, use cases, and recommendations for IS research. Int J Inf Manag 2023;70. https://doi.org/10.1016/j.ijinfomgt.2023.102625.
[58] CEBM. Critical appraisal checklist for cross-sectional study. Centre for Evidence-Based Management; 2014. Available at: https://cebma.org/assets/Uploads/Critical-Appraisal-Questions-for-a-Cross-Sectional-Study-July-2014-1-v2.pdf. Accessed July 2, 2024.
[59] Mancuso CA, Charlson ME. Does recollection error threaten the validity of cross-sectional studies of effectiveness? Medical Care 1995AS77−88.
[60] Margetts B, Vorster H, Venter C. Evidence-based nutrition-review of nutritional epidemiological studies. S Afr J Clin Nutr 2002. Available at: http://www.sajcn.co.za/index.php/SAJCN/article/view/23. Accessed July 2, 2024.
[61] Migliavaca CB, Stein C, Colpani V, Munn Z, Falavigna M. Prevalence estimates reviews - systematic review methodology G. Quality assessment of prevalence studies: a systematic review. J Clin Epidemiol 2020;127:59−68.

[62] Modesti PA, Reboldi G, Cappuccio FP, Agyemang C, Remuzzi G, Rapi S, et al. Panethnic differences in blood Pressure in europe: a systematic review and meta-analysis. PLoS One 2016;11:e0147601.
[63] Moskalewicz A, Oremus M. No clear choice between Newcastle-Ottawa Scale and Appraisal Tool for Cross-Sectional Studies to assess methodological quality in cross-sectional studies of health-related quality of life and breast cancer. J Clin Epidemiol 2020; 120:94−103.
[64] Murray C, Marshall M, Rathod T, Bowen CJ, Menz HB, Roddy E. Population prevalence and distribution of ankle pain and symptomatic radiographic ankle osteoarthritis in community dwelling older adults: a systematic review and cross-sectional study. PLoS One 2018;13:e0193662.
[65] Myers EF, Parrott JS, Splett P, Chung M, Handu D. Using risk of bias domains to identify opportunities for improvement in food- and nutrition-related research: an evaluation of research type and design, year of publication, and source of funding. PLoS One 2018;13:e0197425.
[66] Nudelman G, Otto K. The development of a new generic risk-of-bias measure for systematic reviews of surveys. Methodology 2020;16:278−98.
[67] Ohadike CO, Cheikh-Ismail L, Ohuma EO, Giuliani F, Bishop D, Kac G, et al. Systematic review of the methodological quality of studies aimed at creating gestational weight gain charts. Adv Nutr 2016;7:313−22.
[68] Patra J, Bhatia M, Suraweera W, Morris SK, Patra C, Gupta PC, et al. Exposure to second-hand smoke and the risk of tuberculosis in children and adults: a systematic review and meta-analysis of 18 observational studies. PLoS Med 2015;12:e1001835. [discussion e].
[69] Perera S, Eisen R, Bawor M, Dennis B, de Souza R, Thabane L, et al. Association between body mass index and suicidal behaviors: a systematic review protocol. Syst Rev 2015;4:52.
[70] Powell DJ, Liossi C, Moss-Morris R, Schlotz W. Unstimulated cortisol secretory activity in everyday life and its relationship with fatigue and chronic fatigue syndrome: a systematic review and subset meta-analysis. Psychoneuroendocrinology 2013;38:2405−22.
[71] Protogerou C, Hagger M. A checklist to assess the quality of survey studies in psychology. Methods Psychol 2020;20:e100031.
[72] Puzzolo E, Stanistreet D, Pope D, Bruce N, Rehfuess E. Factors influencing the large-scale uptake by households of cleaner and more efficient household energy technologies. Evidence for Policy & Practice Information (EPPI) Centre at University College London; 2013.
[73] Rodriguez AJ, Nunes Vdos S, Mastronardi CA, Neeman T, Paz-Filho GJ. Association between circulating adipocytokine concentrations and microvascular complications in patients with type 2 diabetes mellitus: a systematic review and meta-analysis of controlled cross-sectional studies. J Diabet Complications 2016;30:357−67.
[74] Rojanaworarit C. Misleading epidemiological and statistical evidence in the presence of Simpson's paradox: an illustrative study using simulated scenarios of observational study designs. J Med Life 2020;13:37.
[75] Ross AB, Shertukde SP, Staffier KL, Chung M, Jacques PF, McKeown NM. The relationship between whole-grain intake and measures of cognitive decline, mood, and anxiety−a systematic review. Adv Nutr 2023;14:652−70.
[76] Sedgwick P. Bias in observational study designs: cross sectional studies. BMJ 2015;350:h1286.
[77] Sedgwick R, Epstein S, Dutta R, Ougrin D. Social media, internet use and suicide attempts in adolescents. Curr Opin Psychiatry 2019;32:534−41.
[78] Seehra J, Pandis N, Koletsi D, Fleming PS. Use of quality assessment tools in systematic reviews was varied and inconsistent. J Clin Epidemiol 2016;69:179−184 e5.
[79] Shamliyan T, Kane RL, Dickinson S. A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. J Clin Epidemiol 2010;63:1061−70.
[80] Shamliyan TA, Kane RL, Ansari MT, Raman G, Berkman ND, Grant M, et al. Development quality criteria to evaluate nontherapeutic studies of incidence, prevalence, or risk factors of chronic diseases: pilot study of new checklists. J Clin Epidemiol 2011;64: 637−57.
[81] Silva LC, Ordúñez P, Rodríguez MP, Robles S. A tool for assessing the usefulness of prevalence studies done for surveillance purposes: the example of hypertension. Rev Panam Salud Publica 2001;10: 152−60.
[82] Stone JC, Glass K, Clark J, Munn Z, Tugwell P, Doi SAR. A unified framework for bias assessment in clinical research. Int J Evid Based Healthc 2019;17:106−20.
[83] Takahashi N, Hashizume M. A systematic review of the influence of occupational organophosphate pesticides exposure on neurological impairment. BMJ Open 2014;4:e004798.
[84] Tonia T, Buitrago-Garcia D, Peter N, Mesa-Vieira C, Li T, Furukawa TA, et al. A tool to assess risk of bias in studies estimating the prevalence of mental health disorders (RoB-PrevMH). medRxiv 2023. https://doi.org/10.1101/2023.02.01.23285335.
[85] van der Windt DA, Thomas E, Pope DP, de Winter AF, Macfarlane GJ, Bouter LM, et al. Occupational risk factors for shoulder pain: a systematic review. Occup Environ Med 2000;57: 433−42.
[86] van Dijk GM, Maneva M, Colpani V, Dhana K, Muka T, Jaspers L, et al. The association between vasomotor symptoms and metabolic health in peri- and postmenopausal women: a systematic review. Maturitas 2015;80:140−7.
[87] Viswanathan M, Ansari MT, Berkman ND, Chang S, Hartling L, McPheeters M, et al. Assessing the risk of bias of individual studies in systematic reviews of health care interventions. Rockville (MD): Methods Guide for Effectiveness and Comparative Effectiveness Reviews; 2008.
[88] Viswanathan M, Berkman ND. Development of the RTI item bank on risk of bias and precision of observational studies. J Clin Epidemiol 2012;65:163−78.
[89] Viswanathan M, Berkman ND, Dryden DM, Hartling L. Assessing Risk of Bias and Confounding in Observational Studies of Interventions or Exposures: Further Development of the RTI Item Bank [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013. Available at: https://www.ncbi.nlm.nih.gov/books/NBK154461/. Accessed July 2, 2024.
[90] Wagg E, Blyth FM, Cumming RG, Khalatbari-Soltani S. Socioeconomic position and healthy ageing: a systematic review of cross-sectional and longitudinal studies. Ageing Res Rev 2021;69: 101365.
[91] Ward H. A Primer for Evaluating the Quality of Studies on Environmental Health: Critical Appraisal of Cross-Sectional Studies. Vancouver, BC, Canada: National Collaborating Centre for Environmental Health; 2011.
[92] Ximenes RA, Araujo TV. [Internal validity in cross-sectional studies: comments based on an investigation on the association between socioeconomic factors and schistosomiasis]. Cad Saúde Pública 1995;11:118−27.
[93] Yang JC, Yang ZR, Yu SQ, Zhan SY, Sun F. [Introduction on 'assessing the risk of bias of individual studies' in systematic review of health-care intervention programs revised by the Agency for Healthcare Research and Quality]. Zhonghua Liuxingbingxue Zazhi 2019;40:106−11.
[94] Yu IT, Tse SL. Workshop 6–sources of bias in cross-sectional studies; summary on sources of bias for different study designs. Hong Kong Med J 2012;18:226−7.
[95] Yusuf E, Kortekaas MC, Watt I, Huizinga TW, Kloppenburg M. Do knee abnormalities visualised on MRI explain knee pain in knee osteoarthritis? A systematic review. Ann Rheum Dis 2011;70:60−7.

[96] Zeng X, Zhang Y, Kwong JS, Zhang C, Li S, Sun F, et al. The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. J Evid Based Med 2015;8:2−10.
[97] Ross LE, Grigoriadis S, Mamisashvili L, Koren G, Steiner M, Dennis CL, et al. Quality assessment of observational studies in psychiatry: an example from perinatal psychiatric research. Int J Methods Psychiatr Res 2011;20:224−34.
[98] Perez-Rios M, Rey-Brandariz J, Galan I, Fernandez E, Montes A, Santiago-Perez MI, et al. Methodological guidelines for the estimation of attributable mortality using a prevalence-based method: the STREAMS-P tool. J Clin Epidemiol 2022;147:101−10.
[99] Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J. Methodological index for non-randomized studies (minors): development and validation of a new instrument. ANZ J Surg 2003;73:712−6.
[100] Ximenes RadA, Araújo TvBd. Validade interna em estudos de corte transversal: reflexões a partir de uma investigação sobre esquistossomose mansônica e condições socioeconômicas. Cad Saúde Pública 1995;11:118−27.
[101] Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. The Ottawa Hospital Research Institute; 2000.
[102] Kelly SE, Greene-Finestone LS, Yetley EA, Benkhedda K, Brooks SPJ, Wells GA, et al. NUQUEST-NUtrition QUality Evaluation Strengthening Tools: development of tools for the evaluation of risk of bias in nutrition studies. Am J Clin Nutr 2022;115:256−71.
[103] Statistics Canada. Statistics: Power from Data! 2021. Available at: https://www150.statcan.gc.ca/n1/edu/power-pouvoir/toc-tdm/5214718-eng.htm. Accessed January 2, 2024.
[104] Babic A, Pijuk A, Brázdilová L, Georgieva Y, Raposo Pereira MA, Poklepovic Pericic T, et al. The judgement of biases included in the category "other bias" in Cochrane systematic reviews of interventions: a systematic survey. BMC Med Res Methodol 2019;19:77.
[105] Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet 2009;374:86−9.
[106] Crowe M, Sheppard L. A review of critical appraisal tools show they lack rigor: alternative tool structure is proposed. J Clin Epidemiol 2011;64:79−89.
[107] Stewart GL, Barrick MR. The folly of using research lacking rigor as a call to action. Acad Manag Learn Educ 2012;11:278−84.
[108] Glasziou P, Chalmers I. Research waste is still a scandal—an essay by Paul Glasziou and Iain Chalmers. BMJ 2018;363:k4645.
[109] Wang Y, Ghadimi M, Wang Q, Hou L, Zeraatkar D, Iqbal A, et al. Instruments assessing risk of bias of randomized trials frequently included items that are not addressing risk of bias issues. J Clin Epidemiol 2022;152:218−25.
[110] Campbell F, Tricco AC, Munn Z, Pollock D, Saran A, Sutton A, et al. Mapping reviews, scoping reviews, and evidence and gap maps (EGMs): the same but different- the "Big Picture" review family. Syst Rev 2023;12:45.
[111] Hirt J, Ewald H, Briel M, Schandelmaier S. Searching a methods topic: practical challenges and implications for search design. J Clin Epidemiol 2024;166:111201.

NUQUEST

NUtrition QUality Evaluation Strengthening Tools

Risk of bias tools for nutrition studies

Shannon Kelly, MSc, PhD (c)
University of Ottawa, School of Epidemiology and Public Health

Amanda MacFarlane, PhD
Director, Agriculture, Food and Nutrition Evidence Center, Texas A&M Agrilife Research

With special thanks to the NUQUEST Working Group:
Linda Greene-Finestone, Karima Benkhedda, Stephen Brooks, Beth Yetley and George Wells

Conflict of interest

- No relevant declarations

Indigenous affirmation

We pay respect to the Algonquin people, who are the traditional guardians of this land.

We acknowledge their longstanding relationship with this territory, which remains unceded.

We pay respect to all Indigenous people in this region, from all nations across Canada, who call Ottawa home.

We acknowledge the traditional knowledge keepers, both young and old.

And we honour their courageous leaders: past, present, and future.

Learning Objectives

- Understand broadly the key risk of bias issues in nutrition studies
- Know the key features of the NUQUEST checklists
- Understand when and how to apply the NUQUEST checklists

What is NUQUEST?

NUtrition QUality Evaluation Strengthening Tools

- Tools to assess the risk of bias in nutrition studies

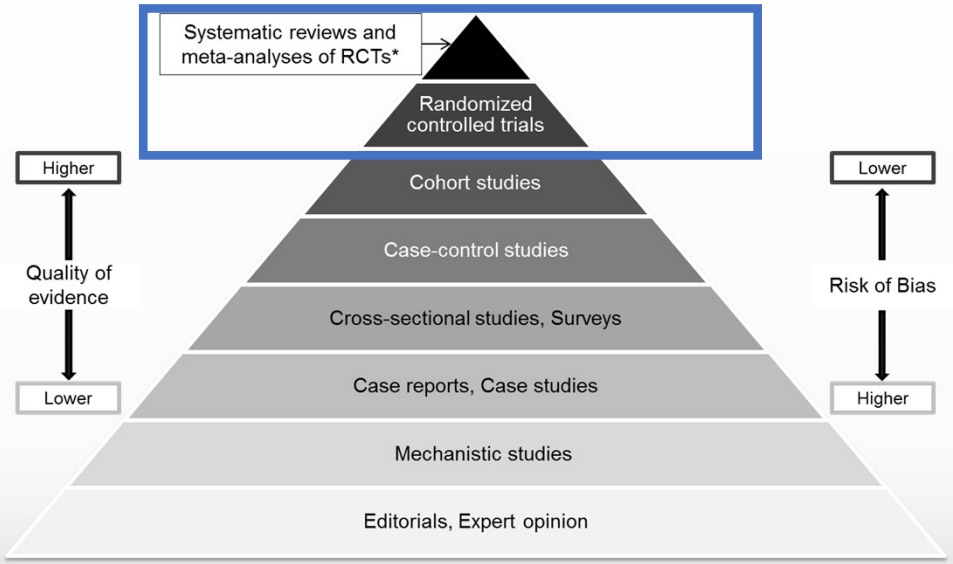
Key risk of bias issues

in nutrition studies

Nutrition studies make an impact!

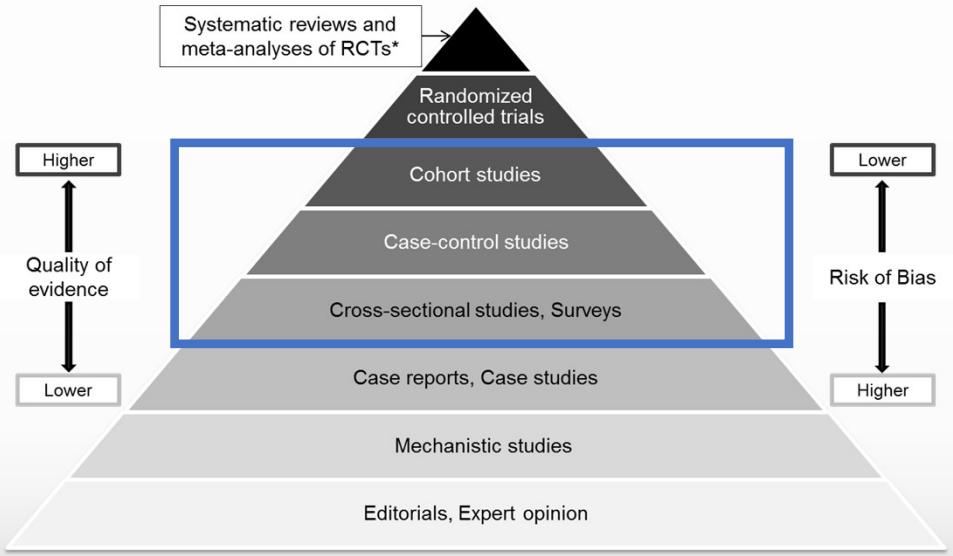
- Nutrition studies advance knowledge and form the basis of:
  - Dietary Reference Intakes (DRIs)
  - Clinical practice recommendations
  - Health policies
  - Guidelines
- High quality nutrition studies will:
  - Increase the degree of scientific rigour available for the development of lasting nutrition policies, standards and dietary guidance
  - Have greater impact

Causal relationship between nutrient intake and endpoint



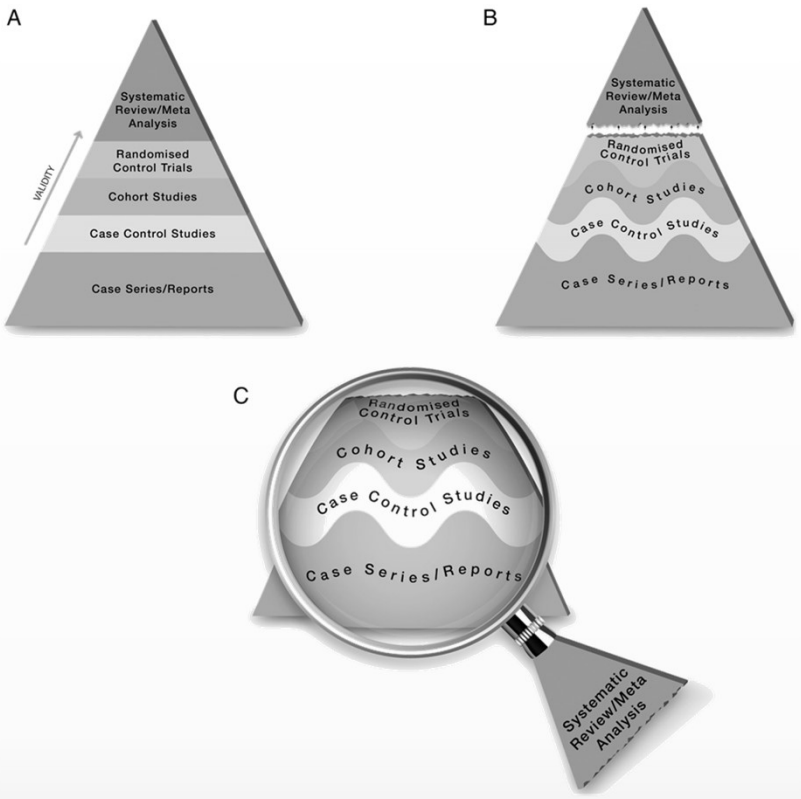
- Establishing causality and/or Intake-Response
  - Ideally RCTs and SR/MAs
  - ≥3 doses (for Intake-Response)

The nature of the evidence differs for chronic disease endpoints



- Nutrient-chronic disease evidence is mostly associational
  - How to establish causality and/or intake-response without RCTs
- Inherent systematic errors and bias with these study designs
  - Confounding and selection bias
  - Self-reported intake

9



Must examine each study for specific design features and conduct

A 'good' quality study at a lower level could yield better evidence than a 'poor' quality study at a higher level

What constitutes "better evidence" varies

M Hassan Murad et al. Evid Based Med 2016;21:125-127

Frequent Types of Human Nutrition Studies

Experimental studies:

- subjects allocated to different 'groups'
- collect and analyze data on the subjects
- examples:
  - Randomized controlled trial (RCT)
  - Non-randomized controlled trial

Observational studies:

- data on subjects collected and analyzed as they "naturally" divide themselves by variables into different 'groups'
- examples:
  - Cohort study (prospective, retrospective)
  - Case-control study (including nested case-control)
  - Cross-sectional study

They all have strengths and weaknesses

Why is risk of bias assessment important for nutrition studies?

- Quality assessment instruments (QAI) and Risk of Bias (RoB) tools
  - QAI assess the quality of a study from conception to interpretation as a whole
  - RoB tools assess the accuracy of estimates of benefit and risk
- Generic tools exist but:
  - May not address covariates, confounders and sources of error specific to nutrition studies
  - Often adapted or misapplied
- Nutritional intake assessment can add major uncertainty to judgements about
  - Intake/exposure
  - Intake–health relationships
- Observational studies are particularly vulnerable to random and non-random intake errors

Importance of intake/exposure assessment

- Accuracy of intakes impacts estimation of:
  - Intake/outcome dose-response relationships
  - Prevalence of population groups with adequate/inadequate intake
- Accuracy of rank-order intakes:
  - Important for assessing the existence, direction and strength of relationships between intakes and outcome

- **Different methods have different strengths and weaknesses, bigger or smaller impact on the Risk of Bias**
- **Impacts interpretation and application**
- **Can be nutrient-dependent**

13

Dietary intake assessment

- Self-reported
  - Diet record: Individuals may alter behavior
  - 24-hr recall: Under-reporting; Need multiple to estimate usual intakes
  - Food frequency: Under-reporting; Generally, greater bias than 24-hr recall
- Other considerations
  - Method used to validate instrument?
  - Validated for the population of interest?
  - Data useful for continuous or categorical comparisons?
  - Food composition database accurate?

Biomarkers of exposure may improve exposure estimates

Factors to consider

- Validated biomarker accurately reflects intakes
- Accuracy of assay methodology
  - Gold standard method; standard reference materials; QA/QC data
- Accurate across subgroups & time
  - e.g., similar validation results for population subgroups
- Confounders

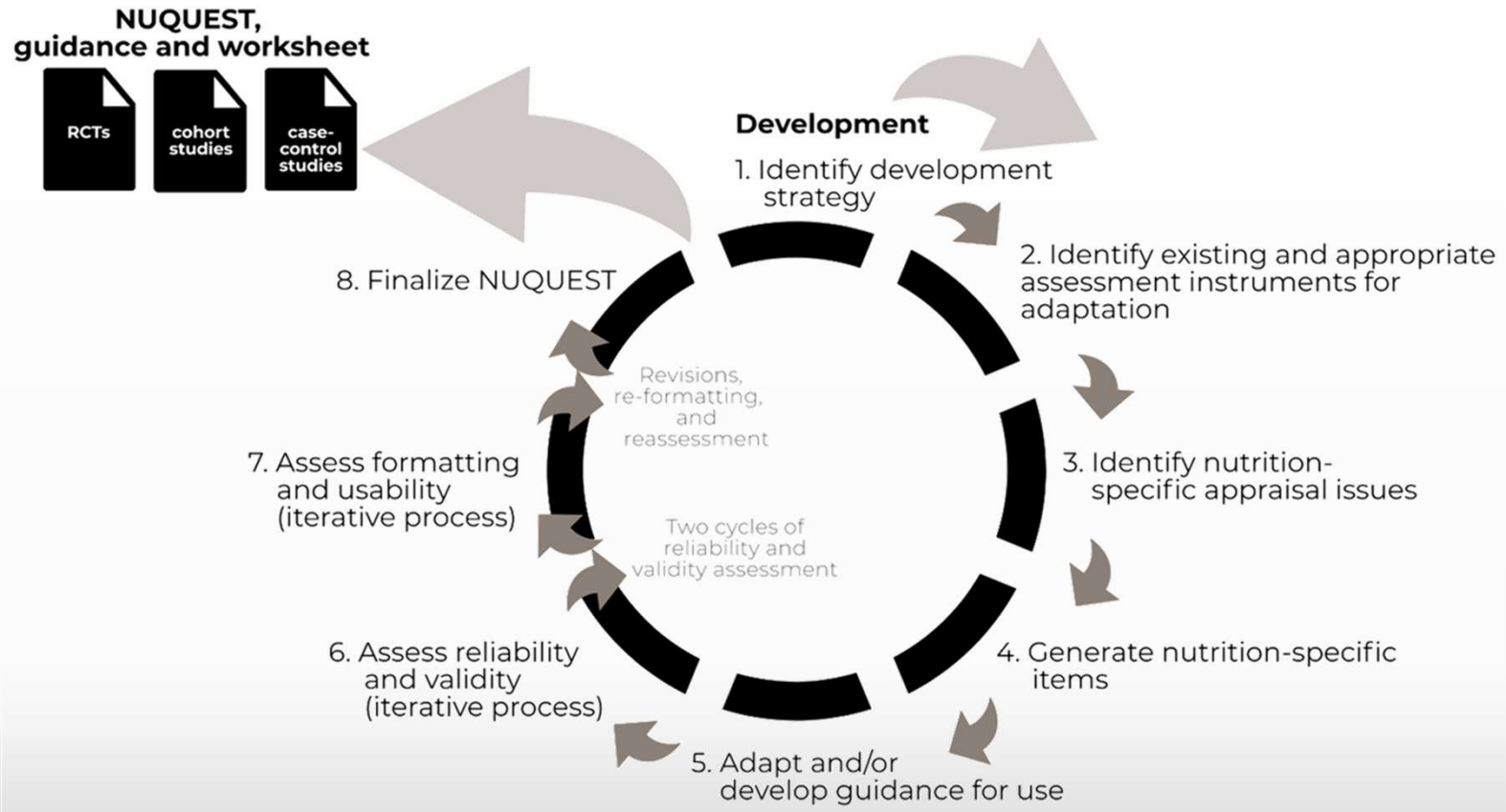
Factors for urinary biomarkers

- Completeness of collection (24 hr, overnight, spot)
- Recovery (% of total recovered)
- Multiple days
  - day-to-day variability within individuals

These nutrition-specific issues warrant special attention when assessing risk of bias!

NUQUEST

Development overview



Kelly, Greene-Finestone, Yetley, Benkhedda, Brooks, Wells, MacFarlane. NUQUEST-NUtrition QUality Evaluation Strengthening Tools: Development of tools for the evaluation of risk of bias in nutrition studies. Am J Clin Nutr. 2021 Oct 4:nqab335. doi: 10.1093/ajcn/nqab335.

Development strategy

Goals

1. Develop tools useful to:
   - epidemiological experts with limited knowledge of nutritional methodologies
   - nutritional scientists with limited knowledge of human study design and conduct issues

Approach

1. Adapt an existing generic assessment instrument
   - maintain scientific concepts/qualities of the tool
2. Add "bolt-on" nutrition-specific items to address concepts not addressed by generic items
3. Focus on common study designs employed in nutrition science
   - controlled trials
   - cohort studies
   - case-control studies.

Identifying an existing and appropriate tool for adaptation

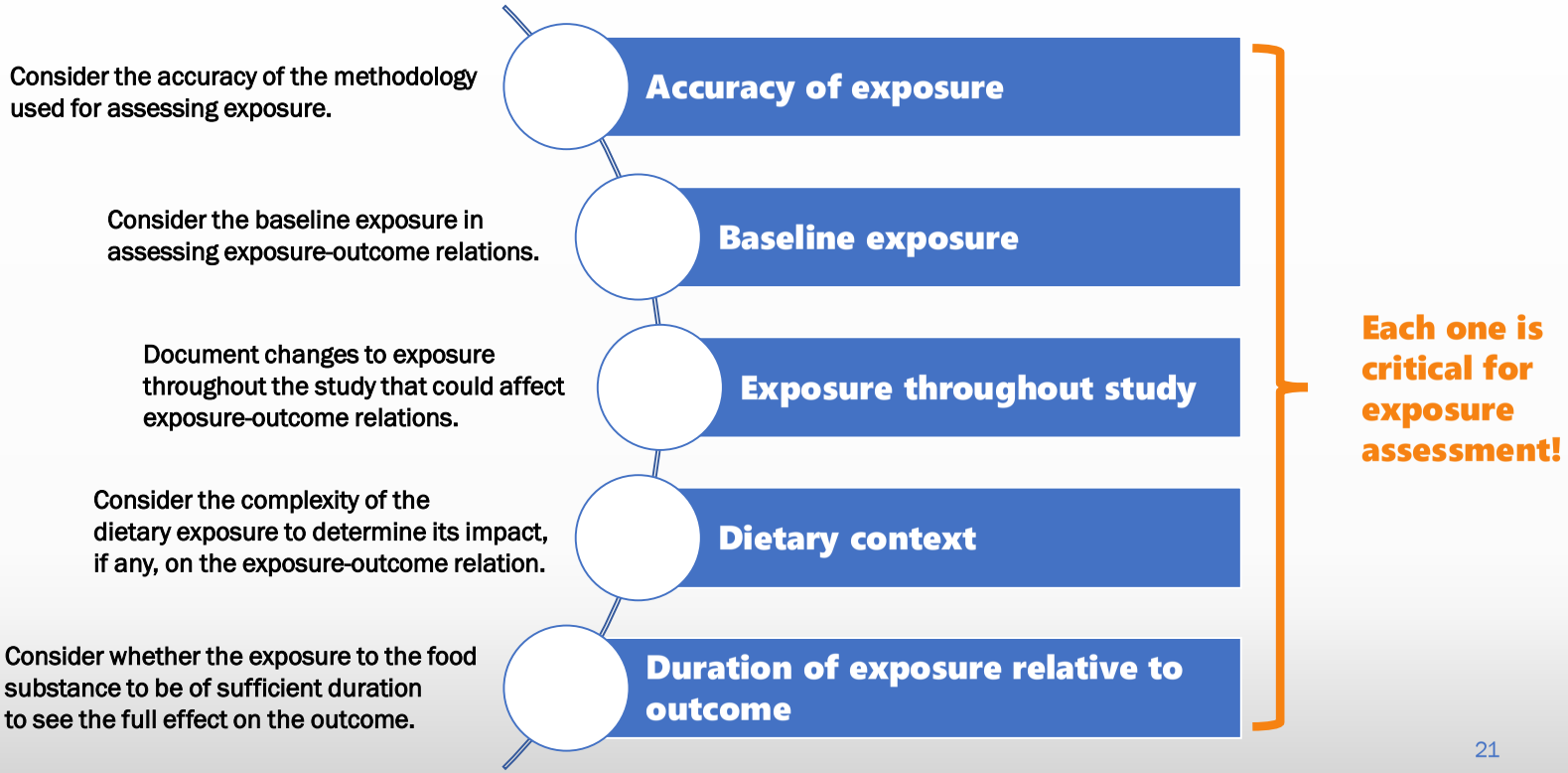
- Literature review to identify generic instruments

❖**SIGN Methodology checklists chosen for adaptation**

  - checklists tailored for assessment of multiple types of study designs (i.e., controlled trials, cohort studies, and case-control studies)
  - account for methodological considerations pertaining to key domains of interest i.e., participant selection, comparability of groups, and outcome and exposure ascertainment.
  - simple to use and clear guidance for their application and interpretation
  - used extensively

SIGN = Scottish Intercollegiate Guidelines Network

Identifying nutrition-specific appraisal issues



21

Nutrition-specific appraisal issues → Nutrition-specific items in NUQUEST

Study design	Nutrition-specific items
Randomized controlled trial	• The frequency and quantity of the intervention exposure under study are accurately and reliably measured. • The relevant baseline exposure is measured and taken into account. • The baseline exposures of all groups have been maintained over the course of the study. • Adherence to the intervention and control intakes is monitored in a reliable and accurate manner. • The interval between the intervention and outcome is of sufficient duration to observe an effect, if there is one.
Cohort study	• The frequency and quantity of the exposure under study are accurately and reliably measured. • The relevant exposure at baseline is measured and taken into account. • The baseline exposure differences between the groups have been maintained over the course of the study. • The interval between the exposure and outcome is of sufficient duration to observe an effect, if there is one.
Case-control study[1]	• The exposure under study is accurately and reliably measured. • The baseline exposure differences between the groups have been maintained over the course of the study. • The interval between the exposure and outcome is of sufficient duration to observe an effect, if there is one.

! **Nutrition-specific items were tailored for each study design**

Adaptation and development of guidance

- Detailed information, context to assist with assessment and ratings for each item

1. Guidance for existing generic items:
   - We added nutrition-specific examples and context
2. For nutrition-specific items:
   - We developed de novo guidance to accompany

! Each NUQUEST checklist has unique guidance

Highlights: reliability and validity

Study Design	No. of Studies	Degree of Agreement	Percentage in Agreement - NUQUEST Rating (%)				
			Selection	Comparability	Ascertainment	Nutrition	Overall
RCT	15	Near agreement	87	80	93	100	87
		Agreement	40	47	93	80	60
	10	Agreement without worksheet	40	40	80	100	40
	5	**Agreement with worksheet**	**40**	**60**	**80**	**80**	**100**
Cohort	15	Near agreement	100	80	80	93	93
		Agreement	67	47	33	87	53
	10	Agreement without worksheet	60	30	30	80	40
	5	**Agreement with worksheet**	**80**	**80**	**40**	**100**	**80**
Case-control	15	Near agreement	93	87	93	100	100
		Agreement	60	47	40	60	60
	10	Agreement without worksheet	70	20	40	40	40
	5	**Agreement with worksheet**	**40**	**100**	**40**	**100**	**100**

! Addition of a completed worksheet in second validation round improved inter-rater reliability

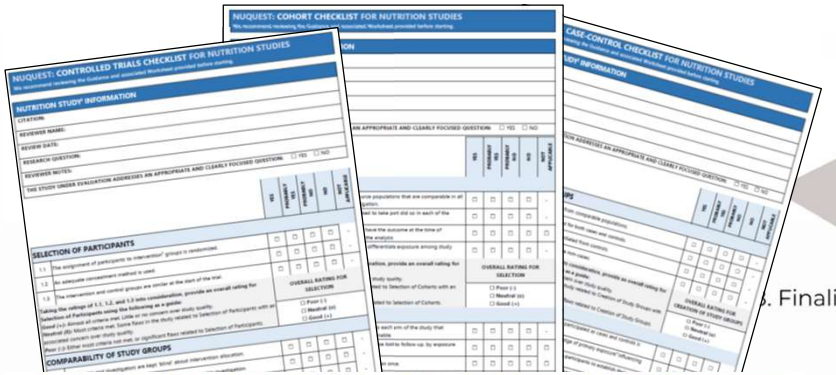
Highlights: reliability and validity con't

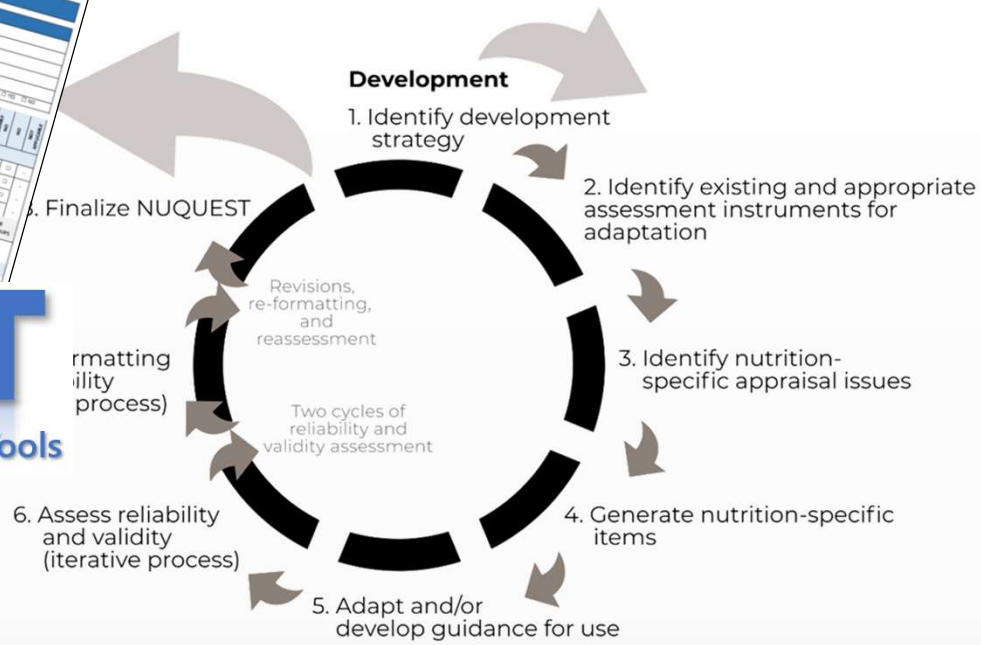
- NUQUEST ratings in agreement with previous Systematic Review rating (concurrent validity)
- Disagreements (up or down grading) driven by the nutrition-specific item rating (content validity)

STUDY ID	NUQUEST ASSESSMENT: Selection[8]	Comparability[9]	Ascertainment[10]	Nutrition[11]	Overall[12]	SR: Overall
16	●	●	●	●	●	●[4]
17	●	○	●	●	●	○[4]
18	●	○	●	●	○	○[4]
19	●	○	○	●	○	○[4]
20	○	○	●	○	○[2]	○[4]
21	●	○	●	●	●	●[4]
22	●	○	●	●	●	●[4]
23	●	●	●	●	●	●[4]
24	●	●	●	●	●	●[4]
25	●	●	●	●	●[2]	●[4]
26	●	○	●	●	●	○[4]
27	○	○	○	○	○[2]	○[4]
28	○	○	●	○	○[2]	○[4]
29	●	○	○	●	●	○[4]
30	○	○	○	○	○[2]	●[4]

Validity (example from cohort study validation)

Kelly, Greene-Finestone, Yetley, Benkhedda, Brooks, Wells, MacFarlane. NUQUEST-NUtrition QUality Evaluation Strengthening Tools: Development of tools for the evaluation of risk of bias in nutrition studies. Am J Clin Nutr. 2021 Oct 4:nqab335. doi: 10.1093/ajcn/nqab335.





Kelly, Greene-Finestone, Yetley, Benkhedda, Brooks, Wells, MacFarlane. NUQUEST-NUtrition QUality Evaluation Strengthening Tools: Development of tools for the evaluation of risk of bias in nutrition studies. Am J Clin Nutr. 2021 Oct 4:nqab335. doi: 10.1093/ajcn/nqab335.

Applying the NUQUEST checklists

When and how

A Tour of NUQUEST

1. NUQUEST checklists
   - Sections
   - Items
   - Response options
   - Section rating
   - Overall rating
2. NUQUEST Guidance
3. NUQUEST Worksheet

A Tour of NUQUEST: Sections

- **Selection, Comparability, Ascertainment of outcomes/exposure, Nutrition-specific**
- **Overall rating**

NUQUEST: RANDOMIZED CONTROLLED TRIALS CHECKLIST FOR NUTRITION STUDIES
We recommend reviewing the Guidance and the Worksheet provided before starting.

NUTRITION STUDY[1] INFORMATION

CITATION:
REVIEWER NAME:
REVIEW DATE:
RESEARCH QUESTION:
REVIEWER NOTES:
THE STUDY UNDER EVALUATION ADDRESSES AN APPROPRIATE AND CLEARLY FOCUSED QUESTION: ☐ YES ☐ NO

		YES	PROBABLY YES	PROBABLY NO	NO	NOT APPLICABLE
SELECTION OF PARTICIPANTS						
1.1	The assignment of participants to intervention[1] groups is randomized.	☐	☐	☐	☐	◦
1.2	An adequate concealment method is used.	☐	☐	☐	☐	◦
1.3	The intervention and control groups are similar at the start of the trial.	☐	☐	☐	☐	◦
Taking the ratings of 1.1, 1.2, and 1.3 into consideration, provide an overall rating for Selection of Participants using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Selection of Participants. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Selection of Participants. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Selection of Participants. High RoB.						**OVERALL RATING FOR SELECTION OF PARTICIPANTS** ☐ Good (+) ☐ Neutral (0) ☐ Poor (-)
COMPARABILITY OF STUDY GROUPS						
1.4	Participants and investigators are kept 'blind' about intervention allocation.	☐	☐	☐	☐	◦
1.5	The only difference between groups is the intervention under investigation.	☐	☐	☐	☐	◦
1.6	The percentage of the individuals or clusters recruited into each arm of the study that dropped out before the study was completed is reasonable.	☐	☐	☐	☐	◦
1.7	All the participants are analyzed in the groups to which they were randomly allocated (often referred to as intention to treat analysis).	☐	☐	☐	☐	☐
1.8	Where the study is carried out at more than one site, results are comparable for all sites.	☐	☐	☐	☐	☐
Taking the ratings of 1.4, 1.5, 1.6, 1.7, and 1.8 into consideration, provide an overall rating for Comparability of Groups using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Comparability of Study Groups. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Comparability of Study Groups. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Comparability of Study Groups. High RoB.						**OVERALL RATING FOR COMPARABILITY** ☐ Good (+) ☐ Neutral (0) ☐ Poor (-)

NUQUEST: RANDOMIZED CONTROLLED TRIALS CHECKLIST FOR NUTRITION STUDIES
We recommend reviewing the Guidance and the Worksheet provided before starting.

		YES	PROBABLY YES	PROBABLY NO	NO	NOT APPLICABLE
ASCERTAINMENT OF OUTCOMES						
1.9	All relevant outcomes are measured in a valid and reliable way.	☐	☐	☐	☐	◦
Taking the rating of 1.9 into consideration, provide an overall rating for Ascertainment of Outcomes using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Ascertainment of Outcomes. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Ascertainment of Outcomes. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Ascertainment of Outcomes. High RoB.						**OVERALL RATING FOR ASCERTAINMENT** ☐ Good (+) ☐ Neutral (0) ☐ Poor (-)
NUTRITION-SPECIFIC						
2.1	The frequency and quantity of the intervention under study are accurately and reliably measured.	☐	☐	☐	☐	◦
2.2	The relevant baseline exposure is measured and taken into account.	☐	☐	☐	☐	◦
2.3	The baseline exposures of all groups have been maintained over the course of the study.	☐	☐	☐	☐	◦
2.4	Adherence to the intervention and control intakes is monitored in a reliable and accurate manner.	☐	☐	☐	☐	◦
2.5	The interval between the intervention and outcome is of sufficient duration to observe an effect, if there is one.	☐	☐	☐	☐	◦
Taking the ratings of 2.1, 2.2, 2.3, 2.4 and 2.5 into consideration, provide an overall rating for Nutrition-Specific using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Nutrition-specific issues. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Nutrition-specific issues. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Nutrition-specific issues. High RoB.						**OVERALL RATING FOR NUTRITION** ☐ Good (+) ☐ Neutral (0) ☐ Poor (-)

OVERALL STUDY RATING

ASSESSMENT SUMMARY: Record the overall ratings from each sections above in the boxes below:

SELECTION OF PARTICIPANTS:	☐ Good (+) ☐ Neutral (0) ☐ Poor (-)	COMPARABILITY OF STUDY GROUPS:	☐ Good (+) ☐ Neutral (0) ☐ Poor (-)	ASCERTAINMENT OF OUTCOMES:	☐ Good (+) ☐ Neutral (0) ☐ Poor (-)	NUTRITION-SPECIFIC:	☐ Good (+) ☐ Neutral (0) ☐ Poor (-)

Taking the overall ratings for each section into consideration, provide an overall study rating: ***Good (+):*** Almost all criteria met. Little or no concern over study design. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern over study design. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to key aspects of the study design. High RoB.	**OVERALL STUDY RATING** ☐ GOOD (+) ☐ NEUTRAL (0) ☐ POOR (-)

A Tour of NUQUEST: Items

- **Structured individual risk of bias judgments within each section**

		YES	PROBABLY YES	PROBABLY NO	NO	NOT APPLICABLE
ASCERTAINMENT OF OUTCOMES						
1.9	All relevant outcomes are measured in a valid and reliable way.	☐	☐	☐	☐	-
Taking the rating of 1.9 into consideration, provide an overall rating for Ascertainment of Outcomes using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Ascertainment of Outcomes. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Ascertainment of Outcomes. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Ascertainment of Outcomes. High RoB.		**OVERALL RATING FOR ASCERTAINMENT** ☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**				
NUTRITION-SPECIFIC						
2.1	The frequency and quantity of the intervention under study are accurately and reliably measured.	☐	☐	☐	☐	-
2.2	The relevant baseline exposure is measured and taken into account.	☐	☐	☐	☐	-
2.3	The baseline exposures of all groups have been maintained over the course of the study.	☐	☐	☐	☐	-
2.4	Adherence to the intervention and control intakes is monitored in a reliable and accurate manner.	☐	☐	☐	☐	-
2.5	The interval between the intervention and outcome is of sufficient duration to observe an effect, if there is one.	☐	☐	☐	☐	-
Taking the ratings of 2.1, 2.2, 2.3, 2.4 and 2.5 into consideration, provide an overall rating for Nutrition-Specific using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Nutrition-specific issues. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Nutrition-specific issues. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Nutrition-specific issues. High RoB.		**OVERALL RATING FOR NUTRITION** ☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**				

A Tour of NUQUEST: Response options

- **Risk of bias rated for each item**
- **yes/probably yes/probably no/no** and for certain items **not applicable**

		YES	PROBABLY YES	PROBABLY NO	NO	NOT APPLICABLE
ASCERTAINMENT OF OUTCOMES						
1.9	All relevant outcomes are measured in a valid and reliable way.	☐	☑	☐	☐	-
Taking the rating of 1.9 into consideration, provide an overall rating for Ascertainment of Outcomes using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Ascertainment of Outcomes. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Ascertainment of Outcomes. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Ascertainment of Outcomes. High RoB.		**OVERALL RATING FOR ASCERTAINMENT** ☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**				
NUTRITION-SPECIFIC						
2.1	The frequency and quantity of the intervention under study are accurately and reliably measured.	☐	☐	☐	☐	-
2.2	The relevant baseline exposure is measured and taken into account.					

Yes	Probably Yes	Probably No	No	Not Applicable[a]
Based on something factual or known from evidence in the study publication.	Based on something the assessor judges likely to have occurred in the study.	Based on something the assessor judges likely to have not occurred in the study.	Based on something factual or known from evidence in the study publication.	This should be determined and recorded *a priori*. It would not be considered in the overall assessment.

a – the NOT APPLICABLE rating is only relevant for certain items in the checklist.

A Tour of NUQUEST: Section ratings

- **Good, neutral, poor**
- **Carried forward to the OVERALL RATING**

NUTRITION-SPECIFIC						
2.1	The frequency and quantity of the intervention under study are accurately and reliably measured.	☐	☑	☐	☐	-
2.2	The relevant baseline exposure is measured and taken into account.	☐	☐	☑	☐	-
2.3	The baseline exposures of all groups have been maintained over the course of the study.	☐	☐	☑	☐	-
2.4	Adherence to the intervention and control intakes is monitored in a reliable and accurate manner.	☐	☐	☐	☑	-
2.5	The interval between the intervention and outcome is of sufficient duration to observe an effect, if there is one.	☑	☐	☐	☐	-
Taking the ratings of 2.1, 2.2, 2.3, 2.4 and 2.5 into consideration, provide an overall rating for Nutrition-Specific using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Nutrition-specific issues. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Nutrition-specific issues. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Nutrition-specific issues. High RoB.		**OVERALL RATING FOR NUTRITION** ☐ **Good (+)** ☑ **Neutral (0)** ☐ **Poor (-)**			

A Tour of NUQUEST: Overall study rating

- **Good, neutral, poor**
- **Four section ratings are carried forward to the OVERALL RATING for the study**
- **Consider relative importance of specific items within the context of the study design**

OVERALL STUDY RATING							
ASSESSMENT SUMMARY: Record the overall ratings from each sections above in the boxes below:							
SELECTION OF PARTICIPANTS:	☑ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**	COMPARABILITY OF STUDY GROUPS:	☐ **Good (+)** ☐ **Neutral (0)** ☑ **Poor (-)**	ASCERTAINMENT OF OUTCOMES:	☐ **Good (+)** ☐ **Neutral (0)** ☑ **Poor (-)**	NUTRITION-SPECIFIC:	☐ **Good (+)** ☑ **Neutral (0)** ☐ **Poor (-)**
Taking the overall ratings for each section into consideration, provide an overall study rating: ***Good (+):*** Almost all criteria met. Little or no concern over study design. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern over study design. Moderate RoB ***Poor (-):*** Most or all criteria not met, or significant flaws related to key aspects of the study design. High RoB.					**OVERALL STUDY RATING** ☐ **GOOD (+)** ☐ **NEUTRAL (0)** ☑ **POOR (-)**		

A Tour of NUQUEST: RCT Items by section

SELECTION OF PARTICIPANTS

		YES	PROBABLY YES	PROBABLY NO	NO	NOT APPLICABLE
SELECTION OF PARTICIPANTS						
1.1	The assignment of participants to intervention[ii] groups is randomized.	☐	☐	☐	☐	-
1.2	An adequate concealment method is used.	☐	☐	☐	☐	-
1.3	The intervention and control groups are similar at the start of the trial.	☐	☐	☐	☐	-
Taking the ratings of 1.1, 1.2, and 1.3 into consideration, provide an overall rating for Selection of Participants using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Selection of Participants. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Selection of Participants. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Selection of Participants. High RoB.				☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**		

A Tour of NUQUEST: RCT Items by section

COMPARABILITY OF STUDY GROUPS

		YES	PROBABLY YES	PROBABLY NO	NO	NOT APPLICABLE
COMPARABILITY OF STUDY GROUPS						
1.4	Participants and investigators are kept 'blind' about intervention allocation.	☐	☐	☐	☐	-
1.5	The only difference between groups is the intervention under investigation.	☐	☐	☐	☐	-
1.6	The percentage of the individuals or clusters recruited into each arm of the study that dropped out before the study was completed is reasonable.	☐	☐	☐	☐	-
1.7	All the participants are analyzed in the groups to which they were randomly allocated (often referred to as intention to treat analysis).	☐	☐	☐	☐	☐
1.8	Where the study is carried out at more than one site, results are comparable for all sites.	☐	☐	☐	☐	☐
Taking the ratings of 1.4, 1.5, 1.6, 1.7, and 1.8 into consideration, provide an overall rating for Comparability of Groups using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Comparability of Study Groups. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Comparability of Study Groups. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Comparability of Study Groups. High RoB.		**OVERALL RATING FOR COMPARABILITY** ☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**				

A Tour of NUQUEST: RCT Items by section

OUTCOME ASCERTAINMENT

		YES	PROBABLY YES	PROBABLY NO	NO	NOT APPLICABLE
ASCERTAINMENT OF OUTCOMES						
1.9	All relevant outcomes are measured in a valid and reliable way.	☐	☐	☐	☐	-

Taking the rating of 1.9 into consideration, provide an overall rating for Ascertainment of Outcomes using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Ascertainment of Outcomes. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Ascertainment of Outcomes. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Ascertainment of Outcomes. High RoB.	**OVERALL RATING FOR ASCERTAINMENT** ☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**

A Tour of NUQUEST: RCT Items by section

NUTRITION-SPECIFIC

		YES	PROBABLY YES	PROBABLY NO	NO	NOT APPLICABLE
NUTRITION-SPECIFIC						
2.1	The frequency and quantity of the intervention under study are accurately and reliably measured.	☐	☐	☐	☐	-
2.2	The relevant baseline exposure is measured and taken into account.	☐	☐	☐	☐	-
2.3	The baseline exposures of all groups have been maintained over the course of the study.	☐	☐	☐	☐	-
2.4	Adherence to the intervention and control intakes is monitored in a reliable and accurate manner.	☐	☐	☐	☐	-
2.5	The interval between the intervention and outcome is of sufficient duration to observe an effect, if there is one.	☐	☐	☐	☐	-
Taking the ratings of 2.1, 2.2, 2.3, 2.4 and 2.5 into consideration, provide an overall rating for Nutrition-Specific using the following as a guide: ***Good (+):*** Almost all criteria met. Little or no concern related to Nutrition-specific issues. Low RoB. ***Neutral (0):*** Most criteria met. Some flaws in the study with an associated concern related to Nutrition-specific issues. Moderate RoB. ***Poor (-):*** Most or all criteria not met, or significant flaws related to Nutrition-specific issues. High RoB.		**OVERALL RATING FOR NUTRITION** ☐ **Good (+)** ☐ **Neutral (0)** ☐ **Poor (-)**				

A Tour of NUQUEST: Guidance

- **Unique guidance for each checklist**
- **Before you start, instructions, item- and section-specific guidance**

SELECTION OF PARTICIPANTS	
1.1	**The assignment of participants to intervention or control groups is randomized.**
	Random allocation of participants by a valid randomization method to receive one or other of the interventions under investigation, or to receive either intervention or placebo, is fundamental to this type of study.
1.2	**An adequate concealment method is used.**
	Allocation concealment refers to the process used to ensure that researchers are unaware which group participants are being allocated to at the time they enter the study. Research has shown that where allocation concealment is inadequate, investigators can significantly overestimate the effect of interventions (e.g., by up to 40%).
1.3	**The intervention and control groups are similar at the start of the trial.**
	Participants selected for inclusion in a trial must be as similar as possible. The study should report any significant differences in the composition of the study groups including factors such as gender, age, social background, ethnic origin, stage of disease (if appropriate), clinical biomarkers, and co-morbid conditions. Baseline intakes and nutritional status should also be described (see 2.2). These factors may be covered by inclusion and exclusion criteria. **Failure to address this item could lead to the study being downgraded.** Reviewers should identify *a priori* the most important factors for evaluation of the study. If a baseline difference exists, the variable should be considered as a covariate in any of the planned analyses. **If this has not been done, then the study could be downgraded.** **If several relevant variables at baseline are different, then a randomization failure may be an issue and the study could be downgraded**. Relevant variables should be identified a *priori* and used as a guide for your assessment of baseline differences and/or the likelihood of randomization failure.

2.3	**The baseline exposures of all groups have been maintained over the course of the study.**
	All participants have baseline intakes of food substances, foods or dietary patterns of interest. Changes in the baseline intake over the duration of the study for any or all of the groups could make it more difficult to detect differences in outcome due to the effect of the intervention. Has the potential for this been monitored? If a change in baseline intake has occurred, does the study have appropriate methods to deal with it? **If it is possible/likely that this has occurred, and has not been addressed, then the study could be downgraded.** A higher rating would be given to studies where there are multiple measures of intake over time and especially where this is confirmed/assessed with biomarker data. A higher rating would also be given if baseline intake changes over time have been addressed in the analysis and/or interpretation. Section 2.1 offers guidance on how to assess the RoB of the method to monitor intake/status or supplement consumption.

NUQUEST: CHECKLIST GUIDANCE FOR RANDOMIZED CONTROLLED NUTRITION TRIALS

1. BEFORE YOU START

Prior to completing the NUQUEST randomized controlled trial checklist, ensure that:

(1). The study under evaluation is a randomized controlled trial.
If in doubt, make sure you have the correct checklist for your study design.

(2). The study is relevant to the key research question of the review.
Inclusion or exclusion of individual studies into an evaluation will depend on the key question(s) and the *a priori* criteria, including the **PICO** (Participants/Population, Intervention/Exposure, Comparator, and Outcome(s)) components that are guiding the evaluation. Decisions should be documented and in place prior to beginning the checklist. Documenting definitions of exclusion/inclusion criteria prior to initiation of the evaluation is important for establishing the applicability of the evaluation. The worksheet can be used to document these criteria.

(3). The worksheet can be used to identify important study criteria and evaluate the risk of bias associated with them. The worksheet can be referred to, as appropriate, to facilitate the study assessment.
One purpose of an assessment tool is to identify and take into account the potential for risk of bias (RoB).The worksheet can be used to identify the importance of each component of the study (intervention, confounding factors, co-intervention factors) for the study assessment.

(4). General risk of bias sources are considered.
While the type of bias is not specified within the NUQUEST checklist for each item, items in each section identify potential issues associated with facets of a study that are related to different sources of bias.

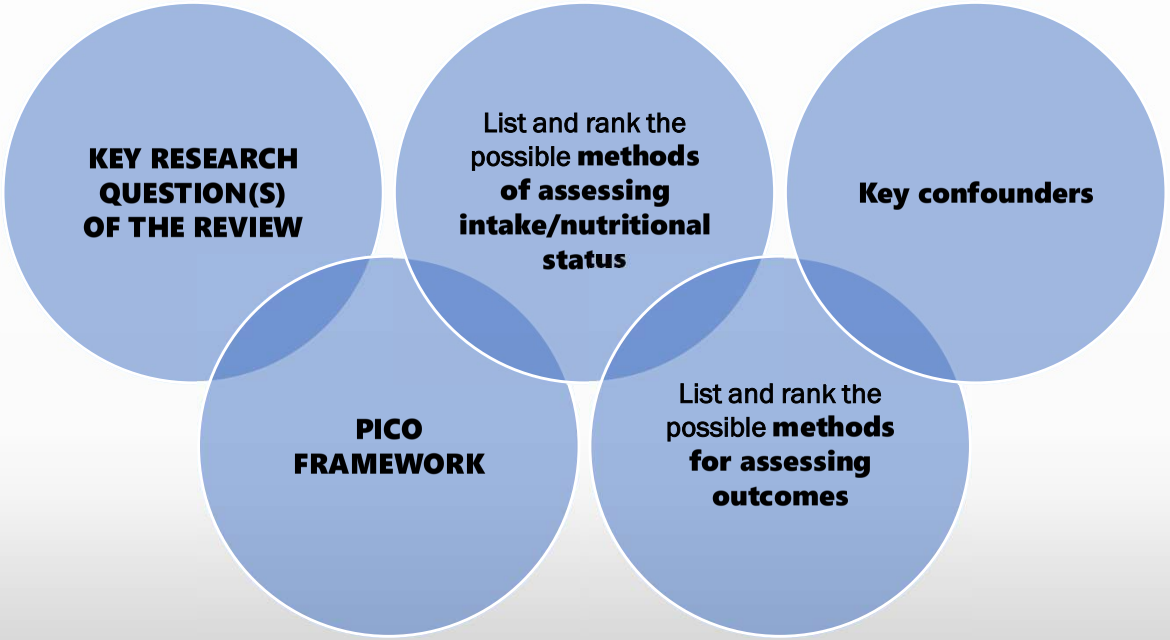
Below are examples of important general bias domains that may be commonly observed in nutrition randomized controlled trials. Note that this is not an exhaustive list of potential bias domains and others may apply to your specific study assessment (see Resources following guidance).

Performance bias	Systematic differences between groups in the care provided or exposure to factors other than the interventions of interest.
Attrition bias	Systematic differences between groups in withdrawals from the study.
Detection bias	Systematic differences between groups in how exposure/status and outcomes are determined.
Selection bias	Systematic differences between groups on baseline characteristics.
Dietary exposure assessment bias	Error associated with the use of methodologies for assessing dietary intakes. A subcategory for self-reporting methodologies is **recall bias**, which refers to systematic error due to differences in completeness or accuracy of recall. Self-reported dietary intakes are at risk of this bias.
Misclassification bias	Systematic error due to inaccurate measurements or classifications of participants' exposure or outcome; error may be related to the risk of outcome. If the error is unrelated to the risk of outcome, the effect is usually biased to the null.

38

A Tour of NUQUEST: Worksheet

- **Developed to accompany the NUQUEST checklists and guidance**
- Provide a mechanism for discussing methodological issues before the formal rating process begins
- Equalize knowledge-base across reviewers
- Focus reviewers on the range of bias for a particular study facet
- Supported by guidance



39

When to use or apply NUQUEST

- We know what nutrition-specific issues are important, and;
- We now have tools to assess risk of bias in nutrition studies.

394

387

Rating risk of bias in individual studies	Rating risk of bias in studies included in a systematic review	Study planning and design
• Nutrition risk assessments for approval of foods	• Guideline and policy development/ food health claims substantiation	• IDs key aspects that should be considered when planning a study

Using NUQUEST for rating risk of bias in a study

Planning

- Fill in the worksheet template (ideally)
- Check the NUQUEST "Before you start" guidance

Risk of bias ratings by item and section

- Rate items in each section
- Complete section ratings

Overall risk of bias rating

- Carry section ratings forward to the overall rating section
- Complete the overall rating for the study

Consult the NUQUEST guidance as needed

What's next for NUQUEST?

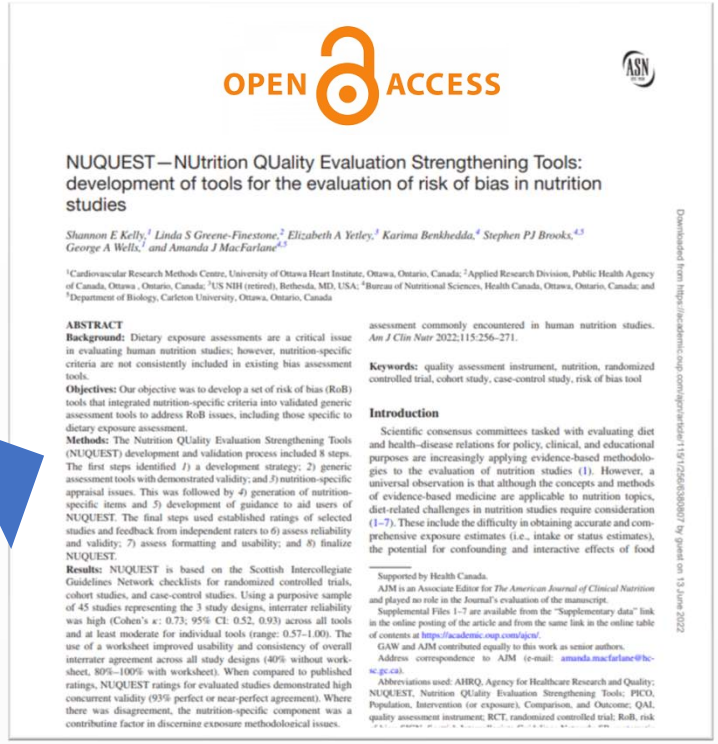
Next up! NUQUEST for cross-sectional studies under development

Register here: https://cns-scn.ca/events-activities/nuquest-virtual-workshop-series/nuquest-workshop-series

TAKEAWAY MESSAGES

- We developed NUQUEST for assessing risk of bias in the results of RCTs, cohort studies and case-control studies
- NUQUEST can be applied to individual studies and in the context of systematic reviews for a variety of purposes.
- NUQUEST checklists are accompanied by detailed guidance and a worksheet to facilitate risk of bias ratings

The American Journal of
CLINICAL NUTRITION
A Publication of the American Society for Nutrition
JANUARY 2022 • VOLUME 115 • NUMBER 1 http://ajcn.nutrition.org/ ISSN 0002-9165



OPEN ACCESS

NUQUEST—NUtrition QUality Evaluation Strengthening Tools: development of tools for the evaluation of risk of bias in nutrition studies

Shannon E Kelly,[1] Linda S Greene-Finestone,[2] Elizabeth A Yetley,[3] Karima Benkhedda,[4] Stephen PJ Brooks,[4,5] George A Wells,[1] and Amanda J MacFarlane[4,5]

[1]Cardiovascular Research Methods Centre, University of Ottawa Heart Institute, Ottawa, Ontario, Canada; [2]Applied Research Division, Public Health Agency of Canada, Ottawa , Ontario, Canada; [3]US NIH (retired), Bethesda, MD, USA; [4]Bureau of Nutritional Sciences, Health Canada, Ottawa, Ontario, Canada; and [5]Department of Biology, Carleton University, Ottawa, Ontario, Canada

ABSTRACT

Background: Dietary exposure assessments are a critical issue in evaluating human nutrition studies; however, nutrition-specific criteria are not consistently included in existing bias assessment tools.

Objectives: Our objective was to develop a set of risk of bias (RoB) tools that integrated nutrition-specific criteria into validated generic assessment tools to address RoB issues, including those specific to dietary exposure assessment.

Methods: The Nutrition QUality Evaluation Strengthening Tools (NUQUEST) development and validation process included 8 steps. The first steps identified *1*) a development strategy; *2*) generic assessment tools with demonstrated validity; and *3*) nutrition-specific appraisal issues. This was followed by *4*) generation of nutrition-specific items and *5*) development of guidance to aid users of NUQUEST. The final steps used established ratings of selected studies and feedback from independent raters to *6*) assess reliability and validity; *7*) assess formatting and usability; and *8*) finalize NUQUEST.

Results: NUQUEST is based on the Scottish Intercollegiate Guidelines Network checklists for randomized controlled trials, cohort studies, and case-control studies. Using a purposive sample of 45 studies representing the 3 study designs, interrater reliability was high (Cohen's κ: 0.73; 95% CI: 0.52, 0.93) across all tools and at least moderate for individual tools (range: 0.57–1.00). The use of a worksheet improved usability and consistency of overall interrater agreement across all study designs (40% without worksheet, 80%–100% with worksheet). When compared to published ratings, NUQUEST ratings for evaluated studies demonstrated high concurrent validity (93% perfect or near-perfect agreement). Where there was disagreement, the nutrition-specific component was a contributing factor in discernine exposure methodoloeical issues.

assessment commonly encountered in human nutrition studies. *Am J Clin Nutr* 2022;115:256–271.

Keywords: quality assessment instrument, nutrition, randomized controlled trial, cohort study, case-control study, risk of bias tool

Introduction

Scientific consensus committees tasked with evaluating diet and health–disease relations for policy, clinical, and educational purposes are increasingly applying evidence-based methodologies to the evaluation of nutrition studies (1). However, a universal observation is that although the concepts and methods of evidence-based medicine are applicable to nutrition topics, diet-related challenges in nutrition studies require consideration (1–7). These include the difficulty in obtaining accurate and comprehensive exposure estimates (i.e., intake or status estimates), the potential for confounding and interactive effects of food

Supported by Health Canada.

AJM is an Associate Editor for *The American Journal of Clinical Nutrition* and played no role in the Journal's evaluation of the manuscript.

Supplemental Files 1–7 are available from the "Supplementary data" link in the online posting of the article and from the same link in the online table of contents at https://academic.oup.com/ajcn/.

GAW and AJM contributed equally to this work as senior authors.

Address correspondence to AJM (e-mail: amanda.macfarlane@hc-sc.gc.ca).

Abbreviations used: AHRQ, Agency for Healthcare Research and Quality; NUQUEST, Nutrition QUality Evaluation Strengthening Tools; PICO, Population, Intervention (or exposure), Comparison, and Outcome; QAI, quality assessment instrument; RCT, randomized controlled trial; RoB, risk

https://doi.org/10.1093/ajcn/nqab335

All checklists, guidance and the worksheet included with the publication!

Funding

- The development of NUQUEST was supported by Health Canada A-Base funding and in-kind contributions from the Public Health Agency of Canada and University of Ottawa
- Shannon Kelly is supported through funding from the University of Ottawa (Excellence Scholarship) and the Province of Ontario (Ontario Graduate Sch398olarship)

Health Canada Santé Canada

Ontario

uOttawa

Public Health Agency of Canada Agence de la santé publique du Canada

Acknowledgments

- SIGN for their permission to adapt SIGN checklists in the development of the NUQUEST checklists
- Raters for validation: Joan Peterson, Dr. Zemin Bai, Amy Johnston, Dr. Shu-Ching Hsieh, Dr. Matthew Parrot, Alexandre Martel, Dr. Jesse Elliott and Cynthia Boudrault
- Drs. Stephanie Atkinson, Alice Lichtenstein and Maria Makrides for insight on the nutrition-specific items and guidance, and/or individual study quality

NUQUEST

NUtrition QUality Evaluation Strengthening Tools

A Tour of NUQUEST: Worksheet con't

Example: Relationship between sodium intakes and blood pressure.

Key Research Question	Among all adults, what is the relationship (benefits and harms) between dietary sodium intake and blood pressure?

Example: Relationship between sodium intakes and blood pressure.

Participants/Population	All adults excluding persons with end stage renal disease, heart failure, HIV, or cancer.
Intervention/Exposure	Interventions to reduce dietary sodium (e.g., lower sodium diet or salt substitute).
Comparator	Higher dietary sodium intakes (e.g., usual diet) vs. lower sodium intakes (e.g., dietary salt reductions or use of salt substitutes).
Outcome(s)	Blood pressure (e.g., systolic and/or diastolic blood pressure, percent participants meeting a blood pressure goal, or changes in blood pressure).

Example: Usual Sodium Intake.

Research objectives	Method of assessment	Risk of Bias[1]
Determining difference between 2 points in time or between 2 groups	One or more 24 hr urine collections with quality control on whole sample at each point in time	Low
	One or more 24 hr urine collections without quality control on whole sample at each point in time	Moderate
	One or more 24 hr diet recalls on whole sample at each point in time	Moderate
	FFQ on whole sample at each point in time.	High
Determining change between 2 points in time or between 2 groups over time, in proportion of individuals above/below some threshold	Multiple 24 hr urine collections with quality control on whole sample at each point in time	Low
	Multiple 24 hr urine collections without quality control on whole sample at each point in time	Moderate
	Multiple 24 hr diet recalls on whole sample at each point in time	Moderate
	Single 24 hr recall on whole sample plus repeat on large subsample	Moderate
	FFQ or single 24 hr recall without repeat subsample	High
Examining the association between diet as an independent variable and a dependent variable (outcome)	Multiple 24 hr urine collections with quality control on whole sample at each point in time	Low
	Multiple 24 hr diet recalls on whole sample at each point in time	Moderate
	Single 24 hr recall on whole sample plus repeat on large subsample	Moderate

Example: Blood pressure measurements.

Research objectives	Method of assessment	Risk of Bias[1]
Systolic and diastolic blood pressure for classification of individuals	Multiple day, 3-repeat measure by trained personnel in clinic setting according to accepted and referenced guidelines.	Low
	Multiple day, 3-repeat measure at home with a validated automated sphygmomanometer with accepted guidelines (4).	Moderate
	Single measure at home with an automated sphygmomanometer without training or guidelines (5).	High

Example: Confounders related to the relationship between sodium intakes and blood pressure.

Key confounders	Method of assessment	Risk of Bias[1]
Sex	Self-reported	Low
Age	Birth certificate (documented)	Low
Use of blood pressure medication	Self-reported	Moderate
Changes in sodium intakes unrelated to the intervention	Dietary assessment method conducted more than once during study (see Table 1 on intake/exposure methods)	Low to High
Pre-existing medical conditions (e.g., hypertension, conditions that affect the relationship between intakes and outcome)	Self-reported	Moderate
Potassium intake	Dietary assessment method conducted more than once during study (see Table 1 on intake/exposure methods)	Low to High
Physical activity	Self-reported	High
Physical activity	Activity monitor	Moderate