uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Sarah Kathleen Connor Gorber
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Population Health)
GRADE / DEGREE

Department of Population Health
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

The relationship between subjective and objective measures of health: a series of systematic reviews and analyses of data from national surveys

TITRE DE LA THÈSE / TITLE OF THESIS

Ian McDowell
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Christine Dallaire

Mark Tremblay

Beth Potter

Paul Veugelers

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

The relationship between subjective and objective measures of health: a series of systematic reviews and analyses of data from national surveys

Sarah Connor Gorber

Thesis by articles submitted to the
Faculty of Graduate and Postdoctoral Studies

In partial fulfillment of the requirements
For the PhD degree in Population Health

Population Health
Faculty of Graduate and Postdoctoral Studies
University of Ottawa

Your file Votre référence
ISBN: 978-0-494-51809-0
Our file Notre référence
ISBN: 978-0-494-51809-0

Acknowledgements

To Benson and Allison Boyd, who deserve this degree as much as I do.
Thank you for your patience, support and endless love.

This dissertation has truly been a group effort; I consider myself very blessed to have had so much support from so many people.

I would like to thank the following people:

Benson and Allison Boyd, for so unselfishly giving me the time I needed for my schooling

Mom and Dad, for their encouragement, unconditional love and never ending support

My friends and family, who each helped in their own special ways

Ian, for encouraging me to think outside my comfort zone

Mark, for being a mentor and good friend

And finally, all those who encouraged me to pursue my studies and who were with us on this journey, you know who you are, thank you for believing in me.

Abstract

The goal of this research was to consolidate and expand on current knowledge about the use of objective and subjective measures to assess the health of Canadians. Specifically, this dissertation examined the relationship between self-report and direct measures of health to determine the extent of discrepancies between the two methods. This was accomplished through a series of five research papers, including three systematic reviews that examined the relationship between self-reported and directly measured estimates of height, weight and Body Mass Index (BMI), hypertension and smoking. An in-depth examination of the discrepancies in reported obesity followed. It included two additional papers, one examining the feasibility of establishing correction factors to increase the accuracy of reported estimates of obesity and a second examining the trends in the reporting bias over time and between two different population groups (Canada and the U.S.).

Results clearly documented a bias in self-reported estimates of height, weight, BMI, hypertension and smoking; reported estimates of height were consistently overestimated while reported weight, BMI, hypertension and smoking were consistently underestimated in both men and women. Statistical adjustment of reported values generated estimates of height, weight and obesity that were more accurate than those generated by self-report alone. Historical analysis with Canadian and American datasets indicated that the bias in reported obesity is increasing, implying that the generalizabilty of the correction equations may be limited and that in Canada, monitoring trends in obesity based solely on self-reported height and weight may produce inaccurate estimates because of the diverging discrepancy between self-reported and measured data. This work has made an important contribution to

understanding the bias in current surveillance practices that are grounded in the collection of self-reported data and presents a series of recommendations to improve population health surveillance in Canada.

Table of Contents

Introduction

Measuring Population Health

Measuring the state of health within a population forms a crucial starting point for population health research. It provides a current picture of a population's status, allows for monitoring changes over time and indicates inequities between population sub groups and among countries. Adequate measurement strategies are essential to ensure that evidence upon which resources will be allocated and interventions designed is reliable and valid.

There are different techniques for measuring health and these may provide discrepant results. The population health challenge posed by excess body weight, for instance, can be estimated based on indices such as Body Mass Index (BMI), or measures such as the waist to hip ratio or waist circumference, and each may provide a different assessment of the prevalence of weight related health risks in the population. Subjective (often self-reported) and objective (measured) indicators of health are often drawn into this debate because of their tendency to provide differing estimates of the same health outcomes (e.g., measured and reported obesity). For instance, using BMI as a summary indicator, the prevalence of obesity in Canada has been estimated to be 15% based on self-reported responses, but this rises to 23% when BMI is based on directly measured values[1].

Because population surveys are one of the most common methods used to collect data about a population's health and traditionally surveys have mostly used self-reported data, most of our health information is based on subjective measures of health. In Canada there have been sporadic initiatives to collect direct physical measures (as in the Nutrition Canada and

Canada Health Surveys in the 1970s, the Heart Health surveys in the 1980s, and more recently the Canadian Study of Health and Aging[2]) but it has been more than 30 years since a national direct measures survey has been undertaken. Recently there has been a push to use objective measures to collect health data based on the belief that direct measurement of core health indicators such as height, weight or disease status will lead to more accurate prevalence estimates. This, in turn, could lead to more appropriate interventions and resource allocation.[i]

Objective measures are favoured by many researches because the data they generate are considered to be more reliable and valid than subjective alternatives[3], and in many cases they are a more appropriate tool. When people are not aware of their health status (because diseases are not yet diagnosed or patients are not showing symptoms), when there is no other way to measure the outcome (e.g., levels of environmental toxins in the body) or when personal or psychological factors may influence the accuracy or truthfulness of a self-reported response, objective measures are preferred.

There are, however, situations where subjective measures provide insight that could not be gained from objective measurement alone. The classic self-reported health question "how would you rate your health compared to others your age?" provides compelling evidence that there are aspects of health into which we have insight that can only be captured by self-

[i] For the purpose of this dissertation, objective measures are defined as those that require some type of physical assessment or examination to determine the condition of a person's health or functioning as compared to subjective measures that focus on a person's interpretation of a state or event. While it is acknowledged that there may be an element of subjectivity inherent even in objective measures (e.g., in interpreting the measured data) it was necessary to adopt an operational definition of the term for this research. The terms 'direct', 'physical' and 'objective' measures are used interchangeably.

report. This single question, albeit with different wording and response categories in different studies, has been shown to be reliable in predicting mortality and morbidity[4] and applicable in different ethnic groups,[5] as well as in developed and less developed countries[6].

This dissertation examines the relationship between self-report and direct measures of health, to determine the extent of discrepancies between the two methods. It focuses on health indicators for which both objective and subjective measures exist, to provide insight into the degree of correspondence between the two and to analyze the circumstances in which the correspondence may be greatest. Three indicators are examined in this thesis: body height, weight and the resulting BMI; hypertension; and smoking status. These are all examples where the ideal or gold standard is the objective measure, but where subjective measures are most often employed as a convenient proxy indicator in population health research. This work is designed to explore whether systematic biases exist in self-reported measures of these health indicators and to assess their extent, to determine whether we can correct for such biases, and whether the extent of bias has changed over time.

Objectives

The purpose of this research is to consolidate and expand our knowledge about the use of objective and subjective measures to assess the health of Canadians. First, a systematic review of the literature is completed to summarize current knowledge regarding the agreement between selected subjective and objective measures of health. Height, weight and BMI, hypertension and smoking status are chosen as examples. A systematic review is an effective way to summarize the evidence in a manner that minimizes bias without having to

(re)examine existing data. In order to ensure that researchers are employing the most appropriate measures in their research it is important to determine if certain approaches to measurement consistently under- or over-estimate results, which can be done through a comprehensive examination of past work.

Next, the research focuses on gaining insight into the nature of these discrepancies and their implications for health research and surveillance using empirical analyses of relevant North American data. Insights from the systematic reviews assisted in generating hypotheses about the factors that may influence agreement between the objective and subjective measures, which were tested in the empirical analyses in Part I of the thesis. Although there is currently no comprehensive physical measures data set in Canada that would allow an exploration of a variety of health indicators, the Canadian Community Health Survey (CCHS)[7] has collected both measured and reported values for height, weight and BMI. Part II of the study therefore focused on obesity, examining whether the discrepancies between reported and measured values could be statistically corrected. Because measured and reported values of obesity also exist in historical surveys, this work also examined whether the error in reporting obesity has varied over time, perhaps reflecting changing awareness of, and attitudes towards, obesity.

As insight is gained into reporting discrepancies relating to obesity this insight can be tested against other indicators as new surveys such as Statistics Canada's Canadian Health Measures Survey (CHMS) become available[8]. The CHMS collects self-reported measures of health as well as physical measures such as anthropometry, cardio- respiratory fitness,

physical activity and blood and urine measurements and data for this survey will be available beginning in 2010.

Research Questions

This dissertation is made up of a series of 5 research papers that are designed to answer three main research questions:

1. What range of agreement exists between subjective and objective measures of selected health outcomes? Three systematic reviews were conducted to compare direct measures versus self-report measures for assessing indicators of i) anthropometry (body composition - height, weight and BMI), ii) hypertension and iii) smoking in observational and experimental studies of adult populations.

The indicators were chosen to represent a possible range in the degree of correspondence expected between the measured and reported values, with a greater correspondence expected for indicators of smoking, a moderate correspondence for obesity measures and a low correspondence for hypertension. This hypothesized range of correspondence reflects the level of complexity involved in assessing each of the outcomes. People are aware, for instance, of their smoking status, whereas awareness of hypertension is dependent on having your blood pressure taken, having it interpreted by a health professional according to a set cut-off value (which can vary over time) and then receiving and understanding this information as it is fed back to you. This research question will be answered by the systematic reviews.

2. Assuming that disagreement exists between the measured and reported values, it is important to examine certain key indicators (in this case obesity) in more detail to determine the extent of disagreement using recent Canadian datasets and the predictors associated with the discrepancies. The second question has two parts:

    a. How does the prevalence of obesity differ when assessed with self-report compared to direct measures?

    b. Can correction factors be developed to adjust for the discrepancies when direct measurement is not feasible?

This research question will be answered via empirical analysis of the Canadian Community Health Survey data.

3. The third set of questions is answered through an empirical analysis of historical data sets. Understanding whether the discrepancy has changed as social acceptance of obesity has evolved may provide insight into the factors that influence the discrepancy and provide further insight for assessing trends.

    a. Has the discrepancy between reported and measured obesity increased or decreased over time?

    b. Does the discrepancy follow the same pattern in Canada and the United States?

This research question is justified because the burden of obesity on the health care system is increasingly being recognized; the health costs of obesity represent 2.2% of total health care

costs[9]. It has been suggested that a change in awareness and attention to issues such as obesity could affect the way individuals respond to self-report questions[3]. This analysis will allow a determination of whether reporting error has increased as society has grown more aware of obesity and its consequences and felt more pressure to present themselves in a socially desirable manner or whether the increase in awareness had led to more accurate reporting.

The dissertation begins by outlining the theoretical framework for the research and presenting a detailed review of the literature to summarize the current state of knowledge about the relationship between self-reported and directly measured indicators of health. This is followed by the presentation of each of the 5 research papers, as they were published (or submitted, in the case of the final paper) in national and international health journals. The final chapter includes a general discussion and summary of the conclusions and the implications of this work for future health surveillance in Canada.

Theoretical Framework

There is a long standing discussion among researchers about which methods are the most appropriate to study and explain human action. This discussion is based in the broader philosophy of truth and how we view the nature of reality[10]. Objectivism and subjectivism are two opposing approaches to establishing truth. Objectivism is most closely tied to the tradition of positivism. Objectivists maintain that attributes *can* be measured, that there is agreement about what *should* be measured, and that these measurements are repeatable. The process of measurement under this paradigm should not affect the attribute under study, so

results are independent of the influence of the researcher. Within this approach gold standards exist and the results of any measurement can be compared to these standards. Subjectivism, on the other hand, holds that the ultimate purpose of measurement is to record actual experience and perception rather than empirically verifiable data. It suggests that there are differing ideas about which attributes of a system should be measured, that there are no gold standards and no single truth[11,12]. Consider measures of social support as an example. Which is more important in predicting health: the number of friends a person has or the subjective feeling she has of being supported? Because health (and life itself) is not a purely mechanical series of chemical and physical reactions, it is possible that objective measures may be accurate in measuring current health status whereas future prognosis may be better captured by a more subjective measure. Certainly, subjective measures seem most appropriate for studying indicators that are themselves more subjective such as quality of life or well being.

The theoretical perspective subsequently influences the type of data collection and measurement that is undertaken[11]. Quantitative methods are generally favoured by objectivists and are popular in the physical and social sciences, while subjectivists tend to prefer qualitative methods. The debate between proponents of reported or measured values of health outcomes (discussed below) is influenced by this theoretical discourse.

Population Health, as a transdisciplinary field, is guided by the theories that underpin the disciplines that are engaged in its study[13] and thus incorporates elements of both objectivism and subjectivism in determining reality and in explaining health. Such an approach is

appropriate as there is increasing agreement that it is unlikely that any one method is entirely sufficient[10]. The reality, therefore, is that truth likely lies somewhere between the two perspectives, with each being able to contribute depending on what question is being posed and the purpose of the investigation. For example, accurate measures of obesity are important for diagnosing health risks and estimating treatment needs associated with being overweight so objective measures would be favoured in these circumstances. Subjective measures of body size or obesity may be more appropriate if the purpose is to gain information about an individual's psychological well being and their risk of eating disorders or issues associated with self-worth or self-esteem. While not denying the relevance of both perspectives, this thesis is set within an objectivist paradigm, and will treat measured values of variables as the reality to which self-reported measures should ideally correspond.

When discrepancies between the measures occur, it is important to understand the underlying contributing factors and whether discrepancies are due to random error or a more predictable response pattern influenced by personality dispositions, individual characteristics (e.g., general health, self-confidence) or factors in the broader social and cultural context. Social desirability, for example, is known to affect responses to subjective questions but there may be factors (e.g., gender, social roles) that can influence an individual or group's likelihood of responding to questions in a socially acceptable manner. A systematic study of these aspects could lead to the development of correction factors to adjust for their effects.

Theories of social desirability and social approval[14,15] provide insight into why indicators such as height and weight are susceptible to inaccurate reporting. Self-reported weight may

be influenced by cultural norms that emphasize thinness. In an attempt to present themselves in a more positive light, for instance, women have been shown to underestimate their weight[16]. Larson[17] studied the relationship between social desirability and error in reporting weight and found that women with higher tendencies to act in a socially desirable manner were more likely to underreport their weight, while those who had lower scores on the social desirability scale were less likely to underreport their weight. Their study concluded that younger female populations may be more weight sensitive than other groups and thus more prone to erroneous self-report. Social desirability has also been suggested as a potential explanation for the "flat slope syndrome" – the tendency for individuals who are at the extremes of the weight ranges to provide estimates that are more in line with their socially desirable rather than their true weights[17]. With height, it has been suggested that discrepancies in self-reported measurements may simply be due to error because height is not measured as frequently as weight[18], and is less under the control of the individual. Errors in reporting may be more common in older populations because of shrinkage that is associated with the aging process[19]. With hypertension it may be more an issue of awareness than biased reporting.

As part of this research, it is anticipated that the required blend between the two perspectives can be established by determining the optimal balance between self-report and direct measures for selected indicators in health survey research. It is likely that there will be situations where there is high correspondence between the two measures, and in such cases, choice of the appropriate alternative should be based on factors such as cost, feasibility and respondent burden. In cases where there is little correspondence, insight can be gained from

studying the determinants of discordance and examining whether theories of personality and social desirability can help to explain these differences. Understanding these determinants can help to inform decisions about which method is preferable in which circumstances.

Review of the Literature

Self-report and Direct Measures

Self-report measures (replies to questions of interest that are posed through personal or group interviews or questionnaires[20]) are one of the most common ways to collect social and health data. Self-report measures of health have been consistently shown to be accurate in predicting mortality and morbidity[21-23]. Idler and Benyamini[24] reviewed 27 studies that examined the relationship between self-reported health and future mortality and found that in nearly all studies self-reported health was an independent predictor of mortality even after controlling for other factors that are known to predict mortality.

Self-reports have often been employed because of their practicality, efficiency, low cost and because they are the only way to measure certain indicators such as pain or emotions. They are quick and easy to administer and are a good way to sample large numbers of individuals. However, the tools are subject to certain limitations: questions may be misunderstood, participants may not accurately recall past events, may not have knowledge of the health condition under study, and response bias e.g., social desirability or response acquiescence is not uncommon[20]. Memory error can depend on the demand of the recall task or the frequency and salience of an event[25] with events that are more salient and less frequent being more accurately recalled[26]. The length of the recall period may also affect memory recall[27].

Social desirability is particularly problematic with sensitive questions (e.g., about weight or obesity) and can be influenced by the method of data collection[28-30]. For example, in the CCHS, one of Canada's largest national health surveys, the proportion of respondents reporting being obese (based on self-reported values of height and weight) was estimated to be 18% for those who were interviewed in person compared to an estimate of only 13% for those who were asked the same questions by phone[31].

If, for operational and budgetary reasons, survey modes do not stay consistent over time the comparability of estimates can be compromised. Similarly, language and translation issues, changes in question wording or respondent characteristics among surveys or within a survey over time, also have the potential to damage the comparability and quality of self-reported estimates. In addition, in order to be implemented on surveys or questionnaires, the questions need to be standardized, which limits their ability to pick up subtle differences among respondents, and may cause problems if the standardardized wording is not consistently interpreted by all respondents.

Self-report measures, then, are not appropriate for all purposes and all types of data collection, and objective or direct measures have been recommended to improve measurement precision. With objective measures a direct assessment is provided for the indicator in question. Examples of direct measures include recording movement patterns with motion sensing devices such as accelerometers or pedometers, measuring disease status with blood or urine testing (to diagnose diabetes or kidney disease) or measuring

physiological changes such as blood pressure and heart rate. These measures are considered to be more robust, reliable and valid than their self-reported counterparts and can be used to measure conditions when self-report is not possible (e.g., undiagnosed cases). Analysis of the National Health and Nutrition Examination Survey (NHANES) in the United States for instance, found that although the prevalence of diabetes in the population was 7.8%, approximately 2.7% of that was undiagnosed[32]. The prevalence of diabetes in Canada (based on self-reported data) is also likely underestimated as a result of undiagnosed cases[33]. A related study by Diaz and colleagues[34] found the prevalence of undiagnosed diabetes, hypertension and hypercholesterolemia to be 11%, 16%, and 28% respectively. Undiagnosed cases are an important source of error in self-reported estimates because measures of population health often sample the general population and even if patients themselves may not be aware of their conditions, this information is routinely used to generate prevalence estimates for the conditions in question, thereby leading to an underestimate.

Despite its advantages, the costs of direct measurement are often prohibitively high, particularly in population research where the samples are large. The intrusive nature of the measures also has the potential to decrease response rates, thereby reducing external validity. As well, such measures often require special training of interviewers for accurate test administration and classification.

Discrepancy Between Measures

Many studies have examined the relationship, predictive ability and accuracy of self-report and direct measures but their results have been conflicting. Studies of variables as diverse as

sleep, back pain, memory, social and functional status and tobacco use have been conducted to compare the discrepancy between reported and measured values with some finding strong correlations between the objective and subjective instruments and others finding no relationship[35-40]. In a study comparing subjective and objective measures of energy expenditure Wareham and colleagues,[41] for instance, found that a simple self-report index was valid and reliable compared to an objective assessment. By contrast, Hill and Davies,[42] in a smaller study of female athletes, found that most women tended to underreport their energy intake. The issue, therefore, is not simply whether self-report and objective measures provide different results but also whether the discrepancies are random or more systematically associated with specific variables or the socio-demographic characteristics of the respondents.

Examining the determinants of error in reporting obesity suggests that the differences may be systematically related to respondent characteristics rather than randomly distributed in the population. For instance, men who are shorter or older[43-44] are more likely to overestimate their height. In women some studies found that as age increased the error in estimates of height increased[18,19,45] while others found that error decreased with age[45-46]. Lower socio-economic status has also been related to increased reporting error[43,47].

Heavier individuals are more likely to underestimate their weight[44,48] (possibly due to regression toward socially acceptable values) and estimates of weight have also been shown to vary among ethnic groups[49]. In women the effects of education show contradictory results; higher education is associated with less error in some studies[50,51] and more in

others[18,52,53]. Error has also been found to be increased in women who are less active, are preoccupied with their weight or live in rural areas[48].

Although many studies have examined the factors associated with the error in obesity, the results among studies are not consistent and when this dissertation began, no national Canadian data had been analysed to determine the predictors of error in this population. This is partially because the precise relationship between objective and subjective measures has been difficult to ascertain because it requires simultaneous collection of both types of measures on the same population at the national level. The CCHS collected information on reported and measured obesity but for other health indicators (diabetes, infectious disease) no such data source exists in Canada.

Compounding the problem is the fact that the influence of social desirability on self-reports has the potential to change over time as social and cultural norms about health conditions such as weight and obesity change. To date and to the author's knowledge no study has examined the relationship between reported and measured values over time.

References

1. Tjepkema M. Adult obesity. *Health Rep* 2006;17(3):9-25.
2. Tremblay MS, Wolfson M, Connor Gorber S. Canadian Health Measures Survey: Background, Rationale and Overview. *Health Rep* 2007;18(Suppl.):7-20.
3. Tremblay M. The need for directly measured health data in Canada. *Can J Public Health* 2004;95:165-8.
4. Lundberg O, Manderbacka K. Assessing reliability of a measure of self-rated health. *Scand J Soc Med* 1996; 24:218-24.
5. Chandola T, Jenkinson C. Validating self-rated health in different ethnic groups. *Ethn Health* 2000;5:151-9.
6. Kuhn R, Rahman O, Menken J. *Relating self-reported and objective health indicators to adult mortality in Bangladesh.* Boulder, CO: University of Colorado at Boulder, Institute of Behavioral Science Population Aging Center; 2004.
7. Statistics Canada. Population health surveys [Online]. 2005 [cited 2007 Jan]. Available from: www.statcan.ca/english/concepts/hs/index.htm#content.
8. Statistics Canada. Canadian Health Measures Survey [Online]. 2004 [cited 2007 Jan]. Available from: www.statcan.ca/english/concepts/hs/measures.htm.
9. Katzmarzyk, PT, Janssen, I. The economic costs associated with physical inactivity and obesity in Canada: An update. *Can J Appl Physiol* 2004;29:90-115.
10. Patton MQ. *How to use qualitative methods in evaluation.* Newbury Park, CA: Sage Publications Inc.; 1987.

11. Sale JEM, Lohfeld LH, Brazil K. Revisiting the quantitative-qualitative debate: implications for mixed methods research. *Qual Quant* 2002;36:43-53.

12. Guba EG, Lincoln YS. *Competing paradigms in qualitative research.* In Denzin NK & Lincoln YS, editors. Handbook of Qualitative Research. Thousand Oaks, CA: Sage; 1994.

13. Labonte R, Polanyi M, Muhajarine N, McIntosh T, Williams A. Beyond the divides: towards critical population health research. *Crit Public Health* 2005; 15:5-17.

14. Edwards AL. *The social desirability variable in personality assessment and research.* New York: Dryden Press; 1957.

15. Edwards AL. *The social desirability variable: A review of the evidence.* In Berg IA, ed. Response set in personality assessment. Chicago: Aldine; 1967:48-70.

16. Engstrom JL, Paterson SA, Doherty A, Trabulsi M, Speer KL. Accuracy of self-reported height and weight in women: an integrative review of the literature. *J Midwifery Womens Health* 2003;48:338-45.

17. Larson MR. Social desirability and self-reported weight and height. *Int J Obes Relat Metab Disord* 2000;24:663-5.

18. Rowland ML. Self-reported weight and height. *Am J Clin Nutr* 1990;52:1125-33.

19. Ziebland S, Thorogood M, Fuller A, Muir J. Desire for the body normal: body image and discrepancies between self-reported and measured height and weight in a British population. *J Epidemiol Community Health* 1996;50:105-6.

20. Singleton RA, Straits BC, Miller Straits M. *Approaches to social research* 2nd ed. New York: Oxford University Press; 1993.

21. Larsson D, Hemmingsson T, Allebeck P, Lundberg I. Self-rated health and mortality among young men: what is the relation and how may it be explained? *Scand J Public Health* 2002;30:259-66.

22. Af Sillen U, Nilsson J, Mansson N, Nilsson PM. Self-rated health in relation to age and gender: influence on mortality risk in the Malmo Preventive Project. *Scand J Public Health* 2005;33:183-9.

23. Heistaro S, Jousilahti P, Lahelma E, Vartiainen E, Puska P. Self rated health and mortality: A long term prospective study in eastern Finland. *J Epidemiol Community Health* 2001;55:227-32.

24. Idler EL, Benyamini Y. Self-rated health and mortality: a review of twenty-seven community studies. *J Health Soc Behav* 1997;38:21-37.

25. Nisbett R, Ross L. *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs NJ: Prentice-Hall; 1980.

26. Mathews BA, Nattinger AB, Anderson RC. Accuracy and certainty of self-report for colorectal cancer screening among ambulatory patients. *Psychol Health Med* 2005;10:1-15.

27. Schroder KEE, Carey MP, Vanable PA. Methodological challenges in research on sexual risk behavior: II Accuracy of self-reports. *Ann Behav Med* 2003;104-23.

28. Adlaf EM, Mates D. *A comparison of telephone and personal interview surveys for estimating alcohol and other drug use*. Addiction Research Foundation: Toronto, Canada; 1994.

29. Midanik LT, Greenfield TK, Rogers JD. Reports of alcohol-related harm: Telephone versus face-to-face interviews. *J Stud Alcohol* 2001 Jan;62:74-8.

30. Wiener L, Riekert K, Ryder C, Wood LV. Assessing medication adherence in adolescents with HIV when electronic monitoring is not feasible. *AIDS Patient Care STDs* 2004;18:527-38.

31. Béland Y, St-Pierre M. Mode effects in the Canadian Community Health Survey: a comparison of CATI and CAPI. In: Lepkowski J, Tucker C, Brick JM, et al, eds. *Advances in Telephone Survey Methodology*. New York, NY: Wiley; 2008: 297-314.

32. Harris MI, Flegal KM, Cowie CC, Eberhardt MS, Goldstein DE, Little RR et al. Prevalence of diabetes, impaired fasting glucose, and impaired glucose tolerance in U.S. adults. The Third National Health and Nutrition Examination Survey, 1988-1994. *Diabetes Care* 1998;21:518-24.

33. Health Canada. *Diabetes in Canada. National statistics and opportunities for improved surveillance, prevention, and control.* Health Canada Cat H49-121-1999. Ottawa, ON: Minister of Public Works and Government Services Canada; 1999.

34. Diaz VA, Mainous III AG, Koopman RJ, Geesey ME. Undiagnosed obesity: Implications for undiagnosed hypertension, diabetes, and hypercholesterolemia. *Fam Med* 2004;36:639-44.

35. Alloway CED. Self-report sleepiness measures, objective sleepiness measures and simulated driving performance in controls and patients with obstructive sleep apnea. *Dissertation Abstracts International: Section B: The Sciences and Engineering* 2002;63(6b):3057.

36. Baker FC. A comparison of subjective estimates of sleep with objective polysomnographic data in healthy men and women. *J Psychosom Res* 1999;47:335-41.

37. Brouwer S, Dijkstra PU, Stewart RE, Goeken LNH, Groothoff JW, Geertzen JHB. Comparing self-report, clinical examination and functional testing in the assessment of work-related limitations in patients with chronic low back pain. *Disabil Rehabil* 2005;27:999-1005.

38. Brown FH, Dodrill CB, Clark T, Zych K. An investigation of the relationship between self-report of memory functioning and memory test performance. *J Clin Psychol* 1991;47:772-7.

39. GoveroverY, Kalmar J, Gaudino-Goering E, Shawaryn M, Moore NB, Halper J, DeLuca J. The relation between subjective and objective measures of everyday life activities in persons with multiple sclerosis. *Arch Phys Med Rehabil* 2005;86: 2303-8.

40. Klesges RC, Debon M, Ray JW. Are self-reports of smoking rate biased? Evidence from the second National Health and Nutrition Examination Survey. *J Clin Epidemiol* 1995;48:1225-33.

41. Wareham NJ, Jakes RW, Rennie KL, Schuit J, Mitchell J, Hennings S, Day NE. Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. *Public Health Nutr* 2003;6:407-13.

42. Hill RJ, Davies PSW. Energy intake and energy expenditure in elite lightweight female rowers. *Med Sci Sports Exerc* 2002;34:1823-9.

43. Bostrom G, Diderichsen F. Socioeconomic differentials in misclassification of height, weight and body mass index based on questionnaire data. *Int J Epidemiol* 1997;26:860-6.

44. DelPrete LR, Caldwell M, English C, Banspach SW, Lefebvre C. Self-reported and measured weights and heights of participants in community-based weight loss programs. *J Am Diet Assoc* 1992;92:1483-6.

45. Kuskowska-Wolk A, Rossner S. The "true" prevalence of obesity. A comparison of objective weight and height measures versus self-reported and calibrated data. *Scand J Prim Health Care* 1989;7:79-82.

46. Pirie P, Jacobs D, Jeffery R, Hannan P. Distortion in self-reported height and weight data. *J Am Diet Assoc* 1981;78:601-6.

47. Nawaz H, Chan W, Abdulrahman M, Larson D, Katz DL. Self-reported weight and height: Implications for obesity research. *Am J Prev Med* 2001;20:294-8.

48. Engstrom JL. Assesment of the reliability of physical measures. *Res Nurs Health* 1988;11:383-9.

49. Gillum RF, Sempos CT. Ethnic variation in validity of classification of overweight and obesity using self-reported weight and height in American women and men: The Third National Health and Nutrition Examination Survey. *Nutr J* 2005;4:27.

50. Stewart, AL. The reliability and validity of self-reported weight and height. *J Chronic Dis* 1982;35:295-309.

51. Villanueva EV. The validity of self-reported weights in US adults: a population based cross-sectional study. *BMC Public Health* 2001;1:11.

52. Jalkanen L, Tuomilehto J, Tanskanen A, Puska P. Accuracy of self-reported body weight compared to measured body weight. A population survey. *Scand J Soc Med* 1987;15:191-8.

53. Smith GT, Hohlsteing LA, Atlas JG. Accuracy of self-reported weight: Covariation with binger or restrainer status and eating disorder symptomatology. *Addict Behav* 1992;17:1-8.

Paper 1: A comparison of direct versus self-report measures for assessing height, weight and body mass index: A systematic review

The paper has been published in:

Connor Gorber S, Tremblay M, Moher D, Gorber B. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obesity Reviews* 2007;8(4):307-26.

This paper was presented at:

Connor Gorber S, Tremblay MS, Moher D, Gorber B. A comparison of direct versus self-report measures for assessing height, weight and body mass index: a systematic review. Poster presentation at the *Canadian Public Health Association Conference* (Ottawa, Canada), 2007.

A comparison of direct versus self-report measures for assessing height, weight and body mass index: a systematic review

Sarah Connor Gorber[1], Mark Tremblay[1], David Moher[2], Benson Gorber

1. Physical Health Measures Division, Statistics Canada
2. Chalmers Research Group, Children's Hospital of Eastern Ontario Research Institute

Key Words: self-report, measurement error, height, weight

Running Title: Comparison of self-report and direct measures

Correspondence: Sarah Connor Gorber, Chief - Research and Analysis, Statistics Canada, Main Building Room 005, 150 Tunney's Pasture Driveway, Ottawa,ON, K1A 0T6,

Abstract

Obesity is a rapidly increasing public health problem, with surveillance most often based on self-reported values of height and weight. We conducted a systematic review to determine what empirical evidence exists regarding the agreement between objective (measured) and subjective (reported) measures in assessing height, weight and body mass index (BMI). Five electronic databases were searched to identify observational and experimental studies on adult populations over the age of 18. Searching identified 64 citations that met the eligibility criteria and examined the relationship between self-reported and directly measured height or weight. Overall, the data show trends of under-reporting for weight and BMI and over-reporting for height, although the degree of the trend varies for men and women and the characteristics of the population being examined. Standard deviations were large indicating that there is a great deal of individual variability in reporting of results. Combining the results quantitatively was not possible because of the poor reporting of outcomes of interest. Accurate estimation of these variables is important as data from population studies such as those included in this review are often used to generate regional and national estimates of overweight and obesity and are in turn used by decision makers to allocate resources and set priorities in health.

Introduction

Height and weight are important indicators of population health because of their role in calculating the Body Mass Index (BMI), a common measure of obesity[1]. Obesity rates have been on the rise in many countries over the last several decades, and more than doubled in Canada over the 13 year period from 1985 to 1998[2]. Internationally, the World Health Organization has declared that obesity has reached epidemic proportions[3] and that it will be one of the greatest challenges and risk factors for chronic disease in this century[4]. Recent evidence indicates that the prevalence of obesity in Canada is 23% in adults[5] and 8% in children[6]. The health costs of obesity in Canada represent 2.2% of total health care costs[7].

Self-report measures are one of the most common ways to collect data on height and weight and have consistently been shown to be associated with mortality and morbidity[8-10]. Self-reports have the advantages of practicality and low cost, they are quick and easy to administer and are a good method for sampling large numbers of individuals. However, the tools are subject to certain limitations: questions may be misunderstood, participants may not accurately recall past events and response bias e.g., social desirability or response acquiescence are not uncommon[11].

Social desirability is particularly problematic with sensitive questions (e.g., about weight or obesity) and can be influenced by the method of data collection[12-14]. For example, in the Canadian Community Health Survey (CCHS), one of Canada's largest national health surveys[15], the proportion of respondents reporting being obese (based on self-reported values of height and weight) was estimated to be 18% for those who were interviewed in person

compared to an estimate of only 13% for those who were asked the same questions by phone[16]. These inconsistencies can be even more pronounced when for operational and budgetary reasons survey modes do not stay consistent over time. Similarly, changes in question wording or respondent characteristics among surveys, or within a survey over time, also have the potential to impact the quality of self-reported estimates. These issues make it difficult to get accurate self-reported estimates of height, weight and obesity over time. Compounding the problem is the fact that the influence of social desirability on self-reports has the potential to change over time as social and cultural norms about weight and obesity change.

Because of the limitations associated with self-report measures, objective or direct measures have been recommended to improve measurement precision. Despite the benefits, the costs of this type of measurement are often prohibitively high, particularly in population research where the samples are large. The intrusive nature of the measures also has the potential to impact response rates and contribute to attrition bias. As well, such measures often require special training of interviewers for accurate administration and classification.

In order to ensure that researchers are employing the most appropriate tools in their research and to ensure that decision making is based upon the best available evidence, the degree of accuracy of self-reported measures needs to be determined. We conducted this systematic review to determine what empirical evidence exists regarding the concordance of objective and subjective measures in assessing height and weight. The primary objective was to compare direct vs. self-report measures for assessing height and weight in observational and experimental studies of adult populations. A secondary objective was to examine how

the use of self-report and direct measures of height and weight impact on estimates of the prevalence of obesity and overweight, as measured by the BMI. Although it is recognized that BMI is only one of many methods used to measure obesity, it is employed in this review because it can be directly calculated from weight and height measurements.

Methods of the Review

Criteria for Considering Studies for this Review

Any observational or experimental studies that included a direct comparison of self-report and objective measures for the outcomes of interest were eligible for inclusion. All study designs were eligible (e.g., retrospective, prospective, case control) and both published and unpublished literature was examined.

Only studies on adult populations over the age of 18 years were considered. Children were excluded as they are still in an important stage of development where variables such as weight may change over short periods of time. It has also been suggested that reporting error in children and adolescents may be of a different nature than that in adults[17].

Any form of weight or height measurement whether by portable or balance beam scale, stadiometer, tape measure or other technique was eligible for inclusion. Likewise any form of self-reporting technique could be included (e.g., interviewer administered, self-complete). Proxy reporting (when one individual responds for another) was not eligible for inclusion.

Search Strategy for Identification of Studies

The following electronic bibliographical databases were searched:

- MEDLINE – from 1966 to January Week 4 2006;
- EMBASE – from 1980 to Week 04 2006;
- CINAHL – from 1982 to December Week 2 2005;
- PsycINFO – from 1806 to January Week 4 2006;
- Health and Psychosocial Instruments (HAPI) – from 1985 to September 2005;
- SPORTDiscus – from 1830 to January 2006

The search strategy used for searching MEDLINE is provided as an example (see Table 1) but was modified according to the indexing systems of other databases. No language of publication restrictions were imposed. The Ovid interface was used for all electronic searches. Grey literature such as published abstracts/proceedings, published lists of theses and dissertations, and government reports were also identified where possible.

Selecting Citations

Titles and abstracts of potentially relevant citations were first reviewed to assess adherence to the inclusion criteria. The full text of all articles that met the criteria was then further screened by two independent reviewers (SCG, BG). Searches of the bibliographies of texts were also performed in order to identify additional studies, which were subsequently retrieved. The two reviewers examined the full text versions of remaining citations and selected those that met the inclusion criteria for the study. There were no disagreements about inclusion.

Data Extraction

Standardized data extraction forms were completed by one reviewer (SCG) and verified by the other (BG). Information was extracted on the type of study, participant characteristics, sample size, methods of weight, height and BMI measurement, time between self-report and direct measurement, measurement setting, tools and equipment, order of measures, and results (means and mean differences between measures, measures of variance). Reviewers were not blinded to the authors or journals when extracting data. The data extraction form is found in Appendix A.

Quality Assessment

Because both experimental and observational study designs were included in this review, two tools were used to assess quality. Jadad's scale[18] was used to assess the quality of randomized controlled trials and the Downs and Black[19] tool was used to assess the quality of non-randomized designs. The Jadad scale includes three items assessing the quality of randomization generation, blinding and dropouts and withdrawals. The scale ranges from 0 to 5 with higher scores indicating higher quality. The Downs and Black instrument was one of six instruments recommended[20] for use in systematic reviews of non-randomized studies. The Downs and Black instrument was selected as it contained the highest number of relevant items for the needs of this review. However, as not all items were relevant to the studies included in this review (many of which were based on surveys), a modified version of the checklist was employed with the following items omitted: items 5 and 8 in the reporting scale, items 14, 15 and 19 in the section on bias, items 21–26 relating to confounding and item 27 addressing power. The final checklist was made up of 15 items

with a maximum score of 15 points (with higher points indicating superior quality) rather than the original 32 points. The checklist covered the categories of reporting, external validity and internal validity (bias).

Results

Description of Studies

A total of 328 citations were identified based on the preliminary search of electronic databases, reference lists, and grey literature (see Figure 1). Of these 137 were identified in MEDLINE, an additional 35 in EMBASE, 117 in CINAHL, 16 in PsycINFO and one in SPORTDiscus. Searching HAPI did not uncover any citations that had not been previously identified. Searching bibliographies of key articles, mostly Engstrom[21] who had conducted a review examining the correspondence between self-reported and directly measured height and weight in women yielded an additional 18 citations. The three reviewer nominated citations were discovered during searching. The one citation that was included from the grey literature search was published by Statistics Canada.

In total, 115 full text articles were retrieved for detailed assessment. Of these 64 met the criteria for study inclusion (see Table 2). Common reasons for excluding studies included a focus on outcomes other than height and weight, a focus on children or adolescents[ii], not having both directly measured and self-reported data within the same study on the same population and examining ideal or perceived weight or height rather than actual values.

[ii] The database search was not restricted to adults, as this would have excluded some studies (e.g., LeJarraga et al..[28]) that examined the characteristics of both children and their parents.

Four of the articles[17, 22-24] analyzed data from the same database, the National Health And Nutrition Examination Survey (NHANES), three of which looked at the data from the NHANES III cohort collected between 1988 and 1994. NHANES is "a program of studies designed to assess the health and nutritional status of adults and children in the United States"[25](p.2). In order to prevent double counting, only one of the NHANES III studies[22] was retained. Kuczmarski was selected because it had the most comprehensive outcome reporting (it provided mean differences with corresponding measures of variance). Rowland's[17] data were based on the NHANES II cohort collected from 1976 to 1980, and therefore, there are no concerns about overlap.

Also requiring clarification is the inclusion of Stunkard and Albaum[26] and Schlichting and colleagues[27], both published in 1981. Stunkard examined data that were collected on subjects at eight different medical and non-medical sites in two countries (Denmark and the United States). Schlichting's study examined the same dataset but focused on analysis of the Danish data only. Hence, Stunkard's mean difference values for weight were used in this review, because they were more comprehensive, but because they did not report any results on height, Schlichting's data for the height outcome were also included.

Of the 64 studies that were retained 60 used different types of observational designs[17,22,26-83] (e.g., case control, cross sectional), two used experimental designs[84,85], one was a letter[86] and there was one unpublished dissertation[87]. Two studies were in languages other than English (Portuguese[81] and Spanish[82]). Studies that were included were published over a 26 year

period between 1979 and 2005. The primary aim of the articles differed but each compared direct measures of height or weight with their self-reported counterparts.

Participants in the studies ranged from 10 to over 60 years, with the oldest participant reported to be 89 years old. Although the focus of the review is on those aged 18 years and over, studies that had a range of ages less than 18 years were not excluded as long as the mean age of the majority of the population was over 18 years. The study by Allison and colleagues[29] for example, had an age range of 10 to 67 years but a mean age of 38.5±12.0 years. A decision was made to include this study because only a small number of cases fell below the 18-year cut-off. Sample size ranged from a low of 44 in the dissertation[87] to a high of over 16,000 in Kuczmarski's NHANES analysis[22].

The majority of studies reported which tools were used to measure the weight and height of participants in the various studies, but 23 studies provided no information about what equipment was used to take the measurements (the two non-English language publications were not assessed on these characteristics). When data on measurement tools were reported, height was most commonly measured by stadiometer, anthropometer, or some type of measuring tape or ruler, and a variety of scales (balance beam, digital, or portable) were used to measure weight. Most studies provided information about what subjects wore during the measurements, which was usually light indoor clothes and no shoes. In six studies[30,47,52,59,60,61] participants wore only undergarments and in another six were measured and weighed in hospital gowns[17,22,55,57,66,75]. In one study focused on height[44], participants were purposely given no instructions about whether shoes should be worn for the

measurement as authors wanted to capture how the participants presented themselves to the world. Sixteen studies did not describe the clothing worn by participants.

Self-report data were most often collected by questionnaire (either by mail, telephone or most often in person) with 16 studies collecting data in an interview. Self-reports were conducted prior to the direct measurement in over 80% of studies. In two studies[36,43] the sample was divided into two groups with one undergoing the direct measurement before the self-report and the other completing the self-report first. Only one study[49] consistently measured participants before the self-report was completed. Five studies[42,53,61,68,83] did not provide enough information to determine which came first. Twenty-five studies collected the self-report with the direct measurement immediately afterwards, with Roberts[66] reporting the longest delay between measures (3 to 8 months).

Quality Assessment

Results of the quality assessment are presented in Table 3. Quality was assessed on 60 of the studies. Zhang's letter[86] was not assessed, because although it presented data it did not provide enough information to determine a quality assessment. Similarly, Wang[83], which was mostly a review article with some new analyses presented, lacked information for an appropriate quality score to be calculated. Quality scores were also not computed for the two non-English language articles. Of the articles that remained, 58 were assessed with the Downs and Black checklist and two using the Jadad scale. The range on the Downs and Black tool was 6 to 15 (maximum possible score was 15) with a mean of 11.4 ±1.9.

Citations were assessed based on the degree to which they presented data relating to the outcomes studied in this review (e.g., height and weight) even if these were only secondary outcomes in the studies. Common gaps in reporting included not clearly describing the intervention of interest (e.g., the method of height or weight measurement), not presenting simple outcome data such as the mean weight or height in the self-reported and measured groups so that major analyses and conclusions could be verified and not reporting the associated measures of random variability. Although most studies carried out some sort of significance testing, many did not report the actual probability values, which resulted in the deduction of a point.

Thirteen percent of studies received no points in the external validity category while 57% received one point or less. Most of the point losses were due to a lack of representativeness of the samples, with subjects who were asked or those who agreed to participate not being representative of the entire population from which they were recruited. No study lost more than one point in the internal validity checklist, with over half receiving perfect scores in this category. Most deductions resulted from not adjusting for different lengths of follow-up of subjects in the analyses (e.g., if the delay between self-report and direct measures was excessively long - over 4 weeks - and was not consistent from subject to subject). Points were also deducted on item 18 when mean differences were not calculated or could not be calculated from the data.

The two experimental studies were assessed with the Jadad scale, which is scored out of a maximum of five points. Black and colleagues[85] received one point for describing the study

as randomized and O'Connell and colleagues[84] received one point for describing the study as randomized, one point for describing withdrawals and dropouts and an additional point because the method of randomization (random numbers table) was an appropriate randomization technique.

Assessing quality, particularly in observational study designs is not without flaws. Even the items on the Downs and Black checklist, which was developed for non-randomized designs, did not always apply to the studies in this review, and at times the interpretation of items had to be modified in order to apply them to the included studies. Yet, it is important to attempt to assess quality, despite its subjectivity and the absence of perfect tools, so that similar studies can be compared on a similar metric, which can lead to improvements in future work.

Height

Fifty three studies contained data on self-reported and measured height and these are displayed in Table 4. The table presents the mean differences between self-reported and measured height (SR-DM) for the overall sample as well as the breakdowns for males and females (in some cases, the actual mean difference was provided and in others it was calculated by subtracting the directly measured value from the self-reported value). What is most apparent from examining the table is the lack of data that are presented on the key outcomes, even in many studies where the goal was to compare measured and reported height. Only 28 of the studies provided both mean differences (either by sex or overall) and a corresponding measure of variance, which are required for pooling data. Only four provided

data for the entire sample as well as according to the sex breakdowns (Nakamura[59] and Nawaz[60] were exceptions as these studies only examined women).

In the 18 studies that had mean error data for the overall sample, height was overestimated in all but one[39]. Doll and Fairburn's examination of healthy controls and individuals with bulimia nervosa found that while healthy controls overestimated their height, there was no significant difference between self-report and directly measured values for those with bulimia. The mean error for the rest of the studies ranged from a low of 0.6[35,62] cm to a high of 7.5 cm[45]. In those reports that had data on men, 29 studies reported that height was overestimated (range 0.1 cm - 5.0 cm) and in two it was underestimated[32,44] (mean error of -1.3 cm and -0.7 cm respectively). Two studies[28,37] had mixed results. LeJarraga[28] examined parents of children attending growth clinics and found that parents in private clinics overestimated their height, while those from public clinics tended to slightly underestimate their height. Chor's[37] study of bank employees found no difference between the self-reported and measured values. Most studies also found that height was overestimated in women except in four studies where height was underestimated[32,44,59,65] and one that had mixed results[67]. Standard deviations of the differences between reported and actual measures were large ranging from 0.1 cm in both men and women[22] to 8.8 cm in men and to 11.2 cm[67] in women. Studies that had data available for the overall sample were also plotted on a forest plot (see Figure 2). The studies are ordered by sample size [smallest to largest with the three studies that had long time lags[34, 45, 72] (mean of over 1 month) plotted at the top of the figure].

Weight

Table 5 displays the mean differences for weight for 56 studies that provided data on self-reported and measured weight. With weight as an outcome, 14 studies provided mean differences and associated variability measures for the overall sample while 24 provided them for either men or women as subgroups.

In examining the overall results the tendency is for weight to be underestimated except in two studies (in Avila-Funes' work[31] in a Mexican population and in the anorexic sample of McCabe's study[57] where weight was overestimated). Similarly in studies reporting data separately for women, all but three found that women underestimated their weight when compared with the directly measured values (-0.1 kg[58] to -6.5 kg[60] in a study of obese women with BMI's between 35 and 40 kg/m^2); Avila-Funes[31] study of Mexican women, and the subgroup of women in Ziebland's study[80] who had BMI's less than 20 reported overestimating their weight by approximately 0.7 kg and 0.8 kg respectively, and Wada's analysis[78] of Japanese women found no difference between the reported and measured values. The trend for men was also in the direction of under-reporting with 27 of the 34 studies that calculated mean differences for men finding that men under-reported their weight (range -0.1 kg to -3.2 kg[80]). The range in standard deviations was slightly smaller in weight than in height from 0.1 kg in a large national American survey[22] to 4.5 kg in a smaller study of veterans[49] in men and in women ranged from 0.1 kg again in the national survey[22] to 10.1 kg in a small study of 96 obese women in the US[60]. These values represent a significant deviation from their measured values for some individuals. Figure 3 depicts a

forest plot of the studies that had data available for the overall sample; the studies with longer time lags are plotted at the top.

BMI

Self-reported BMI is based on self-reports of both height and weight and measured BMI is based on measured values of both of these variables. BMI values from Masheb and Grilo[56] were not included since both their self-report and measured BMI's were calculated from the same directly measured height variable (i.e., self-reported height was not collected). Twenty-nine studies are included in Table 6, 18 reported mean difference and variability estimates (either by sex or overall) to compare self-reported BMI with the directly measured values. Eleven studies are included in the forest plot (Figure 4). All but two studies found that the tendency was for BMI to be underestimated or accurately reported relative to the measured values. The range for men was from no difference[78] to -2.1 kg/m^2 [62] and for women ranged from no difference[78] to -2.2 kg/m^2 in Nawaz and colleague's sample[60] of women who had BMI values over 40. Standard deviations were smaller than for the individual values of height and weight ranging from 0 to 1.6 kg/m^2 for men and from 0 to 2.2 kg/m^2 for women. Bolton-Smith's work[32], which sampled the general population aged 25 - 64 years in Scotland from general practitioner's registries, had results that were inconsistent with this trend, finding instead that BMI tended to be overestimated by a mean difference 0.2±1.4 kg/m^2 in men and 0.2±1.3 kg/m^2 in women. Quiles and Vioque's[82] general population sample of residents 15 and over also found BMI to be overestimated in men by 0.5 kg/m^2 and in women by 0.9 kg/m^2 (standard deviations not provided).

Quantitative Analysis

Quantitative analysis is possible when the data are sufficiently homogeneous in terms of statistical, clinical, and methodological characteristics. Tables 4 - 6 have loosely grouped the studies according to their main populations of interest, and the quality assessment would permit a further subdivision based on study quality. Because of the substantial clinical and methodological inconsistency across the study reports and the large amount of missing data (i.e., mean differences and corresponding measures of variance), we did not combine the data quantitatively. Although all studies should be reporting mean differences and variance estimates, some authors only report proportions or correlation coefficients, which creates difficulties for pooling the estimates. A small subset of authors was contacted in an attempt to acquire missing data, but no data were received to cover the information gaps.

Discussion

The trend in studies where data were available was for height to be overestimated and weight and BMI to be underestimated. This trend was evident for both women and men. The size of the mean errors in weight varied depending on the study and the population being examined. In many cases they were relatively small but in some instances such as , Ziebland's study[80] of individuals with BMI's over 30 and Nawaz's[60] work with populations with BMI's over 35, the mean differences in weight were quite high (underestimated by 3 to 6 kg). In many of the studies the standard deviations were large indicating that there is a great deal of individual variability in reporting of results for all of the outcomes examined. This can have important implications for clinical and population health practice as even a small difference

in body weight has the potential to substantially modify the BMI classification, which could mean that current obesity prevalence rates are underestimated.

The findings in this review are similar to those of Engstrom and colleagues[21] who reviewed the accuracy of height and weight in women and adolescent females. They examined 26 studies on the accuracy of self-reported height and found that in 21 of them height was overestimated. Similarly, they found weight to be underestimated in all of the 34 studies that they reviewed on the accuracy of self-reported weight. In addition, Bowman and Delucia[88] conducted a meta-analysis of the accuracy of self-reported weight and found that there was significant bias in the self-reported measures, with a tendency for self-reported weights to be underestimated. In spite of the trend in the present review of weight and BMI being underestimated and height being overestimated, there were too many gaps in the data to undertake a quantitative analysis or to get a comprehensive understanding of the relationship between self-report and direct measures.

Although studies are for the most part of good quality, Tables 4 - 6 highlight the degree to which there are gaps in reporting; studies lack vital information that allow conclusions to be verified and analyses replicated. Studies in this review have been ongoing since 1979 and would have benefited from the development of some standard reporting criteria to address some of these gaps over this 26 year period. In 1988 Engstrom[89] published a paper outlining the statistics and procedures that should be reported whenever assessing the reliability of physical measures such as weight and height. She suggested that values such as the mean, minimal and maximal differences, standard deviations of the net differences, technical error

of measurement and clinically meaningful indices of agreement be included in all reports. However, as evidenced in the current review almost 20 years later, such data are still not routinely being reported. In addition, the authors of this current review recommend that the elapsed time between self-report and direct measures should also be reported, since changes in variables such as weight can occur over short periods of time and the length of time between measures should be controlled in any analyses that are undertaken. The order of measurement should also be noted as there is evidence to indicate that reporting of weight in women is more accurate when it occurs after actual weight measurement[36]. Measurement conditions such as who is taking the measures, what equipment is being used, what clothing is worn for the weighing as well as what instructions are given when estimating self-report weights (e.g., is this self-reported weight estimated with or without clothes, in the morning or midday and after eating or fasting?) should also be provided.

Adhering to consistent reporting criteria would increase the comparability of results and enable us to further understand the relationship between the measures. With more complete data it may be possible to develop correction factors that could be applied to self-reported data when direct measurement is not feasible. If data are adequately reported by subgroups (e.g., sex, age, education, ethnicity, and immigration status) it may also be possible to make adjustments based on these characteristics, which may be important since studies demonstrate that demographic characteristics can have an influence on the degree of reporting error. For instance, research has shown that self-reports from older populations are less reliable[22], that unemployed, retired or disabled women are more likely to under-report their BMI's than employed women[60], and that men are more likely to over-report their height

than women[38]. Understanding the characteristics of inaccurate reporters is particularly important given the large values of some of the standard deviations, which implies that for some individuals the estimates are extremely inaccurate, which may bias average results and have important implications for health planning for these populations.

Conclusions

To the authors' knowledge this review represents the most comprehensive attempt to examine the relationship between self-reported and directly measured height, weight and BMI in both men and women. Overall the data show trends of underestimating weight and BMI and over estimating height, although the degree of the trend varies for men and women, and between studies; the variability in the estimates is high and no overall effect size could be estimated. Rising obesity rates are an increasing concern for population health and it is therefore essential that the estimates upon which decisions are being made are as accurate as possible. This should be addressed by improving the quality of reporting of future studies in this area, making recommendations based on more comprehensive data about whether (and in what circumstances) self-report measures can be used and then developing correction factors to increase the accuracy of reporting in situations where direct measurement is not possible.

References

1. Health Canada. Body Mass Index (BMI) Nomogram [Online]. 2003 [cited 2006 Apr] Available from: www.hc-sc.gc.ca/fn-an/nutrition/weights-poids/guide-ld-adult/bmi_chart_java-graph_imc_java_e.html.
2. Katzmarzyk PT. The Canadian obesity epidemic, 1985–1998. *CMAJ* 2002;166:1039-1040.
3. World Health Organization. *Obesity and overweight* [Online]. 2006 [cited 2006 Apr] Available from:www.who.int/dietphysicalactivity/publications/facts/obesity/en/prinwt.html.
4. World Health Organization. *World Health Report 2002 Reducing risks, promoting healthy life* [Online]. 2002 [cited 2006 Apr]. Available from: www.who.int/whr/2002/en/
5. Tjepkema M. Adult obesity. *Health Rep* 2006;17(3):9-25.
6. Shields M. Overweight and obesity among children and youth. *Health Rep* 2006;17(3):27-42.
7. Katzmarzyk, PT, Janssen, I. The economic costs associated with physical inactivity and obesity in Canada: An update. *Can J Appl Physiol* 2004;29:90-115.
8. Larsson D, Hemmingsson T, Allebeck P, Lundberg I. Self-rated health and mortality among young men: what is the relation and how may it be explained? *Scand J Public Health* 2002;30:259-66.
9. Af Sillen U, Nilsson J, Mansson N, Nilsson PM. Self-rated health in relation to age and gender: influence on mortality risk in the Malmo Preventive Project. *Scand J Public Health* 2005;33:183-9.

10. Heistaro S, Jousilahti P, Lahelma E, Vartiainen E, Puska P. Self rated health and mortality: A long term prospective study in eastern Finland. *J Epidemiol Community Health* 2001;55:227-32.

11. Singleton RA, Straits BC, Miller Straits M. *Approaches to social research.* 2nd edn. New York, NY: Oxford University Press; 1993.

12. Adlaf EM, Mates D. *A comparison of telephone and personal interview surveys for estimating alcohol and other drug use.* Addiction Research Foundation: Toronto, Canada; 1994.

13. Midanik LT, Greenfield TK, Rogers JD. Reports of alcohol-related harm: Telephone versus face-to-face interviews. *J Stud Alcohol* 2001 Jan;62:74-8.

14. Wiener L, Riekert K, Ryder C, Wood LV. Assessing medication adherence in adolescents with HIV when electronic monitoring is not feasible. *AIDS Patient Care STDs* 2004 Sep;18(9):527-38.

15. Statistics Canada. *Canadian Community Health Survey (CCHS): Cycle 1.1: extending the wealth of health data in Canada* [Online]. 2003 [cited 2006 Feb]. Available from: www.statcan.ca/english/concepts/health/cchsinfo.htm.

16. Béland Y, St-Pierre M. Mode effects in the Canadian Community Health Survey: a comparison of CATI and CAPI. In: Lepkowski J, Tucker C, Brick JM, et al, eds. *Advances in Telephone Survey Methodology.* New York, NY: Wiley; 2008; 297-314.

17. Rowland ML. Self-reported weight and height. *Am J Clin Nutr* 1990;52:1125-33.

18. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1-12.

19. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52:377-84.

20. Deeks JJ, Dinnes, D'Amico R, Sowden AJ, Sakarovitch, C, Song F, et al. Evaluating non randomized intervention studies. *Health Technol Asses* 2003;7(27).

21. Engstrom JL, Paterson SA, Doherty A, Trabulsi M, Speer KL. Accuracy of self-reported height and weight in women: an integrative review of the literature. *J Midwifery Womens Health* 2003;48:338-45.

22. Kuczmarski MF, Kuczmarski RJ, Najjar M. Effects of age on validity of self-reported height, weight, and body mass index: findings from the Third National Health and Nutrition Examination Survey, 1988-1994. *J Am Diet Assoc* 2001 Jan;101:28-34.

23. Villanueva EV. The validity of self-reported weight in US adults: a population based cross-sectional study. *BMC Public Health* 2001;1(11). Epub 2001 Nov 6.

24. Gillum RF, Sempos CT. Ethnic variation in validity of classification of overweight and obesity using self-reported weight and height in American women and men: The Third National Health and Nutrition Examination Survey. *Nutr J* 2005;4:27.

25. US department of Health and Human Services, Centers for Disease Control and Prevention, National Centre for Health Statistics. *National Health and Nutrition Survey 2005-2006 Overview* [Online]. 2005 [cited 2006 Apr]. Available from: www.cdc.gov/nchs/data/nhanes/OverviewBrochureEnglish_May05.pdf.

26. Stunkard AJ, Albaum JM. The accuracy of self-reported weights. *Am J Clin Nutr* 1981;34:1593-9.

27. Schlichting P, Hoilund-Carlsen PF, Quaade F. Comparison of self-reported height and weight with controlled height and weight in women and men. *Int J Obes* 1981;5:67-76.

28. LeJarraga H, Laspuir M, Adamo P. Validity of reported parental height in outpatient growth clinics in Buenos Aires city. *Ann Hum Biol* 1995;22:163-6.

29. Allison DB, Nathan JS, Albu JB, Heymsfield SB, Duprat LJ, Pi-Sunyer FX. Measurement challenges and other practical concerns when studying massively obese individuals. *Int J Eat Disord* 1998;24:275-84.

30. Alvarez-Torices JC, Franch-Nadal J, Alvarez-Guisasola F, Hernandez-Mejia R, Cueto-Espinar A. Self-reported height and weight and prevalence of obesity. Study in a Spanish population. *Int J Obes Relat Metab Disord* 1993;17:663-7.

31. Avila-Funes JA, Gutierrez-Robledo LM, Ponce De Leon RS. Validity of height and weight self-report in Mexican adults: results from the national health and aging study. *J Nutr Health Aging* 2004;8:355-61.

32. Bolton-Smith C, Woodward M, Tunstall-Pedoe H, Morrison C. Accuracy of the estimated prevalence of obesity from self-reported height and weight in an adult Scottish population. *J Epidemiol Community Health* 2000;54:143-8.

33. Booth ML, Hunter C, Gore CJ, Bauman A, Owen N. The relationship between body mass index and waist circumference: implications for estimates of the population prevalence of overweight. *Int J Obes Relat Metab Disord* 2000;24:1058-61.

34. Bostrom G, Diderichsen F. Socioeconomic differentials in misclassification of height, weight and body mass index based on questionnaire data. *Int J Epidemiol* 1997;26:860-6.

35. Brown JK, Feng JY, Knapp TR. Is self-reported height or arm span a more accurate alternative measure of height? *Clin Nurs Res* 2002;11:417-32.

36. Cash TF, Counts B, Hangen J, Huffine CE. How much do you weigh?: Determinants of validity of self-reported body weight. *Percept Mot Skills* 1989;69:248-50.

37. Chor D, Coutinho ES, Laurenti R. Reliability of self-reported weight and height among state bank employees. *Rev Saude Publica* 1999;33:16-23.

38. DelPrete LR, Caldwell M, English C, Banspach SW, Lefebvre C. Self-reported and measured weights and heights of participants in community-based weight loss programs. *J Am Diet Assoc* 1992;92:1483-6.

39. Doll HA, Fairburn CG. Heightened accuracy of self-reported weight in bulimia nervosa. A useful cognitive "distortion". *Int J Eat Disord* 1998;24:267-73.

40. Fonseca MJ, Faerstein E, Chor D, Lopes CS. [Validity of self-reported weight and height and the body mass index within the 'Pro-saude' study]. [Portuguese]. *Rev Saude Publica* 2004;38:392-8.

41. Forster JL, Jeffery RW, Snell MK. One year follow up study to a worksite weight control program. *Prev Med* 1988;17:129-33.

42. Giles E, Hutchison DL. Stature- and age-related bias in self-reported stature. *J Forensic Sci* 1991;36:765-80.

43. Gunnell D, Berney L, Holland P, Maynard M, Blane D, Frankel S, Smith GD. How accurately are height, weight and leg length reported by the elderly, and how closely are they related to measurements recorded in childhood? *Int J Epidemiol* 2000;29:456-64.

44. Hensley WE. The measurement of height. *Adolescence* 1998;33:629-35.

45. Hill A, Roberts J. Body mass index: a comparison between self-reported and measured height and weight. *J Public Health Med* 1998;20:206-10.

46. Imrhan SN, Imrhan V, Hart C. Can self-estimates of body weight and height be used in place of measurements for college students. *Ergonomics* 1996;39:1445-53.

47. Jacobson BH, DeBock DH. Comparison of Body Mass Index by self-reported versus measured height and weight. *Percept Mot Skills* 2001;92:128-32.

48. Jalkanen L;Tuomilehto J;Tanskanen A;Puska P. Accuracy of self-reported body weight compared to measured body weight. A population survey. *Scand J Soc Med* 1987;15:191-8.

49. Kinney EL. Accuracy of self-reported weight in a non-normal population. *Clin Invest Med* 1988;11:347-50.

50. Klag MJ, He J, Mead LA, Ford DE, Pearson TA, Levine DM. Validity of physicians' self reports of cardiovascular disease risk factors. *Ann Epidemiol* 1993;3:442-7.

51. Kuskowska-Wolk A, Bergstrom R, Bostrom G. Relationship between questionnaire data and medical records of height, weight and body mass index. *Int J Obes Relat Metab Disord* 1992;16:1-9.

52. Kuskowska-Wolk A, Rossner S. The "true" prevalence of obesity. A comparison of objective weight and height measures versus self-reported and calibrated data. *Scand J Prim Health Care* 1989;7:79-82.

53. Lackland DT, Wheeler FC. The need for accurate nutrition survey methodology: The South Carolina experience. *J Nutr* 1990;120(11S):1433-6.

54. Larson MR. Social desirability and self-reported weight and height. *Int J Obes Relat Metab Disord* 2000;24:663-5.

55. Lawlor DA, Bedford C, Taylor M, Ebrahim S. Agreement between measured and self-reported weight in older women. Results from the British Women's Heart and Health Study. *Age Ageing* 2002;31:169-74.

56. Masheb RM, Grilo CM. Accuracy of self-reported weight in patients with binge eating disorder. *Int J Eat Disord* 2001;29:29-36.

57. McCabe RE, McFarlane T, Polivy J, Olmstead MP. Eating disorders dieting and the accuracy of self-reported weight. *Int J Eat Disord* 2001;29:59-64.

58. de Araujo DS, de Araujo CG. Self-perception and dissatisfaction with weight does not depend on the frequency of physical activity. *Arq Bras Cardiol* 2003;80:235-49.

59. Nakamura K, Hoshino Y, Kodama K, Yamamoto M. Reliability of self-reported body height and weight of adult Japanese women. *J Biosoc Sci* 1999;31:555-8.

60. Nawaz H, Chan W, Abdulrahman M, Larson D, Katz DL. Self-reported weight and height: Implications for obesity research. *Am J Prev Med* 2001;20:294-8.

61. Niedhammer I, Bugel I, Bonenfant S, Goldberg M, Leclerc A. Validity of self-reported weight and height in the French GAZEL cohort. *Int J Obes Relat Metab Disord* 2000;24:1111-8.

62. Nieto-Garcia FJ, Bush TL, Keyl PM. Body mass definitions of obesity: sensitivity and specificity using self-reported weight and height. *Epidemiology* 1990;1:146-52.

63. Palta M, Prineas RJ, Berman R, Hannan P. Comparison of self-reported and measured height and weight. *Am J Epidemiol* 1982;115:223-30.

64. Payette H, Kergoat MJ, Shatenstein B, Boutier V, Nadon S. Validity of self-reported height and weight estimates in cognitively-intact and impaired elderly individuals. *J Nutr Health Aging* 2000;4:223-8.

65. Pirie P, Jacobs D, Jeffery R, Hannan P. Distortion in self-reported height and weight data. *J Am Diet Assoc* 1981;78:601-6.

66. Roberts RJ. Can self-reported data accurately describe the prevalence of overweight? *Public Health* 1995;109:275-84.

67. Rona RJ, Chinn S, Manning R. The validity of reported parental height in inner city areas in England. *Ann Hum Biol* 1989;16:41-4.

68. Santillan AA, Camargo CA. Body mass index and asthma among Mexican adults: the effect of using self-reported vs measured weight and height. *Int J Obes Relat Metab Disord* 2003;27:1430-3.

69. Schmidt MI, Duncan BB, Tavares M. Polanczyk CA, Pellanda L, Zimmer PM. Validity of self-reported weight--a study of urban Brazilian adults. *Rev Saude Publica* 1993;27:271-6.

70. Smith GT, Hohlsteing LA, Atlas JG. Accuracy of self-reported weight: Covariation with binger or restrainer status and eating disorder symptomatology. *Addict Behav* 1992;17:1-8.

71. Spencer EA, Appleby PN, Davey GK, Key TJ. Validity of self-reported height and weight in 4808 EPIC-Oxford participants. *Public Health Nutr* 2002;5:561-5.

72. Stewart, AL. The reliability and validity of self-reported weight and height. *J Chronic Dis* 1982;35:295-309.

73. Stewart AW, Jackson RT, Ford MA, Beaglehole R. Underestimation of relative weight by use of self-reported height and weight. *Am J Epidemiol* 1987;125:122-6.

74. Tell GS, Jeffery RW, Kramer FM, Snell MK. Can self-reported body weight be used to evaluate long-term follow-up of a weight-loss program? *J Am Diet Assoc* 1987;87:1198-201.

75. Tienboon P, Wahlqvist ML, Rutishauser IH. Self-reported weight and height in adolescents and their parents. *J Adolesc Health* 1992;13:528-32.

76. Vailas LI, Nitzke SA. Self-reported versus measured weight and height in an older adult meal program population. *J Gerontol A Biol Sci Med Sci* 1998;53:M481-M483.

77. van der Voort DJ, Brandon S, Dinant GJ, van Wersch JW Screening for osteoporosis using easily obtainable biometrical data: Diagnostic accuracy of measured, self-reported and recalled BMI, related costs of bone mineral density measurements. *Osteoporos Int* 2000;11:233-9.

78. Wada K, Tamakoshi K, Tsunekawa T, Otsuka R, Zhang H, Murata C, et al. Validity of self-reported height and weight in a Japanese workplace population. *Int J Obes* 2005;29:1093-9.

79. Wing RR, Epstein LH, Ossip DJ, LaPorte RE. Reliability and validity of self-report and observers' estimates of relative weight. *Addict Behav* 1979;4:133-40.

80. Ziebland S, Thorogood M, Fuller A, Muir J. Desire for the body normal: Body image and discrepancies between self-reported and measured height and weight in a British population. *J Epidemiol Community Health* 1996;50:105-6.

81. Clemente L, Moreira P, Oliveira B, Almeida MD Body mass index: Sensitivity and specificity. *Acta Med Port* 2004;17:353-8. Epub 2004 Dec 20.

82. Quiles IJ, Vioque J. [Validity of notified anthropometric data for determining the prevalence of obesity]. [Spanish]. *Medicina Clinica* 1996;106:725-9.

83. Wang DY, DeStavola BL, Allen DS, Fentiman IS, Bulbrook RD, Hayward JL, et al. Breast cancer risk is positively associated with height. *Breast Cancer Res Treat* 1997;43:123-8.

84. O'Connell KJ, Osborne LM, Westhoff C. Measured and reported weight change for women using a vaginal contraceptive ring vs. a low-dose oral contraceptive. *Contraception* 2005;72:323-7.

85. Black DR, Taylor AM, Coster DC. Accuracy of self-reported body weight: stepped approach to model component assessment. *Health Educ Res* 1998;13:301-7.

86. Zhang J, Feldblum PJ, Fortney JA. The validity of self-reported height and weight in perimenopausal women. *Am J Public Health* 1993;83:1052-3.

87. Brunet MEI. Development and validation of the Screening Test for at Risk Individuals for Eating Disorders (STARVED). *Dissertation Abstracts International Section A: Humanities and Social Sciences* 2003 Mar;63(8A).

88. Bowman RL, DeLucia JL. Accuracy of self-reported weight: A meta analysis. *Behavior Therapy* 1992;23:637-55.

89. Engstrom JL. Assesment of the reliability of physical measures. *Res Nurs Health* 1988;11:383-9.

Appendix A: Data Extraction Form

1. Information about reviewer

- Date: - Initials:

2. Information about the study

- Authors:

- Language:

3. Study Design:

Experimental Observational

Specify type:

Published	Yes	No	Can't Tell
Dissertation or thesis	Yes	No	Can't Tell

4. Verification of Study Eligibility

a. Design - observational or experimental:	Yes	No	Can't Tell
b. Participants – 18 and over:	Yes	No	Can't Tell
c. Direct comparison of objective and subjective measures	Yes	No	Can't Tell
d. Proxy reporting	Yes	No	Can't Tell
e. Outcomes			
- Self-reported weight (any collection method) (e.g., telephone, in person)	Yes	No	Can't Tell
- Measured weight (any type of scale)	Yes	No	Can't Tell
- Self-reported height (any collection method) (e.g., telephone, in person)	Yes	No	Can't Tell
- Measured height (any method) (e.g., tape measure, ruler)	Yes	No	Can't Tell

5. Study Characteristics

a. Participants:

b. Exclusion criteria:

c. Sample Size - Enrolled:
- Analyzed:

- Reasons for exclusion:

d. Variables (and definition):
- Method of weight measurement:

- Method of self-reported weight:

- Method of height measurement:

- Method of height self-report:

- Summary measures and method of calculation
(e.g., BMI, waist to hip ratio, skinfold)

c. Setting:

7. Results:

Tables and Figures

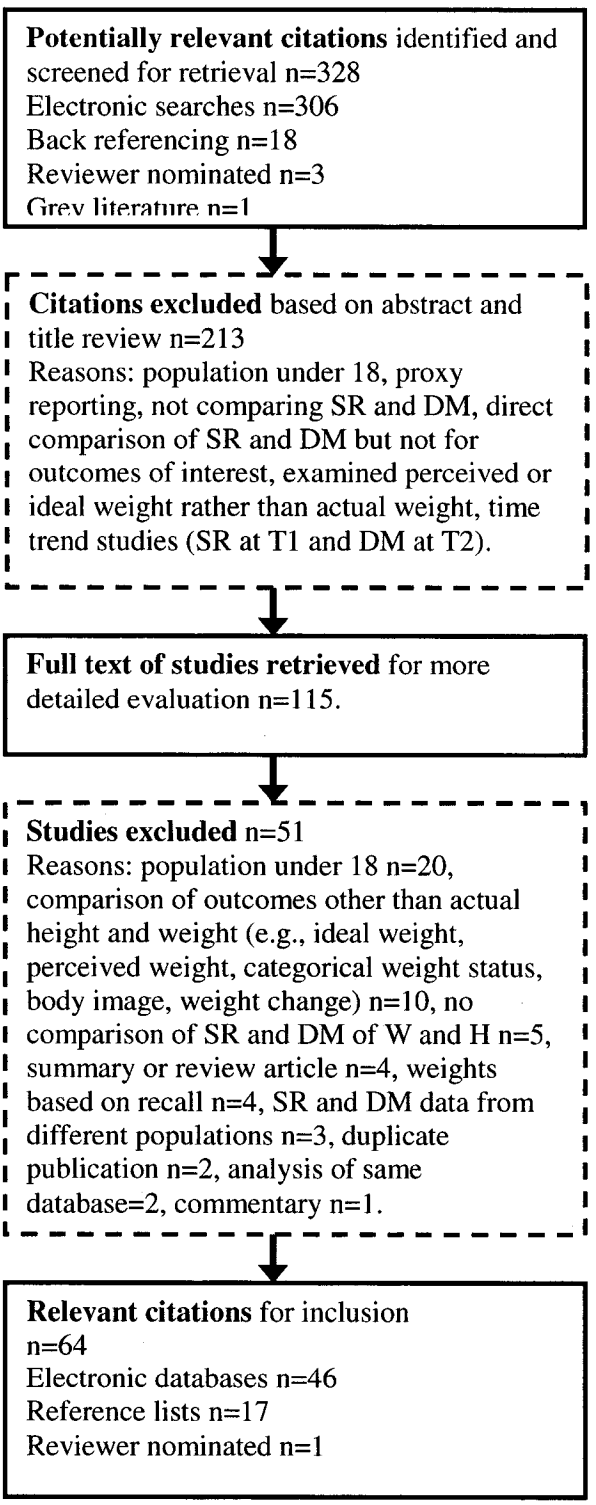

Figure 1: Results of the literature search. DM, direct measure; H, height; SR self-report; W, weight.

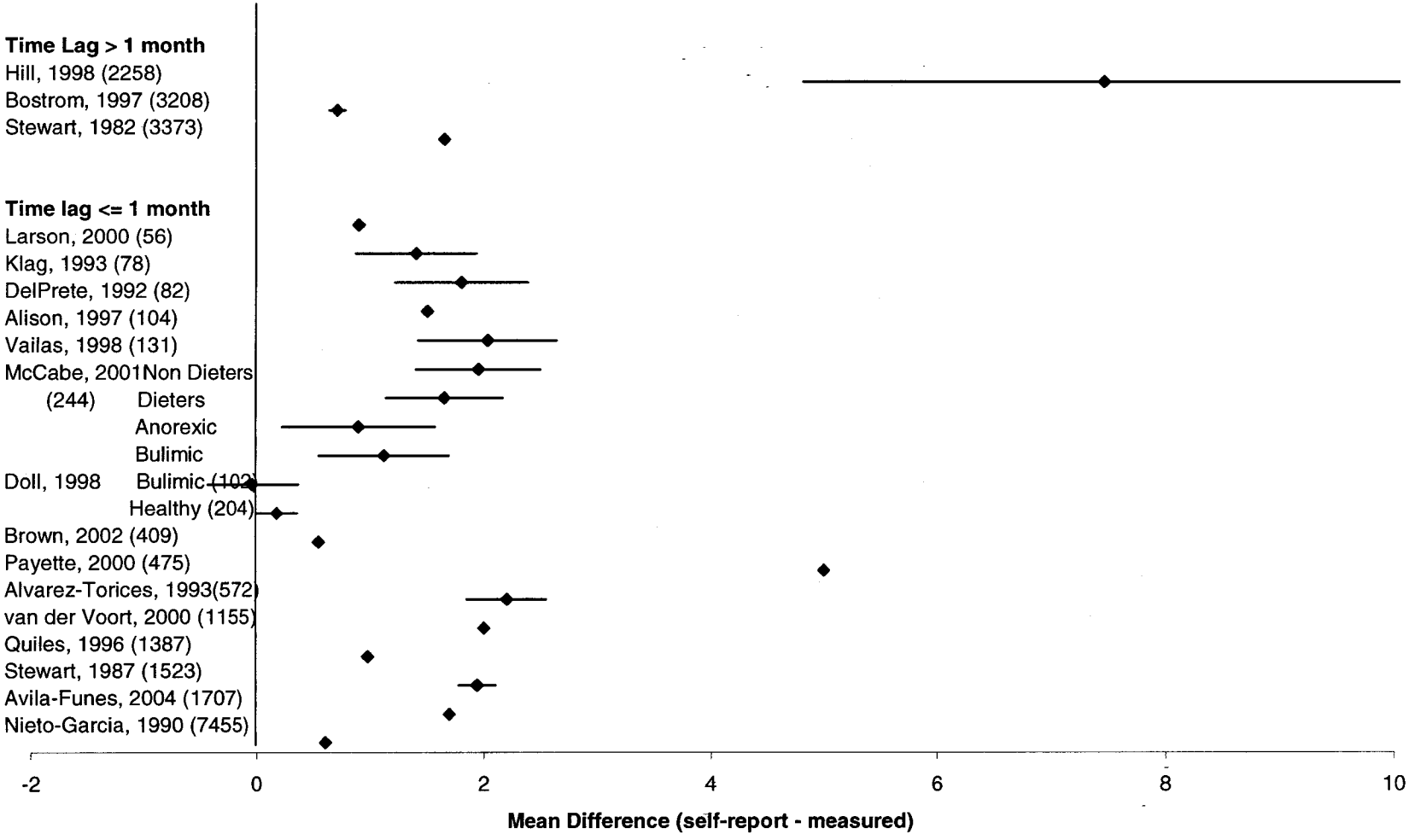

Figure 2: Mean differences in height for studies with available data on the total sample, in ascending order by sample size.

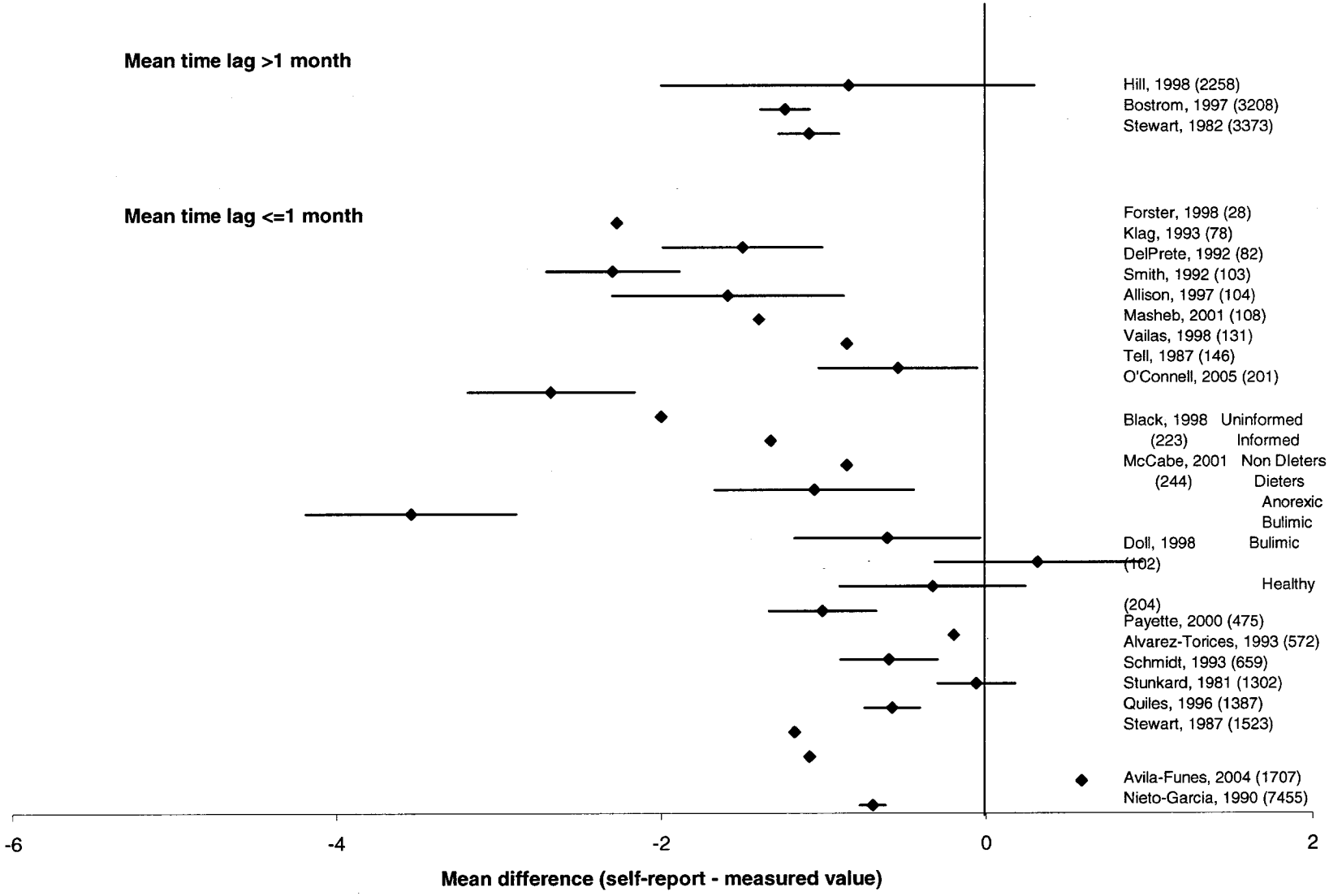

Figure 3: Mean differences in weight for studies with available data on the total sample, in ascending order by sample size.

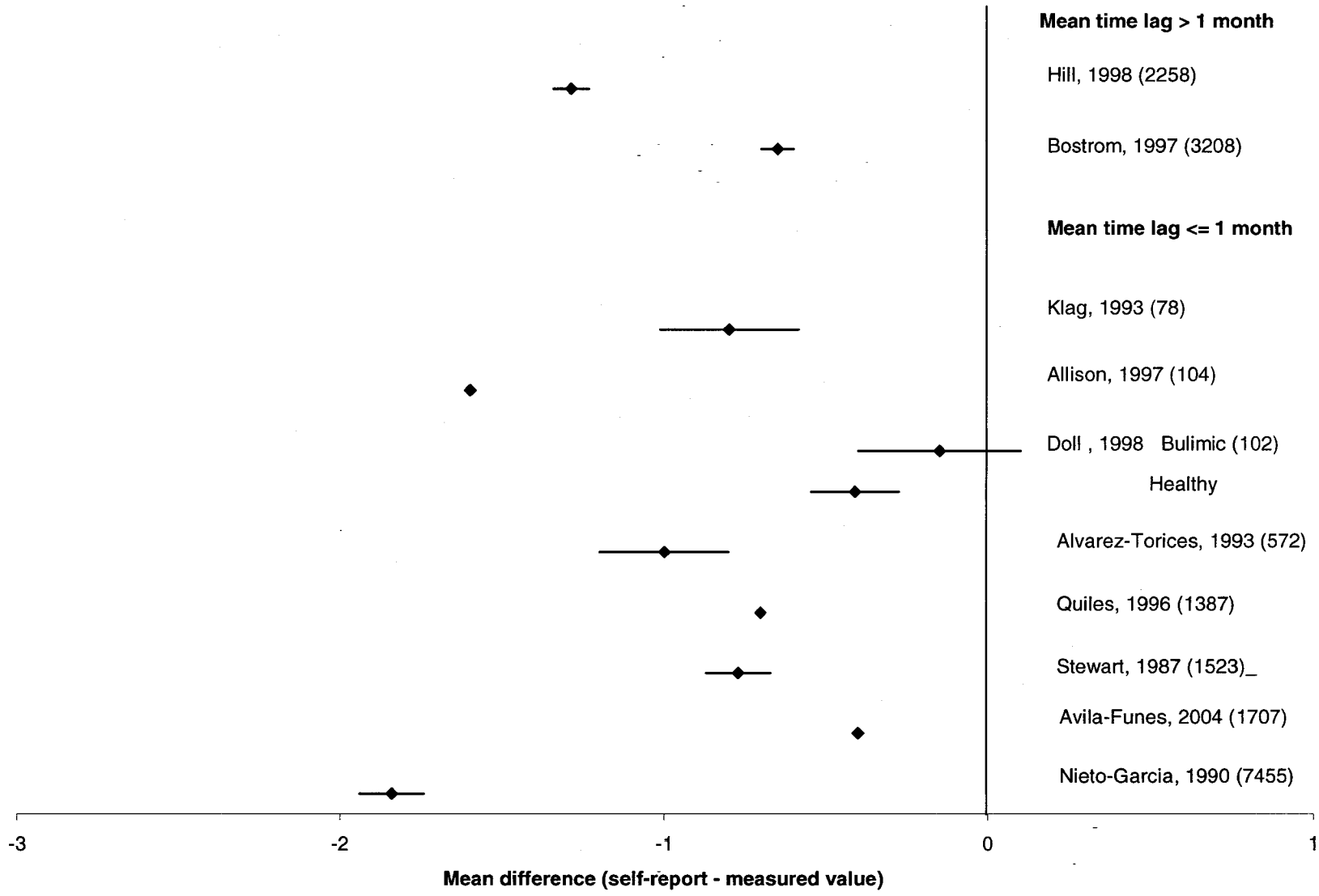

Figure 4: Mean differences in BMI for studies with available data on the total sample, in ascending order by sample size.

Table 1: MEDLINE search strategy

1	objective measure$.mp.
2	direct measure$.mp.
3	physical measure$.mp.
4	measured weight.mp.
5	measured height.mp.
6	exp Self Concept/
7	self report$.mp.
8	body weight/ or body weight changes/ or overweight/ or thinness/
9	Body Height/
10	height.mp.
11	weight.mp. or "Weights and Measures"/
12	1 or 2 or 3 or 4 or 5
13	6 or 7
14	8 or 9 or 10 or 11
15	12 and 13 and 14
16	limit 15 to humans

Table 2: Study and participant characteristics

ID	Study - First Author	Sample Size	Age Range or (Mean ±SD)	Population	Location	Outcomes	Time Lag	Measurement Order
1	Allison 1997	104	10-67	Obese	USA	H,W,BMI	None	SR first
2	Alvarez-Torices 1993	572	18+	General population	Spain	H,W,BMI	NR	SR first
3	Avila-Funes 2004	1707	24-95	Adult population	Mexico	H,W,BMI	None	SR first
4	Black 1998[E]	223	18-82	General population	USA	W	None	SR first
5	Bolton-Smith 2000	1836	25-64	Patients from GP registers	Scotland	H,W,BMI	2 weeks	SR first
6	Booth 2000	1140	18-78	Adults	South Australia	H,W,BMI	2 weeks	SR first
7	Bostrom 1997	3208	18-84	Adults	Sweden	H,W,BMI	4-6 months	SR first
8	Brown 2002	409	19-67	Adults	USA	H	None	SR first
9	Brunet 2003[DU]	44	18-23	Female university athletes	USA	H,W	None	SR first
10	Cash 1989	133	18-52	Female college students	USA	W	10 min.	group 1- SR first; group 2 DM first
11	Chor 1999	322	38.5	Bank employees	Brazil	H,W,BMI	None	SR first
12	Clemente 2004*	380	18-30	University students	Portugal	H,W,BMI	NA	NA
13	de Araujo 2003	193	16-73[#]	Attendees at cardiology congress	Brazil	H,W,BMI	None	SR first
14	DelPrete 1992	82	18+	Adult - Weight loss program participants	USA	H,W	2.5 weeks	SR first
15	Doll 1998	102 BN 204 HC	16-35	General population, BN and healthy	UK	H,W,BMI	NR	SR first
16	Fonseca 2004	3713	22-70	University employees	Brazil	H,W,BMI	None	SR first
17	Forster 1998	28	NR	University employees from worksite weight control program	USA	W	NR	SR first
18	Giles 1991	7999	Male 22.2±4.6 Female: 23.1±5.4	US army personnel	USA	H	NR	NR
19	Gunnell 2000	257	56-78	Elderly	England, Scotland	H,W,BMI	NR	half prior and half after
20	Hensley 1998	59	20.3	College students	USA	H	None	SR first
21	Hill 1998	2258	16-64	Family Health Services Authority register	England	H,W,BMI	1-4 months	SR first
22	Imrhan 1996	469	18-46	College students	USA	H,W	None	SR first

ID	Study - First Author	Sample Size	Age Range or (Mean ±SD)	Population	Location	Outcomes	Time Lag	Measurement Order
23	Jacobson 2001	62	mean 19.9	College students	USA	H,W,BMI	None	SR first
24	Jalkanen 1987	11880	30-64	General population	Finland	W	10 days	SR first
25	Kinney 1988	167	mean 59.0±13.6	Male Veterans Administration patients	USA	W	1 month	DM first
26	Klag 1993	78	49-75	Physicians	USA	H,W,BMI	None	SR first
27	Kuczmarski 2001	16573	20+	General population	USA	H,W,BMI	2-4 weeks	SR first
28	Kuskowska-Wolk 1989	301	16-84	Patients presenting for medical appointment	Sweden	H,W,BMI	None	SR first
29	Kuskowska-Wolk 1992	3390	18-84	Adults	Sweden	H,W,BMI	4-6 months	SR first
30	Lackland 1990	3202	NR	Adults	USA	H,W,BMI	NR	NR
31	Larson 2000	56	18-59	Healthy	USA	H,W	1 week	SR first
32	Lawlor 2002	1310	60-79	Women	England, Scotland, Wales	W	6 weeks	SR first
33	LeJarraga 1995	144	20-66	Parents of children attending growth clinics	Argentina	H	None	SR first
34	Masheb 2001	108	21-61	Adults with binge eating disorder	USA	W	None	SR first
35	McCabe 2001	81 ED, 163 D and ND Controls	mean range 20.0-26.1	Hospital eating disorders program (ED), College students (Controls)	Canada	H,W	2-5 days	SR first
36	Nakamura 1999	354	20-42	Female employees of a computer assembly factory	Japan	H,W,BMI	1 week	SR first
37	Nawaz 2001	95	30-65	Overweight or obese women	USA	H,W,BMI	2 weeks	SR first
38	Niedhammer 2000	7350	35-50	Employees of Electricité de France-Gaz de France	France	H,W,BMI	< 6 months	NR
39	Nieto-Garcia 1990	7455	20-79	60% hyperlipidemic	North America	H,W,BMI	None	SR first
40	O'Connell 2005[L]	201	18-40	Women with regular menstrual cycles	USA	W	None	SR first

ID	Study - First Author	Sample Size	Age Range or (Mean ±SD)	Population	Location	Outcomes	Time Lag	Measurement Order
41	Palta 1982	1344	30-69	Diastolic blood pressure >95mmHG	USA	H,W	3 weeks	SR first
42	Payette 2000	475	mean 81.8	Cognitively intact and impaired elderly	Canada	H,W	None	SR first
43	Pirie 1981	3407	20-59	Caucasians	USA	H,W	1 month	SR first
44	Quiles 1996**	1387	15+	General population	Spain	W,H,BMI	NA	NA
45	Roberts 1995	1622	18-64	Adults	Wales	H,W	3-8 months	SR first
46	Rona 1989	68 couples	NR	Afro-Caribbean, Asian and Caucasian couples	Britain	H	NR	SR first
47	Rowland 1990	11,284	20-74	Adults	USA	H,W	2-6 weeks	SR first
48	Santillon 2003	961	mean 42±14	Asthmatic outpatients, workers and families	Mexico	H,W,BMI	NR	NR
49	Schlichting 1981	752	16-66	Insurance applicants	Denmark	H,W	NR	SR first
50	Schmidt 1993	659	15-64	Adults	Brazil	W	NR	SR first
51	Smith 1992	103	mean 23.1± 6.7	Undergraduate women, 2 with BN	USA	W	None	SR first
52	Spencer 2002	4808	35-76	Middle age adults	Britain	H,W,BMI	few weeks	SR first
53	Stewart 1982	3373	14-61	Families	USA	H,W	< 3 months	SR first
54	Stewart 1987	1523	35-65	Adult Caucasians	New Zealand	H,W	None	SR first
55	Stunkard 1981	1302	mean 40.0±.6 (USA), mean range 30.2-47.5 (Denmark)	Insurance applicants (Denmark), broad range of persons (USA)	Denmark, USA	W	NR	SR first
56	Tell 1987	146	28-63	Former weight control program participants	USA	W	2-64 days	SR first
57	Tienboom 1992	213 Families 135 F 190 M	mean 44.8±5.8 F mean 41.8±5.3 M	Parents of 15 year olds	Australia	H,W,BMI	None	SR first

ID	Study - First Author	Sample Size	Age Range or (Mean ±SD)	Population	Location	Outcomes	Time Lag	Measurement Order
58	Vailas 1998	131	60 and over	Meal program participants	USA	H,W	None	SR first
59	van der Voort 2000	1155	50-80	Postmenopausal women	Netherlands	H	None	SR first
60	Wada 2005	5401	35-64	Public servants	Japan	H,W,BMI	NR	SR first
61	Wang 1997	2731	NR	Females	England	H	NR	NR
62	Wing 1979	S1:78	NR	Psychology undergraduate students	USA	H,W	None	SR first
		S2:118	18-89	Attendees at health fair	USA	H,W		
63	Zhang 1993[L]	352	40-54	Caucasian perimenopausal women	USA	H,W	NR	SR first
64	Ziebland 1996	2205	35-64	Adults	Britain	H,W	NR	SR first

Notes:
[#]This was the range for a larger sample of 844; this study includes only a sub sample of these cases

*Portuguese, **Spanish, [D] Dissertation, [U] Unpublished, [E] Experimental Design, [L] Letter

Abbreviations: GP General Practitioner, BN Bulimia Nervosa, HC Healthy Controls, NR Not Reported, ED Eating Disorder (Anorexia or Bulimia Nervosa), D Dieter, ND Non Dieter, S Study, SR Self Report, DM Direct Measure, NA Not Assessed, F Fathers, M Mothers.

Table 3: Quality assessment of studies

Observational studies: Assessed with Downs and Black quality assessment tool [19]

ID	Study	Quality Scores (value)			
		Reporting (8)	External validity (3)	Internal validity (4)	Total (15)
1	Allison 1997	7	1	4	12
2	Alvarez-Torices 1993	8	3	4	15
3	Avila-Funes 2004	8	2	4	14
5	Bolton-Smith 2000	7	2	3	12
6	Booth 2000	5	2	4	11
7	Bostrom 1997	6	2	4	12
8	Brown 2002	7	1	4	12
9	Brunet 2003	6	1	4	11
10	Cash 1989	5	0	3	8
11	Chor 1999	6*	1	4	11
13	de Araujo 2003	6	1	4	11
14	DelPrete 1992	7	3	4	14
15	Doll 1998	7	1	3	11
16	Fonseca 2004	7	2	4	13
17	Forster 1998	6	2	3	11
18	Giles 1991	5	0	3	8
19	Gunnell 2000	8	2	3	13
20	Hensley 1998	5	1	4	10
21	Hill 1998	4	2	3	9
22	Imrhan 1996	5	1	4	10
23	Jacobson 2001	7	1	4	12
24	Jalkanen 1987	7	2	4	13
25	Kinney 1988	7	1	4	12
26	Klag 1993	8	1	4	13
27	Kuczmarski 2001	6	2	3	11
28	Kuskowska-Wolk 1989	8	3	4	15
29	Kuskowska-Wolk 1992	7	2	3	12
30	Lackland 1990	3	0	3	6
31	Larson 2000	6	0	3	9
32	Lawlor 2002	8	2	4	14
33	LeJarraga 1995	7	2	4	13
34	Masheb 2001	7	1	3	11
35	McCabe 2001	7	1	4	12
36	Nakamura 1999	7	1	4	12
37	Nawaz 2001	8	0	4	12
38	Niedhammer 2000	7	2	3	12
39	Nieto-Garcia 1990	5*	2	3	10
41	Palta 1982	6	1	4	11
42	Payette 2000	8	1	4	13
43	Pirie 1981	6	2	4	12
45	Roberts 1995	5*	0	3	8
46	Rona 1989	7	1	3	11
47	Rowland 1990	7	3	3	13
48	Santillon 2003	5	0	3	8

Observational studies: Assessed with Downs and Black quality assessment tool [19]

ID	Study	Quality Scores (value)			
		Reporting (8)	External validity (3)	Internal validity (4)	Total (15)
49	Schlichting 1981	5	1	3	9
50	Schmidt 1993	7	1	4	12
51	Smith 1992	8	1	4	13
52	Spencer 2002	7	1	4	12
53	Stewart 1982	6	2	3	11
54	Stewart 1987	7	3	4	14
55	Stunkard 1981	5	1	3	9
56	Tell 1987	7	1	4	12
57	Tienboom 1992	7	1	4	12
58	Vailas 1998	7	2	4	13
59	van der Voort 2000	7*	1	4	12
60	Wada 2005	6	3	3	12
62	Wing 1979	5	1	3	9
64	Ziebland 1996	5*	0	3	8

Experimental studies: Assessed with Jadad quality assessment tool [18]

		Quality Score (Max 5)
4	Black 1998[E]	1
40	O'Connell 2005[E]	3

Not assessed

		Reason
12	Clemente 2004	foreign language
44	Quiles 1996**	foreign language
61	Wang 1997	not enough data provided to assess quality
63	Zhang 1993[L]	not enough data provided to assess quality

* These studies reported no p values - and were therefore given a score of 0 for item 10 on reporting probability values.

Table 4: Mean differences in height (self report - direct measure)

ID	Study		Mean Difference (cm)					
			Total	SD (95%CI)	Men	SD (95%CI)	Women	SD (95%CI)
General Population								
2	Alvarez-Torices 1993		2.2	(1.9,2.6)	2.3	(1.9,2.7)	2.2	(1.6,2.7)
3	Avila-Funes 2004		1.7	-	1.2	-	2.2	-
5	Bolton-Smith 2000		-	-	-1.3	2.5	-1.7	2.4
6	Booth 2000		-	-	-	-	-	-
7	Bostrom 1997		0.7	(.6,.8)	0.6	-	0.8	-
8	Brown 2002		0.6	-	-	-	-	-
27	Kuczmarski 2001	20-59 yrs	-	-	1.0	0.1	0	0.1
29	Kuskowska-Wolk 1992		-	-	0.5	-	1.0	-
30	Lackland 1990		-	-	-	-	-	-
31	Larson 2000		0.9	-	-	-	-	-
44	Quiles 1996		1.0	-	0.6	-	1.4	-
45	Roberts 1995		-	-	1.4	2.3	0.7	2.8
47	Rowland 1990		-	-	1.4	2.6	0.6	2.8
52	Spencer 2002		-	-	1.2	2.6	0.6	2.7
54	Stewart 1987		1.9	(1.8,2.1)[+]	2.1	-	1.6	-
61	Wang 1997		-	-	-	-	-	-
64	Ziebland 1996	BMI<20	-	-	0.6	-	0.4	-
		BMI 20-24	-	-	0.9	(.6,1.2)	0.9	(.5,1.2)
		BMI25-29	-	-	1.0	(.7,1.2)	1.1	(.9,1.4)
		BMI>30	-	-	1.6	(1.0,2.1)	1.6	(1.2,2.0)
Overweight or weight loss participants								
1	Allison 1997		1.5	-	1.7	-	1.5	-
14	DelPrete 1992		1.8	2.7	3.0	2.5	1.3	2.6
37	Nawaz 2001	<=25BMI<30	NA	NA	NA	NA	0.1	0.4
		30<=BMI<=35	NA	NA	NA	NA	0	0.9
		35<BMI<=40	NA	NA	NA	NA	0.3	1.0
		BMI>40	NA	NA	NA	NA	0.9	0.9
Special populations								
13	de Araujo 2003		-	-	1.0	-	0.6	-
21	Hill 1998		7.5	(4.8,10.0)	-	-	-	-
26	Klag 1993		1.4	0.3**	-	-	-	-
28	Kuskowska-Wolk 1989		-	-	0.9	-	1.9	-
39	Nieto-Garcia 1990**		0.6	(.6,.6)	0.9	(.9,1.0)	0.3	(.3,.4)
41	Palta 1982		-	-	2.3	2.7	0.9	2.4
42	Payette 2000		5.0	-	5.0	-	5.0	-
43	Pirie 1981		-	-	0.2	-	-0.3	-
48	Santillon 2003		-	-	-	-	-	-
49	Schlichting 1981		-	-	-	-	-	-
58	Vailas 1998		2.0	3.6	1.5	3.8	2.3	3.6

ID	Study	Mean Difference (cm)					
		Total	SD (95%CI)	Men	SD (95%CI)	Women	SD (95%CI)
Parents/families							
33	LeJarraga 1995						
	Private growth clinic	-	-	0.9	2.0	1.1	1.9
	Public growth clinic	-	-	-0.3	2.6	0.9	3.1
46	Rona 1989						
	Afro Caribbean	-	-	0.2	6.5	2.9	11.2
	Asian	-	-	0.6	4.2	4.5	6.2
	Caucasian	-	-	3.9	8.8	-0.1	2.7
53	Stewart 1982	1.7	1.0	-	-	-	-
57	Tienboom 1992	-	-	2.9	2.2	1.7	2.3
Eating Disorders							
15	Doll 1998 Healthy Controls	0.2	(.1,.5)	-	-	-	-
	Bulimia Nervosa	0	(-.5,.4)	-	-	-	-
35	McCabe 2001 Anorexics	0.9	2.0	-	-	-	-
	Bulimics	1.1	2.0	-	-	-	-
	Dieters	1.7	2.3	-	-	-	-
	Nondieters	2.0	2.6	-	-	-	-
Students							
9	Brunet 2003	NA	NA	NA	NA	-	-
12	Clemente 2004	-	-	1.2	2.2	2.7	2.3
20	Hensley 1998	-	-	-0.7	-	-0.9	-
22	Imrhan 1996	-	-	-	-	-	-
23	Jacobson 2001	-	-	1.0	-	0	-
62	Wing 1979	-	-	-	-	-	-
Employees							
11	Chor 1999 Center Branch	-	-	0	-	0	-
	Internal Services Unit	-	-	0	-	0	-
	General Board	-	-	0	-	0	-
	Ilha do Governador Branch	-	-	0	-	0	-
16	Fonseca 2004	-	-	0.2	3.2	1.1	3.0
18	Giles 1991	-	-	2.8	-	1.0	-
36	Nakamura 1999	NA	NA	NA	NA	-0.1	0.7
38	Niedhammer 2000	-	-	0.4	(.4,.4)	0.4	(.5,.3)
60	Wada 2005	-	-	0.1	(0,.1)	0	(0,.1)
Elderly							
19	Gunnell 2000	-	-	2.1	(1.6,2.7)	1.7	(1.3,2.2)
27	Kuczmarski 2001 60+ yrs	-	-	2.7	0.1	2.5	0.1
59	van der Voort 2000	2.0	-	-	-	-	-
63	Zhang 1993[L]	-	-	-	-	-	-

+ represents the 99% CI, ** standard error, *** mean relative errors, - not reported, NA not applicable

Table 5: Mean differences in weight (self-report - direct measure)

ID	Study		Mean Difference (kg) Total	SD (95%CI)	Men	SD (95%CI)	Women	SD (95%CI)
General Population								
2	Alvarez-Torices 1993		-0.6	(-.9,-.3)	-0.2	(-.5,.1)	-0.9	(-1.4,-.5)
3	Avila-Funes 2004		0.6	-	0.4	-	0.7	-
5	Bolton-Smith 2000		-	-	-0.6	3.5	-1.0	2.6
6	Booth 2000		-	-	-	-	-	-
7	Bostrom 1997		-1.2	(-1.4,-1.1)	-0.7	-	-1.6	-
24	Jalkanen 1987		-	-	-0.4	3.0	-0.6	2.0
27	Kuczmarski 2001	20-59yrs	-	-	0.4	0.1	-1.5	0.1
29	Kuskowska-Wolk 1992		-	-	-1.0	-	-1.5	-
30	Lackland 1990		-	-	-	-	-	-
31	Larson 2000		-	-	-0.5	-	-	-
44	Quiles 1996		-1.1	-	-1.0	-	-1.2	-
45	Roberts 1995		-	-	-0.2	2.8	-1.1	2.6
47	Rowland 1990		-	-	0.4	3.0	-1.0	3.0
50	Schmidt 1993		-0.1	3.2	0.3	-	-0.3	-
52	Spencer 2002				-1.9	2.9	-1.4	2.5
54	Stewart 1987		-0.6	(-.8,-.4)+				
55	Stunkard 1981	US site	-1.2	-	-	-	-	-
		Danish site <40	-	-	-1.1	-	-1.6	-
		Danish site>=40	-	-	-1.3	-	-2.5	-
64	Ziebland 1996	BMI<20	-	-	2.6	-	0.8	-
		BMI 20-24	-	-	-0.1	-	-0.5	(-.7,-.2)
		BMI 25-29	-	-	-1.2	(-1.6,-.9)	-1.8	(-2.2,-1.5)
		BMI>30	-	-	-3.2	(-4.3,-2)	-3.3	(-4.2,-2.4)
Overweight or weight loss participants								
1	Allison 1997		-1.4	-	-3.0	-	-0.9	-
14	DelPrete 1992		-2.3	1.9	-2.1	1.5	-2.4	1.9
37	Nawaz 2001	<=25BMI<30	NA	NA	NA	NA	-0.3	4.6
		30<=BMI<=35	NA	NA	NA	NA	-1.6	5.7
		35<BMI<=40	NA	NA	NA	NA	-6.5	0.2
		BMI>40	NA	NA	NA	NA	-5.2	10.1
56	Tell 1987		-2.7	3.2	-2.8	3.4	-2.3	2.5
Elderly								
19	Gunnell 2000		-	-	-1.9	(-2.7,-1.2)	-1.2	(-1.7,-.8)
27	Kuczmarski 2001	60+yrs	-	-	0.5	0.1	-0.6	0.1
32	Lawlor 2002		NA	NA	NA	NA	-1.0	(-1.1,-.8)
63	Zhang 1993		-	-	-	-	-	-

ID	Study		Mean Difference (kg)					
			Total	SD (95%CI)	Men	SD (95%CI)	Women	SD (95%CI)
Eating Disorders								
15	Doll 1998	Healthy Controls	-1.0	(-1.3,-0.7)	-	-	-	-
		Bulimia Nervosa	-0.3	(-.9,.2)	-	-	-	-
34	Masheb 2001		-0.9	-	-	-	-	-
35	McCabe 2001	Anorexics	0.3	1.9	-	-	-	-
		Bulimics	-0.6	2.0	-	-	-	-
		Dieters	-3.5	2.9	-	-	-	-
		Nondieters	-1.1	2.9	-	-	-	-
Students								
9	Brunet 2003		NA	NA	NA	NA	-1.1	1.8
10	Cash 1989		NA	NA	NA	NA	-	-
12	Clemente 2004		-	-	-0.3	3.0	-0.2	2.2
22	Imrhan 1996		-	-	-	-	-	-
23	Jacobson 2001		-	-	0.5	-	-1.9	-
51	Smith 1992		-1.6	3.7	-	-	-	-
62	Wing 1979		-	-	-	-	-	-
Employees								
11	Chor 1999	Center Branch	-	-	-1.1	-	-1.0	-
		Internal Services Unit	-	-	-1.3	-	-0.7	-
		General Board	-	-	-0.7	-	-0.8	-
		Ilha do Governador Branch	-	-	-1.1	-	-0.7	-
16	Fonseca 2004		-	-	-1.0	3.5	-1.1	3.0
17	Forster 1998		-2.3	-	-	-	-	-
36	Nakamura 1999		NA	NA	NA	NA	-0.2	1.8
38	Niedhammer 2000		-	-	-0.5	(-.6,-.5)	-0.9	(-1.0,-.7)
60	Wada 2005		-	-	0	(-.1,0)	0	(-.1,.2)
Parents/Families								
53	Stewart 1982		-1.1	5.5	-	-	-	-
57	Tienboom 1992		-	-	-0.3	3.0	-1.2	2.5
Special Populations								
13	de Araujo 2003		-	-	-0.9	-	-0.1	-
21	Hill 1998		-0.9	(-2.0,.3)	-	-	-	-
25	Kinney 1988		NA	NA	-3.0	4.5	NA	NA
26	Klag 1993		-1.5	0.3**	-	-	-	-
28	Kuskowska-Wolk 1989		-	-	-0.5	-	-0.6	-
39	Nieto-Garcia 1990***		-0.7	(-.8,-.6)	-0.3	(-.5,-.2)	-1.0	(-1.1,-.9)
41	Palta 1982		-	-	-1.5	3.2	-2.4	4.0
42	Payette 2000		-0.2	-	0.2	-	-0.3	-
43	Pirie 1981		-	-	-0.5	3.3	-1.9	3.0
48	Santillon 2003		-	-	-	-	-	-

ID	Study	Mean Difference (kg)					
		Total	SD (95%CI)	Men	SD (95%CI)	Women	SD (95%CI)
58	Vailas 1998	-0.5	3.1	-0.3	3.6	-0.7	2.9
Experimental design							
4	Black 1998 Informed group	-0.9	-	-	-	-	-
	Uniformed group	-1.3	-	-	-	-	-
40	O'Connell 2005	-2.0	-	-	-	-	-

Table 6: Mean differences in Body Mass Index (self report - direct measure)

ID	Study	Mean Difference (kg/m^2)					
		Total	SD (95%CI)	Men	SD (95%CI)	Women	SD (95%CI)
General population							
2	Alvarez-Torices 1993	-1.0	(-1.2,-.8)	-0.8	(-1.0,-.6)	-1.2	(-1.5,-.9)
3	Avila-Funes 2004	-0.4	-	-0.3	-	-0.5	
5	Bolton-Smith 2000	-	-	0.2	1.4	0.2	1.3
6	Booth 2000	-	-	-	-	-	-
7	Bostrom 1997	-0.7	(-.7,-.6)	-0.4	-	-0.9	-
27	Kuczmarski 2001 20-59yrs	-	-	-0.3	0	0	0
29	Kuskowska-Wolk 1992	-	-	-0.4	-	-0.8	-
30	Lackland 1990	-	-	-	-	-	-
44	Quiles 1996	-0.7	-	0.5	-	0.9	-
52	Spencer 2002	-	-	-1.0	1.2	-0.7	1.3
54	Stewart 1987	-0.8	(-.8,-.6)+	-	-	-	-
Overweight or weight loss participants							
1	Allison 1997	-1.6	-	-2.0	-	-1.4	-
37	Nawaz 2001 <=25BMI<30	NA	NA	NA	NA	-0.2	0.7
	30<=BMI<=35	NA	NA	NA	NA	-0.3	1.3
	35<BMI<=40	NA	NA	NA	NA	-1.5	1.9
	BMI>40	NA	NA	NA	NA	-2.2	2.2
Students							
12	Clemente 2004	-	-	-0.4	1.1	-0.8	1.1
23	Jacobson 2001	-	-	-0.2	-	-0.6	-
Employees							
11	Chor 1999 Center Branch	-	-	-0.2	-	-0.3	-
	Internal Services Unit	-	-	-0.6	-	-0.2	-
	General Board	-	-	-0.2	-	-0.3	-
	Ilha do Governador Branch	-	-	-0.4	-	-0.3	-
16	Fonseca 2004			-0.4	1.6	-0.8	1.6
36	Nakamura 1999	NA	NA	NA	NA	-0.1	0.8
38	Niedhammer 2000			-0.3	(-.3, -.3)	-0.4	(-.4,-.5)
60	Wada 2005	-	-	0	(-.1, 0)	0	(-.1,.1)
Elderly							
19	Gunnell 2000	-	-	-1.3	(-1.6,-1.0)	-1.1	(-1.3,-.9)
27	Kuczmarski 2001 60+yrs	-	-	-0.8	0	-0.8	0.0
Special populations							
13	de Araujo 2003	-	-	-	-	-	-
15	Doll 1998 Healthy Controls	-0.4	(-.6,-.3)	-	-	-	-
	Bulimia Nervosa	-0.2	(-.4,.1)	-	-	-	-

ID	Study	Mean Difference (kg/m^2)					
		Total	SD (95%CI)	Men	SD (95%CI)	Women	SD (95%CI)
21	Hill 1998	-1.3	(-1.3,-1.2)	-	-	-	-
26	Klag 1993	-0.8	0.1**	-	-	-	-
28	Kuskowska-Wolk 1989	-	-	-0.4	-	-0.7	-
39	Nieto-Garcia 1990***	-1.8	(-1.9,-1.7)	-2.1	(-2.3,-2.0)	-1.6	(-1.7,-1.4)
48	Santillon 2003	-	-	-	-	-	-
57	Tienboom 1992	-	-	-1.0	1.2	-1.0	1.2

+ represents the 99% CI, ** standard error, *** mean relative errors, - not reported, NA not applicable

Paper 2: The accuracy of self-reported hypertension: A systematic review and meta-analysis

This paper has been published in:

Connor Gorber S, Tremblay M, Campbell N, Hardt J. The accuracy of self-reported hypertension; a systematic review and meta-analysis. *Current Hypertension Reviews* 2008;4:36-62.

The Accuracy of Self-Reported Hypertension: A Systematic Review and Meta-Analysis

Sarah Connor Gorber[1], Mark Tremblay[2], Norm Campbell[3], Jill Hardt[1]

1. Health Information and Research Division, Statistics Canada, Ottawa, Ontario, Canada
2. Physical Health Measures Division, Statistics Canada, Ottawa, Ontario, Canada
3. Departments of Medicine, Community Health Sciences and Pharmacology and Therapeutics, University of Calgary, Calgary, Alberta, Canada

Key Words: hypertension awareness, systematic review, self-report, direct measure, blood pressure

Running Title: Hypertension Awareness

Correspondence: Sarah Connor Gorber, Chief – Health Measures Analysis Section, Health Information and Research Division, Statistics Canada, 24th Floor R.H. Coats Building, 100 Tunney's Pasture Driveway, Ottawa, ON, Canada, K1A 0T6

Tel:
Fax:
Email:

Abstract

Hypertension is a leading cause of morbidity and mortality with surveillance often based on self-reported data. A systematic review and meta-analysis was completed to determine what empirical evidence exists regarding the agreement between measured and self-reported estimates of hypertension. Four electronic databases were searched to identify observational and experimental studies on adult populations. Searching identified 144 studies that examined the relationship between self-reported and directly-measured hypertension. Studies demonstrate that the prevalence of hypertension based on self-report is underestimated and that awareness levels are low and highly variable, ranging from 0 to 97%. Meta-analysis generated overall awareness estimates of 58% at the 140/90 mmHg cut-off and 62% at the 160/95 mmHg cut-off. Awareness levels were consistently higher for females. In many of the reviewed studies, standard guidelines for blood pressure measurement were not being followed or reporting of guideline compliance was poor. Lack of awareness is an important public health issue, yet this review demonstrates that the current practice of surveillance based on awareness alone is insufficient. These data indicate the importance of directly measuring blood pressure according to standardized protocols wherever possible, and increasing population awareness both to increase the validity of self-report and to ensure early diagnosis and treatment.

Background

Hypertension is one of the leading causes of morbidity and mortality in most industrialized countries[1,2] and increasingly in many non-industrialized nations[3]. The World Health Organization (WHO) estimates that hypertension is responsible for about 7.1 million premature deaths and 4.4% of the worldwide disease burden[4].

Effective management of hypertension at the population level depends on the accuracy of information collected on its prevalence, awareness, treatment and control. Although direct measurement of blood pressure is recommended whenever possible, accurate measurement requires specially trained assessors, specialized equipment, appropriate facilities and measurement on more than one occasion, all of which are financially and logistically costly. Population surveys have long relied on self-reports as a practical and cost efficient alternative to direct measurement, but self-reports are limited by an individual's ability to recall and accurately report information about their health. Self-reports depend on a person's awareness of their health condition, which assumes they have been diagnosed, told of the diagnosis, and can recall and report this information on a health survey.

Studies have been examining the validity of self-reported hypertension for decades (e.g., early work on the United States National Health Examination Survey in the 1960s), but to date there has been no systematic assessment of the relationship between self-reported and directly-measured hypertension, and the degree to which people are aware of and able to report their hypertensive status. This information is important, however, in order to ensure that policy makers and clinicians have accurate information upon which to make decisions

and allocate resources for prevention, treatment and research. This paper reviewed 144 studies, representing over 1 million people, to measure the degree of concordance between self-reported and objectively measured hypertension.

Methods

We sought to identify all studies (observational or experimental) that compared direct (objective) measures of hypertension to self-report (subjective) measures in adult populations (18 years and over). Any form of self-reporting technique could be used to solicit the self-reported value (e.g., interviewer administered, self-completed), but data extracted from medical records or provided by a proxy were not included. Direct measures of hypertension included manual or automated devices on any number of occasions in any environment. The maximum allowable time between the reported and measured estimates was two months. Although no language restrictions were imposed in the search, only English language articles were included in the review. Abstracts and letters were included if they provided sufficient details to meet inclusion criteria.

Search Strategy

The following bibliographical databases were searched: Medline (2006 December week 4); EMBASE (2007 week 1); CINAHL (2006 December week 2); and PsycINFO (2007 January week 2). The search strategy used for searching MEDLINE is provided in Table 1, and was modified according to the indexing systems of other databases. The OVID interface was used for all searches. The Cochrane Library was also searched with the terms 'self-report'

and 'blood pressure'. Grey literature (works that have not been peer reviewed), such as published abstracts and conference proceedings, theses, dissertations and government or organizational reports were also identified where possible. One of the authors' (NC) personal databases of hypertension literature was also examined. The bibliographies of key studies selected for the review were examined to identify further studies.

Two independent reviewers screened the titles and abstracts of all studies to identify potentially-relevant articles. Duplicates were removed. The full texts of all studies that met the inclusion criteria were then obtained and reviewed. When disagreements between reviewers occurred, consensus was achieved through discussion.

Standardized data abstraction forms were completed by one reviewer and verified by the other. Information was extracted on the type of study, participant characteristics, sample size, methods of hypertension measurement (device, number of measures and occasions of measurement, assessor, location, positioning, cut-off used, inclusion of medication, type of cuff, restrictions imposed prior to measurement, and the time between the reported and measured estimates) and results including the proportion of hypertensives based on self-report and direct measurement, awareness of hypertensive status (sensitivity) and associated measures of variance. Thirty-seven authors representing 39 studies with missing information were contacted and asked to provide further details about their studies. Information was received from only 9 authors representing 11 studies. Reviewers were not blinded to the authors or journals when extracting data.

Quality Assessment

Two checklists were used to assess study quality. The Downs and Black[5] instrument has been recommended for assessing quality in observational studies in a recent systematic review[6] and other assessments[7] and was employed in this review to assess reporting, external validity and internal validity (bias). The Downs and Black checklist consists of 27 items with a maximum score of 32 points. A modified version of the checklist was employed with items that were not relevant to the studies in this review removed. The adapted checklist consisted of 16 items, including 1-4, 6, 7, 9-13, 15-18 and 20 from the original list, with a maximum possible score of 16 points (higher scores indicate superior quality).

To evaluate the quality of blood pressure measurement in surveys, a specific quality assessment tool was developed (see Table 2). A content expert (NC) developed a draft tool to assess important components of blood pressure measurement. The technical components were in part based on a work group report of minimum standards for assessing blood pressure in physical measures surveys[8] and assessed studies according to the observer training (2 items), quality of equipment (3 items), participant preparation (4 items), and measurement technique (7 items). The draft tool was circulated to two additional content experts and revised based on their input. The quality assessment was undertaken independently by two assessors.

Data Synthesis

Awareness of hypertension (equivalent to sensitivity in epidemiologic terms) was used as the main outcome for this analysis; it is the proportion of all who were measured as hypertensive who self-reported hypertension. In calculating sensitivity values for this review, individuals

who were undergoing treatment for hypertension (medication or other therapy) were considered to be aware of their hypertensive status regardless of whether they reported it or not. Individuals taking medications were also considered to be true hypertensives regardless of their actual blood pressure measurement.

Forest plots (graphical displays of the awareness levels in the individual studies) were constructed to present sensitivity values for each study. In many studies awareness data were not reported, but enough information was provided in order to allow the values to be calculated. Few studies reported confidence intervals; where possible, these intervals were calculated based on the following formula to calculate variances of proportions [$\hat{P}$ (1-$\hat{P}$)/n] which was then used to estimate the 95% confidence intervals. Pooled estimates and their 95% confidence intervals were obtained using the random effects estimator of DerSimonian-Laird[9].

Cochrane's Q was used to test for heterogeneity among studies and the I^2 (squared) index[10] was used to determine the degree of heterogeneity. Sub-group analyses were undertaken to examine results according to sex and hypertension cut-off (140/90 or 160/95 mmHg), which are posited to affect hypertension awareness. Some studies provided awareness data for both hypertension cut-offs. Where studies reported awareness estimates based on only one cut-off, they are included only in the relevant subgroup analyses. Similarly, while some studies reported aggregate as well as sex-disaggregated sensitivity values, many studies provided only an overall estimate or only sex-disaggregated estimates. Pooled sensitivities were calculated for each of the sub-group analyses. Sensitivity analysis was used to determine if

study quality affected results. Funnel plots, which are scatter plots of some measure of study effect (in this case sensitivity levels) against some measure of study information such as sample size or variance (in this case standard errors) were used to assess publication bias. When the resulting funnel shape is symmetrical, concerns that publication or other biases are present in the review are minimized.

Results

Description of Studies

After duplicates were removed 1,107 studies were identified based on the preliminary search of electronic and personal databases, reference lists and grey literature (see Figure 1). Of these, 309 were identified in MEDLINE, 470 in EMBASE, 129 in CINAHL, and 12 from PSYCHINFO. The list of studies was sent to three key informants, two of whom suggested a total of four additional articles and a final article was suggested by an author as part of the author contact process. The full text of 393 articles was retrieved for more detailed assessment of which 144 met the criteria for inclusion[11-152]. Common reasons for excluding studies were a focus on populations under 18 years, not having both directly-measured and self-reported data on the same population, use of medical records as a proxy for blood pressure measurement, and not providing any information on awareness of hypertension status.

Searching identified several articles that analyzed data from large population surveys such as the National Health and Nutrition Examination Survey (NHANES) in the United States, the Canadian Heart Heath Surveys and the WHO Monica Surveys. In cases where several

publications reported analyses from the same data source, only one study per data source was retained in order to avoid double counting. In some cases[23,153] the same author had written two publications using the same source of data and in other cases more than one author was examining the same data source. For example, 15 studies[11-13,154-165] were identified that used NHANES data. Of these Burt and colleagues[11] was kept for the period of NHES to NHANES III, Glover and colleagues[12] for NHANES 1999-2002 and Pappas and colleagues[13] for the Hispanic Health and Nutrition Survey (HHANES) that was conducted in 1982-84. These studies were selected as they had the most comprehensive data and reporting for the outcomes of interest.

Studies that were included were published over a 33 year period from 1973 to 2006. All of the studies used observational designs (e.g., case control, cross-sectional) and all were published as journal articles except for two letters[81,124] and three government publications[59,67,103]. The primary aim of the articles differed but all compared direct measures of hypertension to self-reported awareness of hypertension status. Most studies reported results using unstandardized data but eight studies presented standardized (by age and/or sex) estimates[37,38,43,57,81,119,121,134]. Two studies[14,31] asked respondents to estimate actual blood pressure values as well as their hypertensive status. Correlations between reported and measured systolic and diastolic pressure were poor with an overall correlation of 0.35[14] and correlations of 0.26 and 0.25 (systolic and diastolic respectively) in men and of 0.62 and 0.43 in women[31].

Participant Characteristics

Study and participant characteristics are described in Tables 3 and 4. This discussion excludes studies where required information was missing or could not be inferred from the paper. This review covered a broad international scope with research from over 50 countries represented (Table 3). Sample sizes ranged from a low of 42[77] to a high of over 1 million respondents[133]. Although the focus of the review was on those aged 18 years and over, studies that had a range of ages less than 18 years were not excluded as long as the mean age of the population was over 18 years.

The delay between reported and measured hypertension for the studies that were included ranged from a few minutes to a month (if multiple visits were used to measure blood pressure, the time recorded was the time between the reported value and the last visit). In most studies blood pressure was measured in the home (n = 44) or in a clinical or research setting (n = 43). Fourteen studies collected blood pressure in multiple settings (home, work, clinic or community location).

Most studies used data collected at only one visit to classify hypertensive status, and only thirteen studies had participants return for more than one visit, although only eight of these used measures from both visits to determine the blood pressure value (Table 4). All but ten studies that reported data took more than one measurement of blood pressure, but again not all used more than one measure to determine hypertension status. Of the studies that reported pre-testing conditions, 98% (n= 97 of 99) required that participants rest prior to having their BP measured, although not all rested for 5 minutes as recommended by standard

protocols[166,167]. Of these, 26% (25/97) also imposed additional restrictions on eating, drinking, smoking or exercising, which are also recommended[168]. In almost all studies measurements were taken with the participants in a seated position. Eleven studies specifically indicated that they did not use appropriately-sized cuffs to adjust for different arm circumferences. A combination of manual (usually standard mercury or random zero sphygmomanometers) and automatic (Omron, BpTRU, Dinamap) devices were used but only 14 of the studies described the calibration and accuracy testing of these devices. Phase V Korotkoff sound was used to measure diastolic pressure in 74% of the studies that reported that they employed manual devices (75/101).

Most studies used standard cut-offs of 140/90 mmHg or 160/95 mmHg to determine hypertensive status although not all studies defined hypertension in the same manner. For most studies participants were classified as hypertensive if their blood pressure was above a set cut-off or if they were being treated for high blood pressure (medication or another treatment) regardless of the measured blood pressure value. In 12 studies hypertension was classified based only on the measured blood pressure value and in an additional 11 studies differing combinations of treatment, measured blood pressure and self-reported responses were used to determine hypertensive status.

Quality Assessment

All 144 studies were included in the quality assessment, and all studies were assessed according to the two checklists. Two studies require clarification: Burt[11] presented results from five health examination surveys (NHES I, NHANES I, NHANES II, HHANES, and NHANES III) and an overall assessment was based on whether three or more of the included

surveys met the checklist criteria; Wilber[146] described two studies, but only the reporting relevant to the Baldwin county study was included in this assessment since this study was described in greater detail.

Scores on the Downs and Black checklist ranged between 8 and 15 (out of a possible 16) with a mean of 12.1 ± 1.4. Studies were assessed based on the degree to which they presented data relating to the outcomes studied in this review (self-report and direct measure blood pressure), even when these were secondary outcomes in the studies.

Most studies reported their objectives, main outcomes, the characteristics of participants and the main study findings in good detail. Reporting was less complete on the characteristics of participants who did not provide both self-reported and measured data and in about 31% of studies estimates of variability (standard errors or confidence intervals) were not provided for any outcomes.

In order to score maximum points in external validity, studies must have reported that study participants, both those who were recruited and those who agreed to participate, were representative of the population from which they were recruited and that the staff, places and facilities where participants were interviewed and examined were representative of the conditions in which the majority of patients would be assessed. Credit was given if measures were taken in the home or clinic/hospital environment, but not for measures taken in malls or at health fairs. Seventeen per cent of studies received no points for external validity. Only 15 per cent (22/144) of studies demonstrated that the subjects who were prepared to participate

in the study were representative of the source population (i.e., demonstrated that the main confounding factors were the same in the study sample and the source population).

All studies received three or more of a possible five points in the internal validity category. However, only four[14,15,65,149] received credit for blinding those measuring the main outcomes from the participants' self-reported hypertensive status.

Scores on the checklist of technical elements ranged from 0 to 13 (maximum possible score was 16) with a mean of 5.5 ± 2.8. Most studies (61%) reported that measurements were taken by specifically trained observers, but only 28% stated that the accuracy of the observers was tested. While most studies reported using appropriate measurement devices, in only 14 studies was equipment calibration and accuracy testing described.

Information on blood pressure measurement technique was not consistently reported. Although most studies reported that participants were seated during measurements, only 16 per cent (23/144) noted that the participants' arms were supported at heart level. Three studies[18,96,120] reported assessing both arms in order to determine if one arm recorded a higher blood pressure reading. Most studies also failed to employ, or report, a cuff deflation rate of 2 mmHg per cardiac cycle (or per second). Seven studies received no points on this checklist as they provided little detail on the blood pressure measurement methods.

Quantitative Syntheses

Twenty-five studies reported the prevalence of hypertension as determined through both self-report and direct measurement (see Table 5). In most studies hypertension was underestimated when measured by self-report, with differences between self-reported and measured values (reported-measured) ranging from 1 to 39 percentage points. In four studies[14,75,83,88] reported hypertension was overestimated compared to the measured values for some subgroups (overall estimates and estimates for women). Men consistently underestimated hypertension prevalence compared to the measured values.

Studies were originally grouped according to hypertension cut-off and participant sex in order to decrease clinical and methodological heterogeneity. The difference between the sexes was small so differences between self-reported and directly-measured hypertension for males and females were grouped for each of the standard cut-offs (140/90 and 160-95 mmHg). Studies that did not assess hypertension according to these cut-offs were excluded from the remainder of the analyses. Figure 2 shows the awareness levels for the individual studies with a 140/90 mmHg cut-off. Higher awareness levels indicate that people are able to report their hypertensive status accurately and ideally these values should be as close to 100 as possible. Awareness levels varied widely from no awareness[30] to levels in the mid 90s[14,25,27,28]. Forty-seven studies reported unstandardized awareness values for men and 48 for women. Women had slightly higher awareness levels than men (proportion aware ranged for 0 to 95% for women and 0 to 88% for men) (data not shown).

As expected, awareness levels were higher at the 160/95 mmHg cut-off as these individuals have more serious hypertension and therefore may be more likely to have been diagnosed

and treated for it (Figure 3). Awareness levels were between 31-95% overall (n = 28 studies), between 33 and 92% for men (n = 22) and between 27 and 97% for women (n = 21) (sex-specific data not shown).

A qualitative examination of the forest plots reveals substantial heterogeneity among studies. The heterogeneity was confirmed by Cochran's Q tests for both the 140/90 mmHg (Q = 57099, $p<0.001$) and the 160/95 mmHg cut-off (Q=26053, $p<0.001$). The I^2 statistic revealed that 99% of the total variability among effect sizes was due not to sampling error or chance, but to true heterogeneity among studies for both the 140/90 mmHg and 160/95 mmHg cut-offs.

In order to take the heterogeneity among the sensitivity estimates into account we used a random effects model to account for between-study variability in our meta-analysis. Although we also performed additional sub-group analyses (subdividing studies according to who took the blood pressure measurement, the place of measurement and the population studied) there were no trends in the data that would help to explain the heterogeneity (data not shown). Sample sizes were too small to undertake further sub-group analyses.

Since useful quantitative meta-analysis requires the comparability of studies based on relatively consistent methodological characteristics, we included only studies that provided estimates of awareness and associated measures of variance (or information enabling us to calculate these values), those that presented unstandardized data, those that defined

awareness as the percentage of directly-measured hypertensives that self-reported as hypertensive and those that included patients taking medication as true hypertensives.

Pooled awareness values and 95% confidence intervals are provided in Figures 2 and 3. For the 140/90 mmHg cut-off the pooled estimate for both sexes combined was 58.4% (52.4 - 64.4; n = 63), for males it was 48.8% (43.5 – 54.2; n = 47) and the value was 61% (56.9 - 65.1; n = 48) for females. At the higher cut-off of 160/95 mmHg awareness was higher at 62.1% overall (56.1 – 68.2; n = 28); 57.9% for males (53.4 – 62.4; n = 22) and 71.5% for females (68.2 – 74.8; n = 21). More females had been previously identified as hypertensive by their doctors and informed of, and recalled at time of study, their hypertension status compared to males in both of the cut-offs used in this review.

Sensitivity analysis, which was undertaken on the studies that used the 140/90 mmHg cut-off, revealed that when high and low quality studies were examined independently, there were differences in overall awareness levels. High quality studies (those that scored above 65% on our checklists) had a higher pooled awareness estimate of 60.3% (52.6 – 68.0) whereas for lower quality studies (those scoring less than 65%) the pooled estimate was 54.8% (47.5 – 62.1). Thus, when only high quality studies were included in the meta-analysis, the pooled sensitivity estimate was slightly greater than when all studies were included regardless of quality (60% versus 58%).

Funnel plots graph sensitivity against standard error for the overall data for each of the 140/90 mmHg and 160/95 mmHg cut-offs; these are shown in Figure 4. Neither of the plots

is perfectly symmetrical (the points do not follow an inverted V shape) although for the 140/90 mmHg cut-point this is partially because many of the studies had large sample sizes (smaller standard errors) and they are therefore clustered at the top of the plot. One possible explanation for the asymmetry is the presence of publication bias.

Discussion

Studies in this review consistently demonstrate that the prevalence of hypertension based on self-report is underestimated compared to the directly-measured values (if a cut-off of 140/90 is used, just over half (58%) of hypertensives seem aware of their condition). This trend was evident for both males and females although women have higher awareness levels than men, possibly because women are more likely to seek medical attention and are in general more aware of their health[169]. This higher awareness in women has been found by other researchers[99,159]. The low and highly variable rates of awareness indicate that surveillance based on awareness alone is insufficient, regardless of gender. Lack of awareness is an important public health issue as intervention is difficult without awareness of the condition.

Lack of awareness may be a result of not recalling or never having received a diagnosis of hypertension. As hypertension is often asymptomatic in its early phases, identification relies on adequate screening for the condition, yet studies have shown that screening may not be consistently undertaken[170-171]. A study from the United States found that the screening rate in different clinics ranged from 60 to 100%[170]. Other research has shown that even among hypertensives, one quarter had not had their blood pressure checked in the previous year[171]

and that in some population sub-groups over 40% of the population had not had their blood pressure checked in the last year[75] or two[172]. Indeed, Vargas and colleagues[159] found that the validity of self-reported hypertension was increased in those who had a medical appointment in the past year. Other factors associated with awareness include age[61,88], being overweight[61,159,173] having a family history of hypertension or heart disease[61,121,173] and being married[99,121].

Measured hypertension is considered to be the gold standard for diagnosis but the reliability of blood pressure measurements can vary according to the equipment and conditions of measurement. Measurement on several occasions is also recommended as the diagnosis should be made based on a series of readings with persistent blood pressure elevations. Several guidelines exist that provide recommendations on the best way to measure and treat hypertension[8,166,167]. The Joint National Committee on Detection, Evaluation and Treatment of High Blood Pressure, for instance, stresses the importance of equipment calibration, patient positioning, restrictions prior to measurement and multiple measures[167]. Our quality assessment was based, in part, on the degree to which studies adhered to these guidelines and our analysis indicated that higher study quality is associated with higher awareness levels. In many of the reviewed studies, standard guidelines were not being followed or studies did not report enough information about their protocols to assess the degree to which they complied with standard recommendations, which may partially explain why even our higher quality studies had overall awareness estimates of only 60%. More detailed reporting on guideline adherence is required in order to understand further the causes of heterogeneity among studies.

Standard quantitative indicators of blood pressure should also be consistently reported to increase comparability among studies and allow for monitoring of trends and changes over time. The European Health Risks Monitoring Project[174] has recommended that mean and standard deviations of systolic and diastolic blood pressure, prevalence of actual and potential hypertension, prevalence of antihypertensive drug treatment, awareness of elevated blood pressure and the proportion of those with blood pressure measurement in the past 5 years all be reported. We also recommend that raw data outlining the numbers of people with reported and measured hypertension (standard 2 by 2 tables) be presented so that diagnostic values such as sensitivities, specificities and positive and negative predictive values can be calculated to provide a further understanding of the nature of the relationship between self-reported and directly-measured estimates of hypertension. Standardized protocols for assessing blood pressure in surveys are needed and it would be beneficial if internationally accepted protocols could be developed.

Conclusions

Hypertension is a leading cause of the global burden of disease with two-thirds of stroke and one-half of ischemic heart disease being attributable to sub-optimal blood pressure levels[175]. High blood pressure can be treated, but adequate treatment and control relies on accurate diagnosis. Self-reported data are most often used at the population level to estimate hypertension prevalence, but reported values result in underestimates of true prevalence. Direct measurement, when undertaken, uses diverse methods and variable reporting, which

is not useful for national or international comparisons or for monitoring trends over time. Data from this review indicate the importance of directly measuring blood pressure according to standardized protocols wherever possible, and increasing population awareness both to increase the validity of self-report and to ensure early diagnosis and treatment. In cases where direct measurement is not possible, correction factors should be estimated to adjust the reported estimates to more closely approximate their measured values.

Acknowledgements: The authors acknowledge Lyne Cloutier, Martin Myers and Don McKay for their assistance in developing the quality assessment checklist. We would also like to acknowledge Ian McDowell for insightful comments on earlier drafts of this manuscript and Sean Schofield-Hurwitz and Benson Gorber for their research assistance.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Green MS, Peled I. Prevalence and control of hypertension in a large cohort of occupationally-active Israelis examined during 1985-1987: the Cordis Study. *Int J Epidemiol* 1992; 21: 676-82.
2. Gu D, Reynolds K, Wu X, et al. The International Collaborative Study of Cardiovascular Disease in ASIA. Prevalence, awareness, treatment, and control of hypertension in China. *Hypertension* 2002; 40: 920-7.
3. Jo I, Ahn Y, Lee J, Shin KR, Lee HK, Shin C. Prevalence, awareness, treatment, control and risk factors of hypertension in Korea: the Ansan study. *J Hypertens* 2001; 19: 1523-32.
4. World Health Organization. The World Health Report 2002, Reducing Risks, Promoting Healthy Life. Geneva: World Health Organization Geneva; 2002.
5. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998; 52: 377-84.
6. Saunders DL, Soomro GM, Buckingham J, Jamtvedt G, Raina P. Assessing the methodological quality of nonrandomized intervention studies. *West J Nurs Res* 2003; 25: 223-37.
7. Deeks JJ, Dinnes J, D'Amico R, et al. Evaluating non randomized intervention studies. *Health Technol Assess* 2003; 7: 1–173.
8. Campbell NRC, Joffres MR, McKay DW. Hypertension Surveillance in Canada. Minimum Standards for Assessing Blood Pressure in Surveys. *Can J Public Health* 2005; 96: 217-20.

9. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; 7: 177-88.

10. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; 21: 1539-58.

11. Burt VL, Cutler JA, Higgins M, et al. Trends in the prevalence, awareness, treatment, and control of hypertension in the adult US population. Data from the health examination surveys, 1960 to 1991. *Hypertension* 1995; 26: 60-9.

12. Glover MJ, Greenlund KJ, Ayala C, Croft JB. Racial/ethnic disparities in prevalence, treatment, and control of hypertension – United States, 1992-2002. *JAMA* 2005; 293: 923-5.

13. Pappas G, Gergen PJ, Carroll M. Hypertension prevalence and the status of awareness, treatment, and control in the Hispanic Health and Nutrition Examination Survey (HHANES), 1982-84. *Am J Public Health* 1990; 80: 1431-6.

14. Alonso A, Beunza JJ, Delgado-Rodriguez M, Martinez-Gonzalez MA. Validation of self reported diagnosis of hypertension in a cohort of university graduates in Spain. *BMC Public Health* 2005; 5.

15. Abernethy JD. The Australian National Blood Pressure Study. *Med J Aust* 1974; 1: S21-4.

16. Astagneau P, Lang T, Delarocque E, Jeannee E, Salem G. Arterial hypertension in urban Africa : an epidemiological study on a representative sample of Dakar inhabitants in Senegal. *J Hypertens* 1992; 10: 1095-1101.

17. Altun B, Arici M, Nergizoglu G, et al. Prevalence, awareness, treatment and control of hypertension in Turkey (the PatenT study) in 2003. *J Hypertens* 2005; 23: 1817-23.

18. Achananuparp S, Suriyawongpaisal P, Suebwonglee S, et al. Prevalence, detection and control of hypertension in Thai population of a central rural community. *J Med Assoc Thai* 1989; 72: 66-75.

19. Amoah AGB. Hypertension in Ghana: A cross-sectional community prevalence study in Greater Accra. *Ethn Dis* 2003; 13: 310-15.

20. Arroyo P, Fernandez V, Loria A, et al. Hypertension in urban Mexico: The 1992-93 national survey of chronic diseases. *J Hum Hypertens* 1999; 13: 671-5.

21. Azizi F, Ghanbarian A, Madjid M, Rahmani M. Distribution of blood pressure and prevalence of hypertension in Tehran adult population: Tehran Lipid and Glucose Study (TLGS), 1999-2000. *J Hum Hypertens* 2002; 16: 305-12.

22. Banegas JR, Rodriguez-Artalejo F, de la Cruz Troca JJ, Guallar-Castillon P, del Rey Calero J. Blood pressure in Spain: Distribution, awareness, control, and benefits of a reduction in average pressure. *J Hypertens* 1998; 32: 998-1002.

23. Banegas JR, Rodriguez-Artalejo F, Ruilope LM, et al. Hypertension magnitude and management in the elderly population of Spain. *J Hypertens* 2002; 20: 2157-64.

24. Bharucha NE, Kuruvilla T. Hypertension in the Parsi community of Bombay: a study on prevalence, awareness and compliance to treatment. *BMC Public Health* 2003; 3.

25. Birkett NJ, Donner AP, Maynard M. Assessing hypertension control in the community: the need for follow-up measurements to ensure clinical relevance. *CMAJ* 1987; 136: 595-600.

26. Birkett NJ, Donner AP, Maynard M. Prevalence and control of hypertension in an Ontario county. *CMAJ* 1985; 132: 1019-24.

27. Birkett NJ, Evans CE, Haynes RB, et al. Hypertension control in two Canadian communities: evidence for better treatment and overlabelling. *J Hypertens* 1986; 4: 369-74.

28. Birkett, NJ. The effect of alternative criteria for hypertension on estimates of prevalence and control. *J Hypertens* 1997; 15: 237-44.

29. Brindel P, Hanon O, Dartigues J-F, et al. Prevalence, awareness, treatment, and control of hypertension in the elderly: The Three City study. *J Hypertens* 2006; 24: 51-8.

30. Brown DJ, Metiko EB. Prevalence of hypertension in a sample of black American adults using JNC 7 classifications. *J Natl Black Nurses Assoc* 2005; 16: 1-5.

31. Bowlin SJ, Morrill BD, Nafziger AN, Jenkins PL, Lewis C, Pearson TA. Validity of cardiovascular disease risk factors assessed by telephone survey: The behavioral risk factor survey. *J Clin Epidemiol* 1993; 46: 561-71.

32. Campbell NR, Jeffrey P, Kiss K, Jones C, Anton AR. Building capacity for awareness and risk factor identification in the community: the blood pressure assessment program of the Calgary Fire Department. *Can J Cardiol* 2001; 17: 1275-79.

33. Cappuccio FP, Micah FB, Emmett L, et al. Prevalence, detection, management, and control of hypertension in Ashanti, West Africa. *Hypertension* 2004; 43: 1017-22.

34. Carter-Edwards L, Jackson SA, Runaldue MJ, Svetkey LP. Diet- and blood pressure-related knowledge, attitudes, and hypertension prevalence among African Americans: the KDBP Study. Knowledge of Diet and Blood Pressure. *Ethn Dis* 2002; 12: S1-82.

35. Chadha SL, Radhakrishnan S, Ramachandran K, Kaul U, Gopinath N. Prevalence, awareness and treatment status of hypertension in urban populations of Delhi. *Indian J Med Res* 1990; 92: 233-40.

36. Chen CH, Lin HC, Kuo HS, Chang, MS, Chou P. Epidemiology of hypertension in Kin-Hu, Kinmen. *Am J Hypertens* 1995; 8: 395-403.

37. Chen R, Tunstall-Pedoe H, Morrison C, Connaghan J, A'Brook R. Trends and social factors in blood pressure control in Scottish MONICA surveys 1986-1995: the rule of halves revisited. *J Hum Hypertens* 2003; 17: 751-9.

38. Choi KM, Park HS, Han JH, et al. Prevalence of prehypertension and hypertension in a Korean population: Korean National Health and Nutrition Survey 2001. *J Hypertens* 2006; 24: 1515-21.

39. Chua B, Rochtchina E, Mitchell P. Temporal changes in the control of blood pressure in an older Australian population. *J Hum Hypertens* 2005; 19: 691-6.

40. Cifkova R, Skodova Z, Lanska V, et al. Prevalence, awareness, treatment, and control of hypertension in the Czech Republic. Results of two nationwide cross-sectional surveys in 1997/1998 and 2000/2001, Czech Post-MONICA study. *J Hum Hypertens* 2004; 18; 571-9.

41. Colhoun HM, Dong W, Poulter NR. Blood pressure screening, management and control in England: results from the health survey for England 1994. *J Hypertens* 1998; 16: 747-52.

42. De Backer G, Myny K, De Henauw S, et al. Prevalence, awareness, treatment and control of arterial hypertension in an elderly population in Belgium. *J Hum Hypertens* 1998; 12: 701-6.

43. De Henauw S, De Bacquer D, Fonteyne W, Stam M, Kornitzer M, De Backer G. Trends in the prevalence, detection, treatment and control of arterial hypertension in the Belgian adult population. *J Hypertens* 1998; 16: 277-84.

44. Deprez RD, Miller E, Hart SK. Hypertension prevalence among Penobscot Indians of Indian Island, Maine. *Am J Pub Health* 1985; 75: 653-4.

45. Dickson BK, Blackledge J, Hajjar IM. The impact of lifestyle behavior on hypertension awareness, treatment, and control in a southeastern population. *Am J Med Sci* 2006; 332: 211-15.

46. Duprez D, Van Helshoecht, Van den Eynde W, Leeman M. Prevalence of hypertension in the adult population of Belgium: report of a worksite study, Attention Hypertension. *J Hum Hypertens* 2002; 16: 47-52.

47. Efstratopoulos AD, Voyaki SM, Baltas AA, et al. Prevalence, awareness, treatment and control of hypertension in Hellas, Greece: The Hypertension Study in General Practice in Hellas (HYPERTENSHELL) national study. *Am J Hypertens* 2006; 19: 53-60.

48. Entwisle G, Scott JC, Apostolides AY, et al. A survey of blood pressure in the state of Maryland. *Prev Med* 1983; 12: 695-708.

49. Fodor JG, Lietava J, Rieder A, et al. Work-site hypertension prevalence and control in three Central European countries. *J Hum Hypertens* 2004; 18: 581-5.

50. Foster C, Rotimi C, Fraser H, et al. Hypertension, diabetes, and obesity in Barbados: findings from a recent population-based survey. *Ethn Dis* 1993; 3: 404-12.

51. Franco LJ, Stern MP, Rosenthal M, Haffner SM, Hazuda HP, Comeaux PJ. Prevalence, detection, and control of hypertension in a biethnic community: the San Antonio Health Study. *Am J Epidemiol* 1985; 121 684-96.

52. Freeman V, Fraser H, Forrester T, et al. A comparative study of hypertension prevalence, awareness, treatment and control rates in St Lucia, Jamaica and Barbados. *J Hypertens* 1996; 14: 495-501.

53. Freeman Jr DH, Ostfeld AM, Hellenbrand K, Richards VA, Tracy R. Changes in the prevalence distribution of hypertension: Connecticut adults 1978-79 to 1982. *J Chron Dis* 1985; 38: 157-64.

54. Furumoto-Dawson AA, Pandey DK, Elliott WJ, et al. Hypertension in women: The women take heart project. *J Clin Hypertens* 2003; 5: 38-46.

55. Garbus SB, Garbus SB. Analysis of mass hypertension screening. *Prev Med* 1976; 5: 48-59.

56. Garcia-Pena C, Thorogood M, Reyes S, Salmeron-Castro J, Duran C. The prevalence and treatment of hypertension in the elderly population of the Mexican Institute of Social Security. *Salud Publica Mex* 2001; 43: 415-20.

57. Gasse C, Hense H-W, Stieber J, Doring A, Liese AD, Keil U. Assessing hypertension management in the community: Trends of prevalence, detection, treatment, and control of hypertension in the MONICA Project, Augsburg 1984-1995. *J Hum Hypertens* 2001; 15: 27-36.

58. Gelberg L, Linn LS. Assessing the physical health of homeless adults. *JAMA* 1989; 262: 1973-9.

59. Gentleman J, Tomiak M. The consistency of various high blood pressure indicators based on questionnaire and physical measures data from the Canada Health Survey. *Health Rep* 1992; 4: 293-311.

60. Getliffe KA, Crouch R, Gage H, Lake F, Wilson SL. Hypertension awareness, detection and treatment in a university community: Results of a worksite screening. *Public Health* 2000; 114: 361-6.

61. Giles WH, Croft JB, Keenan NL. The validity of self-reported hypertension and correlates of hypertension awareness among blacks and whites within the stroke belt. *Am J Prev Med* 1995; 11: 163-9.

62. Gonzalez-Villalpando C, Stern MP, Haffner SM, Gonzalez Villapando ME, Gaskill S, Martinez DR. Prevalence of hypertension in a Mexican population according to the Sixth Report of the Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure. *J Cardiovasc Risk* 1999; 6: 177-81.

63. Goldman N, Lin I-F, Weinstein M, Lin Y-H. Evaluating the quality of self-reports of hypertension and diabetes. *J Clin Epidemiol* 2003; 56: 148-54.

64. Guimaraes AC. Hypertension in Brazil. *J Hum Hypertens* 2002; 16: S7-10.

65. Halabi S, Zurayk H, Awaida R, Darwish M, Saab B. Reliability and validity of self and proxy reporting of morbidity data: A case study from Beirut, Lebanon. *Int J Epidemiol* 1992; 21: 607-12.

66. Handa SP, Roma S, Brewer H. Prevalence of hypertension: a screening survey in an adult population. *Can J Public Health* 1980; 71: 171-5.

67. Health and Welfare Canada. *Main Findings Report of the Canadian Blood Pressure Survey*. Ottawa: Department of National Health and Welfare; 1989.

68. Hennis A, Wu S-Y, Nemesure B, Leske MC. Hypertension prevalence control and survivorship in an Afro-Caribbean population. *J Hypertens* 2002; 20: 2363-9.

69. Holmen J, Forsen L, Hjort PF, Midthjell, Waaler HT, Bjorndal A. Detecting hypertension: screening versus case finding in Norway. *Br Med J* 1991; 302: 219-22.

70. Holtzman E, Goldbourt U, Rosenthal T, Yaari S, Neufeld HN. Hypertension in middle-aged men: associated factors and mortality experience. *Isr J Med Sci* 1983; 19: 25-33.

71. Howteerakul N, Suwannapong N, Sittilerd R, Rawdaree P. Health risk behaviours, awareness, treatment and control of hypertension among rural community people in Thailand. *Asia Pac J Public Health* 2006; 18: 3-9.

72. Hypertension Detection and Follow-up Program Cooperative Group. Blood pressure studies in 14 communities: a two-stage screen for hypertension. *JAMA* 1977; 22: 2385-91.

73. Ibrahim MM, Rizk H, Appel LJ, et al. Hypertension prevalence, awareness, treatment, and control in Egypt: results from the Egyptian National Hypertension Project (NHP). *Hypertension* 1995; 26: 886-90.

74. Jamrozik K, Hockey R. Trends in risk factors for vascular disease in Australia. *Med J Aust* 1989; 150: 14-8.

75. Joffres MR, Ghadirian P, Fodor JG, Petrasovits A, Chockalingam A, Hamet P. Awareness, treatment, and control of hypertension in Canada. *Am J Hypertens* 1997; 10: 1097-1102.

76. Johansson J, Hellenius M-L, Elofsson S, Krakau I. Self-report as a selection instrument in screening for cardiovascular disease risk. *Am J Prev Med* 1999; 16: 322-4.

77. Johnson CC, Powers NR, Bao W, Harsha DW, Berenson GS. Cardiovascular risk factors of elementary school teachers in a low socio-economic area of a metropolitan city: The Heart Smart Program. *Health Educ Res* 1994; 9: 183-91.

78. Jones DW, Kim JS, Kim SJ, Hong YP. Hypertension awareness, treatment and control rates for an Asian population: results from a national survey in Korea. *Ethn Health* 1996; 1: 269-73.

79. Jones RAP, Ricard R, Sefcik E, Miller ME. The health risk of hypertension in South Texas: a demographic profile. *Holist Nurs Pract* 2001; 15: 35-44.

80. Kalatan KA, Mohamed AG, Al-Taweel AA, Ghani HMA. Hypertension among attendants of primary health care centers in Al-Qassim region, Saudia Arabia. *Saudi Med J* 2001; 22: 960-3.

81. Kamadjeu RM, Edwards R, Atanga JS, Unwin N, Kiawi EC, Mbanya J-C. Prevalence, awareness and management of hypertension in Cameroon: Findings of the 2003 Cameroon Burden of Diabetes Baseline Survey. *J Hum Hypertens* 2006; 20: 91-2.

82. Kastarinen MJ, Antikainen RL, Laatikainen TK, et al. Trends in hypertension care in eastern and south-western Finland during 1982-2002. *J Hypertens* 2006; 24: 829-36.

83. Kattainen A, Reunanen A, Koskinen S, Martelin T, Knekt P, Aromaa A. Secular changes in prevalence of cardiovascular diseases in elderly Finns. *Scand J Public Health* 2002; 30: 274-80.

84. Klungel OH, De Boer A, Paes AHP, Nagelkerke NJD, Seidell JC, Bakker A. Influence of correction for within-person variability in blood pressure on the prevalence, awareness, treatment, and control of hypertension. *Am J Hypertens* 2000; 13: 88-91.

85. Lang T, de Gaudemaris R, Chatellier G, Hamici L, Diene E. Prevalence and therapeutic control of hypertension in 30 000 subjects in the workplace. *Hypertension* 2001; 38: 449-54.

86. Leynen F, De Backer H, Pelfrene E, et al. Increased absenteeism from work among aware and treated hypertensive and hypercholesterolaemic patients. *Euj J Cardiovasc Prev Rehabil* 2006; 13: 261-7.

87. Li W, Jiang X, Ma H, et al. Awareness, treatment and control of hypertension in patients attending hospital clinics in China. *J Hypertens* 2003; 21: 1191-7.

88. Lima-Costa MF, Peixoto SV, Firmo JOA. Validity of self-reported hypertension and its determinants (the Bambui study). *Rev Saude Publica* 2004; 38: 637-42.

89. Lin T, Chen C-H, Chou P. A hypertension control program in Yu-Chi, Taiwan: Preliminary results. *J Formos Med Assoc* 1997; 96; 613-20.

90. Lisk DR, Williams DE, Slattery J. Blood pressure and hypertension in rural and urban Sierra Leoneans. *Ethn Dis* 1999; 9: 254-63.

91. M'Buyamba-Kabangu J-R, Fagard R, Lijnen P, et al. Epidemiological study of blood pressure and hypertension in a sample of urban Bantu of Zaire. *J Hypertens* 1986; 4: 485-91.

92. Maahs DM, Kinney GL, Wadwa P, Snell-Bergeon JK, Dabelea D, Hokanson J. Hypertension prevalence, awareness, treatment, and control in an adult type 1 diabetes population and a comparable general population. *Diabetes Care* 2005; 28: 301-6.

93. Macedo ME, Lima MJ, Silva AO, et al. Prevalence, awareness, treatment and control of hypertension in Portugal: the PAP study. *J Hypertens* 2005; 23: 1661-6.

94. Matenga JA, Allain TJ, Wilson AO, et al. Hypertension management in Zimbabwe – awareness, treatment and blood pressure control. *SAMJ* 1997; 87: 1371-3.

95. Mancia G, Parati G, Borghi C, et al. Hypertension prevalence, awareness, control and association with metabolic abnormalities in the San Marino population: The SMOOTH study. *J Hypertens* 2006; 24: 837-43.

96. Meissner I, Whisnant JP, Sheps SG, et al. Detection and control of high blood pressure in the community: do we need a wake-up call? *Hypertension* 1999; 34: 466-71.

97. Metcalf CA, Hoffman MN, Steyn K, Katzenellenbogen JM, Fourie JM.. Design and baseline characteristics of a hypertension intervention program in a South African village. *J Hum Hypertens* 1996; 10: 21-6.

98. Mosca L, Edelman D, Mochari H, Christian AH, Paultre F, Pollin I. Waist circumference predicts cardiometabolic and Global Framingham Risk among women screened during National Woman's Heart Day. *J Womens Health* 2006; 15: 24-34.

99. Muntner P, Gu D, Wu X, et al. Factors Associated with Hypertension Awareness, Treatment, and Control in a Representative Sample of the Chinese Population. *Hypertension* 2004; 43: 578-85.

100. Nan L, Tuomilehto J, Dowse G, et al. Prevalence and medical care of hypertension in four ethnic groups in the newly-industrialized nation of Mauritius. *J Hypertens* 1991; 9: 859-66.

101. Nieto FJ, Alonso J, Chambless LE, et al. Population awareness and control of hypertension and hypercholesterolemia: the Atherosclerosis Risk in Communities Study. *Arch Intern Med* 1995; 155: 677-84.

102. Nissinen A, Tuomilehto J, Korhonen HJ, Piha T, Salonen JT, Puska P. Ten-year results of hypertensioncare in the community: follow-up of the North Karelia Hypertension Control Program. *Am J Epidemiol* 1988; 127: 488-99.

103. Nova Scotia Department of Health. *The Nova Scotia Health Survey, 1995*. Nova Scotia: Communications Nova Scotia; 1996.

104. de Pablos-Velasco P, Martinez-Martin FJ, Perez FR, Urioste LMR, Robles RG. Prevalence, awareness, treatment and control of hypertension in a Canarian population. Relationship with glucose tolerance categories. The Guia study. *J Hypertens* 2002; 20: 1965-71.

105. Pan WH, Chang HY, Yeh WT, Hsiao SY, Hung YT. Prevalence, awareness, treatment and control of hypertension in Taiwan: Results of Nutrition and Health Survey in Taiwan (NAHSIT) 1993-1996. *J Hum Hypertens* 2001; 15: 793-98.

106. Panagiotakos DB, Pitsavos CH, Chrysohoou C, et al. Status and management of hypertension in Greece: Role of the adoption of a Mediterranean diet: The Attica study. *J Hypertens* 2003; 21: 1483-89.

107. Paschal AM, Lewis RK, Martin A, Dennis-Shipp D, Simpson DS. Baseline assessment of the health status and health behaviors of African Americans participating in the activities-for-life program: A community-based health intervention program. *J Community Health* 2004; 29: 305.

108. Pavlik VN, Hyman DJ, Vallbona C, et al. Hypertension awareness and control in an inner-city African-American sample. *J Hum Hypertens* 1997; 11: 277-83.

109. Pica LA, Boucher MM, Grignon R. The prevalence, awareness and control of hypertension in Laval, Quebec, 1986. *Can J Public Health* 1990; 81: 427-30.

110. Plans P, Pardell H, Salleras LI. Epidemiology of cardiovascular disease risk factors in Catalonia (Spain). *Eur J Epidemiol* 1993; 9: 381-89.

111. Prencipe M, Casini AR, Santini M, Ferretti C, Scaldaferri N, Culasso F. Prevalance, awareness, treatment and control of hypertension in the elderly: Results from a population survey. *J Hum Hypertens* 2000; 14: 825-30.

112. Primatesta P, Brookes M, Poulter NR. Improved hypertension management and control: results from the Health Survey for England 1998. *Hypertension* 2001; 38: 827-32.

113. Primatesta P, Poulter NR. Improvements in hypertension management in England: results from the Health Survey for England 2003. *J Hypertens* 2006; 24: 1187-92.

114. Psaltopoulou T, Orfanos P, Naska A, Lenas D, Trichopoulos D, Trichopoulou A. Prevalence, awareness, treatment and control of hypertension in a general population sample of 26,913 adults in the Greek EPIC study. *Int J Epidemiol* 2004; 33: 1345-52.

115. Puras A, Sanchis C, Artiago LM, Divison JA. Prevalence, awareness, treatment and control of hypertension in a Spanish population. *Eur J Epidemiol* 1998; 14: 13-36.

116. Ramirez MOA, Pino CT, Furiasse LVA, Lee AJ, Fowkes FGR. Paraguayan National Blood Pressure study: prevalence of hypertension in the general population. *J Hum Hypertens* 1995; 9: 891-97.

117. Reiff M, Schwartz S, Northridge M. Relationship of depressive symptoms to hypertension in a household survey in Harlem. *Psychosom Med* 2001; 63: 711-21.

118. Ruixing Y, Limei Y, Yuming C, et al. Prevalence, awareness, treatment, control and risk factors of hypertension in the Guangxi Hei Yi Zhuang and Han populations. *Hypertens Res* 2006; 29: 423-32.

119. Rywik SL, Davis CE, Pajak A, et al. Poland and U.S. collaborative study on cardiovascular epidemiology hypertension in the community: prevalence, awareness, treatment, and control of hypertension in the Pol-MONICA project and the U.S. atherosclerosis risk in communities study. *Ann Epidemiol* 1998; 8: 3-13.

120. Sarafidis PA, Lasaridis A, Gousopoulos S, et al. Prevalence, awareness, treatment and control of hypertension in employees of factories of Northern Greece: The Naoussa study. *J Hum Hypertens* 2004; 18: 623-29.

121. Satish S, Markides KS, Zhang D, Goodwin JS. Factors influencing unawareness of hypertension among older Mexican Americans. *Prev Med* 1997; 26: 645-50.

122. Schoenberger JA, Stamler J, Shekelle RB, Shekelle S. Current status of hypertension control in an industrial population. *JAMA* 1972; 222: 559-62

123. Schmitz N, Thefeld W, Kruse J. Mental disorders and hypertension: Factors associated with awareness and treatment of hypertension in the general population of Germany. *Psychosom Med* 2006; 68: 246-52.

124. Sekikawa A, Hayakawa T. Prevalence of hypertension, its awareness and control in adult population in Japan (Research Letter). *J Hum Hypertens* 2004; 18: 911-12.

125. Shapiro M, Bleho J, Curran M, et al. Problems in the control of hypertension in the community. *CMAJ* 1978; 118: 37-9.

126. Sharma AM, Wittchen H-U, Kirch W, et al. High prevalence and poor control of hypertension in primary care: Cross-sectional study. *J Hypertens* 2004; 22: 479-86.

127. Shouqi T, Xigui W, Xiufang D, et al. Hypertension prevalence and status of awareness, treatment and control in China. *Chin Med* 1995; 108: 483-9.

128. Silverberg DS, Smith SO, Juchli B, VanDorsser E. Use of shopping centres in screening for hypertension. *CMAJ* 1974; 111: 769-74.

129. Smith WCS, Lee AJ, Crombie IK, Tunstall-Pedoe H. Control of blood pressure in Scotland: the rule of halves. *Br Med J* 1990; 300: 981-3.

130. Sonkodi B, Fodor JG, Abraham G, et al. Hypertension screening in a salami factory: A worksite hypertension study. *J Hum Hypertens* 2004; 18: 567-9.

131. Sonmez HM, Basak O, Camci C, et al. The epidemiology of elevated blood pressure as an estimate for hypertension in Aydin, Turkey. *J Hum Hypertens* 1999; 13: 399-404.

132. Spencer J, Phillips E, Ogedegbe G. Knowledge attitudes, beliefs, and blood pressure control in a community-based sample in Ghana. *Ethn Dis* 2005; 15.

133. Stamler J, Stamler R, Riedlinger WF, Algera G, Roberts RH. Hypertension screening of 1 million Americans: Community Hypertension Evaluation Clinic (CHEC) Program, 1973 through 1975. *JAMA* 1976; 235: 2299-2306.

134. Stein AD, Stoyanovsky V, Mincheva V, et al. Prevalence, awareness, treatment and control of hypertension in a working Bulgarian population. *Eur J Epidemiol* 2000; 16: 265-70.

135. Stergiou GS, Thomopoulou GC, Skeva II, Mountokalakis TD. Prevalence, awareness, treatment and control of hypertension in Greece: The Didima study. *Am J Hypertens* 1999; 12: 959-65.

136. Steyn K, Gaziano TA, Bradshaw D, Laubscher R, Fourie J. Hypertension in South African adults: results from the Demographic and Health Survey, 1998. *J Hypertens* 2001; 19: 1717-25.

137. Takala J. Screening for hypertension in a middle-aged population in Southwest Finland. *Prev Med* 1978; 7: 230-44.

138. Tormo MJ, Navarro C, Chirlaque MD, Barber X, the EPIC Group of Spain. Is there a different dietetic pattern depending on self-knowledge of high blood pressure? *Eur J Epidemiol* 2000; 16: 963-71.

139. Trenkwalder P, Ruland D, Stender M, et al. Prevalence, awareness, treatment and control of hypertension in a population over the age of 65 years: Results from the Starnberg Study on epidemiology of Parkinsonism and hypertension in the elderly (STEPHY). *J Hypertens* 1994; 12: 709-16.

140. Aytekin NT, Pala K, Irgil E, Akis N, Aytekin H. Distribution of blood pressures in Gemlik District, North-West Turkey. *Health Soc Care Community* 2002; 10: 394-401.

141. Vaccarino V, Borgatta A, Gallus G, Sirtori CR. Prevalence of coronary heart disease risk factors in Northern-Italian male and female employees. *Eur Heart J* 1995; 16: 761-9.

142. van Rossum CTM, van de Mheen H, Witteman JCM, Hofman A, Mackenbach JP, Grobbee DE. Prevalence, treatment, and control of hypertension by sociodemographic factors among the Dutch elderly. *Hypertension* 2000; 35: 814-21.

143. Wagner EH, James SA, Beresford SAA, et al. The Edgecombe County high blood pressure control program: correlates of uncontrolled hypertension at baseline. *Am J Public Health* 1984; 74: 237-42.

144. Wander GS, Khurana SB, Gulati R, et al. Epidemiology of coronary heart disease in a rural Punjab population--prevalence and correlation with various risk factors. *Indian Heart J* 1994; 46: 319-23.

145. Welty TK, Rhoades DA, Yeh F, et al. Changes in cardiovascular disease risk factors among American Indians: the Strong Heart Study. *Ann Epidemiol* 2002; 12: 97-106.

146. Wilber JA. The problem of undetected and untreated hypertension in the community. *Bull NY Acad Med* 1973; 49: 510-20.

147. Wilson TW, Dubois M, Ramsden V, Raina P, Tan LK. The Saskatchewan heart foundation blood pressure survey. *Can Fam Physician* 1991; 37: 623-8.

148. Wolf HK, Andreou P, Bata IR, et al. Trends in the prevalence and treatment of hypertension in Halifax County from 1985 to 1995. *CMAJ* 1999; 161: 699-704.

149. Wolffsohn JS, Napper GA, Ho SM, et al. Improving the description of the retinal vasculature and patient history taking for monitoring systemic hypertension. *Ophthalmic Physiol Opt* 2001; 21: 441-9.

150. Wood DM, Jones KD. Comparing NHANES data in the assessment of hypertension of African Americans: awareness, treatment and control of hypertension in an urban Maryland city. *Md Med* 2002: 57-9.

151. Zdrojewski T, Wyrzykowski B, Szczech R, et al. Epidemiology and prevention of arterial hypertension in Poland. *Blood Press* 2005; 14: 10-16.

152. Zrojewski T, Wyrzykowski B, Wierucki L, et al. Attempt to eliminate health inequalities in Poland arising at the time of political and economic transformation: Polish 400 Cities Project. *Eur J Cardiovasc Prev Rehabil* 2006; 13: 832-8.

153. Banegas JR, Guallar-Castillon P, Rodriguez-Artalejo F, Graciani A, Lopez-Garcia E, Ruilope LM. Association Between Awareness, Treatment, and Control of Hypertension, and Quality of Life Among Older Adults in Spain. *Am J Hypertens* 2006; 19: 686-93.

154. Stewart JC, France CR, Sheffield D. Hypertension awareness and pain reports: Data from the NHANES III. *Ann Behav Med* 2003; 26: 8-14.

155. Gillen DL, Coe FL, Worcester EM. Nephrolithiasis and increased blood pressure among females with high body mass index. *Am J Kidney Dis* 2005; 46: 263-9.

156. Geiss LS, Rolka DB, Engelgau MM. Elevated blood pressure among U.S. adults with diabetes, 1988-1994. *Am J Prev Med* 2002; 22: 42-8.

157. Barger SD, Muldoon MF. Hypertension labelling was associated with poorer self-rated health in the Third US National Health and Nutrition Examination Survey. *J Hum Hypertens* 2006; 20: 117-23.

158. Gu Q, Paulose-Ram R, Dillon C, Burt V. Antihypertensive medication use among US adults with hypertension. *Circulation* 2006; 113: 213-21.

159. Vargas CM, Burt VL, Gillum RF, Pamuk ER. Validity of self-reported hypertension in the National Health and Nutrition Examination Survey III, 1988-1991. *Prev Med* 1997; 26: 678-85.

160. Hajjar I, Kotchen TA. Trends in Prevalence, Awareness, Treatment, and Control of Hypertension in the United States, 1988-2000. *JAMA* 2003; 290: 199-206.

161. Roccella EJ. Hypertension prevalence and the status of awareness, treatment, and control in the United States. *Hypertension* 1985; 7: 457-68.

162. Burt VL, Whelton P, Roccella EJ, et al. Prevalence of hypertension in the US adult population results from the third National Health and Nutrition Examination Survey, 1988-1991. *Hypertension* 2005; 25: 305-13.

163. Mulrow PJ. Detection and control of hypertension in the population the United States Experience. *Am J Hypertens* 1998; 11: 744-6.

164. Hyman DJ, Pavlik VN. Characteristics of patients with uncontrolled hypertension in the United States. *N Engl J Med* 2001; 345: 479-86.

165. Ford ES, Harel Y, Heath G, Cooper RS, Caspersen CJ. Test characteristics of self-reported hypertension among the Hispanic population: Findings from the Hispanic Health and Nutrition Examination Survey. *J Clin Epidemiol* 1990; 43: 159-65.

166. O'Brien E, Asmar R, Beilin L, et al. European Society of Hypertension recommendations for conventional, ambulatory and home blood pressure measurement. *J Hypertens* 2003; 21: 821-48.

167. U.S. Department of Health and Human Services. *JNC Express. The Seventh Report on the Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure.* National Institutes of Health, NIH Publication No. 03-5233; 2003.

168. Abbott D, Campbell N, Carruthers-Czyzewski P, et al. Guidelines for measurement of blood pressure, follow-up, and lifestyle counseling. *Can J Public Health* 1994; 85: S29-35.

169. Bertakis KD, Azari R, Helms LJ, Callahan EJ, Robbins JA. Gender differences in the utilization of health care services. *J Fam Pract* 2000; 49: 147-52.

170. Solberg LI, Kottke TE, Brekke ML. Variation in clinical preventive services. *Eff Clin Pract* 2001; 4: 121–6.

171. Heymann AD, Hoch I, Valinsky L, Shalev V, Silber H and Kokia E. Mandatory computer field for blood pressure measurement improves screening. *Fam Pract* 2005; 22: 168–9.

172. Kitai E, Vinker S, Halperin L, Meidan A, Grossman E. Pre-hypertension is a common phenomenon: national database study. *Isr Med Assoc J* 2007; 9: 8-11.

173. Agyemang C, van Valkengoed I, Koopmans R, Stronks K. Factors associated with hypertension awareness, treatment and control among ethnic groups in Amsterdam, The Netherlands: The SUNSET study. *J Hum Hypertens* 2006; 20: 874–81.

174. Tolonen H, Kuulasmaa K, Laatikainen T, Wolf H and the European Health Risk Monitoring Project. *Recommendation for indicators, international collaboration, protocol and manual of operations for chronic disease risk factor surveys.* Helsinki: The Finnish National Public Health Institute; 2002.

175. Lawes CM, Vander Hoorn S, Law MR, Elliott P, MacMahon S, Rodgers A. Blood pressure and the global burden of disease 2000. Part II: estimates of attributable burden. *J Hypertens* 2006; 24: 423-30.

Tables and Figures

Table 1: MEDLINE search strategy

1	adult/ or aged/ or middle aged/
2	Male/
3	Female/
4	Humans/
5	1 or 2 or 3 or 4
6	physical examination/ or exp blood pressure determination/
7	physical measure$.mp.
8	objective measure$.mp.
9	direct measure$.mp.
10	Monitoring, Physiologic/ or Blood Pressure Monitors/
11	objective assess$.mp.
12	6 or 7 or 8 or 9 or 10 or 11
13	subjective assess$.mp.
14	subjective measure$.mp.
15	self concept/
16	"Self Assessment (Psychology)"/ or self report$.mp.
17	self evaluat$.mp.
18	awareness/ or aware$.mp.
19	13 or 14 or 15 or 16 or 17 or 18
20	Hypertension/ or hypertension$.mp.
21	blood pressure.mp. or Blood Pressure/
22	high blood pressure.mp.
23	primary hypertension.mp.
24	secondary hypertension.mp.
25	essential hypertension.mp.
26	diastolic blood pressure.mp.
27	systolic blood pressure.mp.
28	20 or 21 or 22 or 23 or 24 or 25 or 26 or 27
29	5 and 12 and 19 and 28

Table 2: Assessment of quality of blood pressure measurement

1) **Training**

a. Specifically (re)trained. Y/N

b. Accuracy of observer tested. Y/N

2) **Equipment**

a. Calibration / accuracy tested. Y/N

b. Appropriate blood pressure cuffs used. Y/N

c. Mercury or validated electronic (AAMI[a], IP[b] or BHS[c] protocol) or calibrated and regularly re-calibrated aneroid. Y/N

3) **Participant Preparation**

a. Assessment in a quiet comfortable temperature controlled environment. Y/N

b. Participant instructed to not have heavy physical activity prior to assessment. Y/N

c. Participant instructed to not use caffeine or nicotine or vasopressor substances (e.g. decongestants) prior to assessment. Y/N

d. Rested 5 minutes (or more). Y/N

4) **Technique**

a. Seated with back support, legs uncrossed. Y/N

b. Arm supported at heart level. Y/N

c. Cuff deflation rate of 2 mmHg per cardiac cycle (or per second) Y/N

d. Phase V used for diastolic blood pressure. Y/N

e. An average of two or more readings used. Y/N

f. Both arms assessed to determine if one arm has higher BP. Y/N

g. An average of two or more visits used. Y/N

Footnotes: [a] Association for the Advancement of Medical Instrumentation; [b] International Protocol; [c] British Hypertension Society protocol; Yes answers were given a score of 1 point

Table 3: Study and participant characteristics

First Author – Year	Sample Size	Age Range or Mean(SD)	Population	Country	Maximum Time Lag	Measurement Setting
Alonso 2005*	127	18+	University graduates	Spain	–	A, C, D
Bowlin 1993*	628	20-69	Rural population	USA	1 month	D
Abernethy 1974	36 489	30-69	General population (Australian National Blood Pressure study)	Australia	–	B
Achananuparp 1989	2 374	15-93	General population	Taiwan	2 wks	A, B
Altun 2005	4 910	18+	General population	Turkey	–	A
Amoah 2003	4 733	25-102	Urban population	Ghana	–	D
Arroyo 1999	14 657	20-69	Urban population	Mexico	–	A
Astagneau 1992	2 300	15+	General population	Africa	no lag	A
Azizi 2002	8 491	20-69	General population under coverage of medical health centres	Iran	no lag	D
Banegas 1998	2 021	35-64	General population	Spain	no lag	D
Banegas 2002	3 272	60+	Elderly	Spain	1 day	A
Bharucha 2003	2 415	20+	General population (Parsis)	India	no lag	A
Birkett 1986	6 258	30-69	Urban population	Canada	no lag	A
Birkett 1997	2 016	18+	General population	Canada	no lag	A
Birkett 1985/87	2 737	18+	General population	Canada	2 wks	A
Brindel 2006	9 090	65+	Elderly	France	no lag	A
Brown 2005	211	25-79	Undiagnosed, African Americans	USA	–	–
Burt 1995	(1) 6 530 (2) 13 654 (3) 12 504 (4) 7 540	18-74	General population (1. NHES; 2. NHANES-I; 3. NHANES-II; 4. NHANES-III)	USA	–	A, D

Campbell 2001	3 477	55(18)	Attendees at a free screening	Canada	no lag	B
Cappucio 2004	1 013	55(11)	Semi-urban/rural	Ghana	no lag	B
Carter-Edwards 2002	179	23-91	African American church members	USA	no lag	B
Chadha 1990	13 723	25-64	General population	India	no lag	A
Chen 1995	3 289	30+	General population	China	no lag	B
Chen 2003	5 985	46(11)	General populatio (Scottish MONICA)	UK	–	D
Choi 2006	6 074	20+	Civilian non-institutionalized	Korea	–	–
Chua 2005	(1) 3 654 (2) 3 509	49+	Urban population (1. BMES I; 2. BMES II)	Australia	–	A
Cifkova 2004	6 528	25-64	General population	Czech Republic	–	D
Coulhoun 1998	12 116	16+	General population (Health Survey of England)	England	–	A
De Backer 1998	2 165	65-89	Follow-up of general population health survey	Belgium	no lag	A
De Henauw 1998	(1) 9 372 (2) 4 901	25-64	General population (1. BRINH; 2. WHO MONICA)	Belgium	no lag	D
Deprez 1985	(1) 100 (2) 1 229	18+	1. Indians on reservation; 2. General population	USA	no lag	A
Dickson 2006	755	52(1)	Participants at outreach health events, oversample of African Americans	USA	no lag	D
Duprez 2002	3 472	16-67	Company employees	Belgium	–	C
Efstratopoulous 2006	11 540	17+	General population	Greece	no lag	D
Entwisle 1983	6 425	18-70	General population	USA	no lag	A
Fodor 2004	(1) 323 (2) 751 (3) 600	35-42	Male "blue collar" workers 1. Austria; 2. Slovakia; 3. Hungary	Austria, Slovakia, Hungary	no lag	C
Foster 1993	464	40-79	General population	Barbados	–	D

Franco 1985	2 217	25-64	Mexican Americans and Anglos varying socio-economic profiles	USA	–	D
Freeman 1996	(1) 817 (2) 807 (3) 1 080	25-34	General population 1. Jamaica; 2. Barbados; 3. St. Lucia	Jamica, Barbados, St.Lucia	no lag	A
Freeman 1985	9 066	18+	General population	USA	–	–
Furumoto-Dawson 2003	5 718	35+	Women without CVD	USA	no lag	D
Garbus 1976	25 284	mostly 20+	General population	USA	no lag	B
Garcia-Pena 2001	4 777	60+	Elderly - covered by social healthcare services	Mexico	no lag	A
Gasse 2001	(1) 4 022 (2) 4 940 (3) 4 856	25-64	General population, 3 time points: 1. 1984-85; 2. 1989-90; 3. 1994-95	Germany	no lag	D
Gelberg 1989	529	34	Homeless from shelters	USA	no lag	B
Gentleman 1992	4 671	15-65	General population (Canada Health Survey)	Canada	–	A
Getliffe 2000	653	<20-60+	University staff/students	UK	no lag	B
Giles 1995	2 914	18+	General population	USA	4 wks	A, C, D
Glover 2005	12 000	20-60+	General population (NHANES 1999-2001)	USA	–	D
Gonzalez-Villalpando 1999	2 282	35-64	Low-income population	Mexico	–	D
Goldman 2003	1 003	54-80+	Follow-up of general population health survey	Taiwan	several wks (median 22 days)	D
Green 1992	5 250	20-64	Factory employees	Israel	–	C
Gu 2002	15 504	35-74	General population	China	–	D
Guimarães 2002	4 613	43 (approx)	General population	Brazil	no lag	B
Halabi 1992	100 cases/100 controls	18+	Cases with chronic disease; controls without	Lebanon	–	A, D
Handa 1980	8 474	20-70	General population	Canada	–	B,C

Health and Welfare Canada 1989	3 092	18+	General population (Canadian Blood Pressure Survey)	Canada	no lag	A
Hennis 2002	4 307	40-84	General population	Barbados	–	D
Holmen 1991	74 977	20+	General population	Norway	no lag	D
Holtzman 1983	5 155	42+	Men, civil service workers	Israel	no lag	D
Howteerakul 2006	527	35-60	Rural population	Thailand	–	–
HDFPCG 1977	158 906	30-69	General population	United States	no lag	A
Ibrahim 1995	6 733	47	General population	Egypt	–	–
Jamrozik 1989	5 627(1980), 7 640(1983)	25-64	General population (Risk Factor Prevalence Studies)	Australia	no lag	D
Jo 2001	4 226	18-92	General population	S. Korea	no lag	A
Joffres 1997	23 129	18-74	General population (Canadian Heart Health Surveys)	Canada	2 wks	A, D
Johansson 1999	249	20-60	General population	Sweden	no lag	D
Johnson 1994	42	24-68	Teachers (low income schools)	USA	no lag	B
Jones 1996	21 242	30+	General population	Korea	no lag	D
Jones 2001	1 138	18+	General population (Hispanics)	USA	no lag	A
Kalantan 2001	1 114	35-85	Attendants at Primary Health centres	Saudi Arabia	2 wks	D
Kamadjeu 2006	10 011	15+	Urban Cameroon	Cameroon	no lag	A
Kastarinen 2006	29 127	25-64	General population (FINRISK, MONICA)	Finland	–	–
Kattainen 2002	1 078(1978/80) 1 113 (1997)	65-74	General population (MiniFinland Survey, FINRISK)	Finland	–	D
Klungel 2000	56 026	20-59	General population	Netherlands	–	–
Lang 2001	29 626	38.8	Working population	France	–	D
Leynen 2006	20 911	35-59	General population (employees)	Belgium	no lag	–
Li 2003	9 703	35+	Clinic outpatients	China	no lag	D
Lima-Costa 2004	970	18+	General populatio	Brazil	–	A

Lin 1997	471 interv., 426 control	40 +	Rural population	Taiwan	no lag	A
Lisk 1999	1 204	15+	General population	Sierra Leone	no lag	A
M'Buyamba-Kabangu 1986	424	20+	General population	Zaire	4 wks	A
Maahs 2005	1 416	19-56	Diabetics and non-diabetics, no coronary disease	USA	no lag	D
Macedo 2005	5 023	18-90	General population	Portugual	no lag	A
Mancia 2006	4 590	40-75	Patients in physicians' offices	San Marino	no lag	D
Matenga 1997	749	34-64+	Heads of households	Zimbabwe	–	A
Meissner 1999	636	45-85+	General population	USA	1 wk	A, C
Metcalf 1996	1 238	15+	General population	South Africa	no lag	B
Mosca 2006	6 938	18-93	Women at free heart health screening	USA	no lag	B
Muntner 2004	14 989	35-74	General population in China, no CVD history	China	–	D
Nan 1991	5 074	25-74	General population	Mauritius	–	B
Nieto 1995	15 739	45-64	General population	USA	–	A, D
Nissinen 1988	3 807(1972), 3 630(1977), 2 499(1982)	30-59	General population (North Karelia Project)	Finland	–	D
Nova Scotia Dept. of Health 1996	3 227	18-65+	General Population	Canada	–	D
de Pablos-Velasco 2002	690	30-70+	General population (Canary Islands)	Spain	no lag	D
Panagiotakos 2003	2 282	18+	General population	Greece	–	–
Pan 2001	4 894	19+	General population (Nutrition and Health Survey in Taiwan)	Taiwan	–	A
Pappas 1990	8 632	18-74	General Hispanic population (HHANES)	USA	–	A
Paschal 2004	134	20-74	Low income African American	USA	1 wk	D
Pavlik 1997	962	18+	African American	USA	no lag	B, D
Pica 1990	1 374	30-69	Laval population	Canada	–	A
Plans 1993	704	15+	General population	Spain	–	–

Prencipe 2000	1 032	65+	General population	Italy	2 wks	A
Primatesta 2001	11 529	46.3	General population (Health Survey for England 1998)	England	–	A
Primatesta 2006	8 834	16+	General population (Health Survey for England 2003)	England	2 wks	A
Psaltopoulou 2004	26 913	20-86	General population	Greece	–	D
Puras 1998	1 322	48(18)	General population	Spain	no lag	–
Ramirez 1995	9 880	18-74	General population	Paraguay	no lag	A
Reiff 2001	695	18-65	African American, Latino (Harlem Household Survey)	USA	no lag	A
Ruixing 2006	2 184	all ages	Village population	China	–	–
Rywik 1998	7 497 (USA), 1 835 (POL)	45-64	General population (ARIC-USA, POL-MONICA)	USA, Poland	–	–
Sarafidis 2004	1 937	15-73	Factory employees	Greece	1 wk	C
Satish 1997	2 727	65+	Hispanic elderly	USA	no lag	A
Schoenberger 1972	22 929	all ages	Industrial employees	USA	–	C
Schmitz 2006	4 149	18-65	General population	Germany	–	–
Sekikawa 2004	8 926	20-69	General population	Japan	–	–
Shapiro 1978	12 055	25-69	General population	Canada	no lag	A, B, C
Sharma 2004	45 125	16-102	Primary care attendees	Germany	no lag	D
Shouqi 1995	950 356	15+	General population	China	–	–
Silverberg 1974	9 591	1-75+	Shopping center consumers	Canada	no lag	B
Smith 1990	10 359	40-59	General population	Scotland	no lag	D
Sonkodi 2004	1 012	37(11)	Factory employees	Hungary	no lag	C
Sonmez 1999	1 466	18+	General population	Turkey	no lag	A, C
Spencer 2005	100	63	Hypertensives	Ghana	no lag	D
Stamler 1976	1 049 225	all ages	General population	USA	no lag	B
Stein 2000	1 618	18-64	Employees	Bulgaria	–	C
Stergiou 1999	655	18+	General population	Greece	14 days	D
Steyn 2001	13 802	15+	General population	South Africa	no lag	A
Takala 1978	2 265	40-64	General population	Finland	no lag	D

Tormo 2000	6 533	30-69	Healthy, mainly blood donors (Spanish EPIC cohort)	Spain	no lag	–
Trenkwalder 1994	982	65+	Rural villagers	Germany	no lag	A
Aytekin 2002	1 992	30+	General population	Turkey	no lag	A, C
Vaccarino 1995	3 401	all ages	Employees at computer firm	Italy	no lag	C
van Rossum 2000	7 983	55+	General population	Netherlands	–	D
Wagner 1984	2 026	18-60+	Rural population	USA	no lag	A
Wander 1994	1 100	30-90	Rural villagers	India	no lag	A
Welty 2002	3 638	45-74	American Indians	USA	–	–
Wilber 1973 Baldwin County	3 085	15+	Rural population	USA	no lag	A
Atlanta	6 012	15+	General population	USA	no lag	A, B
Wilson 1991	77 890	1+	General population	Canada	no lag	B
Wolf 1999	917 (1985) 1 338 (1995)	25-64	General population (Halifax MONICA, Nova Scotia Health Survey)	Canada	–	–
Wolffsohn 2001	101	48-84	Patients at optometry clinic	Australia	no lag	D
Wood 2002	144	18-88	Health fair attendees	USA	no lag	C
Zdrojewski 2005	3 051	18-93	General population (NATPOL PLUS)	Poland	no lag	A
Zdrojewski 2006	36 696	18-98	General population	Poland	no lag	D

Footnotes: * continuous data; - information not provided; A home; B community location (shopping centre, school); C worksite; D medical centre, clinic, research centre; BMES - Blue Mountains Eye Study, Australia; BRINH - Belgian inter-university Research on Nutrition and Health; CVD - Cardiovascular Disease; ARIC - Artherosclerosis Risk in Communities Study

Table 4: Blood pressure measurement characteristics

First Author - Year	Measures used/ No. of visits	Rest prior	Cuff sized to person	Position	Device	Phase V for DBP	SR[c] hypertension question	Hyp.[d] cut-off	Measurement personnel[e]
Alonso 2005	2/1 visit	5 min	–	seated	Automatic	NA	Ever been told/diagnosed	140/90[b]	D
Bowlin 1993	2 of 3/1 visit	5 min	yes	seated	Random zero	–	Ever been told/diagnosed	140/90	D, N, O
Abernethy 1974	2/1 visit	No	–	seated	Random zero	–	–	200/95[b]	non-medical
Achananuparp 1989	2 of 3/2 visits	15 – 30 min	no	seated	Standard mercury	Y	History of HBP[f]	160/95[b]	D, O
Altun 2005	3/ 1visit	30 min[a]	yes	seated	Random zero	Y	Ever been told/diagnosed	140/90[b]	D
Amoah 2003	2/1 visit	10 min[a]	yes	–	Standard mercury	Y	–	140/90[b]	N
Arroyo 1999	–	5 min	yes	seated	Standard mercury	Y	Awareness of HYP	140/90[b]	N
Astagneau 1992	2/1 visit	5 min	yes	seated	Standard mercury	Y	Awareness of HYP	160/95[b]	O
Azizi 2002	2/1 visit	15 min[a]	yes	seated	Standard mercury	Y	Ever been told/diagnosed	140/90[b]	D
Banegas 1998	3/1 visit	5 min[a]	yes	seated	Random zero	Y	Ever been told/diagnosed	140/90[b]	O
Banegas 2002	6/ 1 visit	5 min[a]	yes	seated	Standard mercury/Automatic	Y	Ever been told/diagnosed	140/90[b]	O
Bharucha 2003	2/1 visit	resting	–	–	Random zero	–	–	140/90[b]	D
Birkett 1986	4/3 visits	5 min	yes	seated	Standard mercury	Y	–	DBP 90[b]	I
Birkett 1997	varying	5 min	yes	seated	Standard mercury	Y	–	various[b]	I

Birkett 1985/87	3/3rd visit	–	yes	–	Random zero	Y	–	DBP 89[b]	I
Brindel 2006	2/1 visit	5 min	yes	–	Automatic(Omron)	NA	History of HBP	160/95[b]	I
Brown 2005	2/1 visit	5 min	yes	–	Aneroid	–	Ever been told/diagnosed	140/90[b]	I
Burt 1995	varying/1 visit	–	no for NHES I, child for NHANES I and II, various for rest	seated	Standard mercury, aneroid (NHANES I)	Y/N	–	140/90[b]; 160/95[b]	D
Campbell 2001	2/1 visit	5 min	yes	seated	Standard mercury	Y	Awareness of HYP	140/90[b]	O
Cappucio 2004	2 of 3/1 visit	5 min	yes	seated	Automatic(Omron)	NA	–	140/90[b]	O
Carter-Edwards 2002	2/1 visit	5 min	yes	seated	Random zero	–	Ever been told/diagnosed	140/90[b]	O
Chadha 1990	2/1 visit	10 min	yes	seated	Aneroid	Y	Awareness of HYP	>160/90[b]	O
Chen 1995	3/1 visit	–	–	seated	–	Y	Ever been told/diagnosed	160/95[b]	D students
Chen 2003	2/1 visit	5 min	–	seated	Random zero	Y	Ever been told/diagnosed	140/90[b]; 160/95[b]	N
Choi 2006	2/1 visit	10 min	yes	seated	Standard mercury	–	Ever been told/diagnosed	140/90[b]	N
Chua 2005	1/1 visit	10 min	yes	seated	Standard mercury	–	Ever been told/diagnosed	140/90[b]	I
Cifkova 2004	2 of 3/1 visit	5 min	yes	seated	Standard mercury	Y	Ever been told/diagnosed	140/90[b]	D
Coulhoun 1998	2 of 3/1 visit	5 min	yes	seated	Automatic(Dinamap)	NA	Prior or current diagnosis	140/90[b]; 160/95[b]	N

De Backer 1998	2/1 visit	10 min	–	seated	Standard mercury	Y	Ever been told/diagnosed	160/95[b]	O
De Henauw 1998	2/1 visit	5 min[a]	yes	seated	Random zero	Y	Ever been told/diagnosed	160/95[b]	O
Deprez 1985	2 of 3/1 visit	–	–	–	–	–	Ever been told/diagnosed	DBP 90[b]	–
Dickson 2006	–	–	–	–	–	–	Ever been told/diagnosed	140/90[b]	N or O
Duprez 2002	3/1 visit	5 min[a]	yes	seated	Standard mercury	Y	–	140/90	D
Efstratopoulous 2006	2 on 2nd visit	5 min	yes	seated	Standard mercury	Y	Ever been told/diagnosed	140/90[b]	D
Entwisle 1983	2 of 3/1 visit	–	–	seated	Standard mercury	Y	Awareness of BP	DBP > 90[b]; 95 if >50yrs[b]	I
Fodor 2004	5 of 6/1 visit	5 min[a]	–	seated	Automatic(Bp TRU)	NA	Ever been told/diagnosed	140/90[b]	D, N
Foster 1993	–	–	yes	seated	Standard mercury	Y	Ever been told/diagnosed	160/95[b]	O
Franco 1985	2 of 3/1 visit	5 min	–	seated	Random zero	Y	Ever been told/diagnosed	DBP 95[b]	O
Freeman 1996	3/1 visit	10 min	yes	seated	Standard mercury	Y	Ever been told/diagnosed	140/90[b]; 160/95[b]	I
Freeman 1985	–	–	–	–	–	–	Ever been told/diagnosed	140/90[b]	–
Furumoto-Dawson 2003	–	–	–	supine	–	–	Knowledge of HYP	140/90	O
Garbus 1976	1/1 visit	–	yes	–	Manual, no details	Y	History of HBP	160/90[b]	D, N
Garcia-Pena 2001	2/1 visit	5 min[a]	yes	seated	Standard mercury	Y	Ever been told/diagnosed	160/90	N
Gasse 2001	2 of 3/1 visit	30 min[a]	yes	seated	Random zero	Y	Ever been told/diagnosed	140/90[b]	O
Gelberg 1989	up to 2/1 visit	resting	–	–	–	–	Ever had HYP	140/90	D, D students

Gentleman 1992	up to 2/1 visit	5 min	yes	seated	Standard mercury	Y	Ever been told/diagnosed	140/90	N
Getliffe 2000	up to 2/1 visit	5 min	–	seated	Automatic(Omron)	NA	History of HBP	DBP 89	N students
Giles 1995	2 of 3/1 visit	–	yes	seated	Random zero	Y	Ever been told/diagnosed	140/90[b]	O
Glover 2005	3/1 visit	–	–	–	–	–	–	140/90[b]	–
Gonzalez-Villalpando 1999	2 of 3/1 visit	5 min	yes	seated	Random zero	Y	–	160/95[b]	O
Goldman 2003	2 /1 visit	20 min	–	seated	Standard mercury	–	Ever had or currently have	140/90[b]	N
Green 1992	2 of 3/1 visit	5 min	–	seated	Standard mercury	Y	–	160/95[b]	O
Gu 2002	3/1 visit	5 min[a]	yes	seated	Standard mercury	–	Ever been told/diagnosed	140/90[b]	O
Guimarães 2002	–	–	–	–	Sphygmomanometer, no details	–	–	140/90[b]	D, N, O
Halabi 1992	–	–	–	–	–	–	Current HYP	–	D
Handa 1980	lower of 2/1 visit	5 min	yes	seated	Sphygmomanometer, no details	Y	–	155/95[b]	D, O
Health and Welfare Canada 1989	2/1 visit	–	–	–	Standard mercury	Y	–	DBP 90[b]	N
Hennis 2002	2/1 visit	5 min	yes	seated	Random zero	Y	Ever been told/diagnosed	140/90[b]	O
Holmen 1991	2/1 visit	2 min	no	seated	Standard mercury	Y	–	various	N
Holtzman 1983	2nd of 2/1 visit	–	–	lying down	Manual, no details	Y	–	150/90[b] (1963); 160/95[b] (1965)	–

Howteerakul 2006	2/1 visit	15 min[a]	–	seated	Standard mercury	–	Ever been told/diagnosed	140/90[b]	I
HDFPCG 1977	2 of 3/1 visit	–	yes	seated	Standard mercury	Y	Ever been told/diagnosed	DBP 95[b]	–
Ibrahim 1995	4/1 visit	–	–	–	–	–	–	140/90[b]	–
Jamrozik 1989	2/1 visit	5 min	NS – probably	seated	Standard mercury	Y	Ever been told/diagnosed	160/95[b]	N
Jo 2001	3/1 visit	5 min[a]	–	seated	Standard mercury	–	–	140/90[b]	N students
Joffres 1997	4/2 visits	5 min[a]	yes	seated	Standard mercury	Y	Ever been told/diagnosed	140/90[b]	N
Johansson 1999	–	2 min	–	supine	–	–	Current HYP	DBP 90	N
Johnson 1994	–	–	–	–	–	–	–	140/90	D students
Jones 1996	2/1 visit	5 min	yes	seated	Standard mercury	–	Ever been told/diagnosed	140/90[b]; 160/95[b]	O
Jones 2001	1/1 visit	–	–	–	–	–	Ever been told/diagnosed	140/90	med prof.
Kalantan 2001	4/2 visits	5 min	–	seated	Standard mercury	Y	–	140/90[b]	D
Kamadjeu 2006	3/1 visit	5 min	yes	seated	Automatic	NA	Awareness of HYP	140/90[b]	O
Kastarinen 2006	2/1 visit	5 min	no	seated	Standard mercury	Y	Ever been told/diagnosed	140/90[b]	N
Kattainen 2002	–	–	–	–	–	–	Ever been told/diagnosed	160/95[b]	D
Klungel 2000	2/1 visit	5 min	–	seated	Random zero	Y	Ever been told/diagnosed	140/90[b]; 160/95[b]	O
Lang 2001	3/1 visit	5 – 7min	yes	seated	Automatic(Omron)	NA	–	140/90[b]	D
Leynen 2006	2nd of 2/1visit	–	–	–	–	–	Ever been told/diagnosed	140/90[b]	–
Li 2003	2/1 visit	no	–	–	–	–	Current HYP	140/90[b]	D

Lima-Costa 2004	2 of 3/1 visit	5 min[a]	–	seated	Standard mercury	–	Ever been told/diagnosed	140/90[b]	health workers
Lin 1997	2/1 visit	5 min	–	seated	Sphygmomanometer, no details	Y	Ever been told/diagnosed	160/95[b]	N
Lisk 1999	1/1 visit	10 min	no	seated	Standard mercury	Y	–	160/95[b]	D, N
M'Buyamba-Kabangu 1986	10/2 visits	5 min	no	seated	Aneroid	Y	Awareness of HYP	160/95[b]	O
Maahs 2005	2 of 3/1 visit	resting	–	seated	–	Y	Ever been told/diagnosed	140/90[b]	–
Macedo 2005	3/1 visit	5 min[a]	yes	seated	Automatic(Omron)	NA	Ever been told/diagnosed	140/90[b]	O
Mancia 2006	2/1 visit	10min	–	seated	Sphygmomanometer	Y	Awareness of HYP	140/90[b]	D
Matenga 1997	3/1 visit	5 min	yes	seated	Standard mercury	Y	Ever been told/diagnosed	DBP 95[b]	N
Meissner 1999	6/3 visits	5 min[a]	yes	seated	Standard mercury/random zero	Y	Ever been told/diagnosed	140/90[b]	N
Metcalf 1996	lower of 3/1 visit	–	–	–	Standard mercury	–	Ever been told/diagnosed	160/95[b]	–
Mosca 2006	1/1 visit	1 – 2 min	yes	seated	Automatic	NA	History of HBP	140/90	trained hospital personnel
Muntner 2004	3/1 visit	5 min	–	seated	–	–	Ever been told/diagnosed	140/90[b]	O
Nan 1991	2/1 visit	5 min	no	seated	Standard mercury	Y	–	160/95[b]	O
Nieto 1995	2 of 3/1 visit	5 min[a]	–	seated	Random zero	Y	–	140/90[b]	O
Nissinen 1988	–	5min	no/ yes in 1982	seated	Standard mercury	Y	–	175/100[b]	N
NS Dept. Health 1996	–	–	–	–	–	–	–	140/90[b]	N

de Pablos-Velasco 2002	2/1 visit	–	–	seated	–	–	Ever been told/diagnosed	140/90[b]	D, N
Panagiotakos 2003	3/1 visit	30 min	–	seated	Aneroid	Y	Awareness of HYP	140/90[b]	D
Pan 2001	2 of 3 /1 visit	5 min	yes	seated	Standard mercury	–	–	140/90[b]; 160/95[b]	D, N
Pappas 1990	2/1 visit	5 min	–	seated	Standard mercury	Y	Ever been told/diagnosed	160/95[b]	D
Paschal 2004	–	–	–	–	–	–	Ever been told/diagnosed	–	–
Pavlik 1997	2 of 3/1 visit	–	yes	seated	Standard mercury	Y	Ever been told/diagnosed	140/90[b]	I
Pica 1990	2 of 3/1 visit	–	yes	seated	Standard mercury	Y	–	140/90[b]	I
Plans 1993	lower of 2 /1 visit	10 min	–	–	Standard mercury	Y	Awareness of HYP	160/95[b]	–
Prencipe 2000	6/2 visits	5 min	yes	seated	Standard mercury	Y	Ever been told/diagnosed	140/90[b]	D, I
Primatesta 2001	2 of 3/1 visit	5 min[a]	yes	seated	Automatic(Dinamap)	NA	Ever been told/diagnosed	140/90[b]; 160/95[b]	N
Primatesta 2006	2 of 3/1 visit	5 min[a]	yes	seated	Automatic(Omron)	NA	Ever been told/diagnosed	140/90[b]	N
Psaltopoulou 2004	2/1 visit	5 min	–	seated	Standard mercury	–	Ever been told/diagnosed	140/90[b]	D
Puras 1998	lower of 2/1 visit	15 min[a]	–	seated	Standard mercury	Y	Ever been told/diagnosed	140/90[b]; 160/95[b]	N
Ramirez 1995	1 of 2/1 visit	10 min[a]	–	seated	Aneroid	Y	–	160/105	N
Reiff 2001	2/1 visit	–	–	–	Standard mercury	–	Current HYP	140/90[b]	I
Ruixing 2006	3/1 visit	5 min	yes	seated	Standard mercury	Y	Ever been told/diagnosed	140/90[b]	–

Rywik 1998	2/1 visit	5 min	ARIC yes POL MONICA no	seated	Random zero/standard mercury	Y	–	160/95[b]	–
Sarafidis 2004	3/2nd visit	5 min	no	seated	Standard mercury	Y	Ever been told/diagnosed	140/90[b]	D
Satish 1997	2/1 visit	–	yes	seated	Standard mercury	–	Ever been told/diagnosed	140/90[b]	I
Schoenberger 1972	1/1 visit	–	–	supine	Standard mercury	–	Awareness of HYP	160/95[b]	N
Schmitz 2006	2 of 3/1 visit	3 min	–	seated	–	–	Ever been told/diagnosed	140/90[b]	D
Sekikawa 2004	–	–	–	–	Standard mercury	–	–	140/90[b]	–
Shapiro 1978	2/1 visit	–	–	–	Sphygmomanometer, no details	4th	–	160/95[b]	D, D student
Sharma 2004	–	–	–	–	Automatic/sphygmomanometer	–	History of HBP	140/90[b]	D
Shouqi 1995	3/1 visit	5 min	–	seated	Standard mercury	Y	Ever been told/diagnosed	140/90[b]; 160/95[b]	O
Silverberg 1974	lower of 2/1 visit	–	yes	seated	Standard mercury	Y	–	155/95(<40yrs); 160/95(40 – 64yrs); 165/100 (>64 yrs)	N, O
Smith 1990	2/1 visit	5 min	–	seated	Random zero	Y	Ever been told/diagnosed	160/95[b]	N
Sonkodi 2004	4 of 5/1 visit	5 min[a]	–	seated	Automatic(BpTRU)	NA	Ever been told/diagnosed	140/90[b]	D, N
Sonmez 1999	2/1 visit	5 min	no	seated	Aneroid	Y	Awareness of BP	140/90[b]; 160/95[b]	I

Spencer 2005	2/1 visit	–	yes	seated	Standard mercury	Y	Ever been told/diagnosed	140/90[b]	–
Stamler 1976	1 reading	–	–	seated	Manual, no details	Y	Current HYP	DBP 95[b]	D, N, O
Stein 2000	2/1 visit	5 min	no	seated	Standard mercury	–	Current HYP	140/90[b]	I
Stergiou 1999	3/2nd visit	5 min	no	seated	Standard mercury	Y	–	140/90[b]	D
Steyn 2001	3/1 visit	5 min	–	seated	Automatic(Omron)	NA	–	140/90[b]; 160/95[b]	O
Takala 1978	1/1 visit	3 – 5 mina	–	seated	Standard mercury	Y	–	160/95[b]; DBP 95[b]	N
Tormo 2000	2/1 visit	–	–	seated	Automatic(Accutorr and Boso – Oscillomat)	NA	Ever been told/diagnosed	160/95	–
Trenkwalder 1994	2 of 3/1 visit	–	yes	seated	Standard mercury	–	–	160/95[b]	D
Aytekin 2002	2/1 visit	5 min	–	seated	Standard mercury	–	–	140/90[b]	D students
Vaccarino 1995	2/1 visit	5 min	–	seated	–	Y	Ever been told/diagnosed	140/90[b]	D
van Rossum 2000	2/1 visit	–	–	seated	Random zero	–	Ever been told/diagnosed	160/95[b]	–
Wagner 1984	2 of 3/1 visit	–	–	seated	Standard mercury	Y	–	DBP 90[b]	I
Wander 1994	1/1 visit	–	–	seated	–	–	–	160/95	D
Welty 2002	–	–	–	–	–	–	–	140/90[b]	–
Wilber 1973 Baldwin County	3/1 visit	–	–	seated	Standard mercury	–	Awareness of HYP	160/95[b]	N
Atlanta	–	–	–	–	–	–	–	160/95[b]	–

Wilson 1991	1 of 2/1 visit	5 min	yes	supine	Standard mercury	–	–	140/90b(<40 yrs); 160/95b(40 – 59); 165/100b(>59yrs)	N
Wolf 1999	2/1 visit			seated	Standard mercury Random zero	Y	Ever been told/diagnosed	160/95	I
Wolffsohn 2001	2/1 visit	5 min	–	seated	Automatic(Omron)	NA	Current HYP	160/95[b]	–
Wood 2002	1/1 visit	5 min	yes	seated	–	–	Awareness of BP	140/90	–
Zdrojewski 2005	2nd and 3rd/3 visits	5 min[a]	yes	seated	Automatic(Omron)	NA	Ever told	140/90[b]	N
Zdrojewski 2006	2/1 visit	2 – 5 min	yes	seated	Automatic(Omron)	NA	Ever told or current HYP	140/90[b]	N

Footnotes: [a] additional restrictions such as no coffee, tea, alcohol, smoking, exercise or eating; [b] Medication is considered in defining hypertension; [c] Self-Report; [d] Hypertension; [e] D doctor; N nurse; O other; I interviewer or researcher; [e] High Blood Pressure; – information not provided; NA: Not applicable, DBP: Diastolic Blood Pressure

Table 5: Difference in reported and measured hypertension prevalence in studies that report both values (studies with reported prevalence greater than measured are in italics)

	Hypertension cut-point	Self-Report (%)	Measured (%)	Difference (Reported-Measured)
Overall				
Alonso 2005	*140/90*	*62.2*	*56.7*	*5.5*
Amoah 2003	140/90	12.6	28.4	-15.8
Brown 2005	140/90	0	33.2	-33.2
Garbus 1976	160/90	20	40.1	-20.1
Garcia-Pena 2001	160/90	43	48.3	-5.3
Gelberg 1989	140/90	20.5	28.1	-7.6
Gentleman 1992	140/90	13.8	26.4	-12.6
Getliffe 2000	DBP 89	11	18	-7.0
Giles 1995	140/90	28	29.4	-1.4
Goldman 2003	140/90	30.3	57.3	-27.0
Joffres 1997	140/90	21	22	-1.0
Johnson 1994	140/90	14	38.1	-24.1
Klungel 2000	140/90	9.4	17.1	-7.7
Klungel 2000	160/95	6.7	8.8	-2.1
Lima Costa 2004	*140/90*	*27.2*	*23.3*	*3.9*
Paschal 2004	not stated	56	62	-6.0
Psaltopoulous 2004	140/90	24.2	44.4	-20.2
Reiff 2001	140/90	22	33	-11.0
Sarafidis 2004	140/90	6	30.5	-24.5
Schoenberger 1972	160/95	8.3	11.9	-3.6
Sonmez 1999	140/90	19.4	29.6	-10.2
Sonmez 1999	160/95	19.4	22.5	-3.1
Wolffsohn 2001	160/95	39.6	66.3	-26.7
Men				
Alonso 2005	140/90	21	37	-16.0
Bowlin 1993	140/90	21	37	-16
Brown 2005	140/90	0	39.10	-39.1
Gentleman 1992	140/90	11	30.4	-19.4
Joffres 1997	140/90	20	26	-6.0
Johansson 1999	DPB 90	4	12	-8.0
Kattainen 2002 (1978-80)	160/95	22.5	28.7	-6.2
Kattainen 2002 (1997)	160/95	32.3	37.1	-4.8
Psaltopoulous 2004	140/90	21	45.2	-24.2
Reiff 2001	140/90	17	33	-16.0
Women				
Alonso 2005	140/90	22	36	-14.0
Bowlin 1993	140/90	22	36	-14
Brown 2005	140/90	0	30.60	-30.6
Furumoto-Dawson 2003	140/90	17.4	43.6	-26.2
Gentleman 1992	140/90	16.1	23.1	-7.0
Joffres 1997	*140/90*	*21*	*18*	*3.0*
Johansson 1999	DBP 90	4	12	-8.0

	Hypertension cut-point	Self-Report (%)	Measured (%)	Difference (Reported-Measured)
Kattainen 2002 (1978-80)	160/95	45.7	46.3	-0.6
Kattainen 2002 (1997)	*160/95*	*31.9*	*30.2*	*1.7*
Psaltopoulous 2004	140/90	26.4	43.8	-17.4
Reiff 2001	140/90	25	33	-8.0

DBP = diastolic blood pressure

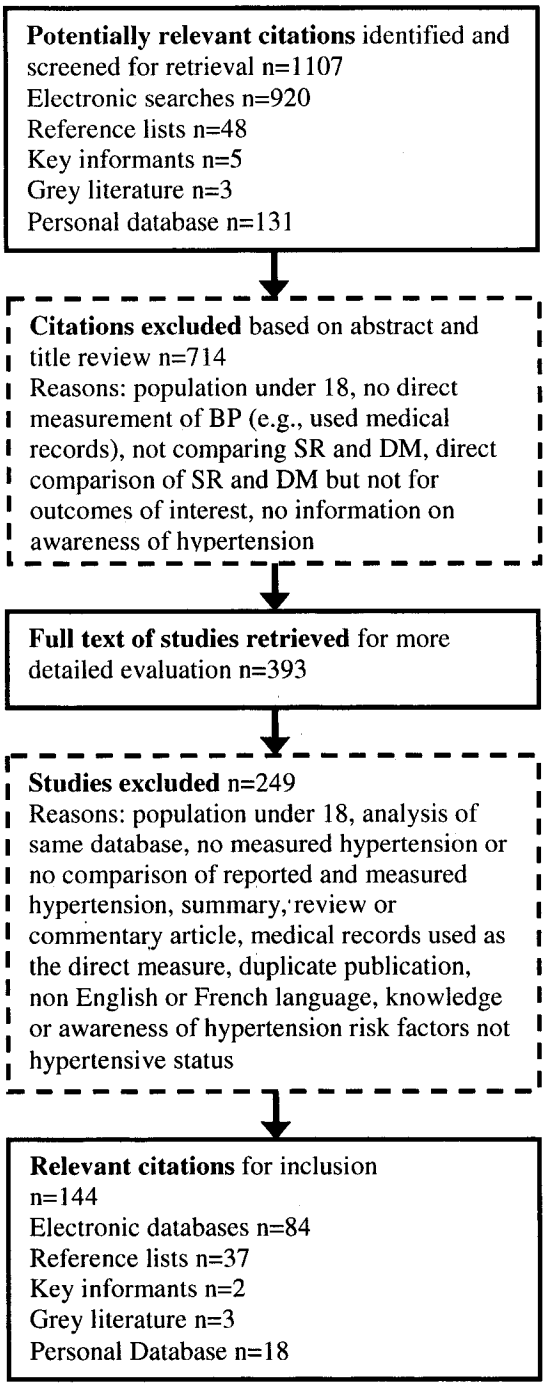

Figure 1: Results of the literature search. BP: Blood Pressure; SR: Self-Report; DM: Direct Measurement.

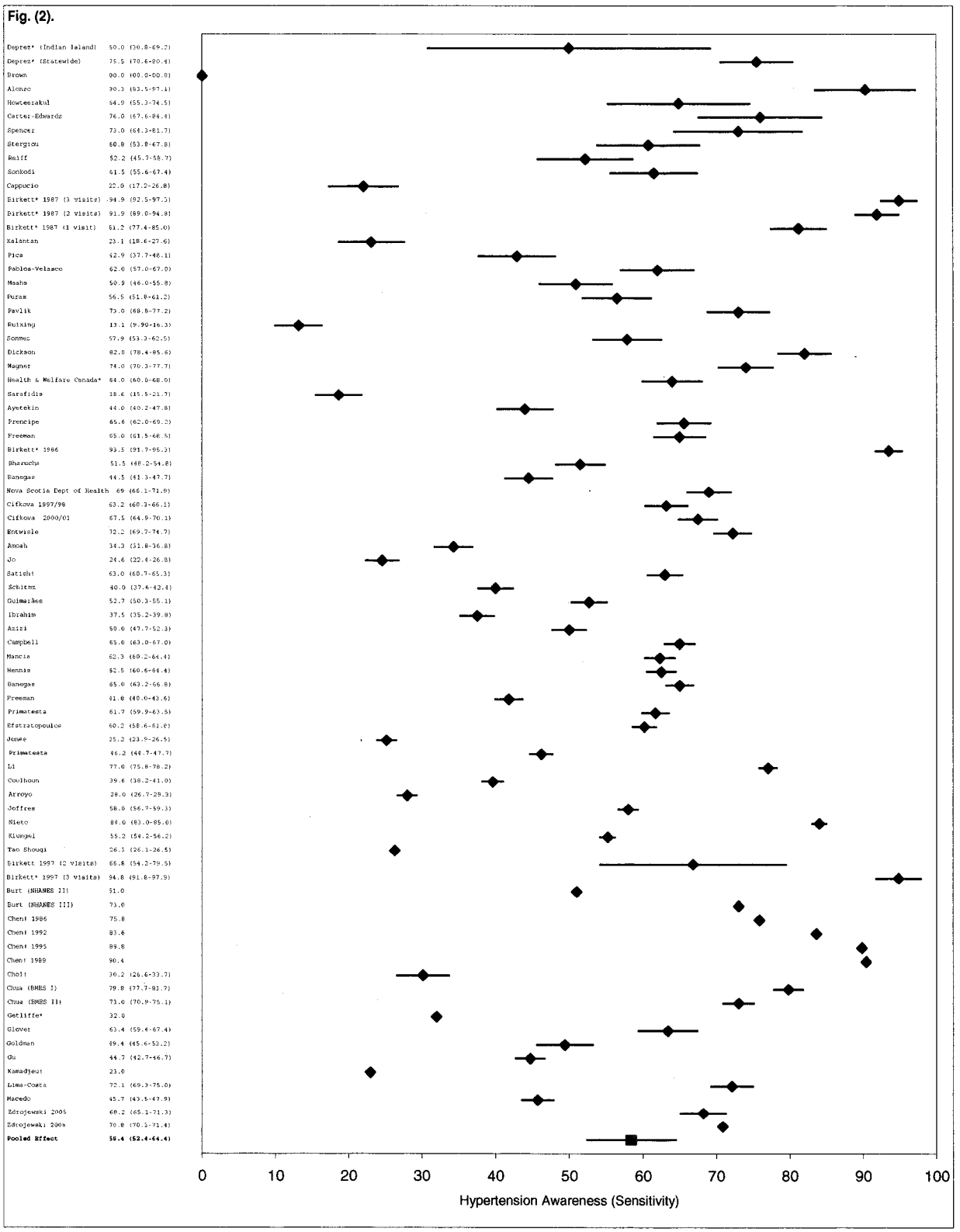

Figure 2: Sensitivity of hypertension awareness among male and female hypertensives for the 140/90 mmHg cut off. Study author and awareness level (95% CI) provided. Studies presented in order of sample size (smallest at top). * based only on DBP > 95; † age and/or sex standardized.

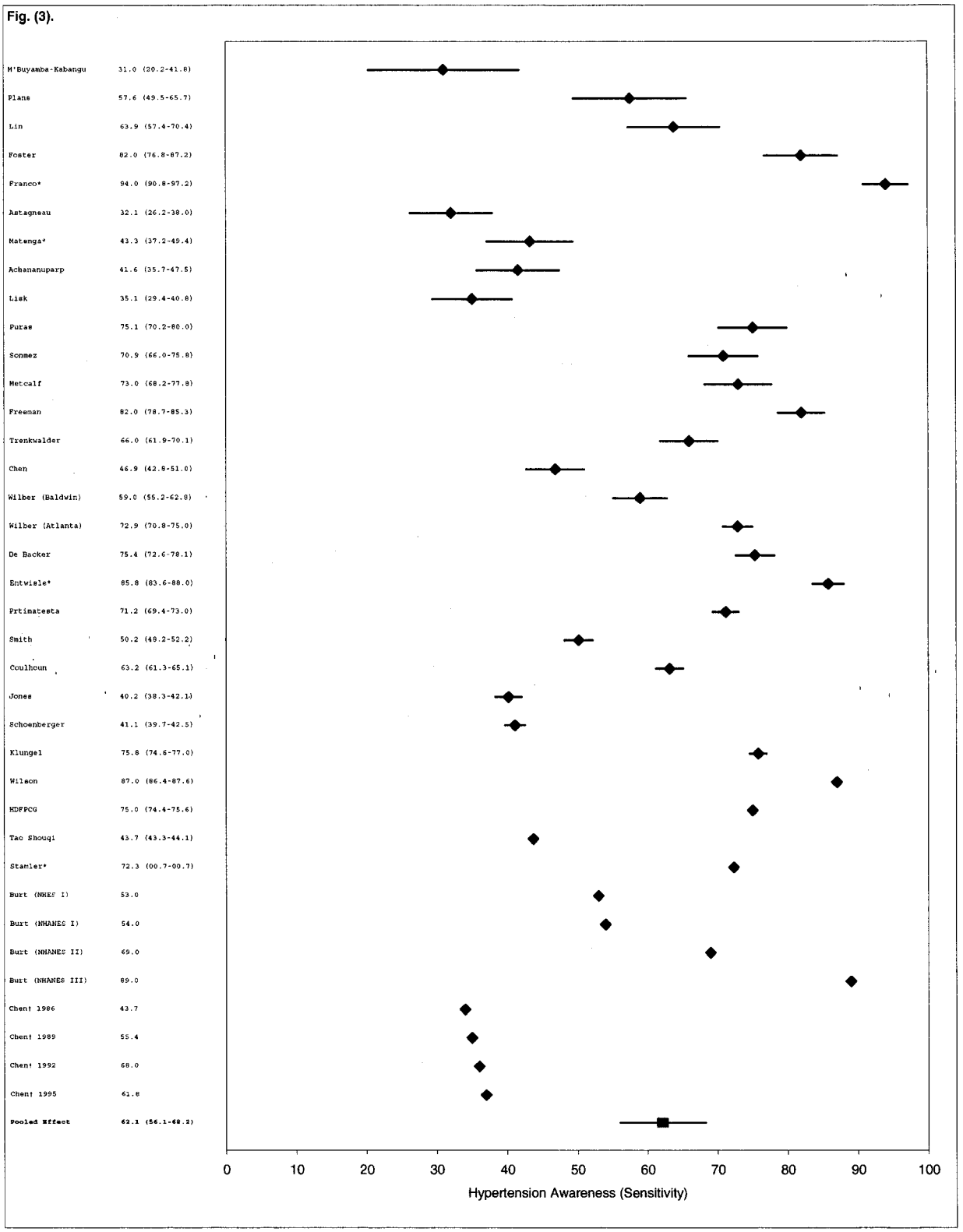

Figure 3: Sensitivity of hypertension awareness among male and female hypertensives for the 160/95 mmHg cut off. Study author and awareness level (95% CI) provided. Studies presented in order of sample size (smallest at top). * based only on DBP > 90; † age and/or sex standardized.

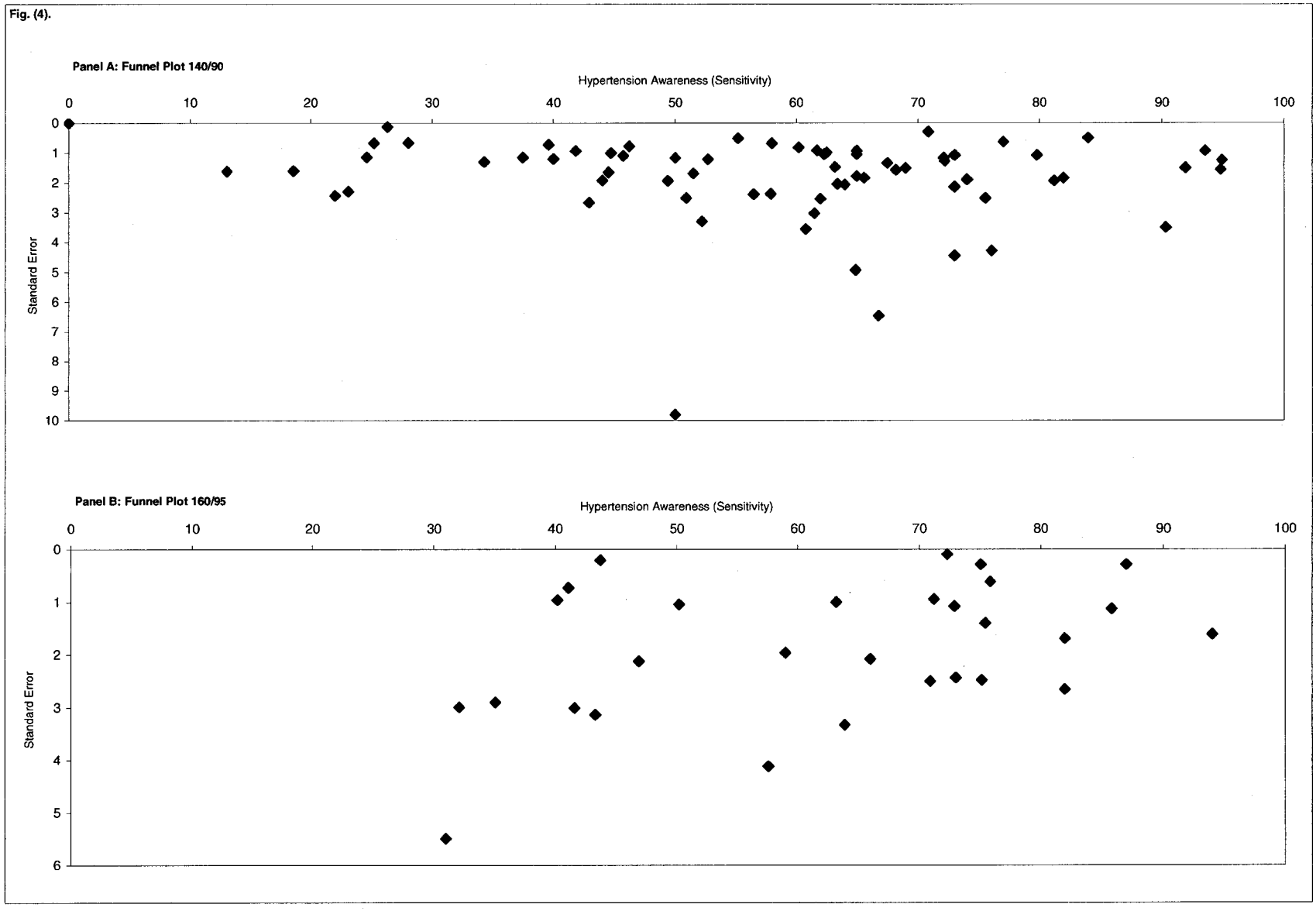

Figure 4: Funnel plots of hypertension awareness by standard error of studies. Panel A: Funnel Plot 140/90;Panel B: Funnel Plot 160/95.

Paper 3: The accuracy of self-reported smoking: A systematic review of the relationship between self-reported and cotinine assessed smoking status

This paper has been published in:

Connor Gorber S, Schofield-Hurwitz S, Hardt J, Tremblay MS, Levasseur G. The accuracy of self-reported smoking: a systematic review of the relationship between self-reported and objectively assessed smoking status. *Nicotine and Tobacco Research* (in press).

The accuracy of self-reported smoking: A systematic review of the relationship between self-reported and cotinine assessed smoking status

Sarah Connor Gorber, Sean Schofield-Hurwitz, Jill Hardt, Geneviève Levasseur, Mark Tremblay

Sarah Connor Gorber, Sean Schofield-Hurwitz and Jill Hardt are from the Health Information and Research Division, Statistics Canada, Ottawa, Ontario, Canada. Mark Tremblay is with the Physical Health Measures Division, Statistics Canada, Ottawa, Ontario, Canada. Geneviève Levasseur is with the Tobacco Control Programme at Health Canada, Ottawa, Ontario, Canada.

The work was performed at Statistics Canada.

Correspondence: Sarah Connor Gorber, Chief – Health Measures Analysis Section, Health Information and Research Division, Statistics Canada, 24th Floor R.H. Coats Building, 100 Tunney's Pasture Driveway, Ottawa, ON, Canada, K1A 0T6

Tel:
Fax:
Email:

Abstract

Smoking is a leading cause of premature mortality and preventable morbidity. Surveillance is most often based on self-reported data, but studies have shown that self- reports tend to underestimate smoking status. This study systematically reviewed the literature to measure the concordance between self-reported smoking status and smoking status determined through measures of cotinine in biological fluids. Four electronic databases were searched to identify observational and experimental studies on adult populations over the age of 18 years. Searching identified 67 studies that met the eligibility criteria and examined the relationship between self-reported smoking and smoking confirmed by cotinine measurement. Overall, the data show trends of underestimation when smoking prevalence is based on self-report and varying sensitivity levels for self-reported estimates depending on the population studied and the medium in which the biological sample is measured. Sensitivity values were consistently higher when cotinine was measured in saliva instead of urine or blood. Meta-analysis was not appropriate because of the substantial heterogeneity among the cut-points used to define smokers and the poor reporting on outcomes of interest. Further research in this field would benefit from the standardization of cut-points to define current smokers and the implementation of standard reporting guidelines to enhance comparability across studies. Accurate estimation of smoking status is important as data from population studies such as those included in this review are used to generate regional and national estimates of smoking status and in turn are used to allocate resources and set health priorities.

Introduction

Smoking is a leading cause of premature mortality and preventable morbidity worldwide, globally responsible for an estimated 26 million years of life lost[1]. The estimated economic burden of tobacco use was almost 18 billion dollars in Canada in 2002[2] and in the United States tobacco was responsible for $75 billion in direct medical costs, and $82 billion in lost productivity[3].

Tobacco surveillance is often based on self-reported information, usually generated from population level surveys, yet studies have suggested that self-reported estimates may underestimate true smoking prevalence[4-6]. Socially undesirable behaviours are particularly prone to under-reporting[7,8] with the degree of bias dependent on the population examined. For example, among populations where smoking is seen as particularly undesirable, such as in pregnant women[9] or individuals suffering from smoking-related diseases[10], the discrepancy between measured and reported smoking rates can be high. Increased public health concerns about the dangers of active and passive smoking have strengthened the perception of smoking as a socially undesirable behaviour, which could further undermine the validity of self-reported estimates[7].

Many studies have investigated the accuracy of self-reports of tobacco use, most often through biochemical validation of cotinine, carbon monoxide, nicotine or thiocyanate levels[11,12] but the conclusions from these studies have not been consistent with sensitivity values ranging from 0[13] to 100%[14]. In order to draw definitive conclusions about the accuracy of self-reported smoking a comprehensive systematic assessment of the

relationship between estimates of smoking status determined through self-report and those based on direct measurement is required. Such an assessment is important because accurate information about the prevalence of tobacco smoking is required in order to monitor population health, assess health risks, and evaluate smoking reduction and cessation interventions[15].

This study presents results from the first known systematic review of the literature to measure the concordance between self-reported smoking status and smoking status determined through measures of cotinine in biological fluids. Cotinine is recognized as the most appropriate indicator of tobacco smoke exposure, except in persons using nicotine replacement therapy, and is widely accepted and used, despite costing more than other biomarkers[12,16-19], This review is limited to studies that used cotinine as an indicator of cigarette smoking, because unlike carbon monoxide and thiocyanate, cotinine is not influenced by diet or pollution exposure. It also has a longer half-life and is less dependent on temporal factors than nicotine[12,16]. In addition as large scale population health surveys such as the National Health and Nutrition Examination Surveys (NHANES) in the United States and the Canadian Health Measures Survey (CHMS) in Canada rely on cotinine as an indicator of smoking status, focusing on cotinine increases the relevance of this review for population health research.

Although there have been previously published reviews assessing the validity of self-reported smoking[17,18,20] none of these reviews were systematic and they were limited by the small number of studies examined, the types of studies included, the mediums of

measurement (e.g., Patrick and colleagues[17] included no urinary measures) and their age (the most recent was published 14 years ago). In addition, given the relationship between reporting bias and social desirability, and the decline in the social acceptability of smoking in recent years, regular updates of the literature are required in order to determine if the bias changes as the social acceptance of smoking changes. Moreover, in 2002 the Society for Research on Nicotine and Tobacco's Subcommittee on Biochemical Verification (SRNT)[19] published a series of recommendations outlining for instance, the most appropriate biochemical markers and cut-points for determining smoking status and the acceptable time lag between self-reported smoking and biochemical verification. They called for future research to further examine the utility of biochemical verification across a range of studies and populations. This review attempted to address this gap in the literature and also assess the degree to which the SRNT recommendations have been addressed in the design, conduct and reporting of validity studies since their report was released.

Methods

Studies were eligible for inclusion if they collected data on, and directly compared, self-reported smoking status and cotinine measures that were taken from a biological specimen. Only studies examining adult populations (mean age of participants was 18 years of age or older) were eligible for inclusion. This review focused on regular cigarette smoking only.

Eligible self-reported measures of cigarette smoking included interviews, questionnaires, surveys, unstructured or informal inquiries. Studies providing self-reported data by proxy were excluded. Cotinine could be measured in any medium including blood (serum, plasma

or whole blood), urine, saliva, or hair samples. Studies that examined either self-reported cigarette smoking or directly measured cotinine levels, but not both, were excluded. Studies were included regardless of study design or publication status. No language restrictions were imposed in the literature searching. Abstracts and letters were included if they provided sufficient details to meet inclusion criteria.

Given that the review focused on studies that directly compared self-reported and cotinine assessed smoking status, some smoking cessation studies of general population smokers were not included. Because cessation studies are aimed at helping smokers to quit, they tend to focus on confirmed smokers and many cessation studies do not specifically examine the consistency between measured and reported estimates, which makes them ineligible for inclusion in this review.

Search Strategy

The following bibliographical databases were searched: Medline (up to 2006 October week 1); EMBASE (up to 2006 week 41); CINAHL (up to 2006 October week 2); and PsycINFO (up to 2006 October week 2). The search strategy used for MEDLINE is provided in Table 1, and was modified according to the indexing systems of other databases. The OVID interface was used for all searches. Grey literature such as published abstracts and conference proceedings, theses and dissertations, and government and organizational reports were also identified where possible. The database of the Tobacco Control Programme at Health Canada (Canada's Federal Department responsible for health) was also searched. This database contains over 13,000 tobacco-related records.

Titles and abstracts of potentially relevant citations were first reviewed to assess correspondence to the inclusion criteria. The full text of all articles that met the criteria was then further screened by two independent reviewers. The bibliographies of texts were also searched in order to identify additional studies, which were subsequently retrieved. Disagreements between reviewers occurred in slightly more than 1% of the studies examined; they were discussed and in all cases a consensus was achieved through these discussions.

Standardized data abstraction forms were completed by one reviewer and verified by the other. Information was extracted on the study design, participant characteristics, sample size, measurement medium, measurement device, cut-point used for defining regular smokers, and the time delay between self-report and direct measurement. The main outcome in this review is the sensitivity of self-reported current smoking status using the biological measure of cotinine as the "gold standard" (i.e., fraction of true smokers who self-identify). Information to permit calculation of sensitivity and associated measures of variance was therefore extracted. Reviewers were not blinded to the authors or journals when extracting data.

When necessary data were not reported in the articles, authors were contacted and asked to provide the missing details. Authors of 34 studies were contacted and additional information was received for six studies.

Quality Assessment

All studies were assessed to determine their methodological quality. The Downs and Black 'checklist for measuring study quality'[21] has been recommended for assessing quality in systematic reviews[22,23] and was used in this review. However, as not all items were relevant to the studies included in this review a modified version of the checklist was employed with the following items omitted: items 5 and 8 in the reporting scale, items 14 and 19 in the section on bias, items 21–26 relating to confounding and item 27 addressing power. The final checklist consisted of 16 items with a maximum score of 16 points (with higher points indicating superior quality) rather than the original 32 points. The checklist covered the categories of reporting, external validity and internal validity (bias). Studies were assessed based on the degree to which they presented data relating to the outcomes studied in this review even if these were only secondary outcomes in the studies. The quality assessment was undertaken independently by two assessors; discrepancies in scoring occurred in approximately 2% of cases. These were discussed and consensus was achieved in all cases.

Results

Description of Studies

Two hundred and ninety-eight citations were identified based on the preliminary search of electronic databases, reference lists, and grey literature (see Figure 1). Of these 63 were identified in MEDLINE, an additional 65 in EMBASE, 75 in CINAHL and 11 in PsycINFO. After a preliminary title and abstract based review, 122 full text articles were retrieved for

further assessment. Of these 67 met the criteria for inclusion[4-7,9-14,16,20,24-78] and their characteristics are described in Table 2.

Studies were published over a 23 year period from 1983 to 2006. The primary aim of the papers differed but each of the included studies collected a biological measure of cotinine and a measure of self-reported smoking status. Of the 67 studies that were retained 55 were of varying observational designs and 12 had experimental or quasi-experimental designs. All studies were written in English.

Participants in the studies ranged from 12 to 89 years of age. Although the focus of the review is on those aged 18 and over, studies that had a range of ages less than 18 years were not excluded as long as the mean age of the sample was over 18 years. Sample sizes ranged from a low of 30[32,38] to a high of almost 10,000 in Woodward's work with the Scottish Heart Health Study[77]. Research from 18 countries is represented.

Different questions were used to measure current reported smoking status. Some studies asked if respondents were current regular smokers, while others asked them to estimate the numbers of cigarettes smoked per day. Most studies measured cotinine in saliva (n=29), urine (n=21) or serum (n=20). One study collected cotinine in semen, one in cord serum and one in infant meconium. No studies that measured cotinine in hair samples met the inclusion criteria. Samples were most often collected in a hospital or clinic setting, although in some cases data were collected in the home or worksites. In the majority of studies there was no delay between the self-reporting of smoking status and cotinine sampling, however, delays

of up to several weeks were reported in some instances. Cut-points used for determining whether an individual was classified as a smoker were highly variable ranging from 50 ng/mL to 500 ng/mL (284 nmol/L to 2837.5 nmol/L) in urine, from 8 ng/mL to 100 ng/mL (45.4 nmol/L to 567.5 nmol/L) in serum and from 7 ng/mL to 44 ng/mL (39.7 nmol/L to 250 nmol/L) in saliva. Nine studies corrected urinary cotinine concentrations for creatinine excretion to account for dilution effects. All cut-points were converted to SI Units (International System of Units) for comparison purposes using the following formula: ng/mL*5.675 = nmol/L[79].

Quality Assessment

All studies scored maximum points for clearly describing study objectives, main outcomes to be measured, participant characteristics and the main findings of the study. Although most studies carried out some sort of significance testing on results, many studies did not report the actual probability values associated with the estimates or their associated measures of random variability (e.g., standard error or confidence intervals). Fifty-four per cent (36/67) of studies did not report the characteristics of participants lost to follow up (for the purposes of this review the follow-up period is defined as the time between the self-report and biospecimen sampling).

In order to score maximum points (three) in the external validity section, studies must have reported that study participants, both those who were recruited and those who agreed to participate, were representative of the population from which they were recruited and that the staff, places and facilities where participants were interviewed and examined were

representative of the assessment the majority of patients would receive. Sixty per cent (40/67) of studies received a point for demonstrating that the subjects asked to participate in the study were representative of the source population. However, 88 per cent (59/67) of studies failed to validate that the subjects who did participate were representative of the broader population.

All studies received three or more of a possible five points in the internal validity category. However, only 14/67 studies blinded those collecting and analyzing the cotinine samples to the reported smoking status of respondents[4,12,14,24,42,43,48,56,62,64,66,68,70,78].

Data Synthesis

Fifty four studies (including sub-studies when more than one study was described in a single publication) compared prevalence estimates of smoking based on self-reported and measured values. Figure 2 displays the difference in prevalence (or bias) when smoking prevalence is based on self-reported versus directly measured values. Negative values indicate that self-reported estimates underestimate prevalence while positive values indicate overestimation (see Figure 2). In most cases self-reported smoking prevalence was lower than the directly measured value with the difference ranging from 1% to 47%. There were five studies in which the absolute difference (reported – measured) was greater than 24%; two of these were studies of pregnant women[6,73], one of participants in a smoking cessation intervention[37] and one of patients with coronary heart disease[50]. In three studies[35,38,62] the self-reported and measured estimates were identical, and self-report overestimated true smoking status in 11 studies. The mean difference between reported and measured prevalence estimates was -4.8

% for studies that measured cotinine in saliva, -6.2 % for those measuring cotinine in serum, blood or plasma and -9.4 % when cotinine was measured in urine. Studies in Figure 2 are ordered according to their year of publication; there does not appear to be a trend of increasing or decreasing bias over time for any of the collection mediums.

In order to examine whether study quality was associated with either reported or measured smoking prevalence, studies that included prevalence estimates were divided into two groups using the median of the quality scores to determine if the study could be considered of "higher" or "lower" quality. Studies scoring 13 points or more on the quality assessment tool were considered to have higher quality for the purposes of this comparison, while those scoring less than 13 points were considered to be of lower quality. The mean self-reported smoking prevalence was 25% for lower quality studies and 39% for the higher quality studies, while the cotinine confirmed smoking rate was 31% and 45% respectively.

Figures 3-5 plot the sensitivity values and 95% confidence intervals for the studies where this information was provided or could be calculated. Cotinine was most commonly collected in serum, plasma or whole blood (Figure 3), urine (Figure 4) or saliva (Figure 5) and therefore the plots are structured according to these mediums. Sensitivity values of 100% indicate that all "true" smokers accurately reported their smoking status; as sensitivity decreases misclassification increases.

Nineteen studies provided sensitivity values judged against cotinine measured in serum, plasma or whole blood. Sensitivity in these studies ranged from 0 in a study of men at high

risk of cardiovascular disease[13] to 98% in Suadicani and colleagues'[65] analysis of data from the 15 year follow-up of the Copenhagen Male Study, a prospective cardiovascular study of men aged 40 to 59 years (see Figure 3).

Ten studies provided sensitivity estimates for cotinine values measured in urine. Sensitivities ranged from 30% in Miwa and colleagues'[50] analysis of coronary heart disease patients and Walsh's 1996 study of pregnant women[71] to 100% in a sample of patients at a mammography clinic[14] (see Figure 4).

The highest and most consistent sensitivities were found when cotinine was measured in saliva (see Figure 5). The 15 studies in Figure 5 show the highest and least variable sensitivities. With the exception of Phillips and colleagues'[55] study primarily concerned with assessing environmental tobacco smoke, the sensitivities in saliva ranged from 78% to 97%, and in almost half (47%) sensitivities were above 90%. Studies that rated higher on the quality assessment tool had higher sensitivities in blood and saliva than lower quality studies (data not shown).

Meta-analysis was not appropriate because of the substantial heterogeneity among the cut-points used to define smokers within each of the measurement mediums, and the significant amount of information that was missing from the studies (which would have allowed us to calculate sensitivity values and associated measures of error for more of the studies). Overall effect sizes, to summarize the magnitude of the discrepancy across the mediums, could therefore not be calculated; however, average unweighted (not weighted by sample size or

variance) sensitivity values of studies with available data in our review were 86% in saliva, 76% in blood and 75% in urine.

Discussion

To the authors' knowledge, this review represents the most up-to-date, comprehensive and systematic attempt to examine the relationship between self-reported and directly measured estimates of smoking status in the international literature. Overall, the data show trends of underestimation when smoking prevalence is based on self-report and varying sensitivity levels depending on the population studied and the medium in which the biological sample is measured.

Sensitivity values were consistently higher, and the discrepancy between reported and measured prevalence was lower, when cotinine was measured in saliva instead of urine or blood. Salivary cotinine is also among the least invasive sampling techniques. Past research has found the cotinine levels in all three mediums are correlated[27,38,80] with decisions about which method to use often dependent on the study setting. A distinction also needs to be made between stimulated and unstimulated saliva sampling because the concentration of cotinine has been shown to be higher in unstimulated saliva samples[27,81]; patient acceptability is also higher for the collection of unstimulated compared to stimulated saliva[27]. Studies in this review were not examined according to whether saliva samples were stimulated because not all studies provided enough information about the sampling technique to determine if the sample was stimulated.

Results from past analyses have suggested that the sensitivity of cotinine (compared to self-report) is high, often greater than 90%[18,19]. Patrick and colleagues analysis[17], which combined data across different measurement mediums, biochemical measures and cut-points indicated that the average unweighted sensitivity for the studies they examined was 88%, leading them to conclude that self-reports are accurate in most cases. Average unweighted sensitivities of studies with available data in our review were lower with sensitivities of 86% in saliva, 76% in blood and 75% in urine. In addition, qualitative observations indicated that studies with higher scores on our quality check list had higher sensitivity values in blood and saliva and higher reported and measured smoking prevalence rates. Although these results must be interpreted cautiously in light of the high degree of missing data, they do suggest that sensitivity levels may be lower than previously believed and that better internal and external validity and data reporting can affect sensitivity and prevalence estimates

A qualitative examination of our data indicates that there is not a clear relationship between the cut-point used to determine smoking status in the biochemical measure and the sensitivity values in any of the mediums examined. The SRNT paper[19] concluded that the sensitivity (and specificity) of cotinine had little variation around a small range of cut-points (10 ng/ml to 20 ng/ml) but cut-points in the studies in our review ranged from a low of 8 ng/ml to a high of 100 ng/ml (in serum). With such a large range it could be expected that cut-points would affect the sensitivity values. However, we would argue that too much data are missing from the studies in order to infer that there is no sensitivity cut-point relationship. Of the 67 studies included in the review only 44 provided sufficient data to

examine sensitivity values and of these the quality of the study design and reporting of study characteristics varied.

A meta-analysis would have allowed us to estimate overall effect sizes for each of the collection mediums and undertake sensitivity analyses to further understand the determinants of high and low sensitivity values. However, inconsistent methods and reporting among included studies made such an analysis methodologically inappropriate.
Further research in this area would benefit from greater consistency in study designs and reporting of results to facilitate comparisons. For instance, the gold standard used for sensitivity calculations in some studies was the self-reported estimate rather than the objective measure[5,14,60,62], and in many cases not enough data were presented to allow us to re-calculate the sensitivity value using the objective estimate as the criterion measure. We recommend that authors present all the relevant data from their studies in standard 2 by 2 tables so that not only sensitivities, but also specificities and positive and negative predictive values can be calculated from their estimates. This enhances their applicability to population health research and practice as different users will then be able to extract the most relevant data to meet their needs.

Cut-points used to identify a "regular smoker" also need to be standardized. The SRNT has recommended that cut-points of 15 ng/ml (85 nmol/L) in plasma or saliva and 50 ng/ml (283 nmol/L) in urine be used in validation studies[19]. Sensitivities for alternative cut-points may be required in special populations such as in pregnant women or members of different racial groups[28,82] and these could be presented in addition to, rather than instead of, the standard

cut-points. Moreover, although some studies continue to adjust urinary cotinine levels for creatinine concentration, there is indication that this adjustment is not necessary[19,83]. When authors choose to use adjusted estimates, both the adjusted and unadjusted numbers should be reported to enhance study comparability.

These recommendations are not new and calls for better reporting in this field have been made by other researchers[17] culminating in the SRNT recommendations that were released in 2002[19]. Yet, methods and reporting have not been standardized in most validation studies. For example, of the 22 studies included in our review that were published after the 2002 SRNT recommendations, 14 provided enough information to calculate sensitivity values and only 5 reported results using the recommended cut-points. More emphasis needs to be placed on ensuring that researchers follow recommended procedures for measuring and reporting data so that we can gain a better understanding of the actual sensitivity and specificity values and their determinants and to examine the appropriateness of existing recommendations.

Although we had hypothesized that the reporting bias may have increased in recent years, there was no trend of increasing or decreasing bias in the accuracy of reported estimates over time in the studies examined in this review. This may be because the differences in study populations, designs and analyses made it difficult to directly compare studies over time or because there has indeed been no systematic change in the bias. Historical analyses of past databases that include both measured and reported smoking data may help to clarify if the bias is shifting.

This review has several limitations; biochemical markers, although seen as the criterion measure for assessment, are not without flaws. For instance, cotinine levels can be elevated in users of chewing tobacco or nicotine containing products such as gum or the transdermal patch[64,67]. While this review attempted to focus only on cigarette smoking, there is a potential for contamination. In addition, although cotinine has been shown to distinguish active smokers from non-smokers and even passive smokers[70,84,85], there is a grey zone where it becomes difficult to differentiate between occasional smokers and those exposed to environmental tobacco smoke[86]. This problem is further compounded by individual variability in metabolism, inhalation patterns and brand of cigarette smoked, each of which can impact the amount of cotinine that is in the body at the time of measurement[70,84]. More research is required to determine how to best validate occasional smoking and environmental tobacco exposure.

Overall, self-reports underestimated true smoking prevalence in most studies in this review but the sensitivities and cut-points used to define them were highly variable and the sensitivities were influenced by the medium of measurement. The costs and benefits of direct measurement need to be considered in any study in order to determine if the added resources required for personnel training and lab analysis justify the increased precision in results. Biochemical verification provides an objective alternative to reported estimates and has the further advantage of quantifying environmental tobacco exposure, which is difficult to accurately self-report[87,88], but may not always be feasible for measurement in large scale population studies. In cases where biochemical verification is not possible, correction factors

could be estimated to adjust the reported estimates to more closely approximate their measured values.

At the population level, underestimates in smoking prevalence have important implications as these data are used to monitor smoking trends, justify spending for research and tobacco control programming and to estimate tobacco-related risks of disease. Regular monitoring of this literature is important to determine if the bias is changing, to further understand the factors associated with the bias, and to determine if the quality of reporting results is improving. The discussion remains relevant because although some population studies (e.g., NHANES) are now collecting biochemical measures to verify smoking status much of our surveillance data continues to be based on self-reported information (because of costs and other constraints). Understanding the degree of bias is important in determining the degree to which self-reported data are providing accurate estimates of smoking prevalence rates. New research is emerging that is indicating that well known relationships, such as between obesity and disease are exaggerated when the risk factor data (obesity) are based on reported versus measured estimates[89,90]. It is possible that similar differences may also emerge in the relationship between smoking and disease if there is a substantial underreporting of smoking rates. Conscientious efforts to adopt robust surveillance procedures and further understand bias in self-report methodologies will help to reduce potential interpretation errors.

Acknowledgements

The authors would like to thank Ian McDowell for insightful comments on earlier drafts of this manuscript and Debra Brody for her content and technical assistance. We would also like to acknowledge the insightful comments from the journal reviewers whose feedback helped us to improve this manuscript. The authors have no competing interests to declare.

References

1. Lopez AD. The evolution of the Global Burden of Disease framework for disease, injury and risk factor quantification: Developing the evidence base for national, regional and global public health action. *Global Health* 2005;1(5).
2. Patra J, Popova S, Rehm J, Bondy S, Flint R, Giesbrecht N. *Economic cost of chronic disease in Canada 1995-2003*. Toronto, ON: Ontario Chronic Disease Prevention Alliance and the Ontario Public Health Association; 2007.
3. Centers for Disease Control. Annual smoking-attributable mortality, years of potential life lost, and economic costs United States, 1995-1999. *MMWR* 2002;51:300-3.
4. Coultas DB, Howard CA, Peake GT, Skipper BJ, Samet JM. Discrepancies between self-reported and validated cigarette smoking in a community survey of New Mexico hispanics. *Am Rev Respir Dis* 1988;137:810-14.
5. Lewis SJ, Cherry NM, Niven RMCL, Barber PV, Wilde K, Povey AC. Cotinine levels and self-reported smoking status in patients attending a bronchoscopy clinic. *Biomarkers* 2003;8:218-28.
6. Tyrpien K, Bodzek P, Manka G. Application of planar chromatography to the determination of cotinine in urine of active and passive smoking pregnant women. *Biomed Chromatogr* 2001;15:50-5.
7. Fendrich M, Mackesy-Amiti ME, Johnson TP, Hubbell A, Wislar JS. Tobacco-reporting validity in an epidemiological drug-use survey. *Addict Behav* 2005; 30:175-81.
8. Pérez-Stable EJ, Marín BV, Marín G, Brody DJ, Benowitz NL. Apparent underreporting of cigarette consumption among Mexican American smokers. *Am J Public Health* 1990;80:1057-61.

9. Ford RP, Tappin DM, Schluter PJ, Wild CJ. Smoking during pregnancy: How reliable are self reports in New Zealand? *J Epidemiol Community Health* 1997;51:246-51.
10. From Attebring M, Herlitz J, Berndt AK, Karlsson T, Hjalmarson A. Are patients truthful about their smoking habits? A validation of self-report about smoking cessation with biochemical markers of smoking activity amongst patients with ischaemic heart disease. *J Intern Med* 2001;249:145-51.
11. Laatikainen T, Vartiainen E, Puska P. Comparing smoking and smoking cessation process in the Republic of Karelia, Russia and North Karelia, Finland. *J Epidemiol Community Health* 1999;53:528-34.
12. Parker DR, Lasater TM, Windsor R, Wilkins J, Upegui DI, Heimdal J. The accuracy of self-reported smoking status assessed by cotinine test strips. *Nicotine Tob Res* 2002;4:305-9.
13. Slattery ML, Hunt SC, French TK, Ford MH, Williams RR. Validity of cigarette smoking habits in three epidemiologic studies in Utah. *Prev Med* 1989;18:11-19.
14. Riboli E, Haley NJ, de Waard F, Saracci R. Validity of urinary biomarkers of exposure to tobacco smoke following prolonged storage. *Int J Epidemiol* 1995;24:354-8.
15. Etter JF, Perneger TV. Measurement of self reported active exposure to cigarette smoke. *J Epidemiol Community Health* 2001;55:674-80.
16. Jarvis MJ, Tunstall-Pedoe H, Feyerabend C, Vesey C, Saloojee Y. Comparison of tests used to distinguish smokers from nonsmokers. *Am J Public Health* 1987;77:1435-38.
17. Patrick DL, Cheadle A, Thompson DC, Diehr P, Koepsell T, Kinne S. The validity of self-reported smoking: A review and meta-analysis. *Am J Public Health* 1994;84:1086-93.

18. Velicer WF, Prochaska JO, Rossi JS, Snow MG. Assessing outcome in smoking cessation studies. *Psychol Bull* 1992;111:23-41.

19. Society for Research on Nicotine and Tobacco Subcommittee on Biochemical Verification. Biochemical verification of tobacco and cessation. *Nicotine Tob Res* 2002;4:149-59.

20. Glasgow RE, Mullooly JP, Vogt TM, Stevens VJ, Lichtenstein E, Hollis JF, et al. Biochemical validation of smoking status: Pros, cons, and data from four low-intensity intervention trials. *Addict Behav* 1993;18:511-27.

21. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52:377-84.

22. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al. Evaluating non randomized intervention studies. *Health Technol Assess* 2003;7:1–173.

23. Saunders DL, Soomro GM, Buckingham J, Jamtvedt G, Raina P. Assessing the methodological quality of nonrandomized intervention studies. *West J Nurs Res* 2003;25:223-7.

24. Abrams DB, Follick MJ, Biener L, Carey KB, Hitti J. Saliva cotinine as a measure of smoking status in field settings. *Am J Public Health* 1987;77:846-8.

25. Arnold CL, Davis TC, Berkel HJ, Jackson RH, Nandy I, London S. Smoking status, reading level, and knowledge of tobacco effects among low-income pregnant women. *Prev Med* 2001;32:313-20.

26. Assaf AR, Parker D, Lapane KL, McKenney JL, Carleton RA. Are there gender differences in self-reported smoking practices? Correlation with thiocyanate and cotinine

levels in smokers and nonsmokers from the Pawtucket Heart Health Program. *J Womens Health* 2002;11:899-906.

27. Binnie V, McHugh S, Macpherson L, Borland B, Moir K, Malik K. The validation of self-reported smoking status by analysing cotinine levels in stimulated and unstimulated saliva, serum and urine. *Oral Dis* 2004;10:287-93.
28. Boyd NR, Windsor RA, Perkins LL, Lowe JB. Quality of measurement of smoking status by self-report and saliva cotinine among pregnant women. *Matern Child Health J* 1998;2:77-83.
29. Buduneli N, Kardesler L, Isik H, Willis CS, Hawkins SI, Kinane DF, et al. Effects of smoking and gingival inflammation on salivary antioxidant capacity. *J Clin Periodontol* 2006, 33, 159-164.
30. Derauf C, Katz AR, Easa D. Agreement between maternal self-reported ethanol intake and tobacco use during pregnancy and meconium assays for fatty acid ethyl esters and cotinine. *Am J Epidemiol* 2003;158:705-9.
31. Etter JF, Vu Duc T, Perneger TV. Saliva cotinine levels in smokers and nonsmokers. *Am J Epidemiol* 2000;151:251-8.
32. Fossum B, Arborelius E, Bremberg S. Evaluation of a counseling method for the prevention of child exposure to tobacco smoke: An example of client-centered communication. *Prev Med* 2004;38:295-301.
33. Gonzales YM, De Nardin A, Grossi SG, Machtei EE, Genco RJ, De Nardin, E. Serum cotinine levels, smoking, and periodontal attachment loss. *J Dent Res* 1996;75:796-802.

34. Greaves R, Trotter L, Brennecke S, Janus E. A simple high-pressure liquid chromatography cotinine assay: Validation of smoking status in pregnant women. *Ann Clin Biochem* 2001;38:333-8.

35. Haddow JE, Palomaki GE, Knight GJ. Use of serum cotinine to assess the accuracy of self-reported non-smoking [Letter to the editor]. *Br Med J* 1986;293:1306.

36. Hald J, Overgaard J, Grau C. Evaluation of objective measures of smoking status: A prospective clinical study in a group of head and neck cancer patients treated with radiotherapy. *Acta Oncol* 2003;42:154-9.

37. Hahn EJ, Rayens MK, Chirila C, Riker CA, Paul TP, Warnick TA. Effectiveness of a quit and win contest with a low-income population. *Prev Med* 2004;39:543-550.

38. Haley NJ, Axelrad CM, Tilton KA. Validation of self-reported smoking behavior: Biochemical analyses of cotinine and thiocyanate. *Am J Public Health* 1983;73:1204-07.

39. Hennrikus DJ, Lando HA, McCarty MC, Klevan D, Holtan N, Huebsch JA, et al. The TEAM project: The effectiveness of smoking cessation intervention with hospital patients. *Prev Med* 2005;40:249-58.

40. Johnston DW, Johnston M, Pollard B, Kinmouth A, Mant D. Motivation is not enough: Prediction of risk behavior following diagnosis of coronary heart disease from the theory of planned behavior. *Health Psychol* 2004;23:533-8.

41. Kendrick JS, Zahniser SC, Miller N, Salas N, Stine J, Gargiullo PM, et al. Integrating smoking cessation into routine public prenatal care: The smoking cessation in pregnancy project. *Am J Public Health* 1995;85:217-222.

42. Klebanoff MA, Levine RJ, Clemens JD, DerSimonian R, Wilkins DG. Serum cotinine concentration and self-reported smoking during pregnancy. *Am J Epidemiol* 1998;148:259-62.

43. Klebanoff MA, Levine RJ, Morris CD, Hauth JC, Sibai BM, Curet LB, et al. Accuracy of self-reported cigarette smoking among pregnant women in the 1990s. *US Government Paediatric and Perinatal Epidemiology* 2001;15:140-43.

44. Lando H, Hennrikus D, McCarty M, Vessey J. Predictors of quitting in hospitalized smokers. *Nicotine Tob Res* 2003;5:215-22.

45. Law KL, Stroud LR, LaGasse LL, Niaura R, Liu J, Lester BM. Smoking during pregnancy and newborn neurobehavior. *Pediatrics* 2003;111:1318-23.

46. Lindqvist R, Lendahls L, Tollbom Ö, Åberg H, Håkansson A. Smoking during pregnancy: Comparison of self-reports and cotinine levels in 496 women. *Acta Obstet Gynecol Scand* 2002;81:240-44.

47. Ma Y, Goins KV, Pbert L, Ockene JK. Predictors of smoking cessation in pregnancy and maintenance postpartum in low-income women. *Matern Child Health J* 2005;9:393-402.

48. Markovic N, Ness RB, Cefilli D, Grisso JA, Stahmer S, Shaw LM. Substance use measures among women in early pregnancy. *Am J Obstet Gynecol* 2000;183:627-32.

49. McDonald SP, Maguire GP, Hoy WE. Validation of self-reported cigarette smoking in a remote Australian Aboriginal community. *Aust N Z J Public Health* 2003;27:57-60.

50. Miwa K, Fujita M, Miyagi Y. Beneficial effects of smoking cessation on the short-term prognosis for variant angina - validation of the smoking status by urinary cotinine measurements. *Int J Cardiol* 1994;44:151-6.

51. Monninkhof E, van der Valk P, van der Palen J, Mulder H, Pieterse M, van Herwaarden C, et al. The effect of a minimal contact smoking cessation programme in out-patients with chronic obstructive pulmonary disease: A pre-post-test study. *Patient Educ Couns* 2004;52:231-6.

52. Murray RP, Connett JE, Lauger GG, Voelker HT. Error in smoking measures: effects of intervention on relations of cotinine and carbon monoxide to self-reported smoking. The Lung Health Study Research Group. *Am J Public Health* 1993;83:1251-57.

53. Nagle AL, Hensley MJ, Schofield MJ, Koschel AJ. A randomised controlled trial to evaluate the efficacy of a nurse-provided intervention for hospitalised smokers. *Aust N Z J Public Health* 2005;29:285-91.

54. Owen L, McNeill A. Saliva cotinine as indicator of cigarette smoking in pregnant women. *Addiction* 2001;96:1001-6.

55. Phillips K, Bentley MC, Howard DA, Alvan G, Huici A. Assessment of air quality in Barcelona by personal monitoring of nonsmokers for respirable suspended particles and environmental tobacco smoke. *Environ Int* 1997;23:173-96.

56. Pichini S, Basagaña X, Pacifici R, Garcia O, Puig C, Vall O, et al. Cord serum cotinine as a biomarker of fetal exposure to cigarette smoke at the end of pregnancy. *Environ Health Perspect* 2000;108:1079-83.

57. Pickett KE, Rathouz PJ, Kasza K, Wakschlag LS, Wright R. Self-reported smoking, cotinine levels, and patterns of smoking in pregnancy. *Paediatr Perinat Epidemiol* 2005;19:368-76.

58. Pierce JP, Dwyer T, Di Giusto E, Carpenter T, Hannam CD, Amin A, et al. Cotinine validation of self-reported smoking in commercially run community surveys. *J Chronic Dis* 1987;40:689-95.

59. Rigotti NA, McKool KM, Shiffman S. Predictors of smoking cessation after coronary artery bypass graft surgery. *Ann Intern Med* 1994;120:287-93.

60. Sandhu S, Humphris G, Whitley S, Cardozo A, Sandhu A. Smoking habits in patients who have been treated for an oral cancer: Validation of self-report using saliva cotinine. *Oral Oncol* 2004;40:576-78.

61. Scherer G, Frank S, Riedel K, Meger-Kossien I, Renner T. Biomonitoring of exposure to polycyclic aromatic hydrocarbons of nonoccupationally exposed persons. *Cancer Epidemiol Biomarkers Prev* 2000;9:373-80.

62. Seccareccia F, Zuccaro P, Pacifici R, Meli P, Pannozzo F, Freeman KM, et al. Serum cotinine as a marker of environmental tobacco smoke exposure in epidemiological studies: The experience of the MATISS project. *Eur J Epidemiol* 2003;18:487-92.

63. Shaffer HJ, Eber GB, Hall MN, Vander Bilt J. Smoking behavior among casino employees. Self-report, validation using plasma cotinine. *Addict Behav* 2005; 25: 693-704.

64. Stookey GK, Katz BP, Olson BL, Drook CA, Cohen SJ. Evaluation of biochemical validation measures in determination of smoking status. *J Dent Res* 1987;66:1597-1601.

65. Suadicani P, Hein HO, Gyntelberg F. Serum validated tobacco use and social inequalities in risk of ischemic heart disease. *Int J Epidemiol* 1994;23:293-300.

66. Van Vunakis H, Tashkin DP, Rigas B, Simmons M, Gjika HB, Clark VA. Relative sensitivity and specificity of salivary and serum cotinine in identifying tobacco-smoking status of self-reported nonsmokers and smokers of tobacco and/or marijuana. *Arch Environ Health* 1989;44:53-8.

67. Vartiainen E, Seppala T, Lillsunde P, Puska P. Validation of self-reported smoking by serum cotinine measurement in a community-based study. *J Epidemiol Community Health* 2002;56:167-70.

68. Vine MF, Hulka BS, Margolin BH, Truong YK, Hu P, Schramm MM, et al. Cotinine concentrations in semen, urine, and blood of smokers and nonsmokers. *Am J Public Health* 1993;83:1335-38.

69. Wagenknecht LE, Burke GL, Perkins LL, Haley NJ, Friedman GD. Misclassification of smoking status in the CARDIA study: A comparison of self-report with serum cotinine levels. *Am J Public Health* 1992;82:33-36.

70. Wall MA, Johnson J, Jacob P, Benowitz NL. Cotinine in the serum, saliva, and urine of nonsmokers, passive smokers, and active smokers. *Am J Public Health* 1988;78:699-701.

71. Walsh RA, Redman S, Adamson L. The accuracy of self-report of smoking in pregnant women. *Addict Behav* 1996;21:675-9.

72. Wang X, Tager IB, Van Vunakis H, Speizer FE, Hanrahan JP. Maternal smoking during pregnancy, urine cotinine concentrations, and birth outcomes. *Int J Epidemiol* 1997;26:978-88.

73. Webb DA, Boyd NR, Messina D, Windsor RA. The discrepancy between self-reported smoking status and urine cotinine levels among women enrolled in prenatal care at four publicly funded clinical sites. *J Public Health Manag Pract* 2003;9:322-5.

74. Whittet HB, Lund VJ, Brockbank M, Feyerabend C. Serum cotinine as an objective marker for smoking habit in head and neck malignancy. *J Laryngol Otol* 1991;105:1036-9.

75. Wilson DK, Wallston KA, King JE, Smith MS, Heim C. Validation of smoking abstinence in newly diagnosed cardiovascular patients. *Addict Behav* 1993; 18:421-9.

76. Windsor RA, Woodby LL, Miller TM, Hardin JM, Crawford MA, DiClemente CC. Effectiveness of agency for health care policy and research clinical practice guideline and patient education methods for pregnant smokers in Medicaid maternity care. *Am J Obstet Gynecol* 2000;182:68-75.

77. Woodward A, Fowles J, Dickson S, Fernando D, Berezowski R, Reid P. Increase in saliva cotinine after three hours' exposure to second-hand smoke in bars. *Aust N Z J Public Health* 2005;29:272-5.

78. Yamamoto Y, Nishida N, Tanaka M, Hayashi N, Matsuse R, Nakayama K, et al. Association between passive and active smoking evaluated by salivary cotinine and periodontitis. *J Clin Periodontol* 2005;32:1041-6.

79. Young DS, Huth EJ. *SI units for clinical measurement,* Philadelphia, Pa: American College of Physicians; 1998.

80. Bernert JT, McGuffey JE, Morrison MA, Pirkle JL. Comparison of serum and salivary cotinine measurements by a sensitive high-performance liquid chromatography-tandem mass spectrometry method as an indicator of exposure to tobacco smoke among smokers and nonsmokers. *J Anal Toxicol* 2000;24:333-9.

81. Schneider NG, Peyton J, Nilsson F, Leischow SJ, Benowitz NL, Olmstead RE. Saliva cotinine levels as a function of collection method. *Addiction* 1997;92:347-51.

82. Caraballo RS, Giovino GA, Pechacek TF, Mowery PD, Richter PA, Strauss WJ, et al. Racial and ethnic differences in serum cotinine levels of cigarette smokers. *JAMA* 1998;280:135-9.

83. Jatlow P, McKee S, O'Malley SS. Correction of urine cotinine concentrations for creatinine excretion: is it useful? *Clin Chem* 2003;49:1932-4.

84. Hill P, Haley NJ, Wynder EL. Cigarette smoking: Carboxyhemoglobin, plasma nicotine, cotinine and thiocyanate vs self-reported smoking data and cardiovascular disease. *J Chronic Dis* 1983;36:439-49.

85. Sepkovic DW, Haley NJ. Biomedical applications of cotinine quantification in smoking-related research. *Am J Public Health* 1985;75:663-5.

86. Etzel RA. A review of the use of saliva cotinine as a marker of tobacco smoke exposure. *Prev Med* 1990;19:190-7.

87. Kawachi I, Colditz GA. Invited commentary: Confounding, measurement error, and publication bias in studies of passive smoking. *Am J Epidemiol* 1996;144:909-15.

88. George L, Granath F, Johansson ALV, Cnattingius S. Self-reported nicotine exposure and plasma levels of cotinine in early and late pregnancy. *Acta Obstet Gynecol Scand* 2006;11:1331-7.

89. Shields M, Connor Gorber S, Tremblay MS. Effects of measurement on obesity and morbidity. *Health Rep* 2008;19(2):77-84.

90. Chiolero A, Peytremann-Bridevaux I, Paccaud F. Associations between obesity and health conditions may be overestimated if self-reported body mass index is used. *Obes Rev 2007;8:*373-374.

Tables and Figures

Table 1: MEDLINE search strategy

1	Adult/
2	Female/
3	Male/
4	Middle Aged/
5	Aged/
6	Human$/
7	1 or 2 or 3 or 4 or 5 or 6
8	Direct measure$.mp
9	Physical measure$.mp
10	Objective measure$.mp
11	Blood/
12	Hair/
13	Urine/
14	Saliva
15	exp Biological Markers/
16	8 or 9 or 10 or 11 or 12 or 13 or 14 or 15
17	Self concept/
18	Self report.mp
19	Subjective measure$.mp
20	Subjective assess$.mp
21	Self evaluat$.mp
22	Self assess$.mp
23	Truth Disclosure/
24	17 or 18 or 19 or 20 or 21 or 22 or 23
25	Smoking
26	Cotinine
27	Nicotine
28	Tobacco Smoke Pollution
29	Smoke$/
30	Environmental Exposure
31	Cigar$.mp
32	25 or 26 or 27 or 28 or 29 or 30 or 31
33	7 and 16 and 24 and 32

Table 2: Study and participant characteristics

First Author Year	Sample Size (Analyzed)	Age Range or Mean(SD)	Population	Country	Self-Report (Q/I)	Method	Medium	Cut-off	SI Unit (nmol/L) cut-off	Setting	Time Lag
Abrams 1987	91	39	Hospital employees	USA	Q	RIA	saliva	10 ng/mL	56.75	Workplace	SS
Arnold 2001	599	12-45	Pregnant women	USA	I	Cotinine Enzyme Immunoassay	urine	100 ng/mL	567.5	Clinic	SS
Assaf 2002	798	18-64	General population	USA	I	GC-NPD	serum	58 ng/mL (male); 30 ng/mL (female)	329.2 (male); 170.2 (female)	Home	SS
Binnie 2004	80	16-75	General population	UK	Q	M-ELISA	1) saliva; 2) serum; 3) urine#	1) 15 ng/mL; 2) 15 ng/mL	1) 85.1; 2) 85.1 ; 3) 50 ng/mmol	Clinic	SS
Boyd 1998*	548	25	Pregnant women	USA	I	RIA	saliva	30 ng/mL	170.2	Clinic	SS
Buduneli 2006	40	25-58	Gingivitis patients; healthy volunteers	Turkey/ USA	I	GC-NPD	saliva	15 ng/mL	85.1	Dental clinic	-
Coultas 1988	1317	13-65+	General Population (Hispanics)	USA	I	RIA	saliva	20 ng/mL	113.5	-	SS
Derauf 2003	426	-	New mothers	USA	Medical records	ELISA	infant meconium	25 ng/g		Hospital	SS
Etter 2000	319	19-61	University students/ staff	Switzer-land	Q	GC	saliva	7 ng/mL; 13 ng/mL	39.7; 73.8	Home (mail)	Variable
Fendrich 2005	536	18-40	General population	USA	Q	GC	saliva	20 ng/mL	113.5	Home	SS

Ford 1997	1399	30(5)	Pregnant women	New Zealand	Q	ELISA	serum	15 ng/mL (light) and 100 ng/mL (heavy)	85.1 and 567.5	Hospital	-
Fossum 2004*	30	-	Mothers	Sweden	I	Enzyme immunoassay	saliva	-	-	Child Health Center	SS
From Attebring 2001	206	62	Cardiology outpatients	Sweden	Q	GC	plasma	14 ng/mL	79.45	Hospital	SS
Glasgow 1993*	1) 2707; 2) 1119; 3) 1097; 4) 576	1) 40; 2) 44; 3) 41; 4) 33	1) outpatients; 2) hospital patients; 3) state employees; 4) dental patients	USA	Q	GC-MS	saliva	25 ng/mL	141.9	Medical offices, hospital, worksite, dental office	2wks-2mos
Gonzalez 1996	79	25-64	Periodontitis patients	USA	Q	ELISA	serum	-	-	Clinic	SS
Greaves 2001	173	-	Pregnant women	Australia	-	LC	urine#	28 µg/mmol		-	-
Haddow 1986	296	-	General population	USA	Q	-	serum	10 ug/L	56.8	-	-
Hald 2003	104	46-80	Cancer patients	Denmark	I	ELISA	serum	10 ng/mL; 50 ng/mL	56.8; 283.8	-	SS
Hahn 2004	51	18+	Low-income tobacco users	USA	I	Immunoassay (Accutest NicoMeter)	urine	Accutest scale (1+)	-	-	-
Haley 1983	30	-	Smokers/ non-smokers	USA	Q	RIA	1) saliva; 2) plasma	-	-	-	SS
Hennrikus 2005*	241	18-75	Inpatient smokers	USA	I	-	saliva	15 ng/mL	85.1	Hospital	9-12 days

Jarvis 1987	211	55	Hospital outpatients	England	Q	GC	1) saliva; 2) plasma; 3) urine	1) 14.2; 2)13.7; 3) 49.7 ng/ml	1) 80.6; 2) 92.5; 3) 282	Hospital	SS
Johnston 2004	130	63(10)	Heart disease patients	UK	Q	-	serum	13.7 ng/mL	77.7	Hospital or clinic	SS
Kendrick 1995*	2310	-	Pregnant smokers	USA	Q	ELISA	urine#	85 ng/mL	482.4	Clinic	SS
Klebanoff 1998	448	-	Pregnant women	USA	NS	M-ELISA	serum	10 ng/mL	56.8	Clinic	-
Klebanoff 2001	104	-	Pregnant women	USA	Q	M-ELISA	1) urine; 2) serum	1) 500 ng/mL; 2)10 ng/mL	1) 56.8; 2) 283.8	Clinic	SS
Laatikainen 1999	825 (Pitkaranta) 1431 (N. Karelia)	25-64	General population	Finland	Q	GC-MS	serum	10 ng/mL	56.8	Home/ Survey Centre	small lag
Lando 2003*	1477	18-75	Hospitalized smokers	USA	I	GC	saliva	15 ng/mL	85.1	Home (mail)	mailed samples
Law 2003	55	18-35	New mothers	USA	Q/I	GC-MS	saliva	10 ng/mL	56.8	Hospital	1-3 days
Lewis 2003	166	22-89	Respiratory patients	UK	I	GC-MS	whole blood	20 ng/ml	113.5	Clinic	SS
Lindqvist 2002	496	median 29	Pregnant women	Sweden	I	GC-NPD	serum	17.5 ng/mL	99.3	Clinic	SS
Ma 2005*	216	-	Pregnant women (low-income)	USA	I	-	saliva	20 ng/mL	113.5	Health center	SS
Markovic 2000	845	23(5)	Pregnant women (innercity)	USA	I	Mi-ELISA	urine	500 ng/mL	2837.5	Hospital	-
McDonald 2003	184	Median 38	Remote Aboriginal community	Australia	Q	Enzymatic immunoassay	urine#		539; 610 CCR	-	SS

Miwa 1994	37	56	Angina patients (smokers)	Japan	Q	RIA	urine	50 ng/mL	283.8	Hospital	SS
Monninkhof 2004	188	63(7)	COPD patients	Nether-lands	I	GC	saliva	20 ng/mL	113.5	Home	SS
Murray 1993*	5887	35-60	Smokers (early lung disease)	CAN/ USA	Q	Assay	saliva	20 ng/mL	113.5	Clinic	SS
Nagle 2005*	1422	18-80	Hospital patients	UK/USA	I	-	saliva	50 nmol/L	50	Hospital	SS
Owen 2001	1001	28	Pregnant women	England	I	GC	saliva	14 ng/ml	79.4	Home	SS
Parker 2002	256	-	Hospital staff/visitors	USA	Q	GC/test strip	urine	100 ng/ml; 250 ng/ml	567.5; 1418.8	Hospital	SS
Phillips 1997	154	20-64	Housewives/ office workers	Spain	Q/I	RIA	saliva	25 ng/mL	141.9	Study Centre (hotel)	SS
Pichini 2000	429	Median 29	Pregnant women	Spain	Q	RIA	cord serum, maternal and newborn urine#	50 ng/mL	283.8	Hospital	SS
Pickett 2005	4838	26(5)	Pregnant women	UK	I	RIA	urine#	200 ng/mL	1135	Clinic	SS
Pierce 1987	975	14-65+	General population	Australia	I	GC	saliva	250 nmol/L	250	Home	SS
Riboli 1995	58	45-60	Mammo-graphy clinic	The Nether-lands	Q	RIA	urine#	200 ng		Prevention Centre	SS
Rigotti 1994*	87	58(8)	Coronary bypass patients	USA	Q/I	Assay	saliva	20 ng/mL	113.5	Hospital (saliva by mail)	-
Sandhu 2004	91	61(11)	Oral cancer patients	UK	Q	LC	saliva	14 ug/L	79.4	Clinic	SS

Scherer 2000	69	18-70	Smokers, non-smokers, ETS exposed	Germany	-	RIA	1) plasma 2) urine#	1) 10 ng/ml; 2) 100 ug/ 24h urine	1) 56.8	Laboratory	SS
Seccareccia 2003	3450	20-79	General population	Italy	Q	RIA	Serum	15 ng/mL	85.1	-	-
Shaffer 2000	3841	17-75	Casino workers	USA	Q	-	Plasma	25 ng/mL	141.9	Workplace	SS
Slattery 1989	1) 112; 2) 512; 3) 157	17-60+	1) male twins; 2) cervical cancer patients; 3) families with high CVD risk	USA	I	-	Serum	15 ng/mL	85.1	House	-
Stookey 1987	1) 213; 2) 20; 3) 102	-	Smokers, non-smokers, quitters	USA	-	GC	Saliva	10 ng/mL	56.8	-	-
Suadicani 1994	3216	53-75	General population (men)	Denmark	Q	RIA	Serum	100 ng/mL	567.5	Hospital	SS
Tyrpien 2001	58	-	Pregnant women	Poland	I	Planar chromato-graphy	urine#	-	-	-	-
Van Vunakis 1989	327	-	Marijuana study participants	USA	I	LC, RIA	1) saliva; 2) serum	1) 25 ng/mL; 2) 8 ng/mL	1) 141.9; 2) 45.4	Laboratory	SS
Vartiainen 2002	5844	25-64	General population	Finland	Q	GC-MS	serum	10 ng/mL	56.8	-	SS
Vine 1993	88	18-35	Males	USA	Q	RIA	urine, blood, semen	none	none	-	SS

Wagenknecht 1992	4984	18-30	General population	USA	I	RIA	serum	14 ng/mL	79.4	American Health Foundation	SS
Wall 1988	98	24-66	Hospital employees,patient spouses	USA	Q	LC	saliva, serum, urine	none	none	Hospital	1 day
Walsh 1996	161	-	Pregnant women	Australia	Q	LC	urine	50 ng/mL; 88 ng/mL	282; 500	Clinic	SS
Wang 1997	740	18+	Pregnant women	USA	Q	RIA	urine#	-	-	Clinic	SS
Webb 2003	74	19-35	Pregnant women (low income)	USA	I	-	urine	80 ng/mL	454	Clinic	SS
Whittet 1991	31	61	Male cancer patients	England	I	LC	serum	100 ng/mL	567.5	Clinic	SS
Wilson 1993*	79	26-76	CVD patients	USA	Q	Assay	saliva	20 ng/mL	113.5	Home	SS
Windsor 2000*	1540	23	Pregnant women	USA	Q	-	saliva	31 ng/mL	175.9	Hospital	SS
Woodward 2005	9740	40-59	General population	Scotland	Q	LC	serum	17.5 ng/mL	99.3	-	SS
Yamamoto 2005	256	18-62	Factory workers	Japan	Q	ELISA	saliva	8 ng/ml	45.4	-	SS

Notes: Q questionnaire, I interview, * experimental or quasi-experimental design, - not stated, # creatinine adjustments were used or reported, COPD chronic obstructive pulmonary disease, CVD cardiovascular disease patients, DM direct measurement, SS same session, GC gas chromatography, GC-MS gas chromatography mass spectrometry, GC-NPD Gas chromatography nitrogen phosphorus detector, LC liquid chromatography, RIA radioimmunoassay, ELISA Enzyme-linked Immunosorbent assay, M-ELISA, Microplate enzyme-linked immunosorbent assay, CCR cotinine creatinine ratio

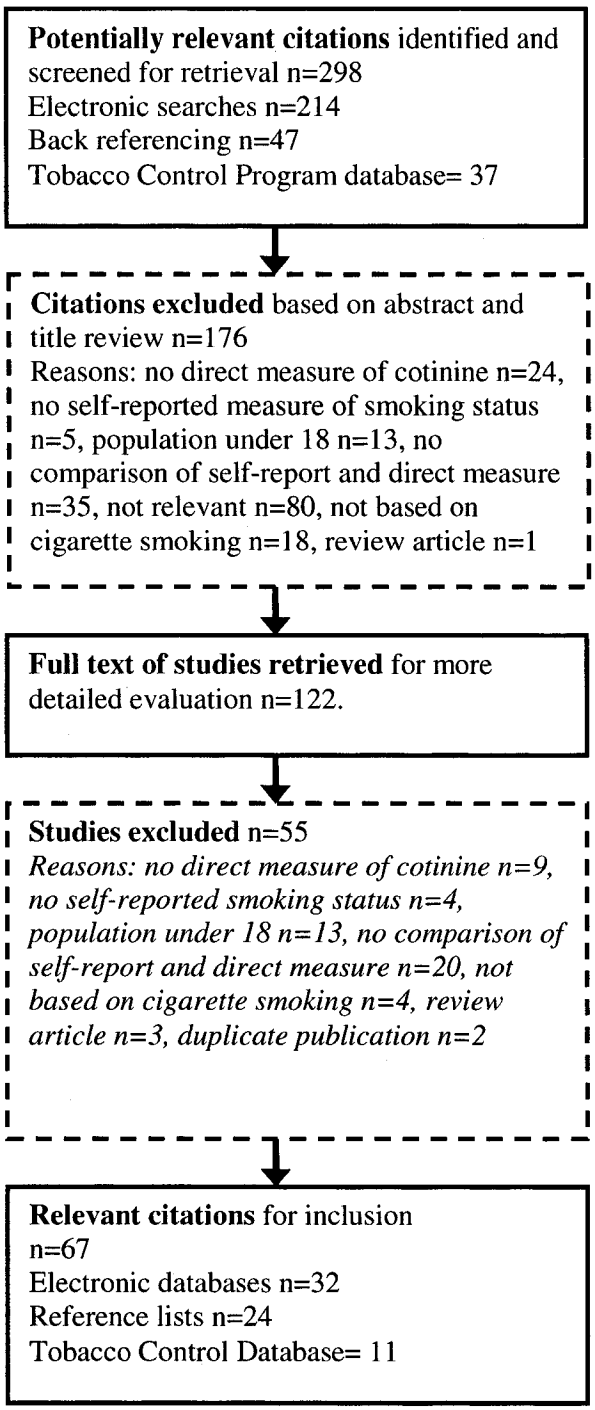

Figure 1: Results of the literature search.

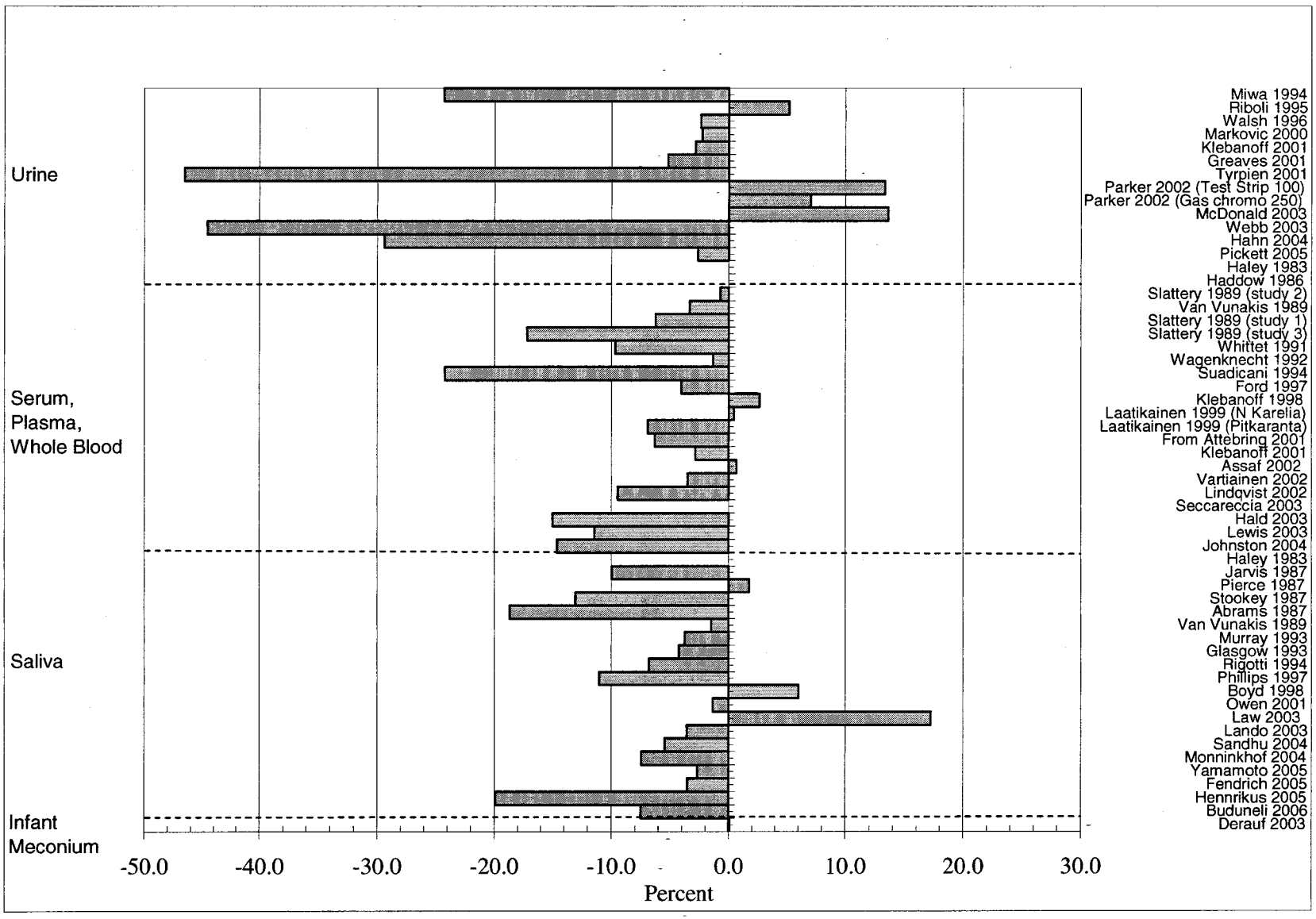

Figure 2: Differences in smoking prevalence (self-reported estimate minus measured estimate) according to medium of measurement. Studies cover the period from 1983 to 2006.

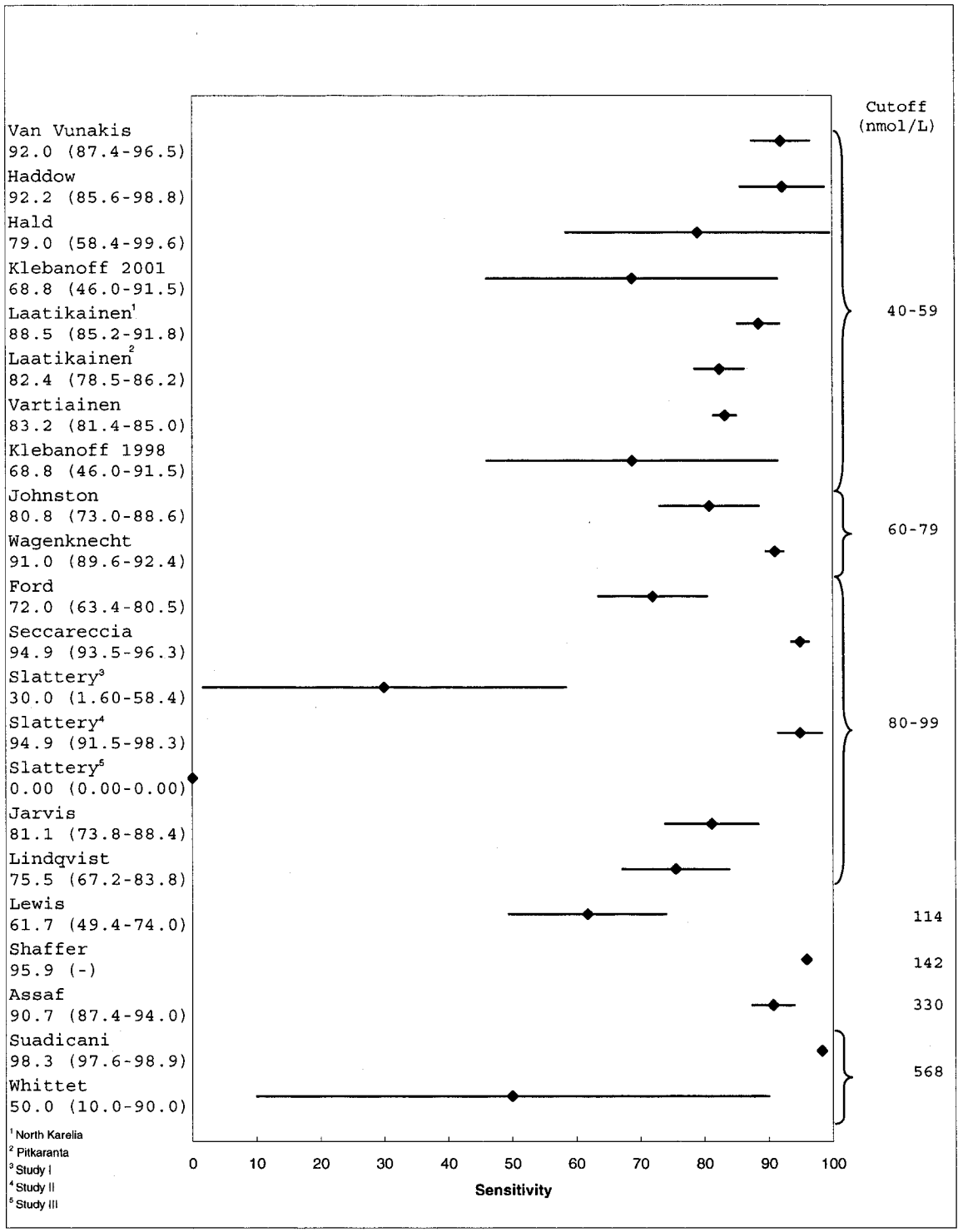

Figure 3: Sensitivity of self-reported smoking versus blood (serum, plasma, whole blood) cotinine among male and female smokers. Study author and sensitivity level (95% CI) provided. Studies presented in order of cotinine cut-off value (smallest at top).

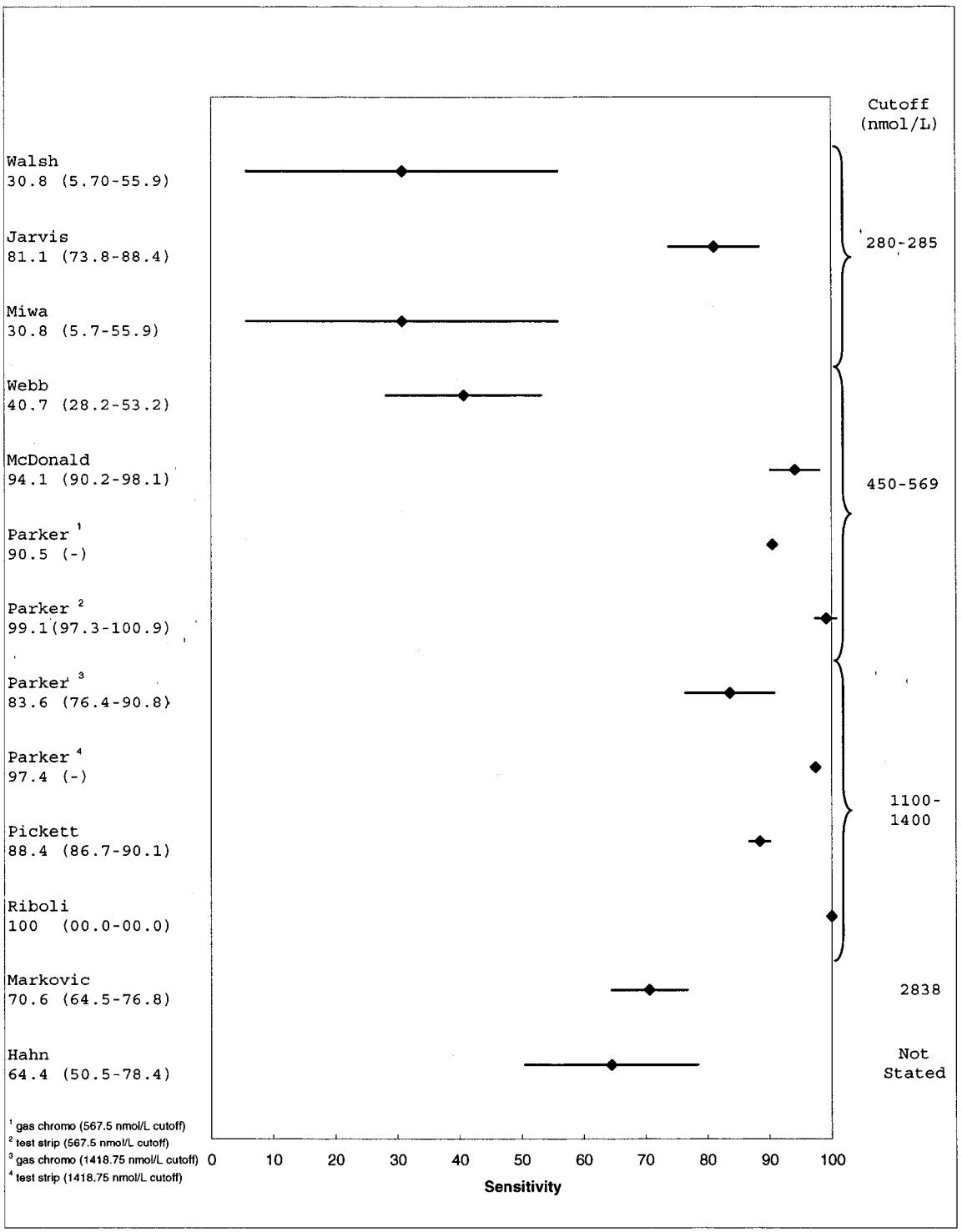

Figure 4: Sensitivity of self-reported smoking versus urine cotinine among male and female smokers. Study author and sensitivity level (95% CI) provided. Studies presented in order of cotinine cut-off value (smallest at top).

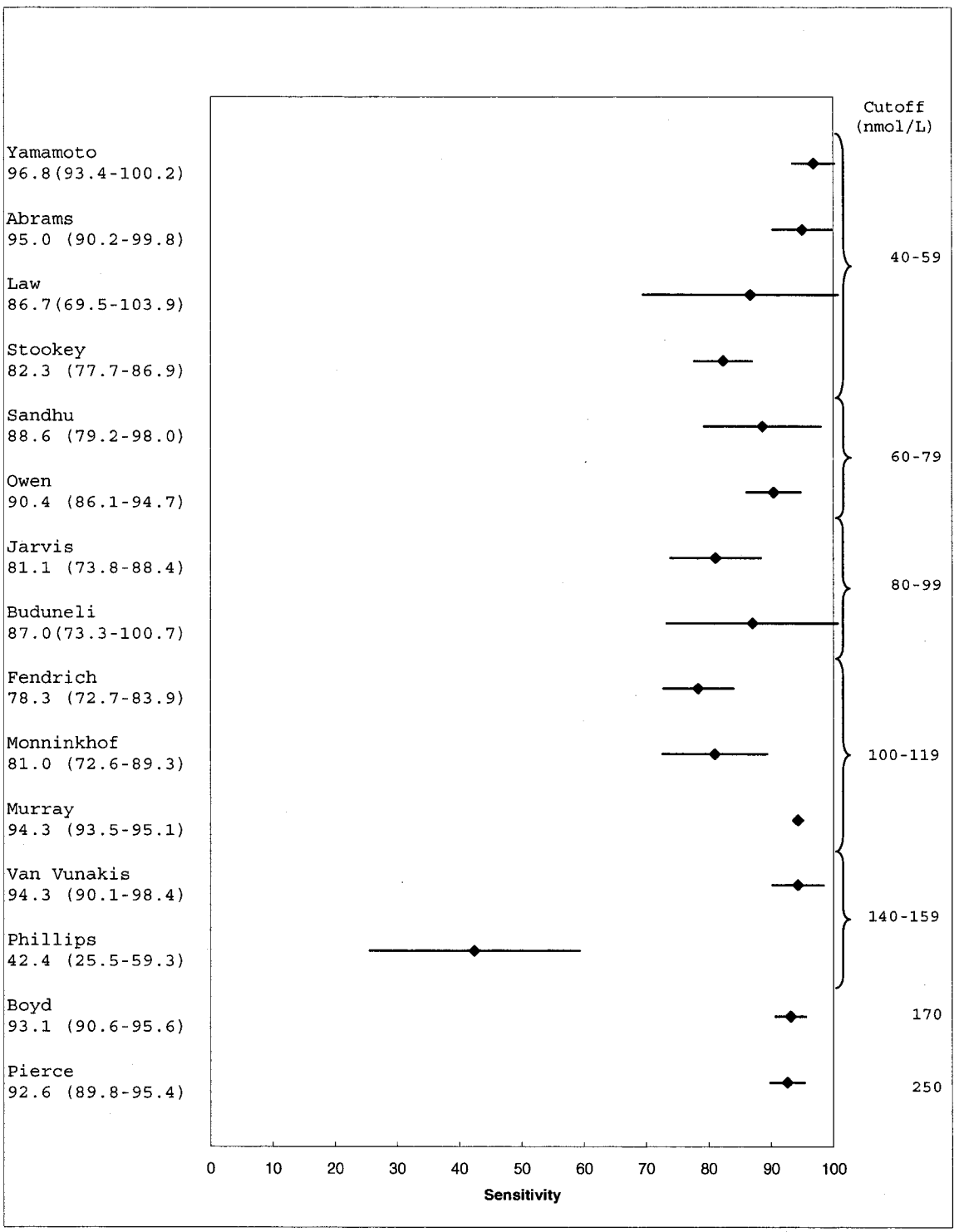

Figure 5: Sensitivity of self-reported smoking versus saliva cotinine among male and female smokers. Study author and sensitivity level (95% CI) provided. Studies presented in order of cotinine cut-off value (smallest at top).

Paper 4: The feasibility of establishing correction factors to adjust self-reported estimates of obesity in the Canadian Community Health Survey

This paper was published in:

Connor Gorber S, Shields M, Tremblay MS, McDowell I. The feasibility of establishing correction factors to adjust self-reported estimates of obesity in the Canadian Community Health Survey. *Health Reports* 2008;19(3):71-82.

This paper was presented at:

Connor Gorber S, Shields M, Tremblay MS, McDowell I. Correcting self-reported estimates of obesity: can we more closely approximate measured values? Presentation at the *Canadian Public Health Association Annual Conference* (Halifax, Nova Scotia), 2008.

Connor Gorber S, Shields M, Tremblay MS, McDowell I. Correcting self-reported estimates of obesity: can we more closely approximate measured values? Presentation at the *24th International Methodology Symposium* (Gatineau, Québec), 2008

The feasibility of establishing correction factors to adjust self-reported estimates of obesity in the Canadian Community Health Survey

Sarah Connor Gorber[1], Margot Shields[1], Mark S. Tremblay[2, 3], Ian McDowell[4]

1. Health Information and Research Division, Statistics Canada
2. Physical Health Measures Division, Statistics Canada
3. Children's Hospital of Eastern Ontario Research Institute
4. Department of Epidemiology and Community Medicine, University of Ottawa

Key Words: obesity, body mass index, self-report, direct measure, measurement error, bias, overweight, prevalence

Correspondence: Sarah Connor Gorber, Chief – Health Measures Analysis Section, Health Information and Research Division, Statistics Canada, 24th Floor R.H. Coats Building, 100 Tunney's Pasture Driveway, Ottawa, ON, Canada, K1A 0T6

Tel:
Fax:
Email:

Abstract

Background

Self-reports underestimate the true prevalence of obesity. This study examined the feasibility of developing correction factors to adjust self-reported measures of Body Mass Index (a proxy for obesity) to more closely approximate measured values.

Data and Methods

Data are from the 2005 Canadian Community Health Survey (sub-sample 2) where respondents were asked to report their height and weight and were subsequently measured. Regression analyses were used to determine which socio-demographic and health characteristics were associated with the discrepancies between reported and measured values. The sample was then split into two groups. In the first, the self-reported BMI and the predictors of the discrepancies were regressed on the measured BMI. Correction equations were generated using all predictor variables that were significant at the $p<0.05$ level. These correction equations were then tested in the second group to derive estimates of sensitivity, specificity and of obesity prevalence. Logistic regression was used to examine the relationship between measured, reported and corrected BMI and obesity-related health conditions.

Results

Corrected estimates provided more accurate measures of obesity prevalence, mean BMI and sensitivity levels (% correctly classified). Self-reported data exaggerated the relationship between BMI and health conditions, while in most cases the corrected estimates provided odds ratios that were more similar to those generated with the measured BMI.

Interpretation

Corrected estimates, though not perfectly predictive of the measured values of BMI, more closely approximate the measured values than the self-reported estimates, which substantially underestimate obesity prevalence and overestimate the relationship between obesity and disease.

What is already known on this subject?

- Self-reported estimates of obesity underestimate true prevalence and overestimate the relationship between obesity and obesity-related health conditions.
- For fiscal and logistical reasons, population health surveillance in Canada is based mostly on self-reported information.

What does this study add?

- Correction factors can be generated to adjust self-reported data to produce more accurate estimates of obesity.
- Although not perfectly predictive of the measured values of BMI, they are an improvement over the self-reported estimates.
- Corrected estimates of BMI are recommended for use in future analyses of adult populations with the 2005 Canadian Community Health Survey.

Introduction

Obesity is a growing public health problem in both the developed and developing world. Globally it is estimated that approximately 400 million people are obese[1], while in Canada the prevalence is estimated to be 23% in adults[2] and 8% in children[3], with rates expected to rise in the coming years[4-5]. The economic costs of obesity in Canada represent approximately 2% of total health care costs[6].

Because of the costs of collecting measured data, national estimates of the prevalence of obesity and obesity-related conditions tend to be based on self-reported survey data. In most countries the Body Mass Index (BMI) is used as the measure of obesity because it can be easily calculated based on reported estimates of height and weight. However, in both clinical and population samples there is a tendency for self-reports to overestimate height and underestimate weight, which results in a systematic underestimation of obesity[7-10]. This trend has recently been confirmed in a systematic review of 64 international studies[11] as well as in Canadian research[12].

Underestimating the prevalence of obesity is significant both because the condition itself can cause significant social and physical impairment and because it is a risk factor for future disease[13-15]. Research has shown that when estimates of obesity are based on self-reported data, the relationship between obesity and obesity-related health conditions such as diabetes, hypercholesterolemia, hypertension, arthritis and heart disease is substantially exaggerated[16-18].

Given the reality that a great deal of population health surveillance will continue to be based on self-reported data, it has been suggested[19] that reported estimates of obesity could be adjusted to more closely approximate their measured values. Using data from the Canadian Community Health Survey (CCHS) in which both measured and reported height and weight were collected on the same individuals, this study examines the feasibility of developing correction equations to adjust reported estimates and assesses whether they improve the estimation of obesity (when based on BMI).

Methods

Data for this study come from the 2005 CCHS. The CCHS is an ongoing cross sectional survey designed to provide timely cross-sectional estimates of health determinants, health status and health system utilization at a sub-provincial level[20]. Sampling is based on a multi-stage cluster sampling technique that is representative of over 98% of the Canadian population (individuals living on Indian Reserves or Crown lands, residents of institutions, Canadian Armed Forces and certain remote regions are excluded). Three sampling frames were used to select the sample of households for the 2005 survey: 49% of the sample of households came from an area frame, 50% came from a list frame of telephone numbers and the remaining 1% came from a Random Digit Dialing (RDD) sampling frame. More details about the CCHS are available in Béland, 2002[21].

The 2005 CCHS collected data from 132,947 respondents, yielding a response rate of 79%. As part of the 2005 CCHS a sub-sample of 7,376 respondents aged 12 years and older were asked to self-report their height and weight and then, later in the interview, they were

directly measured by Statistics Canada interviewers. These participants were all drawn from the area frame and were selected across the ten provinces in proportion to their population sizes (residents of the territories were excluded). Measures of height and weight were obtained for 4,735 of these respondents, giving a response rate of 64% (the main reason for non-response was refusal). Because of the high rate of non-response, an adjustment was made to minimize non-response bias; a special sampling weight was created by redistributing the sampling weights of non-respondents (to measured height and weight) to respondents using response propensity classes. The variables used to create these classes included region (British Columbia, Prairies, Ontario, Quebec, Atlantic provinces), age, sex, household size, marital status, rural/urban indicator, and quarter of data collection.

The present study included adults aged 18 years or older because children are still in a stage of development where weight and height may change over short periods of time. It has also been suggested that reporting error in children and adolescents may be of a different nature than that in adults[10]. Females who were pregnant (n=47) or breastfeeding (n=58) were also excluded as the BMI is not recommended for use in these groups. Respondents for whom the difference between the reported and measured estimates of height, weight or BMI were more than 3 standard deviations from the mean were considered outliers and were excluded from the analysis (n=43, n=44 and n=39 respectively). This left 4080 respondents with measured and reported values for height and weight.

CCHS interviewers were trained to measure height and weight. Height was measured to the nearest 0.5 cm (without shoes) with a measuring tape attached to the wall. Weight was

measured to the nearest 0.1 kg (without shoes) with a calibrated digital scale (ProFit UC-321 made by Lifesource). The interview was approximately 50 minutes long and took place in the respondent's home. Reported height and weight were collected near the beginning of the interview, while the measurements were taken near the end. Respondents were not told that they would be measured. Self-reported height and weight were collected with the questions: "How tall are you without shoes on?" and "How much do you weigh?". Categories for height in feet and inches were listed on the questionnaire with corresponding metric values in brackets. Interviewers were instructed to round up to the closest inch for respondents who reported half inch measures. If asked, interviewers were instructed to tell respondents to report weight without clothing. After reporting their weight, respondents were asked if they had reported in pounds or kilograms. Most respondents (94%) reported in pounds.

Bias in Reported Estimates

The first step was to use the full sub-sample (n=4080) to determine which factors were associated with the bias between the reported and measured estimates. The bias was calculated by subtracting the measured value from the reported value; negative values indicated underestimation and positive values indicated overestimation. Multiple linear regression was used with the bias as the dependent variable in the model and socio-demographic and health variables, selected based on a review of the literature, entered as independent variables. Separate models were estimated with the bias in weight, height and BMI as outcome variables. All models were estimated separately for males and females because the bias has been shown to differ between these two groups[8,22-24]. Significant variables ($p<0.05$) were used to develop the correction equations.

The sample was then randomly divided into two parts, split sample A and split sample B. Split sample A was used to generate the correction equations using the variables that were significantly associated with the bias in height, weight and BMI that were identified in the first step of the analysis. Split sample B was used to test them. In order to generate the correction equations the measured value was used as the outcome variable and the self-reported value and any variables that were significantly associated with the bias from the first part of the study were used as independent variables. Only significant independent variables (or categorical variables for which at least one category was significant) were retained for the final correction equations.

Four models were tested, two Full Models and two Reduced Models. In model 1 (the first Full Model) estimates of height and weight were first adjusted based on the predictors that were significantly related to the reported-measured bias in height and weight (respectively) in step 1. BMI was then calculated using the adjusted values of height and weight. In model 2 (the second Full Model) BMI was adjusted by regressing the predictors of the bias in BMI from step 1 directly onto measured BMI. The reduced models were similar, except that only the reported estimates of height, weight and BMI were used as independent predictors of the measured values. The models are as follows:

Full Models:

Model 1 (Full - Height and Weight):

$$Weight_{measured} = b_o + b_1(weight_{reported}) + b_2(\text{var } 1) + b_3(\text{var } 2) + b_i(x_i) \ldots + error$$

$$Height_{measured} = b_o + b_1(height_{reported}) + b_2(\text{var } 1) + b_3(\text{var } 2) + b_i(x_i) \ldots + error$$

Model 2 (Full - BMI):

$$BMI_{measured} = b_o + b_1(bmi_{reported}) + b_2(\text{var } 1) + b_3(\text{var } 2) + b_i(x_i) \ldots + \text{error}$$

Reduced Models:

Model 3 (Reduced - Height and Weight):

$$Weight_{measured} = b_o + b_1(weight_{reported}) + \text{error}$$

$$Height_{measured} = b_o + b_1(height_{reported}) + \text{error}$$

Model 4 (Reduced - BMI):

$$BMI_{measured} = b_o + b_1(bmi_{reported}) + \text{error}$$

All analyses were again run separately for males and females. Interactions and quadratic terms were tested as appropriate. In order to estimate the regression equations, all variables were entered into the models simultaneously but only significant variables were retained to generate the final correction equations. Final models were tested to ensure they met the assumptions of independence, linearity, equal variance and normality.

Once the correction equations were generated from split sample A, they were applied to the data in split sample B. Descriptive statistics (means, prevalence of selected categories) were used to compare the reported, measured and corrected estimates of obesity. Sensitivity (the proportion of obese respondents, based on the measured values, who were classified as such based on the reported or corrected estimates) and specificity (the proportion of the non-obese who were correctly classified based on the reported or corrected estimates) were used to

indicate whether the corrected estimates improved BMI classification compared to the reported estimates. Respondents were categorized according to the World Health Organization[25] and Canadian classification guidelines[26] which characterize individuals into underweight (BMI <18.5 kg/m^2), normal weight (BMI 18.5-24.9 kg/m^2), overweight (BMI 25.0-29.9 kg/m^2) and obese (BMI 30.0 kg/m^{2}or over) categories.

Logistic regression was then used in order to determine if the corrected estimates more accurately modeled the relationship between obesity and obesity-related health conditions than self-reported values. All models controlled for age and sex and examined the relationship between BMI (measured, reported and corrected) and one of the following six obesity-related conditions: diabetes, heart disease, hypertension, arthritis, activity limitations and fair or poor self-rated health. This analysis was restricted to those aged 40 years or older because these conditions are more prevalent in these age ranges.

Data were appropriately weighted and all measures of variance were estimated with the bootstrap technique to account for the CCHS's complex survey design[27-28]. SAS (version 9.1) was used for all analyses.

Definitions

The socio-demographic variables included *age* (divided in 7 groups, 18-24; 25-34; 35-44; 45-54; 55-64; 65-74 and 75 years and older); the respondent's *highest level of education* divided into 4 categories: no secondary school diploma, secondary school diploma, some post-secondary studies, post-secondary degree; *geographical region* of the country (Atlantic,

Québec, Ontario, West, British Columbia); *urban or rural area*; *employment status* the week prior to the interview (full time, part-time or not working); *immigrant status* (≤10 years,>10 years in Canada, Canadian born); *ethnicity* (collapsed due to sample size into White, East and South East Asian and Other); and *household income. Household income* groups were derived by dividing the total household income from all sources in the previous 12 months by Statistics Canada's low-income cutoff (LICO) specific to the number of people in the household, the size of the community, and the survey year. These adjusted income quotients were grouped into deciles.

The health conditions included self-declared *health status and mental health status* (dichotomized into fair or poor versus good, very good, or excellent); *activity limitations* imposed by a long term health problem (sometimes/often versus never); *smoking status* (daily or occasional versus non smoker); *self-perceived stress (*most days are quite a bit or extremely stressful versus a bit or not very stressful*)*; *life satisfaction* (dissatisfied or very dissatisfied versus satisfied or very satisfied*)*; *the perception of weight* (overweight, underweight or about right); *number of physician consultations in the past year* (continuous); and *chronic conditions* (asthma, arthritis/rheumatism, hypertension, diabetes, heart disease, cancer, mood disorders). Sample sizes were too small to examine associations with eating disorders. *Leisure-time physical activity level* was based on total energy expenditure (EE) during leisure time. EE was calculated from the reported frequency and duration of all of a respondent's leisure-time physical activities in the three months before the 2005 CCHS interview and the metabolic energy demand (MET value) of each activity, which was independently established[29].

EE = Σ(Ni*Di*METi / 365 days), where

Ni = number of occasions of activity i in a year,

Di = average duration in hours of activity i, and

METi = a constant value for the metabolic energy cost of activity i.

An EE of 3 or more kilocalories per kilogram per day (KKD) was defined as *active*; 1.5 to 2.9 KKD, *moderately active*; and less than 1.5 KKD, *inactive*.

The influence of *end digit preference* (the tendency to round responses to numbers ending in 0 and 5) was also examined for weight as past research has indicated that it has been associated with a reporting bias[8,10,30]. In fact, in the CCHS the majority of respondents (73% of males and 67% of females) reported values for their weight that ended in 0 or 5, although it would be expected that by chance only about 20% of respondents would have end-digits of 0 or 5.

Results

Consistent with past research, mean values of self-reported height were over-estimated while weight and BMI were under-estimated. Males overestimated height by 1.08 cm and underestimated weight by 1.84 kg and BMI by 0.94 kg/m^2. For women height was overestimated by 0.56 cm and weight and BMI were underestimated by 2.47 kg and 1.19 kg/m^2.

In order to estimate the correction equations the significant factors from the linear regressions that were associated with the bias in reported weight, height and BMI were retained (data not shown). The data were then randomly split into two parts; split-sample A and split-sample B, each containing approximately 50% of the respondents (2029 or 49.7% and 2051 or 50.3%, respectively). The regression models used to generate the correction equations were first estimated with all of the independent variables that were associated with the bias, insignificant variables were then removed and only significant variables (or categorical variables for which at least one category was significant) were used to estimate the final correction equations. The regression results derived from split-sample A that were used to establish the correction equations for weight are shown in Table 1a. In the full models for males, self-reported weight, age and the respondent's perception that they were over or underweight were significant predictors of measured weight; those who perceived themselves as overweight tended to underestimate their weight and those who perceived themselves as underweight tended to overestimate their weight and the model adjusted these values up or down as appropriate. The adjusted R^2 was .95 for both the full and reduced models. For females, factors associated with measured weight included the reported weight, the perception of being overweight and end digit preference, which is expressed as a tendency for females to round weights (usually down) to numbers ending in zero or five (and thus the model added a positive adjustment to the reported weight to compensate for this tendency). The adjusted R^2 for females for both the full and reduced models was .97.

Results for height are found in Table 1b. For males, reported height, age and life dissatisfaction were all significant predictors of measured height, with a negative adjustment

related to age and a positive adjustment for those who reported being dissatisfied with their lives. The full model had an adjusted R^2 of .82 while for the reduced model it was .81. For females, adjustments were made at all age groups (except those aged 45-54 years), for those whose ethnicity was a group other than White or East/South East Asian and for those who reported living with an activity limitation caused by a long term health problem.

For BMI (Table 1c) the full models adjusted the reported estimates down for males who were dissatisfied with life and who perceived themselves as underweight while positive adjustments were made for age. With females significant predictors of measured BMI included reported BMI, education, perception of being overweight and end digit preference. The R^2 were higher for the female versus male models, but in both cases were similar for the full and reduced models.

In order to generate the final equations adjustments were made for all of the variables in Tables 1a-c. The final equations were as follows:

Males

Full Model 1

$$\text{Weight}_{(\text{measured})} = -0.30+1.01(\text{weight}_{\text{reported}})+0.54(\text{age25-34})+0.39(\text{age35-44})$$
$$+0.50(\text{age45-54})+1.69(\text{age55-64})+0.83(\text{age65-74})+0.39(\text{75andolder})$$
$$+1.16(\text{overweight})-1.52(\text{underweight})$$

$$\text{Height}_{(\text{measured})} = 12.17+0.93(\text{height}_{\text{reported}})-1.48(\text{age25-34})-0.43(\text{age35-44})-1.23(\text{age45-54})$$
$$-2.44(\text{age55-64})-2.87(\text{age65-74})-2.84(\text{75and older})+2.22(\text{life dissatisfaction})$$

Full Model 2

$$\text{BMI}_{(\text{measured})} = -0.67+1.04(\text{BMI}_{\text{reported}})+0.64(\text{age25-34})+0.31(\text{age35-44})$$

+0.39(age45-54)+1.28(age55-64)+1.16(age65-74)+0.86(75and older)

-0.97(life dissatisfaction)-0.73(underweight)

Reduced Model 3

$Weight_{(measured)} = -2.19+1.05(weight_{reported})$

$Height_{(measured)} = 7.70+0.95(height_{reported})$

Reduced Model 4

$BMI_{(measured)} = -1.08+1.08(BMI_{reported})$

Females

Full Model 1

$Weight_{(measured)} = -1.25+1.04(weight_{reported})+1.25(overweight)+0.52(end\ digit\ preference)$

$Height_{(measured)} = 14.85+0.91(height_{reported})-1.20(age25-34)-0.87(age35-44)-0.59(age45-54)$

$-1.34(age55-64)-1.42(age65-74)-3.79(75andolder)$

$-0.32(ethnicity\ E/SE\ Asian)-0.73(ethnicity\ other)-0.66(activity\ limitation)$

Full Model 2

$BMI_{(measured)} = 1.01+1.01(BMI_{reported})-0.91(high\ school\ educ.)-0.32(some\ post\ secondary\ educ.)-$

$0.53(post\ secondary\ educ.)+0.70(overweight)+0.29(end\ digit\ preference)$

Reduced Model 3

$Weight_{(measured)} = -2.14+1.07(weight_{reported})$

$Height_{(measured)} = 8.05+0.95(height_{reported})$

Reduced Model 4

$BMI_{(measured)} = -0.12+1.05(BMI_{reported})$

These equations were applied to data in split-sample B in order to generate corrected estimates of height, weight and BMI. Mean values of the measured, reported and corrected data from split sample B are displayed in Table 2. In all cases reported estimates were statistically different from the measured values and the modeled estimates were closer than

the reported estimates to the true measured values. In all but one case (the difference in BMI for females in model 3) the modeled and measured means were not statistically different.

Obesity prevalence is reported in Table 3. For males, 23.1% were obese by the measured data, 13.8% by the reported data while the corrected estimates generated prevalence estimates of 19-22%. Measured, reported and modeled rates were similar for the overweight category for males but self-report overestimated the numbers who were in the normal weight range. The corrected estimates decreased this reporting bias by 9 to 11 percentage points so that the modeled and measured estimates were similar. For females, 18.9% were obese by measurement compared to 12.5% for reported values. The modeled estimates generated obesity prevalence estimates of 18.7% (model 1), 18.3% (model 2), 18.2% (model 3) and 18.3% (model 4). Similarly, for the overweight estimates, modeled values were closer than the self-reported values to the measured prevalence, with a slight 1-2% overestimate in the modeled values. Sample sizes were too small in the underweight categories to generate reliable estimates.

Sensitivity and specificity values for reported and corrected data are shown in Table 4. Sensitivity values in the normal weight category for self-reported data were 93.9% for males and 91.8% for females, meaning that in most cases reported estimates correctly classified people of normal weight into the normal weight category. Sensitivities for the overweight and obese categories fell to 71.1% and 58.7% respectively for males and to 62.6% and 68.5% for females. When the data were corrected, the sensitivities increased for both the overweight and obese categories; the corrected numbers accurately classified as many as

86.1% of obese females, 76% of obese males, 79.7% of overweight females and 82.8% of overweight males. Corrected estimates decreased sensitivities for those in the normal weight range, however. Specificities were the highest for the underweight and obese categories, indicating that it is rare for someone to be classified into these groups unless they actually are underweight or obese.

Data in Table 5 display the adjusted odds ratios relating measured, reported and corrected BMI to six obesity-related health conditions. The reported and measured models were previously published in a paper that demonstrated that self-reported BMI exaggerates the relationship between obesity and these obesity-related health conditions[16]. Unique to this analysis is that the models have been re-generated based on the corrected estimates. When compared to the odds ratios from the reported models, the ratios for the corrected models are reduced in most cases (i.e., are more similar to the measured values). Arthritis is an exception, with the corrected estimates inflating the relationship for those who are overweight or obese (class II or II, BMI≥35 kg/m^2) even higher than what they would be if based on self-report. In addition, the odds ratios for obese class I are higher than the reported odds ratios for diabetes in models 1 and 2 and for high blood pressure in models 3 and 4.

The data on measured height and weight were available on only a sub-sample of the 2005 CCHS. However, the ultimate goal of developing correction equations is to be able to apply them to the broader survey. When applied to the full sample of the 2005 CCHS (without different adjustments for phone and in-person interviews) respondents who were 18 years and older and who were not pregnant or breastfeeding (n=118,383), the models generated

estimates of obesity that were similar to, although slightly lower than, the measured values (see Table 6). Based on data from both split sample A and B, measured obesity prevalence was 25.6% in males and 22.3% in females while reported prevalence was 16% for both males and females. The models generated prevalence rates of approximately 23% for males and of 21% for females.

Study Limitations

The response rate on the measured height and weight sub-sample of the CCHS was only 65%. If those who agreed to participate had different height and weight profiles than those who refused, the sample could be biased. The self-reported prevalence of obesity among all respondents selected for measured height and weight was 15.9%; 19.1% among non-respondents and 14% among respondents. However, when the special sampling weight was applied to those who underwent the physical measures, the prevalence of obesity based on reported data decreased to 15.2%, comparable to that for the entire sub-sample[12].

Bias in reported height may be due to inconsistent rounding between the self-reported and the measured height. Interviewers asked respondents to round their height up to the nearest inch when half inches were reported, yet for the measured values height was recorded to the nearest 0.5 cm. Moreover, as reported data were only recorded in meters it was impossible to determine the number of people who reported their height in feet and inches in order to asses the extent of this rounding bias. With weight, it was difficult to determine if respondents were consistently asked to empty pockets and remove footwear prior to being weighed and whether they were reporting weight with or without clothing on as interviewer's instructed

respondents to report without clothing only if they were asked. Although interviewers were trained in the correct procedures for measuring height and weight, and the weigh scales and measuring tapes were calibrated, no research on intra- or inter- interviewer reliability was performed.

BMI is commonly used as a measure of obesity on population surveys because height and weight can easily be self-reported. But it has limitations: BMI cannot distinguish between muscle mass and fat, nor does it consider fat distribution[26].

Finally, the models were limited to the variables that were collected in the CCHS. Although the CCHS collects data on a wide variety of health characteristics, it is possible that additional variables that were not included in the survey could be associated with the bias in weight, height or obesity.

Discussion

BMI calculated from self-reported height and weight underestimates obesity prevalence, which has implications for our understanding of the burden of obesity and the relationship between obesity and obesity-related health conditions. This study examined the feasibility of applying correction factors to the self-reported estimates to determine if they could be adjusted to more closely approximate measured values. In each of the four models tested and in all analyses undertaken, the corrected estimates provided more accurate measures of overweight and obesity than the reported values. However, this was not the case for the normal weight category. The sensitivity values for the normal weight population decreased

to as low as 84% in males (a 10 percentage point decrease) and to 83% in females (a 9% decrease). A reduction in sensitivity for normal weight individuals was also found by Kuskowska-Wolk and colleagues[19]. We hypothesized that this was due to the fact that heavier individuals have a greater reporting bias[12] (i.e., greater tendency to underestimate their BMI) and therefore different adjustments may be required depending on where on the BMI distribution the individual lies. Without these differing adjustments, a loss in sensitivity results when a small proportion of normal weight individuals are erroneously shifted to the overweight category. We attempted to address this by incorporating polynomial regressions (quadratic terms for reported weight) and spline regression to determine whether different slopes could be generated for different weight ranges, but the quadratics and differential slopes were not significant and we were unable to refine the estimates for those in the normal weight range. Therefore, although the adjustments improve the estimates for those who are overweight or obese, the non-adjusted numbers provide better estimates for respondents in the normal weight category as the reporting bias in this group is smaller. Further research is required to better understand how to improve the reported overweight and obesity estimates without decreasing the sensitivity for those in the normal weight range. More research is also required to determine whether differential adjustments are required for respondents who were interviewed by telephone.

Despite this limitation, the improvement in classification for overweight and obese individuals is significant and thus we recommend the use of the corrected estimates in addition to the self-reported values for studies examining overweight and obesity in the adult population (18 years and over) of the 2005 CCHS. Although we attempted to adjust for any

independent variables that were related to the reporting bias, the R^2 of these "Full Models" (models 1 and 2) were either the same as or only slightly higher than those of the "Reduced Models" (models 3 and 4, which used only weight, height or BMI) and in most cases there was no predictive advantage of including the extra variables. Plankey and colleagues[31] also found that more complex models (those including self-reported BMI and additional covariates) only minimally improved their model's predictive ability. Of the models we tested, all four generated similar means, prevalence rates and sensitivity values (i.e., no model stood out as being consistently superior). Model 4 had the further advantage of being the most parsimonious and therefore showing the greatest potential for generalizability if it is determined that the equations are generalizable.

This method of generating corrected estimates (linear regression with measured BMI as the outcome) has been used in the past by several researchers[10,19,31-34], but to our knowledge has never been attempted on data from the Canadian population. Plankey and colleagues[31] concluded that a systematic error was associated with the reporting bias, which was impossible to correct with this method. However, in their work the self-reported sensitivity values for the obese population (BMI≥27.3) were 80% in men and 85% in women and increased only marginally with the corrected models, while in the current study self-reported sensitivity of obesity was much lower, 59% for males and 69% for females, and the correction equations increased these values significantly. Also worth mentioning is that the reporting bias in our study was 2 to 3 times larger than it was in the 1976-1980 NHANES II, which was the basis for their analysis.

The generalizability of these equations has not been determined. Some authors[33] assume transportability while others have demonstrated that the correction equations are only applicable to the population in which they have been established[30]. In one Swedish study[32], researchers demonstrated that because height was under- rather than over-reported in their country their self-reported estimates of BMI did not require calibration. More research using Canadian data is required to determine if these equations are stable across Canadian populations and over time. It is probable that with the increase in obesity in recent years[35] there has been a corresponding increase in reporting bias, which could indicate a temporal instability in the equations. At least one study that has examined the bias over time has found that it has increased[36].

In the interim, surveys that collect reported and measured height and weight would benefit from standardization of protocols to ensure that equipment is regularly calibrated, that respondents are consistently asked to report their weight and are measured in light clothing, without shoes. Rounding should also be minimized, if not eliminated.

Conclusion

Although measured data for height and weight provide the most accurate estimates of the prevalence of obesity based on the BMI, the costs of collecting directly measured data are often prohibitive on large population based studies. Corrected estimates, though not identical to the measured values of BMI, are a significant improvement over the self-reported estimates, which substantially underestimate obesity prevalence and overestimate the relationship between obesity and disease.

Acknowledgements

The authors would like to thank Julie Bernier for her methodological assistance and members of the Health Information and Research Division at Statistics Canada for their input on this research.

References

1. World Health Organization. Obesity and Overweight [Online]. 2006 [cited 2008 May]. Geneva: World Health Organization. Available from: www.who.int/mediacentre/factsheets/fs311/en/index.html.
2. Tjepkema M. Adult obesity. *Health Rep* 2006;17(3):9-25.
3. Shields M. Overweight and obesity among children and youth. *Health Rep* 2006;17(3):27-42.
4. Rigby N, Leach R., James WPT. *Seeking bold solutions for Britain's runaway obesity epidemic. International Obesity Taskforce briefing paper.* London: IOTF; 2003.
5. Sturm R, Ringel JS, Andreyeva T. Increasing obesity rates and disability trends. *Health Aff* 2004;23:199-205.
6. Katzmarzyk PT, Janssen I. The economic costs associated with physical inactivity and obesity in Canada: an update. *Can J Appl Physiol* 2004;29:90–115.
7. Engstrom JL, Paterson SA, Doherty A, Trabulsi M, Speer KL. Accuracy of self-reported height and weight in women: an integrative review of the literature. *J Midwifery Womens Health* 2003;48:338–45.
8. Niedhammer I, Bugel I, Bonenfant S, Goldberg M, Leclerc A. Validity of self-reported weight and height in the French GAZEL cohort. *Int J Obes Relat Metab Disord* 2000;24: 1111–8.
9. Nieto-Garcia FJ, Bush TL, Keyl PM. Body mass definitions of obesity: sensitivity and specificity using self-reported weight and height. *Epidemiology* 1990;1:146–52.
10. Rowland ML. Self-reported weight and height. *Am J Clin Nutr* 1990;52:1125–33.

11. Connor Gorber S, Tremblay M, Moher D, Gorber B. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obes Rev* 2007;8:307-26.

12. Shields M, Connor Gorber S, Tremblay M. Estimates of obesity based on self-report versus direct measures. *Health Rep* 2008;19(2):61-76.

13. Puhl RM, Brownell KD. Psychosocial origins of obesity stigma: toward changing a powerful and pervasive bias. *Obes Rev* 2003;4:213-27.

14. Kumanyika S, Jeffery RW, Morabia A, Ritenbaugh C, Antipatis VJ; The Public Health Approaches to the Prevention of Obesity (PHAPO) Working Group of the International Obesity Task Force (IOTF). Obesity prevention: the case for action. *Int J Obes* 2002;26:425-36.

15. Canadian Institutes for Health Information. *Improving the health of Canadians: promoting healthy weights*. Ottawa: Canadian Institute for Health Information; 2006.

16. Shields M, Connor Gorber S, Tremblay MS. Effects of measurement on obesity and morbidity. *Health Rep* 2008;19(2):77-84.

17. Yannakoulia M, Panagiotakos DB, Pitsavos C, Stefanadis C. Correlates of BMI misreporting among apparently healthy individuals: the ATTICA study. *Obesity* 2006;14(5):894-901.

18. Chiolero A, Peytremann-Bridevaux I, Paccaud F. Associations between obesity and health conditions may be overestimated if self-reported body mass index is used. *Obes Rev* 2007;8:373-4.

19. Kuskowska-Wolk A, Bergstrom R, Bostrom G. Relationship between questionnaire data and medical records of height, weight and body mass index. *Int J Obes* 1992;16:1-9.

20. Statistics Canada. Population health surveys [Online]. 2005 [cited 2008 Jan. Ottawa: Statistics Canada. Available from: www.statcan.ca/english/concepts/hs/index.htm#content.

21. Béland Y. Canadian Community Health Survey – methodological overview. *Health Rep* 2002;13(3):9-14.

22. Ziebland S, Thorogood M, Fuller A, Muir J. Desire for the body normal: body image and discrepancies between self-reported and measured height and weight in a British population. *J Epidemiol Community Health* 1996;50:105–6.

23. Bostrom G, Diderichsen F. Socioeconomic differentials in misclassification of height, weight and body mass index based on questionnaire data. *Int J Epidemiol* 1997;26:860-6.

24. Roberts RJ. Can self-reported data accurately describe the prevalence of overweight? *Public Health* 1995;109:275–84.

25. World Health Organization. *Physical status: the use of and interpretation of anthropometry, report of the WHO expert committee* (WHO Technical Report Series, No. 854). Geneva: World Health Organization; 1995.

26. Health Canada. *Canadian guidelines for body weight classification in adults*. Ottawa: Health Canada; 2003.

27. Rao JNK, WU CFJ, Yue K. Some recent work on resampling methods for complex surveys. *Surv Methodol* 1991;18:209-17.

28. Rust KF, Rao NJK. Variance estimation for complex surveys using replication techniques. *Stat Methods Med Res* 1996;5:281-310.

29. Statistics Canada. *Canadian Community Health Survey Cycle 3.1 Derived Variable Specifications*. Ottawa: Statistics Canada; 2006.

30. Visscher TLS, Viet AL, Kroesbergen HT, Seidell JC. Underreporting of BMI in adults and its effect on obesity prevalence estimations in the period 1998 to 2001. *Obesity* 2006;14:2054-63.

31. Plankey MW, Stevens J, Flegal KM, Rust PF. Prediction equations do not eliminate systematic error in self-reported body mass index. *Obes Res* 1997;5:308-14.

32. Bolton-Smith C, Woodward M, Tunstall-Pedoe H, Morrison C. Accuracy of the estimated prevalence of obesity from self reported height and weight in an adult Scottish population. *J Epidemiol Community Health* 2000;54:143-8.

33. Cawley J. The impact of obesity on wages. *J Hum Resour* 2004;39:451-74.

34. Nyholm M, Gullberg B, Merlo J, Lundqvist-Persson C, Rastam L, Lindblad U. The validity of obesity based on self-reported weight and height: implications for population studies. *Obesity* 2007;15:197-208.

35. Katzmarzyk PT. The Canadian obesity epidemic, 1985–1998. *CMAJ* 2002;166:1039-40.

36. Ezzati M, Martin H, Skjold S, Vander Hoorn S, Murray CJL. Trends in national and state-level obesity in the USA after correction for self-report bias: analysis of health surveys. *J Res Soc Med* 2006;99:250-7.

Tables and Figures

Table 1a:

Regression results for establishing correction equations for weight in males and females, full and reduced models, generated from split-sample A

				95% Confidence Intervals		
Men	**Variable**	**Coefficient**				
Full model	Intercept	-0.30		-2.70	to	2.09
Rsqr= .95	Self-reported weight (kg)	1.01	*	0.98	to	1.05
Rsqr(adj)= .95	Aged 25-34 years+	0.54		-0.53	to	1.61
	Aged 35-44 years+	0.39		-0.68	to	1.46
	Aged 45-54 years+	0.50		-0.51	to	1.51
	Aged 55-64 years+	1.69	*	0.59	to	2.79
	Aged 65-74 years+	0.83		-0.16	to	1.82
	Aged 75 years or older+	0.39		-0.58	to	1.35
	Perceives self as overweight	1.16	*	0.41	to	1.92
	Perceives self as underweight	-1.52	*	-2.92	to	-0.13
Reduced Model	Intercept	-2.19	*	-4.27	to	-0.12
Rsqr=. 95	Self-reported weight (kg)	1.05	*	1.02	to	1.07
Rsqr(adj)= .95						
Women						
Full model	Intercept	-1.25		-3.25	to	0.75
Rsqr= .97	Self reported weight (kg)	1.04	*	1.01	to	1.08
Rsqr(adj)= .97	Perceives self as overweight	1.25	*	0.51	to	1.98
	Preference for end digits (0 and 5)	0.52	*	0.01	to	1.02
Reduced model	Intercept	-2.14	*	-3.86	to	-0.43
Rsqr= .97	Self reported weight (kg)	1.07	*	1.04	to	1.10
Rsqr(adj)= .97						

Notes: Dependent variable is measured weight

+ reference group is ages 18-24 years

* $p<0.05$

Table 1b:

Regression results for establishing correction equations for height in males and females, full and reduced models, generated from split-sample A

	Variable	Coefficient		95% Confidence Intervals		
Men						
Full model	Intercept	12.17	*	5.56	to	18.78
Rsqr= .82	Self-reported height (cm)	0.93	*	0.90	to	0.97
Rsqr(adj)= .82	Aged 25-34 years[+]	-1.48	*	-2.43	to	-0.53
	Aged 35-44 years[+]	-0.43		-1.50	to	0.65
	Aged 45-54 years[+]	-1.23	*	-2.34	to	-0.13
	Aged 55-64 years[+]	-2.44	*	-3.41	to	-1.46
	Aged 65-74 years[+]	-2.87	*	-4.09	to	-1.64
	Aged 75 years or older[+]	-2.84	*	-4.17	to	-1.51
	Disatisfied with life	2.22	*	0.30	to	4.13
Reduced Model	Intercept	7.70	*	0.71	to	14.68
Rsqr= .81	Self-reported height (cm)	0.95	*	0.91	to	0.99
Rsqr(adj)= .81						
Women						
Full model	Intercept	14.85	*	9.25	to	20.45
Rsqr = .83	Self-reported height (cm)	0.91	*	0.88	to	0.95
Rsqr(adj) = .83	Aged 25-34 years[+]	-1.20	*	-2.04	to	-0.36
	Aged 35-44 years[+]	-0.87	*	-1.65	to	-0.08
	Aged 45-54 years[+]	-0.59		-1.58	to	0.40
	Aged 55-64 years[+]	-1.34	*	-2.58	to	-0.10
	Aged 65-74 years[+]	-1.42	*	-2.35	to	-0.48
	Aged 75 years or older[+]	-3.79	*	-5.05	to	-2.53
	Ethnicity is East or South East Asian[t]	-0.32		-1.84	to	1.19
	Ethnicity other[t]	-0.73	*	-1.40	to	-0.06
	Presence of activity limitations	-0.66	*	-1.29	to	-0.03
Reduced model	Intercept	8.05	*	2.53	to	13.56
Rsqr = .81	Self-reported height (cm)	0.95	*	0.91	to	0.98
Rsqr(adj) = .81						

Notes: Dependent variable is measured height

+ reference group is ages 18-24 years

[t] reference group is White

* $p<0.05$

Table 1c:

Regression results for establishing correction equations for body mass index in males and females, full and reduced models, generated from split-sample A

Men						
	Variable	**Coefficient**		**95% Confidence Intervals**		
Full model	Intercept	-0.67		-1.83	to	0.48
Rsqr = .864	Self-reported BMI (kg/m^2)	1.04	*	0.99	to	1.09
Rsqr(adj) = .863	Aged 25-34 years$^+$	0.64	*	0.17	to	1.11
	Aged 35-44 years$^+$	0.31		-0.21	to	0.83
	Aged 45-54 years$^+$	0.39		-0.24	to	1.02
	Aged 55-64 years$^+$	1.28	*	0.69	to	1.88
	Aged 65-74 years$^+$	1.16	*	0.60	to	1.72
	Aged 75 years or older$^+$	0.86	*	0.32	to	1.41
	Disatisfied with life	-0.97	*	-1.61	to	-0.33
	Perceives self as underweight	-0.73	*	-1.32	to	-0.14
Reduced Model	Intercept	-1.08		-2.17	to	0.01
Rsqr = .854	Self-reported BMI (kg/m^2)	1.08	*	1.04	to	1.12
Rsqr(adj) = .854						
Women						
Full model	Intercept	1.01		-0.56	to	2.57
Rsqr = .917	Self-reported BMI (kg/m^2)	1.01	*	0.93	to	1.08
Rsqr(adj) = .917	Highest level of education is high school^t	-0.91	*	-1.53	to	-0.29
	Highest level of education is some post secondary^t	-0.32		-1.30	to	0.65
	Highest level of education is post secondary graduation^t	-0.53	*	-1.04	to	-0.03
	Perceives self as overweight	0.70	*	0.17	to	1.23
	Preference for end digits (0 and 5)	0.29	*	0.00	to	0.58
Reduced model	Intercept	-0.12		-1.53	to	1.28
Rsqr = .913	Self-reported BMI (kg/m^2)	1.05	*	0.99	to	1.11
Rsqr(adj) = .913						

Notes: Dependent variable is BMI based on measured height and weight

+ reference group is ages 18-24 years

^{t} reference group is less than a high school education

* $p<0.05$

Table 2:
Mean values of weight, height and BMI for measured, reported and corrected data, generated from split-sample B

	Sample size	Measured	Reported	Corrected			
				Model 1	Model 2	Model 3	Model 4
Mean height (cm)							
Males	942	175.21	176.35*	175.42	...	175.44	...
Females	1087	161.71	162.28*	161.73	...	161.73	...
Mean weight (kg)							
Males	947	83.24	81.44*	83.26	...	83.27	...
Females	1080	66.91	64.47*	66.76	...	66.75	...
Mean BMI (kg/m^2)							
Males	949	27.09	26.12*	27.00	27.05	26.98	27.03
Females	1080	25.73	24.55*	25.60	25.69	25.58*	25.68

Notes: * significantly different from measured estimate ($p<0.05$)
...Not applicable

Table 3:

Percentage of the population in different weight categories based on reported, measured and corrected data from split-sample B

BMI Category	Measured	Reported	Corrected			
			Model 1	Model 2	Model 3	Model 4
Males						
Underweight	...	...	...	...	...	...
Normal weight	32.2	43.1*	33.6	32.2	32.8	33.8
Overweight	44.0	42.5	44.1	45.6	48.0	45.7
Obese (I-III)	23.1	13.8*	21.9	21.6	18.9*	20.1
Females						
Underweight	3.1[E]	4.7*[E]	2.7[E]	1.5*[E]	2.7[E]	1.9[E]
Normal weight	46.9	58.2*	46.8	47.0	46.6	47.1
Overweight	31.1	24.6*	31.8	33.2	32.4	32.7
Obese (I-III)	18.9	12.5*	18.7	18.3	18.2	18.3

Notes: [E] Coefficient of variation between 16.6% and 33.3% (interpret with caution)

... Coefficient of variation > 33.3%; cannot be reliably estimated.

* Significantly different from measured estimate ($p<0.05$)

Table 4:

Sensitivity and specificity values for reported and corrected data

Sensitivity % true positives (95% CI)		**underweight**	**normal weight**	**overweight**	**obese**	**overall**
Males						
	Reported	...	93.9 (91.7-96.2)	71.1 (66.2-76.0)	58.7 (51.7-65.7)	75.0 (72.0-78.0)
	Model 1	...	87.8 (83.6-91.4)	79.8 (73.2-86.4)	76.0 (67.1-84.9)	81.2 (77.3-85.1)
	Model 2	...	85.5 (80.8-90.1)	81.1 (74.5-87.7)	74.6 (65.5-83.8)	80.7 (76.7-84.7)
	Model 3	...	83.8 (77.6-90.1)	82.8 (76.5-89.0)	70.2 (59.9-80.4)	79.9 (75.6-84.1)
	Model 4	...	85.8 (79.7-91.9)	81.1 (74.7-87.6)	73.8 (64.8-82.8)	80.7 (76.6-84.7)
Females						
	Reported	77.8 (63.2-92.3)	91.8 (88.9-94.8)	62.6 (56.8-68.5)	68.5 (62.3-74.8)	77.8 (74.9-80.7)
	Model 1	66.8 (42.6-91.2)[E]	85.1 (79.0-91.3)	74.3 (67.7-81.0)	86.1 (78.7-93.5)	81.4 (77.4-85.3)
	Model 2	39.3 (18.3-60.4)[E]	83.4 (77.2-89.6)	75.0 (68.4-81.5)	85.1 (77.5-92.7)	79.7 (75.7-83.8)
	Model 3	66.8 (43.2-90.3)[E]	85.6 (79.6-91.7)	77.1 (70.7-83.6)	85.4 (78.0-92.8)	82.4 (78.5-86.2)
	Model 4	45.6 (24.0-67.5)[E]	86.9 (81.4-92.4)	79.7 (73.9-85.5)	86.0 (78.6-93.4)	83.2 (79.6-86.8)
Specificity % true negatives (95% CI)						
Males	Reported	99.6 (99.4-99.9)	83.2 (80.2-86.1)	79.7 (76.3-83.2)	98.3 (96.6-99.7)	
	Model 1	99.8 (99.5-100)	92.2 (88.7-95.6)	84.1 (79.8-88.3)	94.4 (91.5-97.2)	
	Model 2	99.7 (99.3-100)	93.2 (89.8-96.6)	82.3 (77.8-86.7)	94.3 (91.4-97.2)	
	Model 3	99.9 (99.7-100)	91.6 (88.0-95.1)	79.3 (74.0-84.7)	96.5 (94.3-98.7)	
	Model 4	99.8 (99.5-100)	91.0 (87.3-94.7)	82.2 (77.4-87.0)	96.1 (93.7-98.4)	
Females	Reported	97.7 (96.8 -98.5)	78.3 (74.6- 82.0)	88.9 (86.5-91.2)	99.6 (99.3-99.8)	
	Model 1	99.4 (98.8-99.9)	87.1 (82.9-91.3)	87.4 (82.9-91.9)	97.0 (95.5-98.4)	
	Model 2	99.7 (99.4-100)	85.2 (80.8-89.7)	85.6 (81.0-90.2)	97.3 (96.0-98.6)	
	Model 3	99.3 (98.8-99.9)	87.9 (84.0-91.8)	87.7 (83.3-92.2)	97.4 (96.1-98.7)	
	Model 4	99.6 (99.1-100)	88.0 (84.1-91.9)	88.4 (84.3-92.5)	97.5 (96.3-98.8)	

Notes: The reported estimates are based on data from split-samples A and B; modelled estimates are generated from split-sample B; [E] Coefficient of variation between 16.6% and 33.3% (interpret with caution); ... Coefficient of variation > 33.3%; cannot be reliably estimated.

Table 5:

Adjusted odds ratios relating measured, self-reported and corrected Body Mass Index (BMI) to selected health conditions, household population aged 40 years or older

BMI category (range kg/m²)	Based on measured values		Based on self-reported values		Based on corrected values Model 1 - Full Height and Weight		Based on corrected values Model 2 - Full BMI		Based on corrected values Model 3 - Reduced Height and Weight		Based on corrected values Model 4 - Reduced BMI	
	Adjusted odds ratio	95% confidence interval	Adjusted odds ratio	95% confidence interval	Adjusted odds ratio	95% confidence interval	Adjusted odds ratio	95% confidence interval	Adjusted odds ratio	95% confidence interval	Adjusted odds ratio	95% confidence interval
Diabetes												
Normal weight (18.5 to 24.9)	1.0	...	1.0	...	1.0	...	1.0	...	1.0	...	1.0	...
Overweight (25.0 to 29.9)	1.4	0.7 to 2.8	2.6 *	1.5 to 4.3	1.8	0.9 to 3.3	2.0 *	1.1 to 3.8	1.8 *	1.1 to 3.0	2.0 *	1.2 to 3.3
Obese Class I (30.0 to 34.9)	2.2 *	1.0 to 4.5	3.2 *	1.8 to 5.6	3.3 *	1.8 to 6.0	3.9 *	2.1 to 7.0	3.1 *	1.7 to 5.7	3.2 *	1.8 to 5.8
Obese Class II & III (>= 35.0)	5.9 *	2.5 to 14.0	9.0 *	4.5 to 17.9	6.8 *	3.7 to 12.5	7.3 *	3.9 to 13.9	7.6 *	4.0 to 14.2	7.4 *	4.0 to 13.7
High blood pressure												
Normal weight (18.5 to 24.9)	1.0	...	1.0	...	1.0	...	1.0	...	1.0	...	1.0	...
Overweight (25.0 to 29.9)	2.1 *	1.5 to 3.0	2.7 *	1.9 to 3.8	2.3 *	1.6 to 3.2	2.5 *	1.8 to 3.5	2.5 *	1.7 to 3.5	2.4 *	1.7 to 3.3
Obese Class I (30.0 to 34.9)	3.4 *	2.3 to 5.2	4.2 *	2.9 to 6.3	4.0 *	2.8 to 5.9	4.1 *	2.8 to 6.0	4.5 *	3.0 to 6.6	4.7 *	3.2 to 7.0
Obese Class II & III (>= 35.0)	5.2 *	2.9 to 9.3	6.8 *	3.2 to 14.8	6.0 *	3.3 to 10.7	6.0 *	3.4 to 10.5	6.1 *	3.4 to 10.9	5.6 *	3.2 to 9.8
Heart disease												
Normal weight (18.5 to 24.9)	1.0	...	1.0	...	1.0	...	1.0	...	1.0	...	1.0	...
Overweight (25.0 to 29.9)	1.0	0.6 to 1.7	1.4	0.9 to 2.3	1.3	0.8 to 2.2	1.3	0.8 to 2.2	1.2	0.7 to 2.0	1.4	0.8 to 2.2
Obese Class I (30.0 to 34.9)	1.5	0.8 to 2.9	1.6	1.0 to 2.6	1.2	0.7 to 2.0	1.4	0.8 to 2.4	1.3	0.8 to 2.2	1.5	0.9 to 2.5
Obese Class II & III (>= 35.0)	2.1	1.0 to 4.4	3.7 *	1.8 to 7.7	3.3 *	1.8 to 6.2	3.4 *	1.8 to 6.5	2.9 *	1.5 to 5.6	2.8 *	1.5 to 5.5
Arthritis												
Normal weight (18.5 to 24.9)	1.0	...	1.0	...	1.0	...	1.0	...	1.0	...	1.0	...
Overweight (25.0 to 29.9)	1.2	0.8 to 1.7	1.2	0.8 to 1.7	1.5 *	1.1 to 2.0	1.5 *	1.1 to 2.1	1.5 *	1.1 to 2.0	1.4 *	1.0 to 1.9
Obese Class I (30.0 to 34.9)	1.2	0.8 to 1.8	2.0 *	1.3 to 3.0	1.7 *	1.2 to 2.5	1.9 *	1.3 to 2.8	1.9 *	1.3 to 2.8	1.7 *	1.2 to 2.5
Obese Class II & III (>= 35.0)	2.7 *	1.6 to 4.6	3.1 *	1.5 to 6.3	3.5 *	2.0 to 5.8	3.2 *	1.8 to 5.4	3.2 *	1.9 to 5.6	3.4 *	1.9 to 6.0
Activity limitation												
Normal weight (18.5 to 24.9)	1.0	...	1.0	...	1.0	...	1.0	...	1.0	...	1.0	...
Overweight (25.0 to 29.9)	1.2	0.9 to 1.6	1.2	0.9 to 1.6	1.2	0.9 to 1.7	1.1	0.8 to 1.5	1.0	0.8 to 1.4	1.1	0.8 to 1.5
Obese Class I (30.0 to 34.9)	1.5 *	1.1 to 2.2	2.0 *	1.3 to 3.0	1.4	0.9 to 2.0	1.4	1.0 to 2.1	1.5 *	1.0 to 2.1	1.5 *	1.0 to 2.2
Obese Class II & III (>= 35.0)	2.9 *	1.7 to 4.7	4.3 *	2.2 to 8.2	4.2 *	2.6 to 6.8	3.7 *	2.3 to 6.1	3.9 *	2.4 to 6.5	3.1 *	1.8 to 5.2
Fair/poor self-perceived health												
Normal weight (18.5 to 24.9)	1.0	...	1.0	...	1.0	...	1.0	...	1.0	...	1.0	...
Overweight (25.0 to 29.9)	0.8	0.5 to 1.2	1.3	0.9 to 2.0	1.1	0.7 to 1.6	1.1	0.8 to 1.7	1.0	0.7 to 1.5	1.0	0.7 to 1.5
Obese Class I (30.0 to 34.9)	1.7 *	1.0 to 2.7	2.8 *	1.8 to 4.3	1.6 *	1.0 to 2.5	1.7 *	1.1 to 2.7	2.1 *	1.3 to 3.3	2.1 *	1.4 to 3.3
Obese Class II & III (>= 35.0)	2.9 *	1.6 to 5.2	4.5 *	2.0 to 10.2	4.1 *	2.4 to 7.0	4.3 *	2.4 to 7.8	3.5 *	1.9 to 6.5	3.6 *	2.0 to 6.6

Note: Models control for age (continuous) and sex. Odds ratios for underweight group not reported due to low sample sizes.

* Significantly different from estimate for normal weight category ($p < 0.05$)

... Not applicable

Table 6:

Percentage of the population in different weight categories when corrected estimates were applied to the full 2005 Canadian Community Health Survey Sample

BMI Category	Measured*	Reported*	Corrected			
			Model 1 (Full) Height and Weight	Model 2 (Full) BMI	Model 3 (Reduced) Height and Weight	Model 4 (Reduced) BMI
Males						
Underweight	0.9[E]	0.7[E]	1.0	1.2	0.9	1.0
Normal weight	32.4	41.8	31.2	30.0	31.2	33.1
Overweight	41.1	41.2	44.3	45.4	44.9	42.9
Obese (I-III)	25.6	16.3	23.5	23.4	23.0	23.1
Females						
Underweight	2.6[E]	4.3	2.7	2.1	2.5	2.0
Normal weight	46.1	54.0	46.5	46.4	46.8	46.7
Overweight	29.1	26.1	29.9	30.6	30.1	30.6
Obese (I-III)	22.3	15.7	20.9	21.0	20.7	20.8

Notes: [E] Coefficient of variation between 16.6% and 33.3% (interpret with caution)

* Measured and reported values were generated based on the full sample of respondents who had their height and weight measured.

Paper 5: The bias in self-reported obesity from 1976 to 2005, a Canada - U.S. comparison

The bias in self-reported obesity from 1976 to 2005, a Canada - U.S. comparison

Sarah Connor Gorber[1] and Mark Tremblay[2,3]

1. Health Information and Research Division, Statistics Canada
2. Healthy Active Living and Obesity Research Group, Children's Hospital of Eastern Ontario Research Institute
3. Department of Pediatrics, Faculty of Medicine, University of Ottawa

Correspondence: Sarah Connor Gorber, Chief – Health Measures Analysis Section, Health Information and Research Division, Statistics Canada, 24th Floor R.H. Coats Building, 100 Tunney's Pasture Driveway, Ottawa, ON, Canada, K1A 0T6

Tel:
Fax:
Email:

Abstract

Objectives: To determine if the bias in self-reported estimates of obesity has changed over time and followed different trends in Canada and the United States.

Methods: Using age standardized data from 3 waves of the National Health and Nutrition Examination Survey in the United States and the Canadian Community Health Survey and the Canadian Heart Health Surveys in Canada, discrepancies were compared between reported and measured estimates of height, weight and obesity (based on the BMI) from 1976 to 2005.

Results: Obesity increased in both countries but rates were higher in the United States. The discrepancy between self-reported and measured obesity was small in the United States, with reported data underestimating measured prevalence by about 3%; this stayed relatively constant over time. In Canada the discrepancy was larger and doubled in the last decade (from 4% to 8%).

Conclusions: In Canada, monitoring trends in obesity based solely on self-reported height and weight may produce inaccurate estimates because of the increasing discrepancy between self-reported and measured data.

Introduction

Obesity is an important public health issue, with well documented links between excess body weight and diseases such as cardiovascular disease, diabetes, high blood pressure and osteoarthritis [1,2]. In Canada, 59% of adults are overweight or obese[3] and in the United States, 66%[4]. Over a quarter of children and youth in the U.S. and Canada are considered overweight or obese[5] with increasing numbers of these children being under the age 5 years. If current trends continue, the number of obese people in the world is expected to rise to over 1 billion by 2030[6].

Self-reported obesity, often assessed using the Body Mass Index (BMI), is commonly used for surveillance purposes because reported height and weight can readily be collected on national surveys at a fraction of the cost of directly measuring respondents. Research has found, however, that self-reports overestimate height and underestimate weight, which leads to an underestimation of the BMI and consequently of obesity prevalence[7,8]. The consequences of this under-reporting are substantial, not only in terms of underestimating the burden of obesity, but also in distorting our understanding of the relationship between obesity and obesity-related diseases. For instance, recent research has shown that when estimates of obesity are based on self-reported data, the relationship between obesity and obesity-related conditions such as heart disease, diabetes and high blood pressure are exaggerated when compared to estimates based on directly measured values[9-11]. Correction equations that statistically adjust self-reported estimates of obesity to more closely approximate their measured values have been developed with some success[12-15] but the

generalizability of these equations depends on the stability of the self-reporting bias over time and across populations.

Little is known about how the discrepancy between reported and measured obesity has changed over time, but there is some evidence to suspect that the discrepancy is growing. First, there has been a dramatic increase in the prevalence of obesity in recent years[4,16,17] and studies have consistently shown that heavier individuals are more likely to under-report their weight[14,18,19]. Second, in this same period of time, the awareness of obesity and its health consequences has increased[2,16], as has the pressure to be thin[20]. A recent study reported that the prevalence of weight discrimination is also on the rise[21], all of which could lead individuals to feel more pressure to present themselves in a socially desirable manner; it has been suggested that a change in awareness and attention to issues such as obesity could affect the way individuals respond to self-report questions[22]. If the bias is constant or changing systematically over time, then self-reported estimates may still be valuable for monitoring trends and could be statistically adjusted to increase their accuracy (i.e., to more closely approximate measured values). However, if there is no systematic trend in the relationship between reported and measured obesity over time, it may limit the usefulness of reported estimates for monitoring population health, informing policy and program development and understanding the relationships between obesity and disease.

To our knowledge this study presents the first historical and international comparison of the trends in the bias in self-reported obesity in Canada and the United States. Specifically, we examined the discrepancy between reported and measured height, weight and BMI, which

were used to calculate reported and measured obesity, over approximately 30 years between 1976 and 2005. Both Canadian and American data were used to determine if the bias changed and whether the rate of change was consistent in both countries.

Methods

Data Sources

This analysis used national Canadian and American data sets that contained self-reported estimates of height and weight, and also direct measurements of these. For the United States, this included the National Health and Nutrition Examination Surveys (NHANES) from 1976-1980 (NHANES II), 1988-1994 (NHANES III) and the NHANES 2003-2004. In Canada only 2 data sets were available with both measured and reported data: the Canadian Heart Health Surveys (HHS; 1986-1992) and the Canadian Community Health Survey (CCHS; 2005).

NHANES is a "program of studies designed to assess the health and nutritional status of adults and children in the United States."[23(p1)] The surveys began in 1971 and in 1999 they became continuous. Each contains a nationally representative sample of the U.S. non-institutionalized civilian population, collected through a stratified multistage probability cluster sampling design. NHANES combines a household interview with physical examinations undertaken in mobile clinics. Reported height and weight were collected at the home, with the measured values collected in the clinic a few days to several weeks later (see Table 1).

The Canadian Heart Health Surveys were a series of 10 provincial surveys conducted between 1986 and 1992 as part of the Canadian Heart Health Initiative to provide information about cardiovascular disease and its risk factors in Canada. Data on measured and reported height and weight were only collected in 8 provinces (Saskatchewan and Nova Scotia did not collect this information); individuals living in military institutions or Indian reserves were also excluded. Respondents were sampled through a stratified multi-stage probability sampling design. Self-reported information was collected at the home and respondents were measured at a subsequent clinic visit.

The CCHS is designed to provide timely cross-sectional estimates of health determinants, health status and health system utilization in Canada. The CCHS samples its population through a multi-stage cluster sampling technique that is representative of over 98% of the Canadian population (individuals living on Indian Reserves or Crown lands, residents of institutions, Canadian Armed Forces and certain remote regions are excluded). Reported and measured height and weight data were collected as part of a household interview.

The characteristics of each survey are described in Table 1. In both countries, the sample sizes were highest in the earlier surveys and fell substantially in the most recent years. The sample sizes varied from a low of 4,163 to a high of 15,488. Calibrated equipment was used to take the measures in all cases. In all surveys except the 2005 CCHS, the measures were taken in a clinic up to several weeks after the self-reported responses were recorded, while in the CCHS the measures were taken in the home immediately following the interview. The

self-reported height question was similar in all surveys but only the American surveys specified that weight was to be reported without shoes.

Study Population

The study is limited to adults aged 18 to 74 years. Children were excluded as they are still in a stage of development where weight and height may change over short periods of time. It has also been suggested that reporting error in children and adolescents may be of a different nature than that in adults[14]. Women who were pregnant at the time of the survey or those who did not have both self-reported and measured data available were also excluded. Participants who completed a home visit, rather than attending the clinic, a feature offered by the 1988-94 NHANES (n=138) were also excluded to ensure that measurement conditions remained consistent. In the NHANES III and 2003-04 files, biologically implausible values were deleted and in the HHS values for the extreme 0.5% of respondents were capped (assigned the previous highest value). Data that were imputed (n=20 in NHANES II and n=510 in NHANES III) were also excluded to ensure that any observed discrepancies between reported and measured values were not influenced by the accuracy of imputation.

Data were appropriately weighted, and to adjust for the complex survey designs, variances were estimated with a Taylor linearization method (NHANES), the bootstrap technique (CCHS), while for the 1986-92 HHS, the formula for simple random sampling was used to estimate standard errors with the incorporation of a design effect (set at 1.5).
All data were age standardized (using direct standardization) with the 2001 Canadian Census as the standard population to allow for comparison among the surveys and over time. Body

Mass Index (BMI) was calculated as weight in kilograms divided by height in meters squared (data were converted to metric units as required). WHO Guidelines were used to assign BMI cut points to identify overweight (BMI 25.0-29.9 kg/m^2) and obese (BMI 30.0 kg/m^2or over) respondents[24]. All analyses were conducted with SAS (version 9.1) and Sudaan (version 9.0).

Results

In the United States, mean measured weight increased from 72 kg in 1976-80 to 82 kg in 2003-04, with similar patterns observed in males and females (see Table 2). In Canada, mean measured weight was 73 kg in 1986-92 and rose to 77 kg in 2005, with similar increases experienced for males and females. However, the contrast between the self-reported and measured weights (reported – measured) was pronounced when comparing the two countries. In the U.S., the difference between reported and measured weight was –0.4 kg in 1976-80, rose slightly at the second time point (–0.7 kg) before falling back to –0.6 kg in 2003-04. This under-reporting was driven entirely by females, as men over- rather than under-estimated their weight. In Canada, the reported – measured differences were greater; they rose from –2 kg in 1986-92 to –2.3 kg in 2005. Women in Canada underestimated by a greater degree than males, but both Canadian males and females underestimated their weight.

Males and females in both countries consistently overestimated their height, to a greater degree for males. For American males this overestimation appeared to be decreasing slightly over time, but in females the pattern was not consistent, with the overestimation decreasing

from the NHANES II to NHANES III and then increasing again in 2003-04. In Canada, the overestimation was increasing between the two time points in both males and females.

The mean BMI rose in both countries; from 25 kg/m^2 to 28 kg/m^2 in the U.S. and from 25 kg/m^2 to 27 kg/m^2 in Canada. The underestimation in weight and overestimation in height resulted in an underestimation of BMI in both countries and sexes, but this was greater in Canada. The underestimation of BMI is also increasing in Canada over time, rising from an underestimate of 0.8 kg/m^2 in the HHS to 1.1 kg/m^2 in the CCHS. The discrepancy was slightly larger in females. In the United States the discrepancy between reported and measured BMI stayed relatively consistent over time.

Figure 1a illustrates trends in obesity over time, based on measured and self-reported data in the two countries. In the United States the prevalence of measured obesity more than doubled from 1976-80 to 2003-2004, from 15% to 32%, with the reported prevalence rising from 12% to 29% in the same time period. The discrepancy between the reported and measured prevalence (about a 3% difference) has stayed relatively constant over time. In Canada, the rise in the prevalence of obesity from the late 80s and early 90s until 2005 has been of a similar magnitude as the increase in the U.S. from the same time period, but the absolute prevalence remains lower in Canada (rising from 14% to 24%). Despite the lower prevalence of obesity, the discrepancy between the reported and measured values appears to have doubled in Canada in this time, going from a difference of 4% in the late 1980s to a difference of 8% in 2005. Similar patterns were seen in males and females in both countries (data not shown).

The percentage of the population who were considered overweight (not including obesity) based on measured data has changed little in either country, ranging from 32-33% in the U.S. and 34-35% in Canada, whereas the reported prevalence of overweight has increased slightly (see Figure 1b). In the U.S. though, the prevalence of overweight was under-reported in NHANES II by 1.4%, and then in NHANES III and 2003-04 was over-reported (by 2.1% and 0.9% respectively). In Canada the prevalence of overweight was under-reported in both surveys but the degree of under-reporting decreased between then two surveys, from –3.5% to –1.2%. In Canada this decrease was driven almost entirely by the males whose reporting bias went from –5% to almost no difference, whereas the women stayed consistent around –2%.

The reporting bias (reported – measured) in the prevalence of obesity by age is displayed in Figure 2. In Canada the bias was always smaller in the HHS compared to the more recent CCHS, suggesting that the bias across all age groups may be increasing over time in Canada. In the NHANES surveys the bias was consistently greater in all age groups in NHANES III compared to NHANES II. The NHANES 2003-04 showed the greatest bias at the younger age ranges which decreased with increasing age, with the lowest bias seen for those aged 45-74 years. This was the only survey in which the bias was not greatest in the highest age groups. The discrepancy at the oldest ages (65-74 years) ranged from a reported – measured difference in obesity prevalence of –4% (NHANES 2003-04) to –17% (CCHS 2005). The reporting bias in the prevalence of overweight showed no clear pattern (data not shown).

Discussion

Consistent with past research[7], this study found that respondents in national health surveys in both Canada and the United States have a tendency to under-estimate their weight and exaggerate their height, which results in an underestimation of BMI when it is based on these values. This trend has remained consistent since 1976 in the United States and since 1986 in Canada, with one exception: American males did not underestimate their weight in any of the NHANES cycles that we examined. The discrepancy in reported weight appears to be driven mainly by an underestimation of weight in females, whereas the overestimation of height in males seems to be the main driver of the height discrepancy.

In the last 30 years average weights and BMIs have increased substantially in both countries, although both the average BMI and the prevalence of obesity was higher in the U.S at all time points, with Canada seeming to lie approximately a decade behind the U.S. in terms of obesity prevalence. In both countries, the reservoir of overweight people appears to be consistently replenished as many graduate to the obese category. At any given prevalence value, however, the discrepancy between the reported and measured estimates was greater in Canada. For instance in NHANES III (1988-94) , the prevalence of obesity was 23%, corresponding to the Canadian prevalence of 24% from the 2005 CCHS, yet in NHANES III the bias was only 4% and in the CCHS it was twice this at 8%. This indicates that the reporting bias may be changing differently in the two countries. This is an interesting observation which begs further research to investigate the source of this difference and the implications of such differences for interpreting trends based on self-reported height and weight.

Over the 30 years studied in this paper, the bias in reported BMI and obesity has stayed relatively constant in the U.S., whereas in Canada this bias has increased substantially. Therefore, it is likely that in the United States, self-reported data may be more accurate in monitoring trends in obesity over time as the estimates have consistently remained about 3-4% below the measured estimates. More caution is required, however, when examining trends in self-reported data for specific age groups, as the bias tends to be greater for those in the older age ranges.

Although a third data point is required in order to establish a trend in Canada, the increasing discrepancy from the late 1980s to 2005 indicates that the reporting bias may not be constant over time. This makes it difficult to interpret self-reported data in the absence of the measured estimates as it introduces an underestimate, the magnitude of which is difficult to quantify, and which appears to be changing over time. In addition, the age-based discrepancies, especially in the 2005 CCHS are substantial, with the older age groups underestimating their obesity prevalence by approximately 17 percentage points. This appears to be due to both the over-reporting of height (mostly in males) and the under-reporting of weight (in females). We would recommend caution particularly when researchers are using self-reported obesity estimates to quantify obesity prevalence in the older age groups.

The bias in reported obesity is not a new phenomenon, and researchers have attempted to correct for this bias in many ways. One of the most common methods has been to use

regression models to adjust self-reported data based on measured values[13]. Recently this has been successfully attempted with the 2005 CCHS data[15], but the change in the bias over time in Canada suggests that formulas derived from one data set may not be applicable to Canadian data collected at other time points. Visscher's work has also concluded that correction equations may only be applicable to the datasets in which they are generated[25]. This means that in order to generate appropriate adjustments to self-report data, a sub-sample of data with both measured and reported information would be required in all data sets. This may not be feasible in small-scale or telephone surveys.

The smaller bias in the U.S. data is an unexpected finding given the higher prevalence of obesity among the Americans and the fact that the bias in reported obesity has been shown to be greater as BMI increases[14,18,19]. Other factors such as age, education, ethnicity, and activity levels[8,14,18,26] have also been associated with increased discrepancies, but to date there is no research that we are aware of that has compared the discrepancies or their determinants between Canada and the U.S. Further research is required to understand the reasons for the discrepancies between the two countries.

It is possible that cultural differences between the two countries are responsible for the differences, but little evidence exists to help quantify what these differences are or how they operate. There is some evidence within both countries that indicates that ethnic differences can influence perceptions of one's weight. For instance, within Canada foreign-born women are more likely to perceive themselves as overweight than their Canadian-born

counterparts[27] while in the U.S. Caucasians are more likely to perceive themselves as overweight than African or Mexican Americans[28].

It is also conceivable that with rising obesity rates, weight perceptions and individual views of socially acceptable weight may be changing. With approximately two thirds of Americans being overweight or obese, those with BMI's greater than 25 kg/m^2 are becoming the new norm as they now outnumber those in the normal weight range. This "normalizing" of overweight is not without consequences and research is showing that it is affecting people's perceptions of their weight status, which in turn could be influencing how they respond to national surveys. Although we may have expected an increase in the proportion of Americans who perceive themselves as being overweight corresponding with the rise in obesity rates, research with the NHANES data has actually shown that the number of overweight people who perceive themselves as overweight is declining as obesity rates rise[29]. In addition, a 22 country comparison found that women in the United States and the United Kingdom, areas where weight has increased dramatically in recent years, were less concerned about their weight than women from Asian countries such as Japan, where women have actually been getting slimmer over time[30]. A Finnish study has further demonstrated that this trend seems to begin in adolescence, with its finding that between 1979 and 1999 the proportion of adolescents perceiving themselves as overweight had decreased[31]. Although it is counterintuitive that the perception of being overweight has decreased while actual overweight and the idealization of thinness have increased, researchers postulate that it is the shift in our perception of overweight and its normalization

that is changing as we increasingly compare ourselves to our peers, who are themselves heavier, rather than to idealized media images[31].

The implications of these changes both for measuring obesity and addressing the epidemic are not clear, but it does offer one potential explanation for why we are not seeing an increase in the discrepancy between reported and measured estimates in the U.S. More work is needed to determine if this change in overweight perception is related to the way people report data on self-reported questionnaires and whether this is also occurring in the Canadian population. Should Canadian obesity levels continue to climb, it will be interesting over time to see if the reporting bias in Canada decreases as our obesity rates reach the levels in the United States.

There are many limitations inherent in any international or historical research. Surveys have different designs, measurement protocols and methodologies which could lead to small differences in results. Nevertheless, we believe that the use of standardized nationally representative data, with similar analytic designs improves the accuracy of our findings. In addition, a minimum of three time points of data are preferable for establishing trends and patterns over time. Further research will be required when more Canadian data become available to determine if the discrepancy between reported and measured values continues to diverge.

Data for all five surveys were collected through in-person interviews. Telephone interviews have shown a greater bias in self-reported obesity than in-person interviewing[32,33] and the

degree to which the current findings apply to data collected by phone is not yet known. We also know that reporting bias can be increased if respondents are unaware that they will subsequently be measured[32]. In all surveys except the 2005 CCHS we could reasonably suspect that participants may have had an indication that they would be measured, as they were participating in an in-depth health survey that required a follow-up visit to a clinic for physical testing. In the CCHS, a household survey, a sub-sample of the population was selected to have their height and weight measured at the end of the survey; this was the only physical measures data that were collected and no indication was given that the measurements would occur. It is therefore possible that the increase in bias seen in the CCHS may be due in part to the fact that participants were unaware that their self-reported responses would be validated.

To our knowledge this paper is the first to undertake an international comparison of the bias in reported obesity and how it has changed over time. We have found indications of a consistent reporting bias in the U.S. and of an increasing discrepancy in Canada, indicating that self-reports could be more accurately used in the U.S. to monitor trends over time. More research is required to provide insight into the determinants of the bias in the two countries, and possible data transformation procedures to adjust self-reported data to better reflect objective measures.

Acknowledgements

The authors would like to thank Georgia Roberts for her methodological advice and assistance in variance estimation for the historical surveys.

References

1. Health Canada. *Canadian Guidelines for Body Weight Classification in Adults.* Ottawa, Ontario: Health Canada; 2003.
2. US National Institutes of Health. *Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults: The Evidence Report.* Bethseda, MD: National Institutes of Health; 1998.
3. Tjepkema M. Adult obesity. *Health Rep.* 2006;17(3):9-25.
4. Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999-2004. *JAMA.* 2006;295:1549-1555.
5. Shields M. Overweight and obesity among children and youth. *Health Rep.* 2006;17(3):27-42.
6. Kelly T, Yang W, Chen CS, Reynolds K, He J. Global burden of obesity in 2005 and projections to 2030. *Int J Obes.* 2008;[Epub ahead of print]
7. Connor Gorber S, Tremblay M, Moher D, Gorber B. A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review. *Obes Rev.* 2007;8:307-326.
8. Engstrom JL, Paterson SA, Doherty A, Trabulsi M, Speer KL. Accuracy of self-reported height and weight in women: an integrative review of the literature. *J Midwifery Women's Health.* 2003;48:338-345.
9. Yannakoulia M, Panagiotakos DB, Pitsavos C, Stefanadis C. Correlates of BMI misreporting among apparently healthy individuals: the ATTICA study. *Obesity.* 2006;14:894-901.

10. Chiolero A, Peytremann-Bridevaux I, Paccaud F. Associations between obesity and health conditions may be overestimated if self-reported body mass index is used. *Obesity Rev.* 2007;8:373-374.

11. Shields M, Connor Gorber S, Tremblay MS. Effects of measurement on obesity and morbidity. *Health Rep.* 2008;19(2):77-84.

12. Bolton-Smith C, Woodward M, Tunstall-Pedoe H, Morrison C. Accuracy of the estimated prevalence of obesity from self reported height and weight in an adult Scottish population. *J Epidemiol Community Health.* 2000;54:143-148.

13. Kuskowska-Wolk A, Bergstrom R, Bostrom G. Relationship between questionnaire data and medical records of height, weight and body mass index. *Int J Obes.* 1992;16:1-9.

14. Rowland ML. Self-reported weight and height. *Am J Clin Nutr.* 1990;52:1125-1133.

15. Connor Gorber S, Shields M, Tremblay MS, McDowell I. The feasibility of establishing correction factors to adjust self-reported estimates of obesity in the Canadian Community Health Survey. *Health Rep.* 2008;19(3):71-82.

16. World Health Organization. *Obesity: Preventing and Managing the Global Epidemic.* Geneva: World Health Organization; 2000.

17. Shields M, Tjepkema M. Trends in adult obesity. *Health Rep.* 2006;17(3)53-59.

18. Shields M, Connor Gorber S, Tremblay M. Estimates of obesity based on self-report versus direct measures. *Health Rep.* 2008;19(2):61-76.

19. Elgar FJ, Roberts C, Tudor-Smith C, Moore L. Validity of self-reported height and weight and predictors of bias in adolescents. *J Adolesc Health.* 2005;37:371-375.

20. Katzmarzyk PT, Davis C. Thinness and body shape of Playboy centrefolds from 1978 to 1998. *Int J Obes.* 2001;25:590-592.

21. Andreyeva T, Puhl RM, Brownell KD. Changes in perceived weight discrimination among Americans, 1995-1996 through 2004-2006. *Obesity*. 2008;16:1129-1134.

22. Tremblay M. The need for directly measured health data in Canada. *Can J Public Health*. 2004;95:165-168.

23. Centers for Disease Control and Prevention, National Center for Health Statistics. *National Health and Nutrition Examination Survey, 2005-2006 Overview*. Hyattsville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention; 2006.

24. World Health Organization. *Physical Status: The Use of and Interpretation of Anthropometry, Report of the WHO Expert Committee*. Geneva: World Health Organization; 1995.

25. Visscher TLS, Viet AL, Kroesbergen HT, Seidell JC. Underreporting of BMI in adults and its effect on obesity prevalence estimations in the period 1998 to 2001. *Obesity*. 2006;14:2054-2063.

26. Gillum RF, Sempos CT. Ethnic variation in validity of classification of overweight and obesity using self-reported weight and height in American women and men: the third National Health and Nutrition Examination Survey. *Nutr J*. 2005;4:27.

27. Gucciardi E, Wang SC, Badiani T, Stewart DE. Beyond adolescence: exploring Canadian women and men's perception of overweight. *Womens Health Issues*. 2007;17:374-382.

28. Paeratakul S, White MA, Williamson DA, Ryan DH, Bray GA. Sex, race/ethnicity, socioeconomic status, and BMI in relation to self-perception of overweight. *Obesity Res*. 2002;10:345-350.

29. Johnson-Taylor WL, Fisher RA, Hubbard VS, Starke-Reed P, Eggers PS. The change in weight perception of weight status among the overweight: comparison of NHANES III (1988-1994) and 1999-2004 NHANES. *Int J Behav Nutr Phys Act.* 2008;5:9.

30. Wardle J, Haase AM, Steptoe A. Body image and weight control in young adults: international comparisons in university students from 22 countries. *Int J Obes.* 2006;30:644-651.

31. Kaltiala-Heino R, Kautiainen S, Virtanen SM, Rimpelä A, Rimpelä M. Has the adolescents' weight concern increased over 20 years? *Eur J Public Health.* 2003;13(1):4-10.

32. Ezzati M, Martin H, Skjold S, Vander Hoorn S, Murray CJL. Trends in national and state-level obesity in the USA after correction for self-report bias: analysis of health surveys. *J Res Soc Med.* 2006;99:250-257.

33. Béland Y, St-Pierre M. Mode effects in the Canadian Community Health Survey: a comparison of CATI and CAPI. In: Lepkowski J, Tucker C, Brick JM, et al, eds. *Advances in Telephone Survey Methodology*. New York, NY: Wiley; 2008; 297-314.

Tables and Figures

Table 1: Survey characteristics

	Survey				
	NHANES II, 1976-80	**NHANES III, 1988-94**	**NHANES, 2003-04**	**HHS, 1986-1992**	**CCHS, 2005**
Sample size (18-74 years)	12,134	13,599	4,163	15,488	3,687
Equipment used to measure height	Height scale	Stadiometer	Stadiometer (Seca)	Tape measure	Measuring tapes
Equipment used to measure weight	Self-balancing scale (Toledo Scale)	Digital scale (Toledo)	Digital scale (Toledo)	Balance beam scale (Continential Scale Works)	Scales (ProFit UC-321 Lifesource)
Measurement site	MEC	MEC	MEC	Clinic	Home
Reported height question	How tall are you without shoes?	How tall are you without shoes?	How tall are you without shoes?	How tall are you without your shoes?	How tall are you without shoes on?
Reported weight question	How much do you weigh without clothes or shoes?	How much do you weigh without clothes or shoes?	How much do you weigh without clothes or shoes?	How much do you weigh?	How much do you weigh?
Time between reported and measured	Up to several weeks	Up to several weeks	Up to several weeks	Within 2 weeks	Same session
Equipment calibrated	Yes	Yes	Yes	Yes	Yes
Special notes	Weights over 400lbs were self reported n=4; imputed measured height and weight data were excluded n=20	Imputed measured height and weight data were excluded n=510	Weights over 400 lbs were measured	Values for extreme 0.5% of subjects were capped (assigned the previous highest value)	Height data were rounded up to nearest inch

Notes: NHANES National Health and Nutrition Examination Survey; HHS Canadian Heart Health Survey; CCHS Canadian Community Health Survey; MEC Mobile Examination Centre

Table 2: Mean differences between reported and measured values

Weight (kilograms)		**Overall**			**Males**			**Females**	
American Surveys	reported	measured	difference	reported	measured	difference	reported	measured	difference
NHANES II 76-80	71.6	72.0	-0.4	79.1	78.7	0.4	64.6	65.7	-1.1
NHANES III 88-94	75.3	76.0	-0.7	82.6	82.4	0.2	68.1	69.6	-1.5
NHANES 03-04	80.9	81.5	-0.6	88.4	88.2	0.1	73.4	74.6	-1.3
Canadian Surveys									
Heart Health 86-92	70.8	72.8	-2.0	77.9	79.7	-1.8	63.6	65.8	-2.3
CCHS (2005)	74.5	76.8	-2.3	82.1	83.9	-1.8	66.5	69.3	-2.8
Height (centimetres)		**Overall**			**Males**			**Females**	
American Surveys	reported	measured	difference	reported	measured	difference	reported	measured	difference
NHANES II 76-80	169.4	168.4	0.9	176.9	175.5	1.4	162.3	161.9	0.5
NHANES III 88-94	169.9	169.1	0.8	177.1	175.9	1.2	162.8	162.4	0.3
NHANES 03-04	170.7	169.9	0.8	178.0	176.9	1.2	163.4	162.9	0.5
Canadian Surveys									
Heart Health 86-92	169.4	169.0	0.4	176.2	175.6	0.6	162.6	162.4	0.2
CCHS (2005)	169.9	169.1	0.8	176.6	175.5	1.1	162.9	162.5	0.5
BMI ($kg/m^{2)}$		**Overall**			**Males**			**Females**	
American Surveys	reported	measured	difference	reported	measured	difference	reported	measured	difference
NHANES II 76-80	24.9	25.3	-0.4	25.2	25.5	-0.3	24.6	25.1	-0.6
NHANES III 88-94	26.0	26.5	-0.5	26.3	26.6	-0.3	25.7	26.4	-0.7
NHANES 03-04	27.7	28.1	-0.5	27.8	28.2	-0.3	27.5	28.1	-0.7
Canadian Surveys									
Heart Health 86-92	24.6	25.4	-0.8	25.1	25.8	-0.7	24.1	24.9	-0.9
CCHS (2005)	25.7	26.8	-1.1	26.3	27.3	-1.0	25.1	26.3	-1.2

Notes: NHANES National Health and Nutrition Examination Survey; CCHS Canadian Community Health Survey

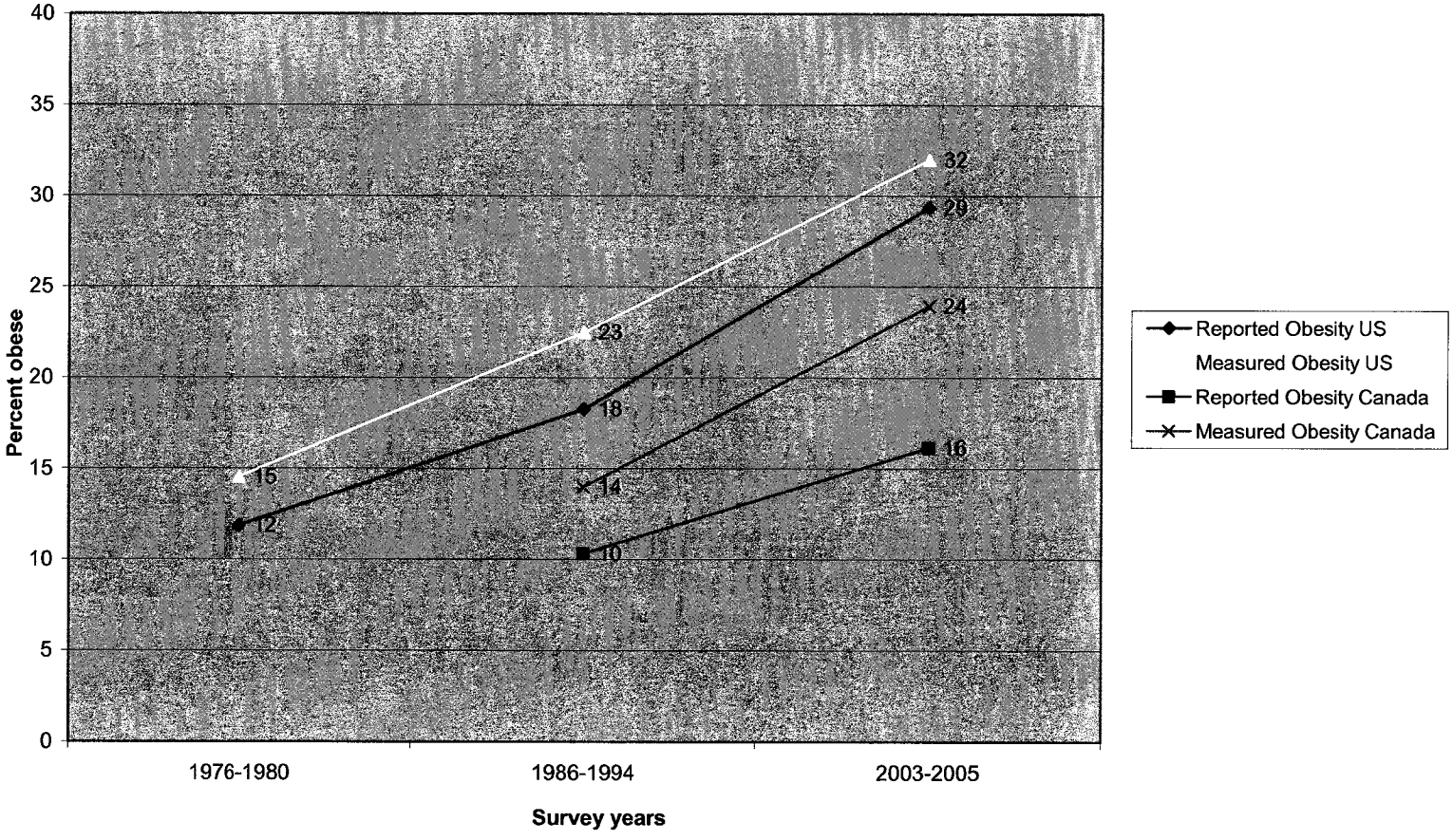

Figure 1a: Measured and reported prevalence of obesity, Canada and the United States, 1976-2005, males and females combined.
Note: Not all surveys covered the full range of years.

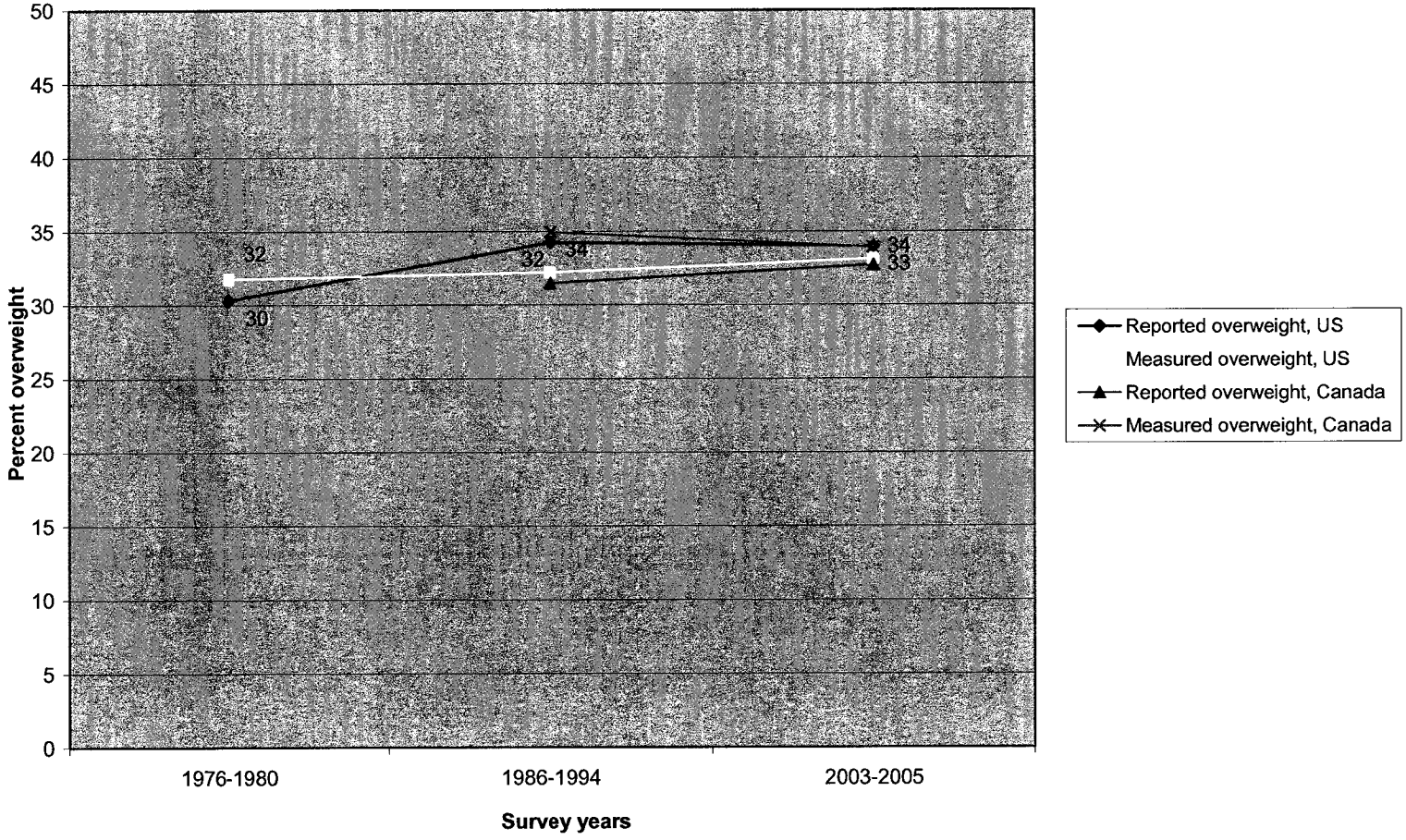

Figure 1b: Measured and reported prevalence of overweight, Canada and the United States, 1976-2005, males and females combined.
Note: Not all surveys covered the full range of years.

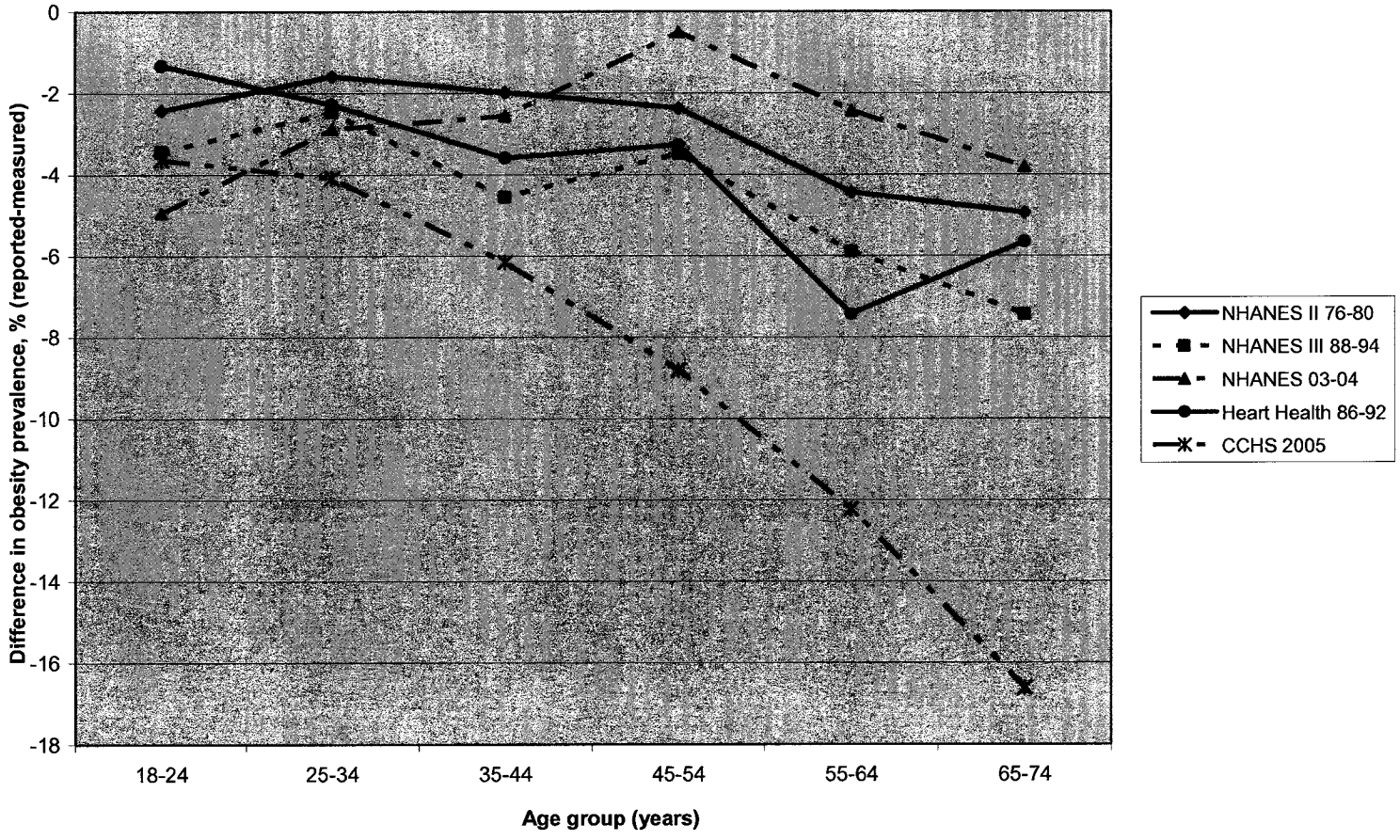

Figure 2: Bias in obesity prevalence (reported-measured) by age group, males and females, 1976-2005.

General Discussion and Conclusions

The goal of this research was to consolidate and expand on current knowledge about the use of objective and subjective measures to assess the health of Canadians. Specifically, this dissertation examined the relationship between self-report and direct measures of health to determine the extent of discrepancies between the two methods. This was accomplished through a series of five research papers, including three systematic reviews that examined the relationship between self-reported and directly measured values for height, weight and BMI, hypertension and smoking. An in-depth examination of the discrepancies in reported obesity followed. It included two additional papers, one examining the feasibility of establishing correction factors to increase the accuracy of reported estimates and a second examining the trends in the bias over time and between two different population groups (Canada and the U.S.).

The first review, which examined 64 studies, found that self-reported height was consistently overestimated while reported weight and BMI were consistently underestimated in both men and women. The second review of 144 studies (including data on over 1 million people) consistently demonstrated that the prevalence of hypertension was under-estimated when based on self-reports. Moreover, if a standard 140/90 mmHg was used to diagnose hypertension, only just over half of respondents were aware of their hypertensive status. The third review examined the relationship between self-reported smoking and cotinine concentration in biological fluids in 67 studies, and found the best correspondence between reported and measured estimates. However, sensitivity did vary depending on the cut-point

used to determine a regular smoker and the medium of measurement (e.g., the highest sensitivities were found when cotinine was measured in saliva).

A detailed analysis of the determinants of the discrepancies was beyond the scope of this dissertation. However, theories of social desirability[1,2] do provide insights into why people underestimate their weight or smoking behaviours, as socially undesirable behaviours are particularly prone to under-reporting[3-5]. Self-reported weight is particularly susceptible to prevalent cultural norms that emphasize thinness, with women who have higher tendencies to act in a socially desirable manner demonstrating greater tendencies toward reporting bias[3]. In populations where smoking is seen as highly undesirable (e.g., pregnant women or those with smoking-related illness) under-reporting has also been shown to be higher than in the general population of smokers[6,7]. Different factors are likely at play with hypertension awareness as it is a much more complex indicator to measure subjectively, being dependent on having been diagnosed with hypertension, being told of the diagnosis, and recalling and reporting it on a health survey. Lack of awareness, therefore, may be the result of not recalling or never having received a diagnosis of hypertension.

The results from all three reviews clearly advocate for a greater role of direct measures in assessing obesity, hypertension and smoking status. Surveillance based only on self-report will lead to an underestimation of the prevalence of these indicators meaning that policy makers and clinicians may not have the most accurate information from which to make appropriate decisions about resource allocation for the prevention, treatment and control of disease. In the case of hypertension, this is particularly concerning at the individual level as lack of awareness may result in a delay in intervention for a treatable condition. Finding

consistent evidence of a reporting bias in all three indicators also stresses the need for further research with new conditions such as diabetes, cholesterol and alcohol use to determine if the reporting bias extends beyond the indicators studied in this dissertation, which would be the expectation.

All three reviews found that reporting of results was poor. In many of the research papers we reviewed, vital information to allow conclusions to be verified and analyses replicated was absent. This despite the fact that international organizations[8,9] have published official recommendations about what should be reported in hypertension and smoking research, and at least one researcher has published recommendations about what should be reported in validity studies that examine obesity[10]. The current dissertation adds some additional suggestions to the official and unofficial recommendations in order to ensure that adequate data are available for meta-analysis and other review work to permit a more comprehensive understanding of the relationship between the reported and direct measures. More complete reporting would enable more detailed analysis of existing research to determine the degree to which the reporting bias changes according to certain subgroups such as age, sex, education, or ethnicity.

Standard protocols for collecting measured data exist for all three indicators that were examined: The Canadian Society for Exercise Physiology[11] has established protocols for measuring height, weight and BMI, as well as other anthropometric data; the Joint National Committee on Detection, Evaluation and Treatment of High Blood Pressure has published clinical practice guidelines that stress the importance of equipment calibration, patient

positioning, restrictions prior to measurement and multiple measures[12]; and the Society for Research on Nicotine and Tobacco published a series of recommendations outlining for instance, the most appropriate cut-points and biomarkers for determining smoking status and the acceptable time lag between self-reported smoking and biochemical verification[9]. Yet, in many of the reviewed studies standard guidelines were not being followed, or studies did not report enough information about their protocols to assess the degree to which they complied with standard recommendations. More emphasis needs to be placed on ensuring that researchers and government or other organizations involved in the development and collection of data follow recommended procedures for measuring and reporting data. This will help to ensure the accuracy of the measured data and the comparability of results with other studies; it will also allow us to examine the appropriateness of existing recommendations and procedures.

All of the studies that were included in the reviews underwent an assessment of quality and of risk of bias. Higher quality studies demonstrated higher levels of hypertension awareness and higher sensitivities in detecting smoking when blood or salivary cotinine was used to determine smoking status. As the quality assessment tools were based in part on the degree to which studies adhered to standard guidelines (with hypertension) and the accuracy and appropriateness of sampling and reporting (hypertension and smoking) this provides some evidence to indicate that adherence to standard protocols makes a difference in the quality of the data. This is important, as cost and logistical constraints often limit the feasibility of collecting directly measured data and adhering to standard protocols may be one way to

improve the quality of self-reported information. Authors, journal editors and peer reviewers all have a responsibility to ensure that protocols were followed and reporting is complete.

The constraints associated with collecting direct measures are acknowledged, which is why the second part of this dissertation focused on trying to statistically adjust the self-reported data so that it would more closely approximate the measured estimates. If reliable correction equations could be generated and if these equations were stable over time and across populations it may lessen the need to regularly collect direct measures data. Obesity was selected as the indicator on which to undertake the more detailed analysis as it was the only indicator for which national Canadian measured and reported data existed. These data were collected as part of the 2005 Canadian Community Health Survey (CCHS) in which a subsample of adults were asked to report their height and weight and subsequently had these values measured.

The analysis in the fourth paper confirmed the findings of the systematic review in which height was overestimated and weight and BMI were underestimated, but did so for the first time with Canadian data. Four models were generated to adjust the self-reported estimates. In all cases the corrected models generated mean values of height, weight and BMI that were closer to the measured values than were the self-reported estimates. Further, in almost all cases the estimates of the prevalence of obesity, overweight and normal weight were generated more accurately with the corrected models than they were with the self-reported values. Sensitivity values for classifying individuals as overweight and obese were improved with the corrected estimates, although for people in the normal weight range corrected

models decreased the sensitivity. This is because people in the normal weight ranges more accurately report their height and weight and, as such, their reported estimates correctly classify them into the normal weight category more often. Although the corrected models are not without flaws, the improvement in the overweight and obesity estimates was great enough to lead to the recommendation that researchers studying overweight or obesity in the 2005 CCHS apply the equations to correct the self-reported obesity data in subsequent analyses. Calls were also made for more research to test the generalizability of the equations over time and across population groups.

One way to determine whether these equations may be appropriate for use in other data sets was to examine whether the bias in reporting obesity was changing over time. If the bias is constant, or at least changing systematically over time, then it is more likely that a statistical adjustment may be successful and it may indicate that the self-reported data may still be valuable for monitoring trends over time. If the bias is not constant, this makes it difficult to know the degree to which the underreporting is present in any given self-reported survey, as it is ever changing.

This work led to the final paper in the series that presented the first known historical and international comparison of the trends in the bias in self-reported obesity in Canada and the U.S.

The goal of the fifth study was to determine if the trend in reporting bias was constant over time and between two culturally and socially distinct population groups. This paper analysed data from three American Surveys and two Canadian data sets between 1976 and 2005 and

found that measured weight, BMI and the prevalence of obesity have increased substantially in both countries over the last 30 years. Mean BMIs and the prevalence of obesity were higher at all time points in the U.S. than in Canada. Given the dramatic increase in the prevalence of obesity in recent years[13-15] and research that has consistently shown that heavier individuals are more likely to under-report their weight[16-18], it was expected that the reporting bias would increase in both countries. While this held true for the Canadian data, where the bias in reported obesity doubled between the 1986-1992 Canadian Heart Health Surveys and the 2005 CCHS, the American data indicated that the reported data have consistently underestimated measured prevalence by about 3% and this bias in reported obesity has stayed relatively constant over the last 30 years.

Potential explanations for this difference between the two countries were suggested, including cultural differences or a "normalizing" of obesity that is corresponding with the increase in prevalence and which in turn may be affecting people's acceptance of their own weight status. It is possible for instance, that if we judge ourselves according to our peers and our peers are themselves much heavier, then we may be less concerned about presenting ourselves in a socially desirable manner. Although this may help to explain why there has been no rise in the bias in obesity in the U.S. in recent years, it does not help to explain why the bias in Canada has continued to climb. Perhaps there is a threshold that is attained when the absolute number of obese individuals reaches a critical point and this "normalization" begins to occur. It may be that in Canada obesity rates have not yet reached this threshold, or it may be that there are fundamental differences in the way Canadians view obesity.

More research is required to fully investigate the causes of this discrepancy and ultimately to determine why the bias is changing at a different rate between the two countries and the implications of these changes for measuring obesity and addressing the epidemic. However, it was clear that the increase in the bias in Canada suggests that the correction equations generated from the 2005 CCHS would likely not be applicable to other data sets over time. It is still possible that the equations may be generalizable within a shorter time frame, such as within a few years, but more research will be required in order to determine whether this is the case.

The new Canadian Health Measures Survey (CHMS), which is Canada's first comprehensive health measures survey in almost 30 years, will be a good vehicle through which to test the applicability of these equations over a shorter time span when it releases data on measured height and weight in early 2010. Without additional work in other Canadian datasets, however, it is not appropriate to recommend the use of the correction equations outside of the 2005 CCHS. Instead, it is recommended that a sub-sample of data with both measured and reported information be collected in all data sets in order to determine the degree of bias that is present and to generate study- or survey-specific correction equations. Although this addresses some of the cost and feasibility issues associated with collecting direct measures in a full population sample, it is acknowledged that sub-sampling still may not be feasible in small clinical or telephone surveys. Researchers must be aware that without a sub-sample of measured data it will be difficult to interpret the self-reported estimates as they introduce an underestimate, the magnitude of which is difficult to quantify and which appears to be changing over time. Therefore, time trends based on self-reported data must be viewed with caution.

The work on historical comparisons and correction factors was only undertaken on obesity because it was the only measure for which recent national measured and reported data were available in Canada, yet it is likely that some of these results may be applicable to other indicators including smoking and hypertension but also diabetes, nutritional status, and even other measures of obesity such as waist circumference or the waist to hip ratio. The CHMS collects both measured and reported data on a variety of health determinants and outcomes and will provide a unique opportunity to expand these analyses.

Implications

The implications of this under-reporting are significant both in terms of underestimating the burden of obesity and other self-reported conditions but also in altering our understanding of the relationship between obesity and obesity-related diseases. For instance, recent research has shown that when estimates of obesity are based on self-reported data, the relationship between obesity and conditions such as heart disease, diabetes and high blood pressure are exaggerated when compared to estimates based on directly measured values[19-21]. Associations between obesity and food insecurity have also been shown to be more pronounced when self-reported rather than measured height and weight data are used[22].

Our work indicated that this occurs because fewer respondents are classified as overweight or obese when the classification is based on reported data, because they erroneously classify themselves into a lower weight category. Yet, the average weights of those who do self-report as being overweight or obese are higher than then average weights of those whose

measured data place them in the overweight or obese categories. As a result we observe a stronger association with morbidity when overweight and obese categories are based on self-reported data because the respondents in these categories are actually heavier[21].

This underestimation also leads to an underestimation in the disease burden of obesity. For instance, of Canadian adults aged 40 years and older who were classified as obese based on self-reported data in the 2005 CCHS 360,000 had diabetes. If, however, measured data are used to classify respondents as obese then 530,000 adults had diabetes[21].

This controversy has led several researchers to recommend using reported estimates of height and weight with caution or not at all[20, 22-24]. One of the strengths of using correction factors to adjust the reported estimates is that they are able to reduce the exaggeration in the relationship between self-reported obesity and morbidity, and thereby represent an improvement over the reported data. However, they are not able to reduce the odds ratios to the levels that are seen when the associations are studied based on measured data.

Undeniably, this research highlights the need for more work to examine the validity of self-reported estimates, but while validity studies on a variety of determinants and outcomes are required, additional work is also needed. This work has clearly documented that estimates generated from self-reported data for at least three indicators (obesity, hypertension and smoking) will be biased, yet for fiscal and logistical reasons in Canada and perhaps internationally, self-report will likely continue to be the main source of population level surveillance data for some time to come. More therefore needs to be done to try to improve

the quality of existing self-reports. As discussed above, adherence to standard guidelines and protocols for collecting not just measured but also self-reported data is important.

Consistently eliminating rounding from self-reported responses is another way to decrease some of the bias and increase the comparability of studies. The CCHS is one of the worst offenders for building bias into questionnaire responses as most respondents report their height in feet and inches, but the values are then rounded up to the closest inch and then converted to a range of metric values, with the mid point in that range being selected as the person's height. It would be much less onerous and more accurate to simply convert the unrounded estimate that the respondent provides into its corresponding metric value.

Where studies do include sub-samples with both measured and reported data they must ensure that the circumstances in which the data are being collected are kept consistent between the two methods. For instance, if clothing or shoes are not worn for the height or weight measurements, participants should be instructed to report their height or weight without shoes or clothes.

While these examples may seem trivial, they were observed regularly in the literature, and small biases at different stages in the collection process can add up to large discrepancies. In order to determine which aspects of the bias are due to the measurements and which are due to true discrepancies among respondents, researchers need to take care in the design and implementation of questionnaires and survey instruments.

In addition, it has been documented that for topics that are susceptible to social desirability that the mode of data collection can also impact on the bias[25-27], with studies showing a greater bias in reported obesity for example, for respondents who were interviewed by telephone compared to those who were interviewed in person[28,29]. Although in-person interviews carry additional costs, advantages in terms of increased accuracy of responses may make these extra costs worthwhile.

With some indicators such as hypertension, self-reported data could be improved both through improved reporting standards but also by public health efforts to increase awareness of blood pressure. As hypertension is often asymptomatic in its early phases, identification relies on adequate screening for the condition, yet studies have shown that screening may not be consistently undertaken[30,31]. Vargas and colleagues[32] found that the validity of self-reported hypertension was increased in those who had a medical appointment in the past year. Efforts to increase screening and reporting of blood pressure results to patients may help to increase the quality of reported estimates as well as to ensure early diagnosis and treatment.

Summary and Recommendations

This research has documented an international problem in how obesity, smoking and hypertension are measured, has confirmed the relevance of the findings relating to obesity for Canadians, and has suggested a potential method in which self-reported data can be corrected to more closely approximate measured values. Furthermore, the historical trends in the reporting bias have been investigated in both Canada and the United States to begin to

assess the generalizability of the correction factors approach and to document the degree of bias over time. This work has made an important contribution to understanding the bias in current surveillance practices that are anchored in the collection of self-reported data and has suggested practical solutions that can continue to move the field forward.

This research has provided a solid base from which future research can build to investigate if the relationships are generalizable to other indicators and could be reproduced with the CHMS data to determine the extent of the discrepancy between reported and measured values for a wider variety of health outcomes. One of the reviews has already sparked a series of commentaries in a well respected journal about the implications of the bias[20,33,34] and Statistics Canada is currently considering including a sub-sample of respondents from whom reported and measured height and weight data would be collected in its health surveys, this sub-sample could then be used to develop survey specific correction equations. It is hoped that this work will lead to the development of standardized tools to measure health indicators so that regional, national, and international comparability will be enhanced. Increasing comparability will only help to consolidate knowledge and build on research conducted in other areas. This will ensure that health professionals, policy makers, researchers and government officials have the most accurate evidence from which to make decisions about the population's health and well-being.

To summarize, this dissertation presents a series of recommendations that would improve the current practice of collecting self-reported data for generating population health estimates of health conditions and enhance population health surveillance in Canada.

They are as follows:

1. Whenever possible collect direct measures data for indicators of obesity, hypertension and smoking. More work needs to be undertaken to determine which indicators are key and for which direct measures data should always be collected. Obesity is an example as it is both a determinant and outcome for a variety of conditions and the collection of height and weight is relatively straightforward.

2. If direct measurement is not possible on an entire survey sample, consider collecting measured and reported data on a sub-sample of respondents, which will permit:

    a. the generation of study-specific correction factors; given the increased accuracy of the corrected estimates of obesity generated from this study, it is probable that study-specific correction factors may also be a means by which to improve the precision of reported estimates;
    b. documentation of the bias in different populations;
    c. monitoring changes in the bias of over time; and
    d. understanding the degree to which our self-reported estimates are underestimating measured prevalence.

3. If direct measurement is not possible on a continuous basis, consider the occasional collection of directly measured data in ongoing data collection initiatives. For example, in data collection initiatives such as the CCHS where data are collected on a yearly basis, direct measurement could be undertaken every second or third year, rather than yearly.

4. Continue methodological research to improve existing methods of adjusting self-reported estimates; continue to test and refine the correction factors presented herein.

5. Conduct further research into the validity of self-reported measures and questions with additional conditions such as diabetes, physical activity or high cholesterol.

6. Undertake additional research to examine how we can further improve existing self-reported measurement tools.

7. Follow protocols and standards for collecting and reporting both measured and reported data, work to build on and improve national and where possible international protocols for data collection and establish protocols in areas where they are lacking.

In any data collection initiative it is always a challenge to achieve an adequate balance between efficiency and precision of measurement. Objective measures provide a more accurate estimate of the prevalence of the indicators that were examined in this research, but this precision does come at a cost, a cost that has both logistical and financial implications. For health surveillance in Canada this work has clearly documented a bias in self-reported estimates of height, weight, BMI, hypertension and smoking yet these estimates are regularly employed to make program, policy and spending decisions. While no known cost-effective analysis has been undertaken to examine the cost and benefits of self-report versus direct measurement, it is clear that the costs of direct measurement are substantial. In this author's

opinion, the benefits of gaining higher quality data for the conditions examined in this dissertation far outweigh the additional costs, as accurate data are required for effective decision making. With other indicators where perhaps the discrepancy is reduced or non-existent, reported estimates may continue to be appropriate. More research is required in order to determine how to continue to advance our knowledge in this area.

References

1. Edwards AL. *The social desirability variable in personality assessment and research.* New York: Dryden Press; 1957.
2. Edwards AL. *The social desirability variable: A review of the evidence.* In Berg IA, editor. Response set in personality assessment. Chicago: Aldine; 1967:48-70.
3. Larson MR. Social desirability and self-reported weight and height. *Int J Obes Relat Metab Disord* 2000 May;24(5):663-5.
4. Pérez-Stable EJ, Marín BV, Marín G, Brody DJ, Benowitz NL 1990. Apparent underreporting of cigarette consumption among Mexican American smokers. *Am J Public Health* 1990;80:1057-61.
5. Fendrich M, Mackesy-Amiti ME, Johnson TP, Hubbell A, Wislar JS. Tobacco-reporting validity in an epidemiological drug-use survey. *Addict Behav* 2005;30:175-81.
6. Ford RP, Tappin DM, Schluter PJ, Wild CJ. Smoking during pregnancy: How reliable are self reports in New Zealand? *J Epidemiol Community Health* 1997;51:246-51.
7. From Attebring M, Herlitz J, Berndt AK, Karlsson T, Hjalmarson A. Are patients truthful about their smoking habits? A validation of self-report about smoking cessation with biochemical markers of smoking activity amongst patients with ischaemic heart disease. *J Intern Med* 2001; 249(2):145-51.
8. Tolonen H, Kuulasmaa K, Laatikainen T, Wolf H and the European Health Risk Monitoring Project. *Recommendation for indicators, international collaboration, protocol and manual of operations for chronic disease risk factor surveys.* Helsinki: The Finnish National Public Health Institute; 2002.

9. Society for Research on Nicotine and Tobacco Subcommittee on Biochemical Verification. Biochemical verification of tobacco and cessation. *Nicotine TobRes* 2002;4:149-59.

10. Engstrom JL. Assesment of the reliability of physical measures. *Res Nurs Health* 1988;11:383-9.

11. Canadian Society for Exercise Physiology. *Canadian Physical Activity, Fitness and Lifestyle Approach (CPAFLA).* 3rd edition. Ottawa, On: Canadian Society for Exercise Physiology; 2003.

12. U.S. Department of Health and Human Services. *JNC Express. The Seventh Report on the Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure.* Washington, DC: U.S. Department of Health and Human Services, NIH Publication No. 03-5233; 2003.

13. Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999-2004. *JAMA* 2006;295:1549-55.

14. World Health Organization. *Obesity: Preventing and Managing the Global Epidemic.* Geneva: World Health Organization; 2000.

15. Shields M, Tjepkema M. Trends in adult obesity. *Health Rep* 2006;17(3)53-9.

16. Rowland ML. Self-reported weight and height. *Am J Clin Nutr* 1990;52:1125-33.

17. Shields M, Connor Gorber S, Tremblay M. Estimates of obesity based on self-report versus direct measures. *Health Rep* 2008;19(2):61-76.

18. Elgar FJ, Roberts C, Tudor-Smith C, Moore L. Validity of self-reported height and weight and predictors of bias in adolescents. *J Adolesc Health* 2005;37:371-5.

19. Yannakoulia M, Panagiotakos DB, Pitsavos C, Stefanadis C. Correlates of BMI misreporting among apparently healthy individuals: the ATTICA study. *Obesity* 2006;14:894-901.

20. Chiolero A, Peytremann-Bridevaux I, Paccaud F. Associations between obesity and health conditions may be overestimated if self-reported body mass index is used. *Obesity Rev* 2007;8:373-4.

21. Shields M, Connor Gorber S, Tremblay MS. Effects of measurement on obesity and morbidity. *Health Rep* 2008;19(2):77-84.

22. Lyons A, Park J, Nelson CH. Food Insecurity and Obesity: A Comparison of Self-Reported and Measured Height and Weight. *Am J Public Health* 2007;97:751-7.

23. Visscher TLS, Viet AL, Kroesbergen HT, Seidell JC. Underreporting of BMI in adults and its effect on obesity prevalence estimations in the period 1998 to 2001. *Obesity* 2006;14(11):2054-63.

24. Yannakoulia M, Panagiotakos DB, Pitsavos C, Stefanadis C. Correlates of BMI misreporting among apparently healthy individuals: the ATTICA study. *Obesity* 2006;14(5):894-901.

25. Adlaf EM, Mates D. *A comparison of telephone and personal interview surveys for estimating alcohol and other drug use*. Addiction Research Foundation: Toronto, On; 1994.

26. Midanik LT, Greenfield TK, Rogers JD. Reports of alcohol-related harm: Telephone versus face-to-face interviews. *J Stud Alcohol* 2001 Jan;62:74-8.

27. Wiener L, Riekert K, Ryder C, Wood LV. Assessing medication adherence in adolescents with HIV when electronic monitoring is not feasible. *AIDS Patient Care STDs* 2004 Sep;18(9):527-38.

28. Ezzati M, Martin H, Skjold S, Vander Hoorn S, Murray CJL. Trends in national and state-level obesity in the USA after correction for self-report bias: analysis of health surveys. *J Res Soc Med* 2006;99:250-7.

29. Béland Y, St-Pierre M. Mode effects in the Canadian Community Health Survey: a comparison of CATI and CAPI. In: Lepkowski J, Tucker C, Brick JM, et al, eds. *Advances in Telephone Survey Methodology*. New York, NY: Wiley; 2008;297-314.

30. Solberg LI, Kottke TE, Brekke ML. Variation in clinical preventive services. *Eff Clin Pract* 2001;4:121–6.

31. Heymann AD, Hoch I, Valinsky L, Shalev V, Silber H and Kokia E. Mandatory computer field for blood pressure measurement improves screening. *Fam Pract* 2005;22:168–9.

32. Vargas CM, Burt VL, Gillum RF, Pamuk ER. Validity of self-reported hypertension in the National Health and Nutrition Examination Survey III, 1988-1991. *Prev Med* 1997;26:678-85.

33. Shields M, Connor Gorber S, Tremblay MS. Associations between obesity and morbidity: effects of measurement methods. *Obesity Rev* 2008;9:501-2.

34. Chiolero A, Peytremann-Bridevaux I, Paccaud F. Biased associations with obesity using self-reported body mass index. *Obesity Rev* 2008; [Epub ahead of print].

Statement of Contributions of Collaborators and Co-Authors

SCG was responsible for the conceptualization and design of the studies and the analysis and interpretation of the data. She conducted all analyses and wrote all papers included in this dissertation. Co-authors/collaborators (MT and IM) assisted with the conceptualization and design of the studies and interpretation of results. Co-author DM assisted with the implementation of Paper 1. Co-author NC provided technical expertise regarding hypertension measurement for Paper 2. Co-author GL provided content expertise on smoking status for Paper 3. Co-author MS assisted with the implementation of Paper 4. Co-authors (SS, BG and JH) were research assistants and second reviewers for the systematic reviews. All authors reviewed the manuscripts and approved the final content