Enabling Multi-site, Tailored Feedback Interventions to Reduce Low-Value Test Ordering: A Systematic Approach.

EYAL PODOLSKY

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
MSc degree in Epidemiology and Community Medicine

School of Epidemiology and Public Health

Faculty of Medicine

University of Ottawa

ABSTRACT

Introduction: Innovations in modern medicine have led to the development of many effective laboratory tests for diagnosis and monitoring of disease. However, research suggests that 20-50% of tests may be inappropriately ordered. Test-ordering intervention development experts i.e. those staff at individual institutions who are tasked with designing different approaches to improving test-ordering practice, often have large datasets on the test-ordering practices within their jurisdiction as part of clinical care, but often have no clear guidance on how to choose which tests are worthy of intervention. These staff can include both clinicians and knowledge translation experts working in separately or in tandem to improve test-ordering behaviours at their institutions. The overall aim of this research is to build a prioritization framework to assist healthcare organizations in deciding which tests are worth targeting for intervention.

Methods: The first stage of this project was a scoping review of the literature, the purpose of which was to determine the factors and processes used by intervention developers to choose which tests to target for intervention. Identification of these articles allowed for the descriptive analysis of the factors and processes reported in making the test-ordering decisions. The second stage involved semi-structured interviews with intervention development experts and contains information on their experiences with measuring and prioritizing among factors, as well as which processes they found to be most effective in making the decision on which tests to target for intervention.

Results: By exploring all relevant guideline statements and related empirical studies, we gathered a wide range of factors to consider when deciding which test(s) to target for intervention. Overall, we identified 18 factors in our review and 30 in our interviews, highlighting the potential complexity of these decisions. While our studies were not the first to demonstrate that targeting tests is complex and that many factors must go into the decision of deciding which tests to target, our work has identified the most comprehensive range of factors available. We also explored processes reported by the studies that led to these decisions. Of the nine processes identified in our review, some were identified in a majority of cases (literature review, followed by clinical standard-implicit and consensus process), and other processes far more infrequently. Our interviews with relevant stakeholders demonstrated that the interviewees used 18 unique processes to decide which tests to target for intervention and helped to prioritize among the processes that are most effective in making these decisions.

Conclusion: The current work prepares the way to develop a framework designed to help intervention developers choose which tests can most efficiently result in improved test-ordering processes. After additional interviews to ensure saturation of themes, we will be able to proceed with framework development, perhaps involving a consensus process of all relevant stakeholders. We hope to widely distribute our framework to assist intervention development experts working in a wide variety of milieus to help them decide which tests are worth targeting for intervention such that their respective institutions can provide the highest quality of care to patients.

ACKNOWLEDGEMENTS

This mountainous task of a thesis would have been impossible if not for the help, encouragement and support of so many others. First, I must extend a huge thanks to my supervisor Dr. Jamie Brehaut. Thank you for your ever-lasting patience, kindness, and assistance and for allowing me the opportunity to take on such an interesting project! Dr. Brehaut's expertise, guidance and wisdom provided me with the support necessary to successfully finish this thesis. Thank you for always making time for me and answering every question with understanding and calmness. I will always cherish your mentorship!

I would also like to thank Dr. Chris McCudden, Dr. Justin Presseau and Dr. David Neilipovitz for their insight at critical stages of my project development. Thank you all for your invaluable feedback. To Natasha Hudek, thank you for your help in screening, coding, interviewing and above all else being a confidant as I progressed through all the difficult stages of this project. To Kelly Carroll, and Dr. Nicola McCleary I greatly appreciate the positive environment that you both fostered which always made it such a pleasure to come to work.

Next, I would like to thank my friends from the MSc. Epidemiology cohort. I could not have asked for a better group of classmates with whom to share this experience. Thank you for always offering a helping hand when needed and making assignments, studying and writing much more enjoyable.

To my family and friends, who have always been the backbone behind anything I have ever pursued thank you for pushing me through this phase of my journey. Especially to my mom and dad, Dekel and Rony, who have asked me practically every day since March if I'm finished my thesis yet, thank you for all your encouragement. And to my brothers, best friends, and partner: Tommy, Ariel, Ethan, Jack, Sammy, JJ, and Esther for supporting me at every stage! Here's to the future!

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

Chapter 2

Chapter 3

Chapter 1: Introduction

Laboratory tests are used extensively in healthcare. In addition to monitoring patient status, tests may be used for checking medication levels, screening for disease, exploration and diagnosis.[1,2] Now ingrained in medical practice for the last 50 years,[1] medical testing has become more targeted, facilitating detection of disease,[3] identification of specific cancers and bacterial strains so that treatments may be customized to the patient.[4,5] Laboratory tests are also essential for patient management decisions; it has been hypothesized that over two thirds of patient care decisions are guided by the outcome of laboratory tests.[6–9] An important distinction to make is between screening tests and diagnostic tests. Screening tests are offered to asymptomatic people who may or may not have early disease or disease precursors and test results are used to guide whether a diagnostic test should be offered. Diagnostic tests are offered to people who have a specific indication of possible illness (a history, symptom, sign or positive screening test result) to determine whether they have the disease in question.[10] As laboratory results can determine treatment course and whether a patient will be admitted to or discharged from the hospital, they also have a great impact on further resource use.[1,6,8,9]

A study published in the 1960's remarked an "explosive" increase in laboratory testing,[10–12], and growth in healthcare testing has since been observed over many years.[13] For example, in the period between 1999-2006, there was a growth of 6-7% observed in American hospital laboratory tests.[14] In the UK, a growth of 24.2% was observed for primary care laboratory tests over the time period of 2005 to 2009.[7] Moreover, a 6-8% annual rise in tests conducted in Calgary, Alberta was noted.[12] In Eastern Ontario (Ottawa included), just over 13 million tests were completed for the fiscal year of 2015/2016, translating into a growth of 2.1% from the previous year's numbers.[15]

Overall, laboratory tests are associated with considerable cost. In 2007, the United States laboratory services industry was expected to yield a profit of 52 billion dollars, with an estimated 6.8 billion tests conducted per year.[14] A report published in 2006 in England; stated that over 697 million laboratory tests were completed over the course of the previous year, leading to an expenditure of approximately 2.5 billion pounds.[6,7] An increase in laboratory service expenditure has been observed in several provinces across Canada. Between the fiscal years of 1996/1997 and 2001/2002, increases in costs of 8%, 8%, 14%, 15% and 34% were seen for Ontario, Manitoba, Alberta, Saskatchewan and British Colombia respectively.[16] Among the provinces, Ontario was found to have the second highest per capita expenditure in 2001/2002 ($90.41).[16] A more recent report estimates that Ontario laboratory facilities conducted over 244 million tests during the fiscal year of 2013/2014, resulting in a cost of 2.05 billion dollars.[17]

Inappropriate Use of Resources in Healthcare

During the late 1990s, overuse in healthcare was identified as one of several urgent quality issues by the Institute of Medicine's "National Roundtable on Health Care Quality".[18] Waste in healthcare is at the forefront at a number of workshops and conferences, such as "the Institute of Medicine (IOM) Roundtable on Value & Science-Driven Health Care" workshops[19] and an "Avoiding Avoidable Care" conference. [20,21] It is believed that among the total expenditures on health services in the USA, as much as 30% is thought to be unnecessary.[19,20,22,23] This value is congruent with a more recent analysis of several Canadian tests and interventions which observed that 30% may have been avoidable.[24] Financial losses are among the most commonly cited reasons behind reducing unnecessary tests.[25] However, other reasons for reducing and eliminating unnecessary medical care include adverse events of unnecessary and that it can affect the care of other patients by reducing the availability of resources.[22,26]

Inappropriate Testing

Unnecessary testing is commonly identified as a focus for cost containment programs, due to the cascading effect of tests on further resource use.[8,27,28] While it can be difficult to articulate a general definition for an 'inappropriate' test, several have been suggested in the literature. One study defined unnecessary testing as "a test that would not change management regardless of its result".[29] Other areas of overuse have been identified: laboratory tests that are "repeated at too frequent intervals," "ordered when clinical assessment is superior," "ordered to confirm an expected response to a routine intervention," "redundant laboratory tests ordered concurrently" and those which "[do] not affect management or prognostication".[30] There are a number of terms often used inter-changeably for unnecessary testing including: overuse, underuse, inappropriate use, overutilization and they will be used in a similar fashion throughout the course of this work. An important point to mention is that inappropriate use is a broad term that can refer to both overuse of resources but can also be used to refer to under-ordering of tests or insufficient use of resources.

One indicator suggested to identify inappropriate ordering of laboratory tests is variation between providers and institutions without discernible differences in patient populations or outcomes.[2,11] Studies have shown substantial variation in provider and institution test ordering (or resource use including laboratory tests) in primary, outpatient, clinic care and inpatient hospital care.[7,31–37] These same studies demonstrated that this variation rarely leads to major differences in patient outcomes such as mortality[31,37] or length of stay.[31] This indicates that some sites and settings are capable of safely providing care with use of fewer tests, opening the door for implementation of interventions to help reduce unnecessary orders.[2,31]

A wide range in inappropriate ordering can be seen across the literature. Additionally, differences in test type and setting across these studies suggests that inappropriate ordering is a widespread issue. Several highly cited studies have estimated expenditures associated with potentially wasteful practices. An Eastern Ontario study estimated that between 1999-2000, the redundant use of several frequently used tests led to $13.9-35.9 million worth of testing.[43] Another study found that in one year, between $0.6-$2.2 million dollars were wasted on potentially unnecessary repeated tests identified among 103,000 patients.[40] A criticism of these types of cost estimations, however, is that it is difficult to determine which tests can truly be classified as unnecessary.[27] Many studies have explored the magnitude of inappropriate ordering for a variety of specific tests and contexts such as a variety of patient populations, clinical settings and different healthcare systems. For example, studies have estimated the proportion of potentially inappropriate or unnecessary ordering of ionized calcium tests to be 97%,[38] arterial blood gas analyses (ABGs) to be 30%,[39] blood tests (cholesterol, hemoglobin A1c, thyroid- stimulating hormone, vitamin B12, vitamin D, ferritin) to be 16%,[40] thyroid tests to be 10%,[41] and serum sodium tests to be 5.1%.[42] A review that covered studies from 1966-1997 reported a range of 4.5%-95% inappropriate utilization in test-ordering.[1,13] A more recent systematic review from 2013, demonstrated a 20.6% mean rate of inappropriate overutilization in 38 studies between 1997 and 2012.[1]

Influences on Test Ordering

Overuse of test-ordering appears to be a challenging and complex issue and testing overuse may be caused by a multitude of related factors. Several of these factors depend on the physician.[44,45] Knowledge of the ordering physician is one such factor, including lack of knowledge regarding price, that a test is obsolete, or that a test should not be repeated within a certain timeframe; all of these could lead to the inappropriate ordering of a test.[2,30,45–48] Fear of medico-legal issues may lead to defensive and potentially inappropriate test ordering.[7,44] Defensive medicine is by definition a deviation from sound medical practice induced mainly by the fear of liability.[44] Furthermore, beliefs regarding the utility of a laboratory test, or uncertainty, may influence a provider's decision to order.[2,44,48] Habit and training have also been suggested by residents to play a role in the ordering of potentially inappropriate tests.[29]

Test-ordering behaviour may also be influenced by the method in which tests are ordered. A commonly used method is computer provider order entry (CPOE) system which allows for more efficient, but potentially wasteful ordering.[2] Furthermore, the specific formatting of a test order form has a demonstrated impact on test-ordering volume. For example, the un-bundling of tests (creating separate entities, such that each test must be individually ordered) has demonstrated to be associated with a decreased number of individual tests ordered.[49] Moreover, standing orders, in which tests are ordered repeatedly through one initial order may result in unnecessary testing when the healthcare provider does not stop the standing order when it is no longer necessary.[42,46,50,51] Physician uncertainty has also been found to contribute to over-testing when the physicians are unsure if they are ordering the correct test as a result of vague nomenclature.[45,52]

Hospital characteristics and other aspects of the environment may impact test-ordering practices as well. For example, one study demonstrated that ICU patients in teaching hospitals as compared to non-teaching hospitals had far more blood tests ordered.[53] Certain clinical specialties such as critical care, may order some tests routinely (based on a time interval) rather than on the basis of medical necessity.[54,55] Moreover, the frequency of blood collection is increased when a patient already has an arterial line placed.[56,57]

Unnecessary test-ordering may be facilitated by social pressures. Pressure from patients, family members or even the perception that patients expect testing to be completed may result in unnecessary or excessive ordering.[7,45,58] Additionally, junior healthcare providers feeling pressure (real or perceived) from their superiors may lead to unnecessary test-ordering.[29,59] Finally, test-ordering is impacted by the payment system,[44,47,60,61] which can vary from country to country or even within different jurisdictions in the same country. For example, in payment systems where an incentive to order more tests exists because of reimbursement, unnecessary testing is more likely.[61] With so many potential factors contributing to overuse identifying and prioritizing among them should be a first step in designing better interventions.

Underutilization

Underutilization of laboratory testing also exists in a variety of situations. A study assessing an intervention to improve compliance with the Surviving Sepsis Campaign resuscitation bundle[62,63] found the average compliance for ordering lactate tests and blood cultures (in advance of treatment with antibiotics) to be between 61.0% and 64.5% respectively.[62] This indicates that there are opportunities for improvement with regards to appropriately ordering tests for severe sepsis patients.[63] However, in a systematic review entitled "the landscape of inappropriate laboratory testing"[1], the results showed that underutilization may be less common (or less commonly explored) than overutilization. This review identified 38 studies which assessed overutilization compared to eight studies involved underutilization.[1]

Initiatives for Change

Several organizations across the world have identified prudent test-ordering as a significant goal in improving quality of care as demonstrated by the establishment of local and national initiatives. Choosing Wisely campaigns initiated by the American Board of Internal Medicine (ABIM), and the Canadian Medical Association/University of Toronto are key examples of such initiatives. The campaigns focus their attention on appropriate use of tests, in addition to proper use of treatments and services.[64,65] "More is not always better" is a key component of the Choosing Wisely Canada campaign.[66] The central message of both campaigns is the importance of discussions between patients and healthcare providers regarding the necessity of testing and/or treatment.[64,65] As of 2019, Choosing Wisely Canada has published 38 lists (227 recommendations) by different associations, ranging across a variety of specialties (anesthesiology, critical care, family medicine, general surgery, palliative care, psychiatry, etc.).[66] Table 1 demonstrates a few examples of Choosing Wisely Canada guidelines related to test ordering.

Table 1. Choosing Wisely Canada guidance statements related to test-ordering.

Society	Example of Guidance
CMA's Forum on General and Family Practice Issues & College of Family Physicians of Canada	"Don't do annual screening blood tests unless directly indicated by the risk profile of the patient."[67]
Canadian Anesthesiologists' Society	"Don't order baseline laboratory studies (complete blood count, coagulation testing, or serum biochemistry) for asymptomatic patients undergoing low-risk non-cardiac surgery."[68]
Canadian Society of Endocrinology and Metabolism	"Don't routinely test for Anti-Thyroid Peroxidase Antibodies (anti – TPO)."[69]
Canadian Federation of Medical Students Fédération médicale étudiante du Québec	"Don't suggest ordering tests or treatments pre-emptively for the sole purpose of anticipating what your supervisor would want."[70]
Canadian Association of Pathologists	"Avoid standing orders for repeat complete blood count (CBC) on inpatients who are clinically/laboratorial stable."[71]

Various healthcare service utilization initiatives can be found across many jurisdictions which further indicates that this is a widespread issue. Outside of North America, sixteen other countries have implemented Choosing Wisely initiatives.[22] The UK's National Institute for Health and Care Excellence (NICE) has also signalled the importance of the reduction of unnecessary testing by establishing guidelines regarding the appropriateness of several laboratory tests prior to elective surgery.[72] Moreover, a 2014 report released by the Canadian Agency for Drugs and Technologies in Health (CADTH) identified a wide variety of other provincial and global initiatives attempted to improve laboratory test practice.[73] Within Ontario, Health Quality Ontario (HQO) published recommendations pertaining to several community-based laboratory tests in order to address their "Appropriateness Initiative", the goal of which was "to develop a systematic framework for the ongoing identification, prioritization, and assessment of health interventions in Ontario for which there is possible misuse, overuse, or underuse."[74]

Changing Practice

Despite the prioritization of reductions of unnecessary testing across many organizations, there appears to be very little published literature detailing the success of these campaigns, guidelines and initiatives or their impacts on test-ordering. A 2014 report by CADTH focused on laboratory utilization initiatives and noted this scarcity in the published literature, stating that this makes it difficult to determine which initiatives are successful.[73] Some studies have compared changes in practice before and after publication of the Choosing Wisely guidelines. It would appear as though the publication of the guidelines is the necessary but not sufficient step because the publication of a guideline presumably triggers the intervention at each site and it is these activities that ultimately influence the drop in tests, not the publication of the guidelines per se. An observational study assessed the ordering of low-value imaging for back pain following the implementation of a Choosing Wisely guidance statement and found a small (3.8%) change in practice.[75] Another study assessed changes in claims data for seven services (including imaging, lab testing, medication and antibiotic prescribing) each of which was associated with a Choosing Wisely guideline.[76] This author found small but significant decreases for some of the resources (1.1%-1.5%).[76] The two papers both discuss the need for other strategies to increase the uptake of these guidelines.[75,76]

Implementation of best test-ordering practices is a difficult endeavour. Multiple surveys demonstrated a wide ranging (21% to 66.7%) awareness/familiarity among medical professionals (primary care physicians, medical and surgical specialists) with the Choosing Wisely campaign.[58,77] Aside from awareness of the guideline, a systematic review indicated that a multitude of barriers to guideline adherence exist such as physician disagreement with the guideline, and discordant patient preferences.[78] A study that surveyed primary care physicians recorded 43.8% or more of participants as believing that the "current medical malpractice system", "patient requests", "[recommendations] by specialists", and "lack of time for shared decision making" were also rated as barriers to addressing overutilization.[79] Additionally, habitual behaviours seem to cause difficulty with regards to changing ordering behaviour.[47,78,80] Because the problem of overutilization often requires the healthcare provider to decrease a specific behaviour that has become habitual, designing interventions to change this practice requires additional consideration.

Interventions to address the challenges mentioned above have been widely studied. A systematic review of reviews concluded that 'active' Continuing Medical Education (CME) interventions (i.e. academic detailing) tended to yield better results than 'passive' interventions (i.e. printed educational material) in terms of increasing the desirable behaviour.[81] The efficacy of clinician behaviour change interventions for the improvement of provider laboratory test ordering has been assessed by a number of reviews. [8,45,82,84] A review by Kobewka et al.[84] assessed laboratory test ordering in all healthcare settings, while systematic reviews by Cadogan et al.[45] and Thomas et al.[8] focused specifically on test ordering in primary care. All reviews identified a variety of potential effective interventions including one or a combination of Audit and Feedback (A&F), education, incentive and penalty, and "system-based" or administrative interventions.[8,45,82,84] Moreover, all reviews[8,45,84] found a variety of intervention strategies to be effective in changing provider behaviour, with Kobewka et al.'s results showing multifaceted interventions to have a greater effect than that of single component interventions.[84] Both Kobewka et al. and Cadogan et al.'s reviews comment on the overall low quality of the included studies, suggesting further work remains to improve study conduct, with use of more rigorous study design methods.[45,84] Additional suggestions for intervention [6] implementation in the literature include gaining "leadership level support" and taking context-specific considerations into account by conducting a barriers assessment. [2,45] As a result, it is clear that the success of interventions to improve test ordering practice has been variably successful and a more systematic approach is needed to ensure greater success.

Designing Effective Test-Ordering Interventions

Test-ordering intervention developers, i.e. people who are tasked with designing different approaches to improving test-ordering practice, can find it difficult to decide which among the hundreds of tests routinely ordered would most benefit care if targeted for intervention. These staff can include both clinicians and knowledge translation experts working in separately or in tandem to improve test-ordering behaviours at their institutions. Yet in order to improve test ordering efficiency, organizations need to choose where to intervene to reduce inappropriate testing and increase appropriate testing. These interventions must be worth the organization's time, effort and resources, in terms of improving efficiency on an organizational level and ensuring that they are making a meaningful impact on patient care. For the purposes of this thesis, the term 'intervention' is used to refer to projects or efforts that are specifically designed to implement change in the processes or behaviours relevant to test-ordering. Intervention developers often have immense amounts of data on the test-ordering practices of the clinicians that they oversee. For instance, conservatively, there are about 700 orderable tests within The Ottawa Hospital (TOH). With so much information available to the intervention developers, it is difficult to decide which information and by extension which tests are worth intervention. Intervention developers rarely have clear directions on how to affect improved test-ordering practices. Even *Choosing Wisely*

recommendations do not always provide such guidance; critical care guidance standards state that testing should not be repeated, but the recommendations make no mention of which specific tests are most likely to be repeated, nor do they describe which tests are most often over-used.[13]

Aim & Objectives

The overall aim of this research is to build a prioritization framework to assist healthcare organizations in deciding which tests are worth targeting for intervention. The framework will provide intervention developers with the range of factors that have guided other test prioritization decisions, as well as those considered most important based on the experience of experts in this area. In the long term, the impact should be an increased efficiency of resource utilisation and improved clinical decision making as a result of reduced use of low-value testing.

Objective I: Conduct a scoping review of the literature to identify factors and processes that guided the decisions by intervention developers to intervene on the test-ordering practices of certain tests over others.

Objective II: Conduct semi-structured interviews with intervention development experts to gain further insight about how they prioritized certain factors over others, measured the factors and which processes they found to be most effective in deciding which tests to target for intervention.

Overview of Methodology & Structure

The research of this thesis has been structured in a manuscript format, which consists of an introductory chapter, two draft manuscripts, and a discussion chapter. The first draft manuscript (Chapter 2) describes a scoping review of the literature, the purpose of which was to determine the factors and processes used by intervention developers to choose which tests to target for intervention. Identification of these articles allowed for the descriptive analysis of the factors and processes reported in making the test-ordering decisions. The draft manuscript provides an overview of the methodology; search strategy development, information sources, screening, data extraction and analysis. The second draft manuscript (Chapter 3) reports on semi-structured interviews with intervention development experts and contains information on their experiences with prioritizing among factors, measuring the factors and which processes they found to be most effective in making the decision on which tests to target for intervention. Chapter 4 contains a discussion and integration of the results from Chapters 2 and 3, as well as a discussion of the implications and significance of these results.

References

1. Zhi M, Ding EL, Theisen-Toupal J, Whelan J, Arnaout R. The Landscape of Inappropriate Laboratory Testing: A 15-year Meta-Analysis. *PLoS One*. 2013;8(11):1-8. doi:10.1371/journal.pone.0078962.

2. Yeh DD. Laboratory Utilization. In: Lewandrowski K, Sluss PM, eds. *Utilization Management in the Clinical Laboratory and Other Ancillary Services.* Cham: Springer International Publishing; 2017:203-210. doi:10.1007/978-3-319-34199-6_18.

3. Aschengrau A, Seage III GR. *Essentials of Epidemiology in Public Health*. 3rd ed. Burlington, MA: Jones & Bartlett Learning; 2014.

4. Manegold C. Treatment algorithm in 2014 for advanced non-small cell lung cancer: Therapy selection by tumour histology and molecular biology. *Adv Med Sci*. 2014;59(2):308-313. doi:10.1016/j.advms.2014.08.008.

5. Bassetti M, Poulakou G, Ruppe E, Bouza E, van Hal SJ, Brink A. Antimicrobial resistance in the next 30 years, humankind, bugs and drugs: a visionary approach. *Intensive Care Med*. 2017:1-12. doi:10.1007/s00134-017-4878-x.

6. Lord Carter of Coles. *Report of the Review of NHS Pathlogy Services in England*.; 2006. http://webarchive.nationalarchives.gov.uk/20130107105354/http:/www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_091984.pdf.

7. Busby J, Schroeder K, Woltersdorf W, et al. Temporal growth and geographic variation in the use of laboratory tests by NHS general practices: Using routine data to identify research priorities. *Br J Gen Pract*. 2013;63(609):256-266. doi:10.3399/bjgp13X665224.

8. Thomas RE, Vaska M, Naugler C, Turin TC. Interventions at the laboratory level to reduce laboratory test ordering by family physicians: Systematic review. *Clin Biochem*. 2015;48(18):1358-1365. doi:10.1016/j.clinbiochem.2015.09.014.

9. Plebani M. Errors in Clinical Laboratories or Errors in Laboratory Medicine? *Clin Chem Lab Med*. 2006;44(6):750-759. doi:10.1515/CCLM.2006.123.

10. Sheringham J, Kalim K, Crayford T. Mastering Public Health: A guide to examinations and revalidation. ISBN-13 978-1-85315-781-3

11. Emson HE. The Growth of a Laboratory. *Can Med Assoc J*. 1965;93:211-218.

12. Naugler C. Laboratory Utilization Management in Canada. In: Lewandrowski K, Sluss P, eds. *Utilization Management in the Clinical Laboratory and Other Ancillary Services*. Springer, Cham; 2017.

13. Naugler C. A Perspective on Laboratory Utilization Management from Canada. *Clin Chim Acta*. 2014;427:142-144. doi:10.1016/j.cca.2013.09.022.

14. van Walraven C, Naylor CD. Do We Know What Inappropriate Laboratory Utilization Is? A Systematic Review of Laboratory Clinical Audits. *JAMA*. 1998;280(6):550-558. doi:10.1309/AJCPX1HIEM4KLGNU.

15. The Lewin Group, Wolcott J, Schwartz A, Goodman C. *Laboratory Medicine: A National Status Report.;* 2008. http://wwwn.cdc.gov/futurelabmedicine/pdfs/2007%5Cnstatus%5Cnreport%5Cnlaboratory_medicine_-_a_national_status_report_from_the_lewin_group_updated_2008-9.pdf.

16. Eastern Ontario Regional Laboratory Association. *Annual Report 2015-16: DRAFT*.

17. Bayne L, Lillian Bayne & Associates. *BC Laboratory Services Review*.; 2003.

18. Ministry of Health and Long-Term Care. *Laboratory Services Expert Panel*.; 2015.

19. Chassin M, Galvin R. The urgent need to improve health care quality: Institute of medicine national roundtable on health care quality. *JAMA.* 1998;280(11):1000-1005. doi:10.1001/jama.280.11.1000.

20. Institute of Medicine (US) Roundtable on Evidence-Based Medicine. *The Healthcare Imperative: Lowering Costs and Improving Outcomes. Workshop Series Summary*. (Yong P, Saunders R, Olsen L, eds.). Washington, DC: National Academies Press; 2010. doi:10.17226/12750.

21. Lenzer J. Unnecessary care: are doctors in denial and is profit driven healthcare to blame? *Bmj*. 2012;345(oct02 3):e6230-e6230. doi:10.1136/bmj.e6230.

22. Lown Cardiovascular Research Foundation, New America Foundation. About The Conference. Avoiding Avoidable Care Web Page. http://avoidablecare.org/about-the-conference/. Accessed October 11, 2017.

23. Wintemute K, McDonald K, Huynh T, Pendrith C, Wilson L. Addressing overuse starts with physicians. Choosing Wisely Canada. *Can Fam Physician*. 2016;62(3):199-200. doi:62/3/199 [pii].

24. Cho HJ. Countering Medicine's Culture of More. *Am Med Assoc J Ethics*. 2015;17(11):1079- 1081.

25. Canadian Institute for Health Information. Canadians have more than 1 million potentially unnecessary medical tests and treatments every year. Canadian Institute for Health Information Web Page. https://www.cihi.ca/en/canadians-have-more-than-1-million-potentially-unnecessarymedical-tests-and-treatments-every-year. Published 2017. Accessed October 11, 2017.

26. Berwick DM, Hackbarth AD. Eliminating Waste in US Health Care. *JAMA*. 2012;307(14):1513-1516. doi:10.1001/jama.2012.362.

27. Canadian Institute for Health Information. Unnecessary Care in Canada: Infographic. Canadian Institute for Health Information Web Page. https://www.cihi.ca/en/unnecessary-care-in-canadainfographic. Published 2017. Accessed October 11, 2017.

28. Lewandrowski K, Baron J, Dighe A. Utilization Management in the Clinical Laboratory: An Introduction and Overview. In: Lewandrowski K, Sluss P, eds. *Utilization Management in the Clinical Laboratory and Other Ancillary Services*. Springer, Cham; 2017.

29. Sluss P. Calculating Costs and Savings in Utilization Management. In: Lewandrowski K, Sluss P, eds. *Utilization Management in the Clinical Laboratory and Other Ancillary Services.* Springer, Cham; 2017.

30. Sedrak MS, Patel MS, Ziemba JB, et al. Residents' Self-Report on Why They Order Perceived Unnecessary Inpatient Laboratory Tests. J Hosp Med. 2016;11(12):869-872. doi:10.1002/jhm.2645.

31. Ezzie ME, Aberegg SK, O'Brien JM. Laboratory Testing in the Intensive Care Unit. *Crit Care Clin*. 2007;23(3):435-465. doi:10.1016/j.ccc.2007.07.005.

32. Garland A, Shaman Z, Baron J, Connors AF. Physician-attributable Differences in Intensive Care 129 Unit Costs: A Single-Center Study. *Am J Respir Crit Care Med.* 2006;174(11):1206-1210. doi:10.1164/rccm.200511-1810OC.

33. Mindemark M, Wernroth L, Larsson A. Costly regional variations in primary health care test utilization in Sweden. *Scand J Clin Lab Invest*. 2010;70(3):164-170. doi:10.3109/00365511003642519.

34. Verstappen WHJM, ter Riet G, Dubois WI, Winkens R, Grol RP, van der Weijden T. Variation in test ordering behaviour of GPs: Professional or context-related factors? *Fam Pract.* 2004;21(4):387-395. doi:10.1093/fampra/cmh408.

35. Cadogan SL, McHugh SM, Bradley CP, Browne JP, Cahill MR. General practitioner views on the determinants of test ordering: a theory-based qualitative approach to the development of an intervention to improve immunoglobulin requests in primary care. *Implement Sci.* 2016;11(1):102. doi:10.1186/s13012-016-0465-8.

36. Schroeder SA, Kenders K, Cooper J, Piemme TE. Use of Laboratory Tests and Pharmaceuticals: Variation Among Physicians and Effect of Cost Audit on Subsequent Use. *JAMA*. 1973;225(8):969-973. doi:10.1001/jama.1973.03220360029007.

37. Signorelli H, Straseski JA, Genzen JR, Walker BS, Jackson BR, Schmidt RL. Benchmarking to Identify Practice Variation in Test Ordering: A Potential Tool for Utilization Management. *Lab Med*. 2015;46(4):356-364. doi:10.1309/LM2JVTWX8TKCURMF.

38. Zimmerman JE, Seneff MG, Sun X, Wagner DP, Knaus WA. Evaluating laboratory usage in the intensive care unit: Patient and institutional characteristics that influence frequency of blood sampling. *Crit Care Med*. 1997;25(5):737-748. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed7&NEWS=N&AN=27249148.

39. Newman DB, Siontis KC, Chandrasekaran K, Jaffe AS, Kashiwagi DT. Intervention to Reduce Inappropriate Ionized Calcium Ordering Practices: A Quality-Improvement Project. *Perm J*. 2015;19(1):49-51. doi:10.7812/TPP/14-108.

40. Melanson SEF, Szymanski T, Rogers SO, et al. Utilization of Arterial Blood Gas Measurements in a Large Tertiary Care Hospital. *Am J Clin Pathol*. 2007;127(4):604-609. doi:10.1309/ELH5BPQ0T17RRK0M.

41. Morgen EK, Naugler C. Inappropriate Repeats of Six Common Tests in a Canadian City: A Population Cohort Study within a Laboratory Informatics Framework. *Am J Clin Pathol*. 2015;144(5):704-712. doi:10.1309/AJCPYXDAUS2F8XJY.

42. Gill J, Barakauskas VE, Thomas D, et al. Evaluation of thyroid test utilization through analysis of population-level data. *Clin Chem Lab Med*. 2017;[Epub ahea(0). doi:10.1515/cclm-2016-1049.

43. Weydert JA, Nobbs ND, Feld R, Kemp JD. A Simple, Focused, Computerized Query to Detect Overutilization of Laboratory Tests. *Arch Pathol Lab Med*. 2005;129(9):1141-1143. doi:10.1043/1543-2165(2005)129[1141:ASFCQT]2.0.CO;2.

44. Van Walraven C, Raymond M, for the Network of Eastern Ontario Medical Laboratories (NEOMeL). Population-based Study of Repeat Laboratory Testing. *Clin Chem*. 2003;49(12):1997-2005.

45. Sood R, Sood A, Ghosh AK. Non-Evidence-Based Variables Affecting Physicians' Test-Ordering 130 Tendencies: A Systematic Review. *Neth J Med*. 2007;65(5):167-177. http://www.ncbi.nlm.nih.gov/pubmed/17519512.

46. Cadogan SL, Browne JP, Bradley CP, Cahill MR. The effectiveness of interventions to improve laboratory requesting patterns among primary care physicians: a systematic review. *Implement Sci*. 2015;10:167. doi:10.1186/s13012-015-0356-4.

47. Hindmarsh JT, Lyon AW. Strategies to Promote Rational Clinical Chemistry Test Utilization. *Clin Biochem*. 1996;29(4):291-299.

48. Brateanu A, Rothberg MB. Why do clinicians continue to order "routine preoperative tests" despite the evidence? *Cleve Clin J Med*. 2015;82(10):667-670. doi:10.3949/ccjm.82a.15118.

49. Miyakis S, Karamanof G, Liontos M, Mountokalakis TD. Factors contributing to inappropriate ordering of tests in an academic medical department and the effect of an educational feedback strategy. *Postgrad Med J*. 2006;82(974):823-829. doi:10.1136/pgmj.2006.049551.

50. Kahan NR, Waitman D-A, Vardy D. Curtailing Laboratory Test Ordering in a Managed Care Setting Through Redesign of a Computerized Order Form. *Am J Manag Care*. 2009;15(3):173-176.http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L70015831%5C nhttp://dx.doi.org/10.1111/j.1524-4733.2009.00537-2.x%5Cnhttp://sfx.library.uu.nl/utrecht?sid=EMBASE&issn=10983015&id=doi:10.1111/j.1524-4733.2009.00537-2.x&atitle=Curtai.

51. Raad S, Elliott R, Dickerson E, Khan B, Diab K. Reduction of Laboratory Utilization in the Intensive Care Unit. *J Intensive Care Med*. 2016:885066616651806. doi:10.1177/0885066616651806.

52. Civetta JM, Hudson-Civetta J a. Maintaining Quality of Care While Reducing Charges in the ICU: Ten ways. *Ann Surg*. 1985;202(4):524-532. doi:10.1097/00132586-198608000-00041.

53. Hickner J, Thompson PJ, Wilkinson T, et al. Primary Care Physicians' Challenges in Ordering Clinical Laboratory Tests and Interpreting Results. *J Am Board Fam Med.* 2014;27(2):268-274. doi:10.3122/jabfm.2014.02.130104.

54. Zimmerman JE, Seneff MG, Sun X, Wagner DP, Knaus WA. Evaluating Laboratory Usage in the Intensive Care Unit: Patient and Institutional Characteristics that Influence Frequency of Blood Sampling. *Crit Care Med*. 1997;25(5):737-748.

55. Cismondi F, Celi LA, Fialho AS, et al. Reducing Unnecessary Lab Testing in the ICU with Artificial Intelligence. *Int J Med Inform*. 2013;82(5):345-358. doi:10.1016/j.ijmedinf.2012.11.017.

56. Critical Care Societies Collaborative. Five Things Physicians and Patients Should Question. Choosing Wisely. http://www.choosingwisely.org/societies/critical-care-societies-collaborativecritical-care/. Published 2014. Accessed November 18, 2016.

57. Flabouris A, Bishop G, Williams L, Cunningham M. Routine Blood Test Ordering for Patients in Intensive Care. *Anaesth Intensive Care*. 2000;28(5):562-565.

58. Jakacka N, Snarski E, Mekuria S. Prevention of Iatrogenic Anemia in Critical and Neonatal Care. *Adv Clin Exp Med*. 2016;25(1):191-197. doi:10.17219/acem/32065.

59. Colla CH, Kinsella EA, Morden NE, Meyers DJ, Rosenthal MB, Sequist TD. Physician 131 Perceptions of Choosing Wisely and Drivers of Overuse. Am J Manag Care. 2016;22(5):337-343.

60. Axt-Adam P, van der Wouden JC, van der Does E. Influencing Behavior of Physicians Ordering Laboratory Tests: A Literature Study. *Med Care*. 2017;31(9):784-794.

61. Schattner A. The unbearable lightness of diagnostic testing: time to contain inappropriate test ordering. *Postgrad Med J*. 2008;84(998):618-621. doi:10.1136/pgmj.2008.072579.

62. Van Der Weijden T, Van Bokhoven MA, Dinant GJ, Van Hasselt CM, Grol RPTM. Understanding laboratory testing in diagnostic uncertainty: A qualitative study in general practice. *Br J Gen Pract*. 2002;52(485):974-980.

63. Levy MM, Dellinger RP, Townsend SR, et al. The Surviving Sepsis Campaign: Results of an international guideline-based performance improvement program targeting severe sepsis. *Intensive Care Med*. 2010;36(2):222-231. http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=20069275.

64. Levy MM, Rhodes A, Phillips GS, et al. Surviving Sepsis Campaign: Association Between Performance Metrics and Outcomes in a 7.5-Year Study. *Crit Care Med*. 2015;43(1):3-12. doi:10.1097/CCM.0000000000000723.

65. Canadian Medical Association, University of Toronto. What is CWC? Choosing Wisely Canada Website. http://www.choosingwiselycanada.org/about/what-is-cwc/. Accessed March 22, 2019.

66. The American Board of Internal Medicine Foundation. About. Choosing Wisely Website. http://www.choosingwisely.org/about-us/. Published 2017. Accessed April 10, 2019.

67. Canadian Medical Association, University of Toronto. Choosing Wisely Canada- The Lists. Choosing Wisely Canada Website. http://www.choosingwiselycanada.org/recommendations/. Accessed April 10, 2019.

68. CMA's Forum on General and Family Practice Issues, College of Family Physicians of Canada. Eleven Things Physicians and Patients Should Question. Choosing Wisely Canada Website. http://www.choosingwiselycanada.org/recommendations/family-medicine/. Published 2014. Accessed April 10, 2019.

69. Canadian Anesthesiologists' Society. Five Things Physicians and Patients Should Question. Choosing Wisely Canada Website. http://www.choosingwiselycanada.org/recommendations/anesthesiology/. Published 2015. Accessed April 10, 2019.

70. Canadian Society of Endocrinology and Metabolism. Five Things Physicians and Patients Should Question. Choosing Wisely Canada Website. http://www.choosingwiselycanada.org/recommendations/endocrinology-and-metabolism/. Published 2014. Accessed April 10, 2019.

71. Canadian Federation of Medical Students, Fédération médicale étudiante du Québec. Six Things Medical Students and Trainees Should Question. Choosing Wisely Canada Website. http://www.choosingwiselycanada.org/recommendations/medical-students-and-trainees/. Published 2015. Accessed April 10, 2019.

72. Canadian Association of Pathologists. Five Things Physicians and Patients Should Question. 132 Choosing Wisely Canada Website. http://www.choosingwiselycanada.org/recommendations/pathology/. Published 2014. Accessed April 10, 2019.

73. National Institute for Health and Care Excellence. *Routine Preoperative Tests for Elective Surgery*.; 2016. https://www.nice.org.uk/guidance/ng45.

74. Verbrugghe S. Initiatives to Optimize the Utilization of Laboratory Tests. Ottawa: Canadian Agency for Drugs and Technologies in Health. 2014;(Environmental Scan; issue 44).

75. Ontario Health Technology Advisory Committee. *Appropriateness Phase 1 OHTAC Recommendations: Annual Health Exams, Aspartate Aminotransferase Testing, Chloride Testing, Creatine Kinase Testing, Ferritin Testing, Folate Testing, and Vitamin B12 Testing.*; 2013. http://www.hqontario.ca/Portals/0/Documents/evidence/reports/recommendation-appropriatenessphase-1-130722-en.pdf.

76. Hong AS, Ross-Degnan D, Zhang F, Wharam JF. Small Decline In Low-Value Back Imaging Associated With The "Choosing Wisely" Campaign, 2012–14. *Health Aff*. 2017;36(4):671-679. doi:10.1377/hlthaff.2016.1263.

77. Rosenberg A, Agiro A, Gottlieb M, et al. Early Trends Among Seven Recommendations From the Choosing Wisely Campaign. *JAMA Intern Med*. 2015;175(12):1913-1920. doi:10.1001/jamainternmed.2015.5441.

78. Grover M, Abraham N, Chang Y-H, Tilburt J. Physician Cost Consciousness and Use of LowValue Clinical Services. *J Am Board Fam Med*. 2016;29(6):785-792. doi:10.3122/jabfm.2016.06.160176.

79. Cabana MD, Rand CS, Powe NR, et al. Why Don't Physicians Follow Clinical Practice Guidelines? A Framework for Improvement. *JAMA*. 1999;282(15):1458-1465. doi:10.1001/jama.282.15.1458.

80. Zikmund-Fisher BJ, Kullgren JT, Fagerlin A, Klamerus ML, Bernstein SJ, Kerr EA. Perceived Barriers to Implementing Individual Choosing Wisely Recommendations in Two National Surveys of Primary Care Providers. *J Gen Intern Med*. 2017;32(2):210-217. Epub 2016. doi:10.1007/s11606-016-3853-5.

81. Nilsen P, Roback K, Broström A, Ellstrøm P-E. Creatures of habit: accounting for the role of habit in implementation research on clinical behavior change. *Implement Sci*. 2012;7:53. doi:10.1186/1748-5908-7-53.

82. Foster, M., Presseau, J., McCleary, N. et al. Audit and feedback to improve laboratory test and transfusion ordering in critical care: a systematic review. *Implementation Sci* 15, 46 (2020). https://doi.org/10.1186/s13012-020-00981-5

83. Mostofian F, Ruban C, Simunovic N, Bhandari M. Changing Physician Behavior: What Works? *Am J Manag Care*. 2015;21(1):75-84.

84. Kobewka DM, Ronksley PE, McKay JA, Forster AJ, Van Walraven C. Influence of educational, audit and feedback, system based, and incentive and penalty interventions to reduce laboratory test utilization: A systematic review. *Clin Chem Lab Med*. 2015;53(2):157-183. doi:10.1515/cclm2014-0778.

Chapter 2 (Draft Manuscript 1):
Factors Reported in Determining Laboratory Tests to Target for Intervention: A Scoping Review

Preface

Chapter 2 describes a scoping review intended to identify publications which report on the factors and processes that were given consideration when deciding which test-ordering practices to target for intervention. Our goal was to combine the results of the scoping review with semi-structured interviews of experts in test ordering guidance development to develop an evidence-based prioritization framework that guides the choice of tests that are most amenable to change.

Contributions: JCB was responsible for the conception of this project and provided guidance and expertise throughout the entire project. EP drafted the manuscript, and JCB, NH, KC, JP and CM provided critical input and aided in the revision of the manuscript. EP and NH completed title, abstract and full-text screening. Data extraction and quality assessment were completed by EP and NH. The guarantor of this review is EP.

Notes: Appendices for the PRISMA checklist, the Medline search strategy, additional details on the search strategy and sample can be found at the end of this chapter. Ethics approval was not required for this study as the systematic review relied solely on published studies.

Publication Citation: Podolsky E, Hudek H, Carroll K, McCudden C, Presseau J, Brehaut J. "Factors Reported in Determining Laboratory Tests to Target for Intervention: A Scoping Review." (In preparation)

Factors Reported in Determining in-hospital Laboratory Tests to Target for Intervention: A Scoping Review

Eyal Podolsky, MSc (candidate)[1,2], Natasha Hudek, PhD[2], Kelly Carroll, MA[2], Christopher McCudden, PhD[2,3,4] Justin Presseau, PhD[1,2], Jamie Brehaut, PhD[1,2*]

[1]School of Epidemiology and Public Health, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, K1H 8M5, Canada

[2]Ottawa Hospital Research Institute, Clinical Epidemiology Program, The Ottawa Hospital, General Campus, 501 Smyth Road, Centre for Practice Changing Research, Box 201B, Ottawa, Ontario, K1H 8L6, Canada

[3]Division of Biochemistry, Department of Pathology and Laboratory Medicine, The Ottawa Hospital, 501 Smyth Road, Ottawa, ON K1H 8L6, Canada

[4]Eastern Ontario Regional Laboratory Association, 501 Smyth Road, Ottawa, ON K1H 8L6

*__Corresponding Author:__ Jamie Brehaut, The Ottawa Hospital, General Campus, 501 Smyth Road, Centre for Practice Changing Research, Box 201B, Ottawa, Ontario, K1H 8L6, Canada

Abstract

Background: It has been suggested that lab testing is a precursor for 70% of all medical decisions. Considering that 20-30% of tests may be inappropriately ordered, reduction of inappropriate use is a clear target for improving efficiency of care. Intervention development experts often have large datasets on the test-ordering practices within their jurisdiction but have limited to no clear guidance on how to use these data.

Methods: Our systematic search strategy was developed by a librarian scientist and reviewed by a second medical librarian as per Peer Review of Electronic Search Strategies (PRESS) guidance. Medical Subject Headings and title and abstract terms were chosen for the two broad categories of terms for 'Laboratory Tests', and 'Clinical Practice Guideline'; both guidance and empirical studies were eligible if they contained these terms. We extracted descriptive study/guidelines details, information on the tests involved, and factors guiding prioritization decisions around which tests should be targeted for intervention.

Results: We identified 80 studies that reported on their processes of identifying test ordering intervention targets, spanning a wide array of clinical disciplines and geographical regions. Overall, we identified 20 factors related to choices about where to put intervention resources, highlighting the complexity of these decisions. The most common factors we identified included test value, cost, prevalence of test, quality of test and actionability of test results.

Discussion: Intervention developers face difficult choices when deciding where to put their intervention resources. Our scoping review identifies a wide range of factors thought relevant to such choices. We see this review as the first step towards a prioritization framework that helps intervention developers decide which testing practices are worth their time, effort and money to attempt to change.

Registration: N/A

Keywords: Laboratory Testing, Low-Value, Scoping Review

Background

Laboratory testing is one of the highest volume activities in health care. Testing accounts for 3- 5% of all medical costs[1] and is thought to guide up to 70% of medical decisions determining subsequent, more costly care.[1-3] The demand for testing is also increasing.[4;5] Against this backdrop, estimates suggest that 20-30% of tests ordered are low-value, i.e. unnecessary, not indicated, or potentially harmful.[3;6] Inappropriate repeat ordering of six of the most common tests alone (cholesterol, hemoglobin A1c, thyroid-stimulating hormone, vitamin B12, vitamin D, and ferritin) is estimated to cost $160 million per year in Canada.[7]

One of the challenges to improving testing practices stems from the large amount of administrative test ordering data available, as well as the huge number of tests and circumstances under which these tests are ordered. Test-ordering intervention developers, i.e. those staff at individual institutions who are tasked with designing different approaches to improving test-ordering practice, can find it difficult to decide which among the hundreds of tests routinely ordered would most benefit care if targeted for intervention. Yet in order to improve test ordering efficiency, organizations need to choose where to intervene to reduce inappropriate testing and increase appropriate testing. These interventions must be worth the organization's time, effort and resources, in terms of improving efficiency on an organizational level. If these initiatives were only seen as cost-saving projects, they would be unlikely to garner engagement from critical stakeholders who order tests and care for patients

Large-scale initiatives such as *Choosing Wisely Canada* and National Institute for Health and Care Excellence (NICE) guidelines have sought to improve many clinical practices including the efficiency of test ordering.[9,10] *Choosing Wisely* guidance statements have developed over 250 recommendations pertaining to reducing low-value care, including unnecessary testing, treatments and procedures. [9] In terms of reducing testing, these recommendation do not always give clear guidance on which tests to focus on for reduction, such as those provided for critical care physicians.[11] In addition, individual organizations often want to be informed by, but not limited to, such general recommendations as those provided by *Choosing Wisely*. Such choices are complex, and many factors are and should be considered when prioritizing tests to target for intervention.[12]

To help inform decisions about where to put test ordering resources, we conducted a scoping review of factors that might inform such decisions. A scoping review often seeks to gain clarity on broader topics where many different study designs might be applicable, in contrast to the usual systematic review which might usually focuses on a well-defined question where appropriate study designs can be identified in advance (usually randomized controlled trials).[17] Additionally, quality assessment of studies is not usually performed in a scoping review whereas in the typical systematic review, this is a foundational step in the process. A key strength of the scoping study is that it can provide a rigorous and transparent method for mapping areas of research in a relatively short space of time (compared with full systematic review). This in turn makes it possible to identify the gaps in the evidence base, as well as summarizing and disseminating research findings quickly. With regards to weaknesses the scoping reviews does not appraise the quality of evidence in the primary research reports in any formal sense. The quantity of data generated can be considerable. This can lead to difficult decisions about how far

breadth (covering all available material) is more important than depth (providing a detailed analysis and appraisal of a smaller number of studies).

Our goal is to inform a generalizable, evidence-based prioritization framework to guide healthcare organizations in determining which tests should be prioritized for intervention. The resulting evidence-based prioritization framework will describe what factors to consider, how to prioritize among factors, and how to measure the factors in the local environment.

METHODS

Design

Based on scoping review guidance established by Levac[13] and Tricco[14] we defined our objectives as:

- **Objective I:** Identify all the factors reported by the authors as being important in determining how they chose to prioritize one test over another when targeting testing for intervention.
- **Objective II:** Gather information on the processes used by intervention developers to determine that target tests were worthy of intervention.

Protocol and Registration

Our results have been reported as per the PRISMA-ScR guidelines[15], and a PRISMA-ScR checklist[15] was completed to document the inclusion of all critical elements of this review (Appendix D). We registered with the University of Ottawa study registration database.

Eligibility Criteria

Inclusion: Studies with the following PICOS characteristics were included in the review:

Population: Studies that were guidelines or recommendation statements pertaining to test-ordering practice. The guidelines could be about decreasing, increasing or improving appropriateness of testing in the clinical setting. Empirical studies were also included if they made clear recommendations about changing test-ordering practice.

Intervention: All forms of guidance statements were accepted if they pertained to improved test-ordering practice. Empirical studies could involve any form of intervention, including but not limited to educational, audit and feedback (A&F), and computer provider order entry (CPOE) changes. Every included article had to make some sort of recommendation about improving test-ordering practice.

Comparator: The intervention of interest was recommendations about improving testing practice and thus as a result, we did not use a comparator in this review.

Outcomes: The articles needed to report on at least one factor (e.g. prevalence of test, cost of test) relevant to guiding decisions about target tests for intervention. In addition, it was preferable that the study discussed a process that was used (i.e. literature review or consensus process) in order to make an effective decision of choosing one test versus another for intervention. The term “unclear” was assigned to a study that did not effectively describe the

processes (i.e. literature review, consensus process) that was followed in order to make the decision to pursue certain tests over others for intervention.

Study design: We included any type of guideline or recommendation statement as well as any empirical study that directly focused on improved test-ordering practice.

Setting: We assessed studies stemming from tertiary care settings including all clinical specialties. Any article that focused on primary care was excluded. There is already a large body of work about testing in the community and inappropriate testing in the community vs the hospital is quite different. As a result, we felt that it was most important to focus on the hospital setting for the purpose of our work.

EXCLUSION: No study time restrictions were imposed. We excluded conference abstracts, commentaries and letters to the editor, as well as studies not published in English to maintain feasibility. Empirical studies not primarily focused on improved test-ordering practice, and studies not reporting any factors involved in choosing one test over another, were excluded.

Search Strategy Development and Information Sources

Our Medline search strategy (Appendix C) was developed with help from an information specialist using the PRESS guidance for Peer Reviewed search strategies.[16] Medline was the only database used primarily because of resource availability, as the initial search yielded an enormous number of articles that would have made the review unfeasible had it been run in other databases as well as a result of sheer size. After the search parameters were finalized, the search strategy was reviewed by a second medical librarian. Medical Subject Headings (MeSH terms) and title and abstract terms ('.tw') were chosen for the two broad categories of terms for 'Laboratory Tests', and 'Clinical Practice Guideline' and empirical studies were eligible if they contained these terms as well. This search included all available years up till May 30th, 2019. The bibliographies of included articles and relevant systematic reviews were also hand searched to identify any further articles meeting the inclusion criteria.

Study Records

Data Management: Citations retrieved from the search were imported into the reference manager software program Mendeley Desktop 1.17.12 (Mendeley Ltd., London, UK) for de-duplication, then imported into Covidence, a web-based platform for screening.

Selection Process: The titles and abstracts of unique citations identified from electronic database searches were screened by two independent reviewers (EP and NH), and screening through reference lists as well as any further snowball searching by one reviewer (EP). During the abstract stage, screeners searched articles that included discussion of improved test-ordering practice through a clinical practice guideline or empirical study to improve testing. Conflicts were resolved through discussion or reference to a third independent reviewer (JCB). Full text articles were screened by two reviewers (EP and NH), and justifications for inclusion or exclusion were agreed upon through consensus discussions. In this stage of review, the articles were required to provide guidance statements about the appropriate use of at least one clinical

test. Additionally, it was necessary for the article to discuss at least one factor guiding the decision to target the test and discussion as to why certain factors were more important than others were considered as important information as well. Conflicts again were resolved through discussion between the two primary reviewers (EP and NH) or third reviewer (JCB) if consensus could not be reached (Details in Appendix C).

Data Collection Process: All data was extracted by two independent reviewers (EP and NH) using a standardized data extraction form (example of form in Appendix F) implemented in Microsoft Excel 2011. Both reviewers piloted the form on the six randomly selected articles and only minor refinements were required. Conflicts between data extraction forms were identified by one reviewer (EP), and consensus was reached between reviewers through discussion (EP and NH). If reviewers were not able to come to an agreement, a third reviewer (JCB) was consulted to reach consensus.

Data Extracted

We extracted three broad categories of information: 1) Descriptive study/guidelines details 2) Test information 3) Factors guiding prioritization decisions. Descriptive details included publication date, journal of publication, funding source, type of study (empirical study, KT study, guidance document or other), intervention type (if empirical). Test information included clinical specialty, number of tests discussed, test names, individual tests or a bundle of tests. Prioritization details included factors considered when deciding which tests to target for intervention (e.g cost, lab workload) and processes used to arrive at these decisions (e.g literature review, consensus process).

Risk of Bias

In this study we did not collect quantitative outcomes and thus an assessment of risk of bias was not considered useful, which is consistent with Arksey & O'Malley[1-] that scoping reviews do not typically include a risk of bias assessment.

Data Synthesis and Analysis

As this was a scoping review and not a systematic review, we deemed it unnecessary to perform a meta-analysis. We created tables of study characteristics as well as tables summarizing the frequency counts of each factor and decision processes reported in our included studies.

RESULTS

Study Selection

Figure 1 describes our screening process. Starting from 8,749 citations (extracted from MEDLINE database on May 30th, 2019 after removal of duplicates and two rounds of screening, 80 unique studies were identified for inclusion.

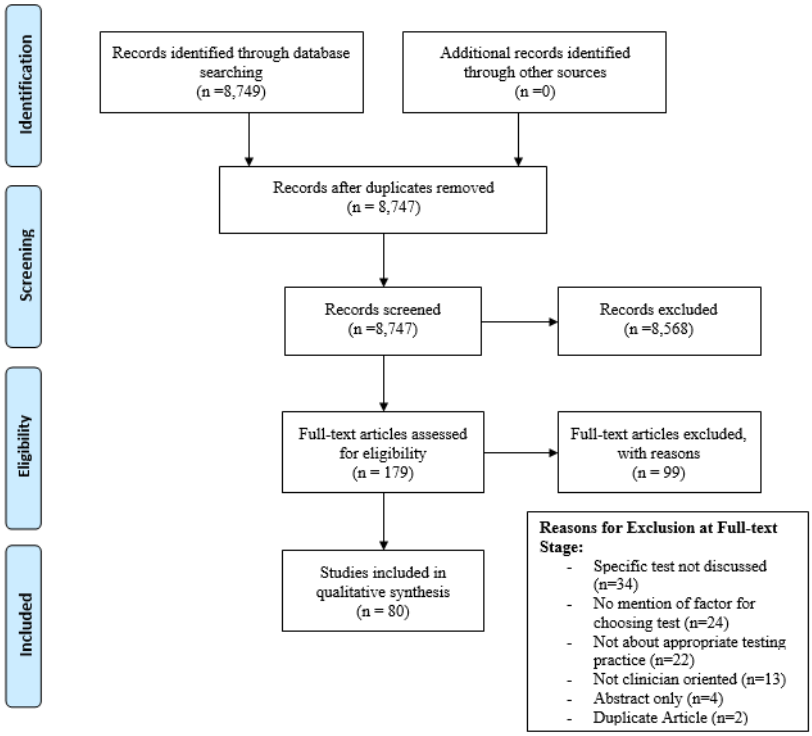

Figure 1. PRISMA Flow Diagram. A list of the excluded full text articles, sorted by reason for exclusion, can be found in Appendix E.

Study Characteristics

Table 1 describes characteristics of the included studies (n=80). Most were from the USA (64%), with smaller percentages coming from the UK, Italy, Canada, Israel, South Africa and France. The bulk of the studies were published after 2000 (83%). Most of the studies did not report on their funding source (74%) but among those that did, privately funded studies represented 11% of the studies and publicly funded made up 15% of studies. The included studies covered a broad range of clinical specialties including oncology (11%), internal medicine (10%), hematology (9%) and rheumatology (9%). Three tests were most commonly targeted: Complete Blood Count (CBC) (29%), Glucose (11%) and Partial Thromboplastin Time (PTT) (11%). About half (48%) were a guideline or recommendation statement, 28% were empirical and the rest were either a review or a combination of guideline and empirical study (25%). Most of the studies (57%) targeted reduction in tests while most of the others (42%) included efforts to decrease some tests and increase others. Out of the 42 studies that included an empirical component three intervention types were most commonly used: system based (31%), educational (29%), and audit and feedback (12%).

Table 2 describes the processes that the articles described as leading to decisions about prioritizing certain tests over others. The processes reported were rarely mentioned alone, meaning that articles that did report the processes that they followed, tended to report a few different processes at once. The most reported process was some form of **literature review** (56%), followed by **clinical standard-internal** (39%) (an internal standard of practice that is not formally established by the governing body) and consensus process (38%). Similarly, **clinical standard-external** (existence of an external guideline or other formalized procedure) was reported in about one-third of the articles (33%). Processes less frequently reported included: consultation with the **local team** (11 **%), chart review** (10%), **vetting of a guideline** (6%), **case review** (4%) and **consulting local data (3%)**. Additionally, an "other" process that could not be well categorized was seen in 11% of the studies. Another option that was very frequently reported was **unclear** (34%) which was indicative that reporting was not sufficiently detailed about the processes they used to choose tests. Appendix F contains a break-down of every process reported in each article.

Table 3 summarizes the frequency with which factors were identified as the reason that certain tests were chosen among our 80 articles. Among our list of 18 factors there were three factors that appeared most commonly: **test value** (80%), **cost** (78%) and **prevalence of test** (65%). Other commonly cited factors included **test quality** (53%), **impact of false positives** (53%), the **quality of evidence for/against using the test** (51%), **the actionability of test results** (46%), **patient care** (41%), **the existence of a guideline** (41%) and **the prevalence of the disease** (40%). Less commonly cited factors included: **relevance to current practice** (33%), **the risk of the test** (26%), **lab workload** (19%), **bundling of multiple tests** (18%), **anecdotal expert experience** (15%) **evidence of inappropriate use** (15%) and the **strain on the health-care system.** Finally, two factors were mentioned very rarely, to the extent that it was surprising how little they were discussed: **the feasibility of changing test-ordering behavior** (11%), the **ease with which the intervention could be implemented** to improve testing practice (6%). In a significant majority of the articles (61%) an additional "other" factor was mentioned that was beyond the list of 18 factors. The various factors came relatively equally from guideline/recommendation papers and empirical studies. Factor labels were identified by consensus as a way to amalgamate over-arching ideas from the wide variety of studies.

Discussion

We performed a scoping review to examine guideline statements and previous experimental interventions implemented to improve test-ordering practice. The review gathered all the factors reported by the authors as being important in determining how they chose to prioritize one test over another for intervention. In our scoping review, we sought to include studies that focused on both de-implementation (reducing the ordering of a certain test) and implementation (increasing the ordering of a certain test), because both were thought to be essential processes to improve efficiency of test ordering. We identified 80 studies, spanning a wide array of clinical disciplines and geographical regions. Taken together, these results indicate that deciding where to put resources to improve test-ordering practice is a complex problem felt across the world and across a wide range of disciplines. Of the 80 articles that we extracted, only two articles focused primarily on implementation (increasing a test), while 78 approached the

problem from a de-implementation lens. Thus, our review primarily emphasized issues with respect to de-implementation.

Quality of reporting was also an issue, with 34% of studies rated as Unclear in their reporting of the factors guiding this decision. It is unclear which approaches may have led to better intervention; better reporting of processes in studies evaluating the effectiveness of interventions would improve this literature.

Our review showed the processes used to arrive at the decision about which tests to prioritize for interventions, when specified, are relatively common across clinical areas. The most common of these were literature reviews, where developers looked to the literature for guidance about some aspect of the decision. Another common process was undergoing a consensus with national/international experts in the specific field, usually meaning gathering the opinions of the experts and synthesizing them into a decision to be followed locally. A less common but perhaps more rigorous process involved the vetting of a guideline, i.e. testing if the guideline was effective in producing the desired test-ordering behaviour by implementation in the local context. Overall, our review suggests room for improved rigour in both the reporting and conduct of the processes leading to these (sometimes involved and expensive) test ordering interventions. Perhaps, if the included studies had used more systematic processes (e.g. systematic reviews instead of literature reviews, building more diverse teams of experts for their consensus processes) more effective interventions could be developed.

Overall, we identified 18 factors indicated as reasons for targeting certain tests for intervention, highlighting the complexity of these decisions. Some of the most common factors we identified as relevant to most situations, included test value (80%), cost (78%) and prevalence of test (65%). Given that these initiatives are often conducted in the context of shrinking budgets, their prevalence is not surprising. In practice change literature, the term “test value” is typically defined as quality/cost, meaning the test value is whether the test in question is the most efficient based on its cost. In the context of this project, the term was used more loosely to refer to a test that impacted patient care decisions such a modifying treatment or aiding a clinician in establishing an official diagnosis and is more often labelled “test utility” in the literature. Additionally, we found that the implications of a false positive (53%) and whether the results of the test were actionable (46%) were also quite commonly considered, highlighting that decision-making considerations often explicitly include not only considerations of the test-ordering practice’s impact on the financial bottom-line but also the quality of care for patients. Our review also suggests that the range of relevant factors can extend well beyond these few, and that at least consideration of a broader range of factors may contribute to intervention development discussions.

Reporting guidelines provide structured advice on what information needs to be included in a research article to allow readers to assess the study methodology, relevance, and validity of presented findings. The most common of which are those published by the EQUATOR Network online Library for Health Research Reporting. However, in our study the use of a completeness of reporting assessment in these papers was not relevant. Even if an article mentioned only one factor, it would still be a useful addition to our study, even if another study might report several factors. One study would not be considered less ‘complete’ than the other it would simply be describing any factor and processes that was considered rather than be considered lacking in its completeness of reporting.

Limitations of Study

Our study had several limitations that warrant consideration. Our search strategy was implemented in only one database (Medline) for resource reasons. While Medline identified all our initial target articles, future work might expand the search to include other common databases (Embase, CINHAL, Psycinfo). We sought to reduce the effects of this limitation by enhancing our review with snowball sampling to ensure that any relevant articles that may not have been in the initial search strategy were included. Medline is a very comprehensive database that includes all of PubMed and thus most of the clinician-oriented literature. However, in missing out on CINHAL, which primarily covers the nursing literature, we may have missed out on interventions that targeted improved test-ordering practice among nurses or other allied-health professionals. While our review identified a wide range of factors potentially relevant to the decision of choosing which test to target for intervention, we could not assess which ones are most important overall, or in which circumstance some will become more important. Our study could have been limited by publication bias as it pertains to our included empirical studies, given that only those that succeeded in showing improvement of test ordering appropriateness might have been published. Finally, when categorizing whether a factor was mentioned or not, we used a rather inclusive approach such that almost any indication that an intervention developer referred to that factor was included. In some cases that factor may not have been instrumental in the planning stage of the study, but rather only given limited consideration or even considered post-hoc. While we tried to limit this by only including factors that were mentioned in the introduction or methods sections of the study, the limitation remains and could be addressed by more detailed discussion/interviews with participants in the decision.

In order to facilitate reporting of these issues in future work, we propose guidance development authors describe their specific rationale for considering certain factors (e.g. costs, quality of evidence) and not considering others. Such reporting would assist others in making decisions on which factors are most important based on similar or dissimilar circumstances. Better reporting of processes that led to the guidance would also benefit readers seeking to decide whether to implement the guidance, and future systematic assessments of whether different processes lead to more effective interventions.

A number of studies (i.e Ament 1993, Alonso-Cerezo 2009, Beland 2003, Burke 1981, Glasziou 1989, Pannall 1996, Rubinstein 2018, Wachtel, 2020) were included during the full-text section of our review. On more detailed examination these studies were excluded because they did not specifically address and/or measure the change of at least one clearly identified test. Thus, because they did not identify specific tests and the experience of changing their ordering in a real-world setting, these papers were seen as more hypothetical rather than practical and were therefore excluded. We do however acknowledge that there can be value in considering and discussing such factors in future, more broadly inclusive work.

Conclusion

Test-ordering intervention developers face difficult choices when deciding where to put their intervention resources. Our scoping review identifies a wide range of factors thought relevant to such choices. We see this review as the first step towards a prioritization framework that helps developers decide which testing practices are worth their time, effort and money to attempt to change. Future work will include interviews with intervention development experts in order to help contextualize the factors and processes identified in this review and develop guidance for developers.

RESULTS TABLES

Table 1. Characteristic of studies in test-ordering interventions scoping review.

Characteristics	Number (%) of studies (n=80)
Country of Article	
USA	51 (63.8)
UK	9 (11.3)
Italy	6 (7.5)
Canada	4 (5.0)
Other	10 (12.5)
Year of Publication	
1980-2000	14 (17.5)
2000-2019	66 (82.5)
Funding	
Not Discussed	59 (73.8)
Publicly Funded	12 (15.0)
Privately Funded	9 (11.3)
Clinical Specialty	
Oncology	9 (11.3)
Internal Medicine	8 (10.0)
Hematology	7 (8.8)
Rheumatology	7 (8.8)
Anesthesiology	6 (7.5)
Gastroenterology	6 (7.5)
Infectious Disease	5 (6.3)
Critical Care	4 (5.0)
Pediatrics	4 (5.0)
Other	24 (30.0)
Most Commonly Targeted Tests	
Complete Blood Count (CBC)	23 (28.8)
Glucose	9 (11.3)
Prothrombin Time	9 (11.3)
Type of Article	
Guideline/Recommendation Statement	38 (47.5)
Empirical Study	22 (27.5)
Other	20 (25.0)
Guideline + Empirical	8 (10.0)
Review	6 (7.5)
Guideline+ Intervention	5 (6.3)
Guidance Endorsement	1 (1.3)
Target of Guidance or Intervention	
Decreasing Testing	45 (56.3)
Both	34 (42.3)
Increasing Testing	1 (1.3)
Type of Intervention (Empirical +Other, N=42)	
System Based	13 (31.0)
Educational	12 (29.0)
Other	12 (29.0)
Audit and Feedback	5 (11.9)

Table 2. Processes reported to inform decisions about which tests to target for intervention.

Processes	Number (%) of articles reporting process (n=80)	Descriptive Example
Literature Review	45 (56.3)	"SHM staff conducted a literature review of the list of tests and treatments…" (Bulger, 2013)
Clinical Standard-Internal	31 (38.8)	"Laboratory testing and diagnostic imaging are routinely used for the management of children with community-acquired pneumonia." (Parikh, 2016)
Consensus Process	30 (37.5)	"Using nominal group technique,4 the ASH CWTF reduced the list of suggested Choosing Wisely items to a short list of 20." (Hicks, 2013)
Clinical Standard-External	27 (33.8)	"Because Hemoccult II is the most studied and most commonly used of the faecal occult blood tests, it is appropriate to use it as the 'criterion standard" against which other faecal occult blood tests can be compared." (Allison, 1998)
Local Team	9 (11.3)	"The Medical Evaluation Committee of the hospital developed clinical guidelines for tumor marker ordering." (Durieux, 2003)
Chart Review	8 (10.0)	"Electronic medical record data containing all thrombophilia tests ordered on inpatients from June 2013 to June 2015 were obtained." (Mou, 2017)
Vetting a Guideline	6 (7.5)	"…to test the effectiveness of newly developed guidelines for obtaining blood cultures in pediatric patients with CAP through retrospective chart review before institutional guideline adoption." (Heine, 2013)
Case Review	5 (6.25)	"We thoroughly reviewed the charts of all patients who had any of the following…" (Tapper, 2013)
Consulting Local Data	3 (3.8)	"We performed a laboratory audit of specimens received from one hospital site that provides inpatient and ambulatory care services for adult and pediatric patients. The audit cycle included an initial audit, the implementation of interventions and a repeat audit" (Huang, 2012)
Pilot Study	1 (1.3)	"The first part was designed to limit certain specific tests by limited justifications, and the second phase broadened the approach to limit the most commonly ordered laboratory tests in both hematology and biochemistry." (Novich, 1985)
Unclear	27 (33.8)	N/A
Other	9 (11.3)	"…the USPSTF commissioned a decision analysis to help clarify the age at which to begin and end screening, the optimal interval for screening, and the relative benefits and harms of different strategies for screening." (Decision analysis) (Moyer, 2012)

Table 3. Factors stated as rationale for choosing certain tests over others.

Factors	Definition	Example	Number (%) of articles reporting factor (n=80)
Test Value	The clinical utility of the test according to the health-care provider.	"We identify 5 common laboratory tests whose use persists in dermatologic practice despite evidence confirming their limited utility." (Ogbechie-Godec, 2018)	64 (80.0)
Cost associated with Test	The amount spent in order to collect, analyze and/or interpret the lab-test and associated fees.	"… blood tests are expensive both in terms of economic costs of laboratory and equipment resources, in addition to increased workload incurred on junior medical staff and phlebotomists." (Faulkner, 2016)	62 (77.5)
Prevalence/Frequency of the Test	The number of times that a test is ordered and/or the volume that the test is ordered patient/day.	"An evaluation of this issue should consider the frequency of abnormal test results within a given population…" (ASGE, 1999)	52 (65.0)
Test Quality	The diagnostic characteristics of the test in question such as sensitivity and specificity.	"Considering the low specificity of the ANA test in the diagnosis of autoimmune diseases…" (Mann, 2013)	42 (52.5)
Implications of a False Positive	Any negative effects associated with an incorrect disease diagnosis (e.g. further testing, unnecessary invasive procedures, unnecessary cost incurred).	"…falsely abnormal test results may unnecessarily delay endoscopy and subject the patient to additional risks…" (ASGE, 1999)	42 (52.5)
Quality of supporting evidence for/against using the test	High caliber evidence provided by clinical practice guidelines, systematic reviews or other peer reviewed publications which point to the utility or lack of utility of a test.	"Arterial blood gas analysis is not supported by strong evidence and seems to be driven by cultural factors." (Martinez Balzano, 2017)	41 (51.3)
Actionability of Test-Results	Tests that directly the treatment or management plan for the patient.	"… "clinicians were likely to act on the results of the test." (Pilon, 1997)	37 (46.3)
Patient Care	Impact of the test on patient experience (e.g. be physical pain due to test or feelings of wasted time or anxiety).	"… confers no benefit and leads to unnecessary surveillance, diagnostic tests, and treatments with the associated harms." (Moyer, 2012)	33 (41.3)

Prior existence of a guideline for the specific test(s)	A well-established protocol backed by a governing body which provides direction on appropriate use of the test.	"In 2006 the National Institute for Health and Clinical Excellence (NICE) published a guideline entitled, 'Anaemia management in people with chronic kidney disease." (Thomas, 2013)	33 (41.3)
Prevalence of target disease	The prevalence rate of the disease which is being detected using the test in question.	"In one study of 5003 patients tested prior to elective cholecystectomy, there were only four detected cases (0.08%) that could not have been anticipated by history and physical examination." (Wu, 1998)	32 (40.0)
Relevance to current practice	Test being considered relates to clinician area of practice (e.g. test is outdated by newer test).	"…the usefulness of these autoantibodies in clinical practice still has to be determined." (Csernok, 2018)	26 (32.5)
Risk/harm of administering test	Any potential negative consequences for the patient in administering the test.	"Blood tests can induce iatrogenic anaemia in patients…" (Merlani, 2001)	21 (26.3)
Laboratory workload	The quantity of human resources required to run, process and/or analyze a specific test within the pathology lab.	"The anti-FXa assay can be carried out even in emergency settings, with little expertise…" (Tripodi, 2018)	15 (18.8)
Test being bundled with other tests	Test combined on paper or electronic ordering systems as bundles or order sets.	"… avoid the predetermined packages offered by various laboratories because, in most instances, they provide a mix of useful tests and irrelevant tests." (Griffin, 1996)	14 (17.5)
Expert experience	Guidance from local or external experts based on personal practice regarding the test.	"The opinion of experts about the appropriateness of use of procalcitonin was assessed in different clinical settings." (Bartoletti, 2018)	12 (15.0)
Evidence of inappropriate use	Data demonstrating that the test in question is used inappropriately compared to what would be expected based on practice guidelines or standard of practice.	"We superimposed the guidelines on levels that were performed and found that 74% of inappropriately ordered inpatient serum AED monitoring was due to a common practice…" (Chen, 2003)	12 (15.0)
Strain on health-care system	Overall negative impact on the health care system resulting from inappropriate use of the test (e.g. financial strain, as well as human-time costs incurred).	"Additional blood coagulation tests would yield nothing in the vast majority of the cases, thus creating unnecessary burdens to the child and his family, as well as to the economy of the health care system." (Toker, 2004)	11 (13.8)
Feasibility of changing test-ordering behaviour	The perception that the test-ordering practice behavior can be changed through	"The feasibility of implementing any nationwide policy." (Balk, 2007)	9 (11.3)

	various intervention strategies.		
Ease of implementation of intervention for changing test-ordering behaviour	The quantity of time, effort and financial as well as human resources that are needed in order to apply an intervention to change ordering behaviours of the test in question.	"…the ease in implementation of avoidance." (Onuoha, 2014)	5 (6.3)
Other	Any other factor listed in the articles that cannot be categorized as one of the factors above.	"Are there medicolegal considerations for performing the investigations?" (Medico-legal factor) (Barnard, 1994)	49 (61.3)

Declarations

Ethics Approval and Consent to Participation

Not applicable

Consent for Publication

Not Applicable

Availability of Data and material

The datasets used and/or analyzed during the current study are available from publicly accessible databases.

Competing Interests

The authors declare that they have no competing interests

Funding

EP received a University of Ottawa Graduate Studies Scholarship and held a graduate studentship with the Ottawa Hospital Research Institute. Funding bodies had no role in the design of the study, collection, analysis, interpretation of data, or in the writing of the manuscript.

Authors' Contributions

JCB was responsible for the conception of this project and provided guidance and expertise throughout the entire project. EP drafted the manuscript, and JCB and NH, provided critical input and aided in the revision of the manuscript. EP and NH completed title, abstract and full-text screening. Data extraction and quality assessment were completed by EP and NH. The guarantor of this review is EP.

Acknowledgements

The authors would like to thank Lindsey Sikora and Risa Shorr for their help in developing and reviewing the search strategy.

Appendix A: Medline Search Strategy

Database: Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Daily <1946 to May 28, 2019>

Search Strategy:

---

1 (chemistry adj3 test*).ti,ab. (1011)

2 (clinical adj3 lab*).ti,ab. (86653)

3 (laboratory adj3 diagnos*).ti,ab. (18441)

4 (laboratory adj3 technique*).ti,ab. (3760)

5 (laboratory adj3 exam*).ti,ab. (15570)

6 (laboratory adj3 test*).ti,ab. (49880)

7 (cytologic* adj3 technic*).ti,ab. (94)

8 (cytologic* adj3 technique*).ti,ab. (700)

9 (genetic adj3 screening*).ti,ab. (7131)

10 (genetic adj3 test*).ti,ab. (27659)

11 (genetic adj3 predisposition).ti,ab. (11722)

12 (genetic adj3 predictive).ti,ab. (1080)

13 (blood adj3 test*).ti,ab. (40458)

14 (hematologic* adj3 screen*).ti,ab. (157)

15 (hematologic* adj3 test*).ti,ab. (1163)

16 (haematologic* adj3 screen*).ti,ab. (101)

17 (haematologic* adj3 test*).ti,ab. (420)

18 (histologic* adj3 technic).ti,ab. (89)

19 (histologic* adj3 technique*).ti,ab. (3409)

20 immunodiagnos*.ti,ab. (3346)

21 (immunologic* adj3 diagnos*).ti,ab. (1677)

22 (immunologic* adj3 test*).ti,ab. (4169)

23 (preanalytic* adj3 phase*).ti,ab. (231)

24 (pre-analytic adj3 phase).ti,ab. (19)

25 (pregnancy adj3 test*).ti,ab. (5678)

26 (radio* adj3 assay*).ti,ab. (16077)

27 (specimen adj3 collection*).ti,ab. (2240)

28 (specimen adj3 handling*).ti,ab. (475)

29 diagnostic services/ or clinical laboratory services/ or diagnostic screening programs/ or "direct-to-consumer screening and testing"/ (2821)

30 (diagnostic adj3 service*).ti,ab. (2657)

31 (diagnostic adj3 screening*).ti,ab. (4933)

32 *Laboratories/ (15471)

33 pathology/ or pathology, clinical/ (37680)

34 (clinical adj3 pathology).ti,ab. (8770)

35 exp Laboratories, Hospital/ (4586)

36 (Carcinoembryonic adj3 test*).ti,ab. (165)

37 (Surveillance adj3 test).ti,ab. (369)

38 *Blood Cell Count/ (3363)

39 (blood adj2 cell adj2 count*).ti,ab. (21459)

40 (Blood adj2 cell adj2 number*).ti,ab. (465)

41 (complete adj2 blood adj2 count*).ti,ab. (7159)

42 *Blood Chemical Analysis/ (13042)

43 (blood adj2 chemical adj2 analys*).ti,ab. (107)

44 *Blood Gas Analysis/ (4600)

45 (blood adj2 gas adj2 analys*).ti,ab. (5882)

46 Antibodies, Antinuclear/ (14445)

47 exp Fluorescent Antibody Technique, Indirect/ (16415)

48 (antinuclear adj2 antibody adj2 test).ti,ab. (218)

49 (Total adj2 testosterone adj2 test).ti,ab. (7)

50 (morning adj2 testosterone adj2 level).ti,ab. (21)

51 (free adj2 testosterone adj2 level).ti,ab. (211)

52 (free adj2 testosterone adj2 test).ti,ab. (5)

53 (ANA adj2 test).ti,ab. (273)

54 (Human adj2 papillomavirus adj2 test).ti,ab. (283)

55 Vaginal Smears/ (21684)

56 Colonoscopy/ (24898)

57 Sigmoidoscopy/ (4666)

58 sigmoidoscopy.ti,ab. (3751)

59 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 (467202)

60 exp Critical Pathways/ (6273)

61 exp Consensus/ (10536)

62 exp Consensus Development Conference/ (11406)

63 exp Consensus Development Conferences as Topic/ (2698)

64 *Practice Guidelines as Topic/ (38663)

65 exp clinical pathway/ (6273)

66 exp Health Planning Guidelines/ (4033)

67 (guideline or practice guideline or consensus development conference or consensus development conference, NIH).pt. (40921)

68 (position statement* or policy statement* or practice parameter* or best practice*).ti,ab,kf,kw. (29188)

69 ((practice or treatment* or clinical) adj guideline*).ab. (35984)

70 (CPG or CPGs).ti. (5454)

71 consensus*.ti,kf,kw. (23465)

72 consensus*.ab. /freq=2 (22717)

73 recommendat*.ti,kf,kw. (37458)

74 (care adj2 (standard or path or paths or pathway or pathways or map or maps or plan or plans)).ti,ab,kf,kw. (51830)

75 Cost-Benefit Analysis/ (76540)

76 *Health Policy/ (33775)

77 *Guideline Adherence/ (14257)

78 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70 or 71 or 72 or 73 or 74 or 75 or 76 or 77 (366569)

79 59 and 78 (10955)

80 limit 79 to (english language and humans) (8748)

Appendix B: List of Target Articles Required for Search String to Proceed

Article #	Citation
1	Meyerhardt JA, Mangu PB, Flynn PJ, et al. Follow-up care, surveillance protocol, and secondary prevention measures for survivors of colorectal cancer: American Society of Clinical Oncology clinical practice guideline endorsement. J. Clin. Oncol. Dec 10 2013;31(35):4465-4470.
2	Qaseem A, et al. Screening for colorectal cancer: A guidance statement from the American College of Physicians. Ann Intern Med. 2012 Mar 6;156(5):378-86.
3	Czoski-Murray C, et al. What is the value of routinely testing full blood count, electrolytes and urea, and pulmonary function tests before elective surgery in patients with no apparent clinical indication and in subgroups of patients with common comorbidities: a systematic review of the clinical and cost-effective literature. Health Technol Assess. 2012 Dec;16(50):i-xvi, 1-159. PMID: 23302507.
4	Garber JR, et al. Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. Endocr Pract. 2012 Nov-Dec;18(6):988-1028
5	Kavanaugh A, Tomar R, Reveille J, Solomon DH, Homburger HA. Guidelines for clinical use of the antinuclear antibody test and tests for specific autoantibodies to nuclear antigens. American College of Pathologists. Arch Pathol Lab Med. 2000 Jan;124(1):71-81.
6	Baron EJ, et al. A guide to utilization of the microbiology laboratory for diagnosis of infectious diseases: 2013 recommendations by the Infectious Diseases Society of America (IDSA) and the American Society for Microbiology (ASM). Clin Infect Dis. 2013 Aug;57(4):e22-e121
7	Bhasin S, et al. Testosterone therapy in adult men with androgen deficiency syndromes: an endocrine society clinical practice guideline. J Clin Endocrinol Metab. 2006 Jun;91(6):1995-2010
8	Moyer VA, et al. Screening for cervical cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2012 Jun 19;156(12):880-91, W312.
9	Attali M, et al. A cost-effective method for reducing the volume of laboratory tests in a university-associated teaching hospital. Mt Sinai J Med. 2006 Sep;73(5):787-94.
10	Canadian Task Force on Preventive Health Care, et al. Recommendations on screening for cervical cancer. CMAJ. 2013 Jan 8;185(1):35-45.

Appendix C: Details on Search Strategy and Sample

Detailed Search Strategy Development:

An initial scoping review search strategy was developed with help from a library information specialist. Medical Subject Headings (MeSH terms) and title and abstract terms ('.tw') were chosen for the two general categories 'Laboratory Tests' and 'Clinical Practice Guideline'. The key words and MeSH terms of relevant primary studies and systematic reviews were assessed during the development process. The initial search strategy was then reviewed by a second, independent library information specialist.

Initially, a list of 35 articles was compiled from the reference lists of the "Choosing Wisely" recommendations that pertained to test-ordering. This list was further reduced to 10 articles based on our belief that these 10 articles were models of the types of articles that would be most effective in helping us answer our research question. These 10 articles represented the target articles for validating our search strategy. It was decided *a priori* that any search strategy needed to yield at least nine of the target articles, and if it did not, small modifications needed to be made to increase the sensitivity of the search. Based on these 10 target articles, the terms used in the search strategy were decided upon. The target articles have been attached in **Appendix B.**

Inclusion Criteria:

A) A term pertaining to lab-based testing such as: *chemistry test, cytological technique, histological test, blood gas analysis.*
B) A term pertaining to clinical practice guidelines such as: *practice guideline, clinical pathway, recommendation, health planning guideline.*

Exclusion Criteria:

A) No time restriction on publication date was applied.
B) In order to be included, the studies were required to be published in English, thus publication in any other language was basis for exclusion.
C) All animal-based studies were excluded.

<u>Abstract Screening:</u>

Inclusion Criteria:

A) The article needed to be a systematic review or systematically developed guideline in order to be included.
B) The article needed to indicate that it was a guidance statement on appropriate test-ordering practice.

Exclusion Criteria:

A) If the study was an empirical study, this was basis for exclusion. Unless the empirical study was expressly designed to improve test ordering practice.
B) If the main focus of the study was on pathological practices, the study was excluded.
C) If the study did not meet any of the other inclusion criteria for the abstract screening level such a stating that it was a guidance statement about appropriate test-ordering practice.
D) If the article was a review of other guideline statements and not an originally proposed guideline.

Before definitively selecting articles, we conducted a calibration exercise to evaluate and increase the reliability of title and abstract screening. This entailed two authors independently screening the first 500 articles. Following this, the two reviewers met to discuss the ease with

which inclusion criteria were applied and how to better define the criteria to ensure congruent application of the criteria. All conflicts were discussed and this allowed for improvement of the specificity of the criteria, with a 95% inter-rate reliability being the goal.

Full-Text Review:

Inclusion Criteria:

A) The article must include an explicit guideline or guidance about appropriate use of a clinical test (i.e complete blood count (CBC), troponin test) by a clinician.

B) The study must provide a description of at least one factor (i.e. relevance to practice, cost, test value) that should be considered when deciding what tests should be targeted for reduction.

C) There must be a discussion of a systematic processes by which it was decided which factors are most important in deciding which tests to target for reduction.

Exclusion Criteria:

A) If the study did not meet *all* the inclusion criteria mentioned above, it was excluded during full-text review.

Appendix D: Completed PRISMA-ScR Checklist

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
TITLE			
Title	1	Identify the report as a scoping review.	15
ABSTRACT			
Structured summary	2	Provide a structured summary that includes (as applicable): background, objectives, eligibility criteria, sources of evidence, charting methods, results, and conclusions that relate to the review questions and objectives.	16
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known. Explain why the review questions/objectives lend themselves to a scoping review approach.	17
Objectives	4	Provide an explicit statement of the questions and objectives being addressed with reference to their key elements (e.g., population or participants, concepts, and context) or other relevant key elements used to conceptualize the review questions and/or objectives.	17
METHODS			
Protocol and registration	5	Indicate whether a review protocol exists; state if and where it can be accessed (e.g., a Web address); and if available, provide registration information, including the registration number.	17-18
Eligibility criteria	6	Specify characteristics of the sources of evidence used as eligibility criteria (e.g., years considered, language, and publication status), and provide a rationale.	18
Information sources*	7	Describe all information sources in the search (e.g., databases with dates of coverage and contact with authors to identify additional sources), as well as the date the most recent search was executed.	18
Search	8	Present the full electronic search strategy for at least 1 database, including any limits used, such that it could be repeated.	19
Selection of sources of evidence†	9	State the process for selecting sources of evidence (i.e., screening and eligibility) included in the scoping review.	30-32
Data charting process‡	10	Describe the methods of charting data from the included sources of evidence (e.g., calibrated forms or forms that have been tested by the team before their use, and whether data charting was done independently or in duplicate) and any processes for obtaining and confirming data from investigators.	19
Data items	11	List and define all variables for which data were sought and any assumptions and simplifications made.	20
Critical appraisal of individual sources of evidence§	12	If done, provide a rationale for conducting a critical appraisal of included sources of evidence; describe the methods used and how this information was used in any data synthesis (if appropriate).	N/A
Synthesis of results	13	Describe the methods of handling and summarizing the data that were charted.	20
RESULTS			
Selection of sources of evidence	14	Give numbers of sources of evidence screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally using a flow diagram.	20

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
Characteristics of sources of evidence	15	For each source of evidence, present characteristics for which data were charted and provide the citations.	21-22
Critical appraisal within sources of evidence	16	If done, present data on critical appraisal of included sources of evidence (see item 12).	N/A
Results of individual sources of evidence	17	For each included source of evidence, present the relevant data that were charted that relate to the review questions and objectives.	22
Synthesis of results	18	Summarize and/or present the charting results as they relate to the review questions and objectives.	22
DISCUSSION			
Summary of evidence	19	Summarize the main results (including an overview of concepts, themes, and types of evidence available), link to the review questions and objectives, and consider the relevance to key groups.	22
Limitations	20	Discuss the limitations of the scoping review process.	22-23
Conclusions	21	Provide a general interpretation of the results with respect to the review questions and objectives, as well as potential implications and/or next steps.	23
FUNDING			
Funding	22	Describe sources of funding for the included sources of evidence, as well as sources of funding for the scoping review. Describe the role of the funders of the scoping review.	29

Appendix E: Excluded Full-Text Articles Sorted by Reason for Exclusion

Title	First Author	Publication Year	Reason for Exclusion	Total Number
Short version of the Guideline: Early Detection of Breast Cancer in Germany. An evidence-, consensus-, and outcome-based guideline according to the German Association of the Scientific Medical Societies (AWMF) and the German Agency for Quality in Medicine (AeZQ).	Albert, U.S.	2004	Specific test not discussed	34
Appropriate utilization of clinical laboratory tests.	Alonso-Cerezo, M.	2009	Specific test not discussed	
Optimal test strategy in the case of two tests and one disease.	Ament, A.	1993	Specific test not discussed	
Screening laboratory evaluation in psychiatric patients: a review.	Anfinson, T.J	1992	Specific test not discussed	
The non-invasive vascular laboratory: who needs it?.	Barnes, R.W.	1995	Specific test not discussed	
Reducing unnecessary blood work in the neurosurgical ICU.	Beland, D.	2003	Specific test not discussed	
Cost-effective laboratory testing.	Burke, M.D.	1981	Specific test not discussed	
The Effects of Computerized Clinical Decision Support Systems on Laboratory Test Ordering: A Systematic Review.	Delvaux, N.	2017	Specific test not discussed	
Cost-effective diagnostic test sequencing.	Eiseman, B.	1989	Specific test not discussed	
NCCN molecular testing white paper: effectiveness, efficiency, and reimbursement.	Engstrom, P.	2011	Specific test not discussed	
Trial of strategy for reducing the use of laboratory tests.	Fowkes, F.G	1986	Specific test not discussed	
Predictive value and efficiency of hematology data.	Galen, R.S	1980	Specific test not discussed	
Test selection measures.	Glasziou, P.	1989	Specific test not discussed	

Effect of feedback to clinical staff of information on clinical biochemistry requesting patterns.	Grivell, A.R	1981	Specific test not discussed	
Do healthy patients need routine laboratory testing before elective noncardiac surgery?	Houchens, N.	2015	Specific test not discussed	
Utilization management in a large urban academic medical center: a 10-year experience.	Kim, J. Y.	2011	Specific test not discussed	
Influence of educational, audit and feedback, system based, and incentive and penalty interventions to reduce laboratory test utilization: a systematic review.	Kobewka, D.	2015	Specific test not discussed	
Impact of CPOE order sets on lab orders.	Mekhjian, H.	2003	Specific test not discussed	
College of American Pathologists Consensus Conference XXXVI: Diagnostic Issues in Thrombophilia.	Olson, J.	2001	Specific test not discussed	
A strategy to promote the rational use of laboratory tests. International Federation of Clinical Chemistry.	Pannall, P.	1996	Specific test not discussed	
A strategy to promote the rational use of laboratory tests. Guideline.	Pannall, P.	1995	Specific test not discussed	
Principles for high-quality, high-value testing.	Power, M.	2013	Specific test not discussed	
Clinical practice and the use of laboratory tests at the May 15 Hospital in Egypt.	Rafeh, N.	1995	Specific test not discussed	
Laboratory Test Utilization Management: General Principles and Applications in Hematopathology.	Reichard, K.	2016	Specific test not discussed	
Effectiveness of Practices to Support Appropriate Laboratory Test Utilization: A Laboratory Medicine Best Practices Systematic Review and Meta-Analysis.	Rubinstein, M.	2018	Specific test not discussed	
Analysis of Daily Laboratory Orders at a Large Urban Academic Center: A Multifaceted Approach to Changing Test Ordering Patterns.	Rudolf, J. W.	2017	Specific test not discussed	
Effect of a Price Transparency	Sedrak, M.S.	2017	Specific test	

Intervention in the Electronic Health Record on Clinician Ordering of Inpatient Laboratory Tests: The PRICE Randomized Clinical Trial.			not discussed	
Routine Postoperative Laboratory Tests Are Unnecessary After Partial Knee Arthroplasty.	Shaner, J.	2016	Specific test not discussed	
The influence of an expert system for test ordering and interpretation on laboratory investigations.	Smith, B.J.	1999	Specific test not discussed	
The economic appropriateness of laboratory tests.	Szucs, T.D.	1997	Specific test not discussed	
A "Top Five" list for emergency medicine: a policy and research agenda for stewardship to improve the value of emergency care.	Venkatesh, A.	2013	Specific test not discussed	
Practice guidelines to reduce testing in the hospital.	Wachtel, T.J.	1990	Specific test not discussed	
Clinically relevant, cost-effective clinical microbiology. Strategies to decrease unnecessary testing.	Wilson, M.L.	1997	Specific test not discussed	
Reducing the inappropriate utilization of clinical laboratory tests.	Wu, A.H.	1997	Specific test not discussed	
Critical review of laboratory investigations in clinical practice guidelines: proposals for the description of investigation.	Aakre, K.	2013	No mention of factor for choosing test	24
Cancer Screening Recommendations and Clinical Management of Inherited Gastrointestinal Cancer Syndromes in Childhood.	Achatz, M.	2017	No mention of factor for choosing test	
Colon Cancer Screening Guidelines 2005: the fecal occult blood test option has become a better FIT.	Allison, J.	2005	No mention of factor for choosing test	
American Gastroenterological Association medical position statement: evaluation of liver chemistry tests.	American Gastroenterological Association	2002	No mention of factor for choosing test	
Statement on use of apolipoprotein E testing for Alzheimer disease. American College of Medical Genetics/American Society of Human Genetics Working	American College of Medical Genetics	1995	No mention of factor for choosing test	

Group on ApoE and Alzheimer disease.				
Initial Diagnostic Workup of Acute Leukemia: Guideline From the College of American Pathologists and the American Society of Hematology.	Arber, D.	2017	No mention of factor for choosing test	
The use of auto-antibody testing in the evaluation of interstitial lung disease (ILD)--A practical approach for the pulmonologist.	Bahmer, T.	2016	No mention of factor for choosing test	
The AGNP-TDM Expert Group Consensus Guidelines: focus on therapeutic monitoring of antidepressants.	Baumann, P.	2005	No mention of factor for choosing test	
Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline.	Bhasin, S.	2010	No mention of factor for choosing test	
Testosterone therapy in adult men with androgen deficiency syndromes: an endocrine society clinical practice guideline.	Bhasin, S.	2006	No mention of factor for choosing test	
Intersociety policy statement on the use of whole-exome sequencing in the critically ill newborn infant.	Borghesi, A.	2017	No mention of factor for choosing test	
AARC clinical practice guideline: blood gas analysis and hemoximetry: 2013.	Davis, M.	2013	No mention of factor for choosing test	
Diagnosis and monitoring of hepatic injury. II. Recommendations for use of laboratory tests in screening, diagnosis, and monitoring.	Dufour, D.R.	2000	No mention of factor for choosing test	
Guideline process improves laboratory use and costs.	Durbin, C.G.	1997	No mention of factor for choosing test	
Can physician laboratory-test requests be influenced by interventions?	Erlingsdottir, H.	2015	No mention of factor for choosing test	
Effect of laboratory testing guidelines on the utilization of tests and order entries in a surgical intensive care unit.	Kumwilaisak, K.	2008	No mention of factor for choosing test	
Implementation of an evidence based guideline reduces blood tests and length of stay for the limping child in a paediatric	McCanny, P.	2013	No mention of factor for choosing test	

emergency department.				
Effects of a computerised protocol management system on ordering of clinical tests.	Nightingale, P.G	1994	No mention of factor for choosing test	
Embedding time-limited laboratory orders within computerized provider order entry reduces laboratory utilization.	Pageler, N. M.	2013	No mention of factor for choosing test	
Impact of clinical guidelines to improve appropriateness of laboratory tests and chest radiographs.	Prat, G.	2009	No mention of factor for choosing test	
Demand management: an audit of chemical pathology test rejections by an electronic gate-keeping system at an academic hospital in Cape Town.	Smit, I.	2015	No mention of factor for choosing test	
Test ordering guidelines can alter ordering patterns in an academic emergency department.	Sucov, A.	1999	No mention of factor for choosing test	
A utilization management intervention to reduce unnecessary testing in the coronary care unit.	Wang, T.	2002	No mention of factor for choosing test	
Impact of Laboratory Test Use Strategies in a Turkish Hospital.	Yilmaz, F.	2016	No mention of factor for choosing test	
Genetics Evaluation Guidelines for the Etiologic Diagnosis of Congenital Hearing Loss. Genetic Evaluation of Congenital Hearing Loss Expert Panel. ACMG statement.	ACMG	2002	Not about appropriate testing practice	22
ACOG practice bulletin. Cervical Cytology screening. Number 45, August 2003.	ACOG	2003	Not about appropriate testing practice	
A framework provided an outline toward the proper evaluation of potential screening strategies.	Adriaensen, W.	2013	Not about appropriate testing practice	
EFNS guidelines on diagnosis and treatment of primary dystonias.	Albanese, A.	2011	Not about appropriate testing practice	
Guidelines for laboratory testing and result reporting of antibody to hepatitis C virus.	Alter, M.	2003	Not about appropriate testing	

Centers for Disease Control and Prevention.			practice	
Consensus statement on the worldwide standardisation of the HbA1c measurement.	American Diabetes Association	2007	Not about appropriate testing practice	
AGA institute guidelines for colonoscopy surveillance after cancer resection: clinical decision tool.	American Gastroenterology Association	2014	Not about appropriate testing practice	
Recommendations for HLA-B*15:02 and HLA-A*31:01 genetic testing to reduce the risk of carbamazepine-induced hypersensitivity reactions.	Amstutz, U.	2014	Not about appropriate testing practice	
Executive summary: a guide to utilization of the microbiology laboratory for diagnosis of infectious diseases: 2013 recommendations by the Infectious Diseases Society of America (IDSA) and the American Society for Microbiology (ASM)(a).	Baron, E.J.	2013	Not about appropriate testing practice	
Guidelines for compatibility testing in hospital blood banks. A joint publication of the British Society for Haematology and the British Blood Transfusion Society.	British Society for Haematology	1987	Not about appropriate testing practice	
Association of apolipoprotein B and nuclear magnetic resonance spectroscopy-derived LDL particle number with outcomes in 25 clinical studies: assessment by the AACC Lipoprotein and Vascular Diseases Division Working Group on Best Practices.	Cole, T.	2013	Not about appropriate testing practice	
Medical screening of emergency department patients. American College of Emergency Physicians.	College of Emergency Physicians	2006	Not about appropriate testing practice	
Detection, evaluation, and management of preoperative anaemia in the elective orthopaedic surgical patient: NATA guidelines.	Goodnough, L.	2011	Not about appropriate testing practice	
Reevaluating the Need for Laboratory Testing in the Treatment of Onychomycosis: Safety and Cost-effectiveness Considerations.	Kanzler, M.	2016	Not about appropriate testing practice	
Clinical Utility of Epstein-Barr	Kim, K. Y.	2017	Not about	

Virus DNA Testing in the Treatment of Nasopharyngeal Carcinoma Patients.			appropriate testing practice	
Cost effectiveness in the intensive care unit.	Kirton, O.C.	1996	Not about appropriate testing practice	
Multinational evidence-based recommendations on how to investigate and follow-up undifferentiated peripheral inflammatory arthritis: integrating systematic literature research and expert opinion of a broad international panel of rheumatologists in the 3E Initiative.	Machado, P.	2011	Not about appropriate testing practice	
Screening for colorectal cancer: a guidance statement from the American College of Physicians.	Qaseem, A.	2012	Not about appropriate testing practice	
Reduction of hospital resources utilization in vascular surgery: a four-year experience.	Roddy, S.P.	1998	Not about appropriate testing practice	
Management of newborns born to mothers with chorioamnionitis: is it time for a kinder, gentler approach?	Shakib, J.	2015	Not about appropriate testing practice	
Laboratory evaluation in the diagnosis of Lyme disease.	Tugwell, P.	1997	Not about appropriate testing practice	
Breast fine needle aspiration biopsy: prevailing recommendations and contemporary practices.	Abati, A.	2005	Not clinician oriented	13
Advantages, disadvantages and optimization of one-stage and chromogenic factor activity assays in haemophilia A and B.	Adcock, D.	2018	Not clinician oriented	
Best practices recommendations in the application of immunohistochemistry in urologic pathology: report from the International Society of Urological Pathology consensus conference.	Amin, M.	2014	Not clinician oriented	
Best practices recommendations in the application of immunohistochemistry in the bladder lesions: report from the	Amin, M.	2014	Not clinician oriented	

International Society of Urologic Pathology consensus conference.				
A guide to utilization of the microbiology laboratory for diagnosis of infectious diseases: 2013 recommendations by the Infectious Diseases Society of America (IDSA) and the American Society for Microbiology (ASM)(a).	Baron, E.J.	2013	Not clinician oriented	
Guidelines for the investigation and management of a reduced level of consciousness in children: implications for clinical biochemistry laboratories.	Bowker, R.	2007	Not clinician oriented	
Guidelines for the laboratory investigation of inherited thrombophilias. Recommendations for the first level clinical laboratories.	Carraro, P.	2003	Not clinician oriented	
Guidelines for biomarker testing in metastatic melanoma: a National Consensus of the Spanish Society of Pathology and the Spanish Society of Medical Oncology.	Martin-Algarra, S.	2014	Not clinician oriented	
Laboratory analyses for poisoned patients: joint position paper.	National Poisons Information Service	2002	Not clinician oriented	
Facts and recommendations about total homocysteine determinations: an expert opinion.	Refsum, H.	2004	Not clinician oriented	
Value judgments for priority setting criteria in genetic testing: a discrete choice experiment.	Severin, F.	2015	Not clinician oriented	
National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines for use of tumor markers in liver, bladder, cervical, and gastric cancers.	Sturgeon, C.	2010	Not clinician oriented	
Guidelines for laboratory analyses for poisoned patients in the United Kingdom.	Thompson, J.P.	2014	Not clinician oriented	
ACOG Committee Opinion #300: Cervical cancer screening in adolescents.	ACOG	2004	Abstract only	4
American Society of Clinical Oncology policy statement	American Society of Clinical Oncology	2003	Abstract Only	

update: genetic testing for cancer susceptibility.				
Utilization management strategies: where to start?	Cherry, P.	2005	Abstract only	
Preoperative tests recommendations in adult patients for ambulatory surgery.	Zaballos, M.	2015	Abstract only	
Practice parameter: the evaluation of distal symmetric polyneuropathy: the role of laboratory and genetic testing (an evidence-based review). Report of the American Academy of Neurology, the American Association of Neuromuscular and Electrodiagnostic Medicine, and the American Academy of Physical Medicine and Rehabilitation.	England, J.D.	2009	Duplicate Article	2
Appropriate use criteria in dermatopathology: Initial recommendations from the American Society of Dermatopathology.	Vidal, C.	2018	Duplicate article	

Appendix F: Processes Reported by Individual Articles Extracted

Last name + date	Lit Review	Consensus Proc	Local Team	Clin Std-Exp	Clin Std-Imp	Vet Guide	Case Review	Unclear	Other
Durieux, 2003	No	No	Yes	No	Yes	No	No	No	No
Pilon, 1997	Yes	Yes	Yes	No	No	No	No	No	No
Hicks, 2013	Yes	Yes	No	No	No	No	No	No	No
Parikh, 2016	No	No	No	No	Yes	No	No	No	No
Man, 2013	No	No	No	No	Yes	No	No	No	No
ASGE, 2014	Yes	Yes	No	Yes	No	No	No	No	No
Chen, 2003	No	No	No	No	Yes	No	No	No	No
Attalii, 2006	No	No	No	No	Yes	No	No	No	No
Zhou, 2018	No	No	No	Yes	No	No	No	No	Yes
Hill, 2018	No	No	No	Yes	No	No	No	No	No
Schuur, 2014	No	Yes	No	No	No	No	No	No	No
Sarkar, 2017	No	No	Yes	No	No	No	Yes	No	Yes
Waldron, 2014	No	No	No	No	Yes	No	No	No	No
Martinez-Balzano, 2017	Yes	No	Yes	No	Yes	No	No	No	No
Tampoia, 2007	No	No	No	Yes	No	No	Yes	Yes	Yes
Task Force, 2019	Yes	Yes	No	No	No	No	No	No	No
Qaseem, 2012	Yes	Yes	No	Yes	No	No	No	No	No
O'Toole, 2014	No	No	No	No	Yes	No	No	Yes	No
ASGE, 1999	Yes	Yes	No	No	Yes	No	No	No	No

Cosmi, 2009	No	Yes	No	Yes	No	Yes	No	No	No
Faulkner, 2016	No	Yes	No	No	No	No	No	No	No
Bulger, 2013	Yes	Yes	No	No	No	No	No	No	No
Onuoha, 2014	Yes	Yes	No	Yes	No	No	No	No	No
Toker, 2004	No	No	No	Yes	No	Yes	No	No	No
Garber, 2012	Yes	Yes	No	No	No	No	No	No	No
Bressler, 2015	No	No	No	No	No	No	No	Yes	No
Nardella, 1995	Yes	Yes	No	No	No	No	No	No	No
Ceccato, 2016	No	No	No	No	Yes	No	No	No	No
Harrison, 1991	No	No	No	No	No	No	No	Yes	No
Meyerhardt, 2013	Yes	Yes	No	Yes	No	Yes	No	No	No
Thomas, 2013	Yes	Yes	No	Yes	No	No	No	No	No
Garcia-Alfonso, 2012	Yes	Yes	No	No	No	No	No	No	No
Kavanaugh, 2000	Yes	Yes	No	No	Yes	No	No	No	No
American College of Rheumatology, 2002	Yes	Yes	No	Yes	No	No	No	No	No
Kavanaugh, 2002	Yes	No	No	No	Yes	No	No	No	Yes
Benito-Garcia, 2004	Yes	No	No	No	Yes	No	No	Yes	No
Ogbechie-Godec, 2018	Yes	No	No	No	No	No	No	No	No
Napierala, 2013	No	No	No	No	No	No	No	Yes	No
Guse, 2014	Yes	Yes	No	No	No	No	No	No	Yes

Huang, 2012	No	No	Yes	Yes	No	No	No	No	Yes
Wu, 1998	No	No	No	No	No	No	No	Yes	Yes
CDC, 2016	No	No	No	No	No	No	No	Yes	No
Agmon-Levin, 2014	Yes	Yes	No	Yes	No	No	No	No	No
Csernok, 2018	Yes	No	No	Yes	No	No	No	No	No
Griffin, 1996	Yes	No	No	No	No	No	No	No	No
Mou, 2017	Yes	No	No	Yes	No	No	No	No	Yes
Hayden, 2000	No	No	No	No	Yes	No	Yes	Yes	Yes
Sturgeon, 2008	Yes	Yes	No	Yes	No	No	No	No	No
Seegmiller, 2013	Yes	No	Yes	No	Yes	No	No	No	No
Tripodi, 2018	No	Yes	No	Yes	No	No	No	No	No
ASGE, 2008	Yes	Yes	No	No	Yes	No	No	No	No
England, 2009	Yes	Yes	No	No	Yes	No	No	No	No
Conn,1994	Yes	No	No	Yes	No	No	No	Yes	No
Narr, 1991	No	No	No	No	No	No	No	Yes	No
Barnard, 1994	Yes	No	No	No	Yes	No	No	No	No
Feely, 2013	Yes	Yes	No	Yes	No	Yes	No	No	No
Bartoletti, 2018	Yes	Yes	No	No	Yes	No	No	No	No
Merlani, 2001	Yes	Yes	No	No	Yes	No	No	No	No
Amicosante, 2010	Yes	No	No	Yes	No	Yes	No	No	Yes
Balk, 2007	No	No	No	No	Yes	No	No	Yes	No
Allison, 1998	Yes	No	No	Yes	No	No	No	Yes	No
Martin, 2016	Yes	Yes	No	No	No	No	No	No	No

Moyer, 2012	Yes	Yes	No	No	Yes	No	No	No	No
Patel, 2017	No	No	No	No	Yes	No	No	No	Yes
Morgan, 1985	No	No	No	No	Yes	No	No	Yes	Yes
Luxton, 2008	Yes	No	No	No	No	No	No	Yes	No
Carvalho, 2013	Yes	No	No	Yes	No	No	No	Yes	Yes
Zuerlein, 1985	No	No	No	Yes	No	No	No	Yes	Yes
Novich, 1985	No	No	No	No	No	No	No	Yes	Yes
Tapper, 2013	No	No	No	No	Yes	No	No	No	Yes
Heine, 2013	Yes	No	Yes	No	Yes	No	No	Yes	Yes
Desai, 2001	No	No	No	No	Yes	No	No	Yes	Yes
Duffy, 2014	Yes	No	No	Yes	No	No	No	Yes	Yes
Duffy, 2007	Yes	No	No	No	Yes	No	No	Yes	No
Adebanjo, 2017	Yes	Yes	No	Yes	No	No	No	Yes	Yes
Wang, 2002	Yes	No	Yes	No	Yes	No	No	No	Yes
Tortella, 1995	No	No	No	Yes	No	No	No	Yes	Yes
Rodriguez-Borja, 2017	No	No	No	No	No	No	No	Yes	Yes
Melanson, 2007	No	No	No	No	Yes	No	No	Yes	Yes
Czoski-Murray, 2012	Yes	Yes	No	No	Yes	No	No	Yes	Yes

References:

1. Thomas RE, Vaska M, Naugler C, Turin TC. Interventions at the laboratory level to reduce laboratory test ordering by family physicians: Systematic review. *Clinical Biochemistry* 2015;48(18):1358-65.

2. Busby J, Schroeder K, Woltersdorf W, et al. Temporal growth and geographic variation in the use of laboratory tests by NHS general practices: using routine data to identify research priorities. *Br J Gen Pract* 2013;63(609):e256-66.

3. Australian Association of Pathology Practices Inc. An Analysis of Pathology Test use in Australia. Family Medicine Research Centre, University of Sydney. 2011.

4. Cadogan SL, Browne JP, Bradley CP, Cahill MR. The effectiveness of interventions to improve laboratory requesting patterns among primary care physicians: a systematic review. *Implement Sci* 2015;10(1):167

5. Canadian Agency for Drugs and Technologies in Health (CADTH). Initiatives to Optimize the Utilization of Laboratory Tests: Environmental Scan. October 2014.

6. Zhi M, Ding EL, Theisen-Toupal J, Whelan J, Arnaout R. The landscape of inappropriate laboratory testing: a 15-year meta-analysis. *PLoS One* 2013;8(11):e78962.

7. Morgen EK, Naugler C. Inappropriate Repeats of Six Common Tests in a Canadian City: A Population Cohort Study within a Laboratory Informatics Framework. *Am J Clin Pathol*. 2015;144(5):704-712. doi:10.1309/AJCPYXDAUS2F8XJY.

8. O'Cathain A, Croot L, Duncan E, et al. Guidance on how to develop complex interventions to improve health and healthcare. BMJ Open. 2019;9(8):e029954. Published 2019 Aug 15. doi:10.1136/bmjopen-2019-029954

9. Choosing Wisely Canada. Clinician Recommendations. Available at http://www.choosingwiselycanada.org/recommendations/. Accessed May 29, 2019.

10. National Institute for Health-Care Excellence. Available at https://www.nice.org.uk/search?q=TEST+ORDERING. Accessed May 29, 2019

11. Choosing Wisely Canada. Critical Care: Recommendations. Available at https://choosingwiselycanada.org/critical-care/. Accessed May 29, 2019.

12. Freedman D. Towards better test utilization - strategies to improve physician ordering and their impact on patient outcomes. *The Journal of International Federation of Clinical Chemistry and Laboratory Medicine* 2015;26:15-30.

13. Levac, D., Colquhoun, H. & O'Brien, K.K. Scoping studies: advancing the methodology. *Implementation Sci* 5, 69 (2010). https://doi.org/10.1186/1748-5908-5-69

14. Tricco, A.C., Lillie, E., Zarin, W. et al. A scoping review on the conduct and reporting of scoping reviews. *BMC Med Res Methodol* 16, 15 (2016). https://doi.org/10.1186/s12874-016-0116-4

15. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMAScR): Checklist and Explanation. Ann Intern Med. 2018;169:467–473. doi: 10.7326/M18-0850.

16. PRESS – Peer Review of Electronic Search Strategies: 2015 Guideline Explanation and Elaboration (PRESS E&E). Ottawa: CADTH; 2016 Jan.

17. Arksey H, O'Malley L. Scoping studies: Towards a Methodological Framework. Int J Soc Res Methodol. 2005:19-32.

Chapter 3 (Draft Manuscript 2): How Do Experts Determine Where to Intervene on Test Ordering? An Interview-Based Study.

Preface

The study described in Chapter 3 addresses two main objectives:

1. Semi-structured interviews with intervention development experts to gain further insight about how they prioritized certain factors over others, measured the factors and which processes they found to be most effective in deciding which tests to target for intervention.
2. Combining the results of the semi-structured interviews with our previously completed scoping review to yield an evidence-based prioritization framework for intervention developers to use when deciding which testing practices are worth their time, effort and money to attempt to change.

Contributions: JCB was responsible for the conception of this project and provided guidance and expertise throughout the entire project. EP drafted the manuscript, and JCB, NH, KC, JP and CM provided critical input and aided in the revision of the manuscript. EP and NH performed the interviews of intervention development experts. Data extraction and quality assessment were completed by EP and confirmed by NH. The guarantor of this review is EP.

Ethics Approval: This study was approved by the Ottawa Health Sciences Network Research Ethics Board (OHSN-REB# 20180626-01H).

Publication Citation: Podolsky E, Hudek H, Carroll K, McCudden C, Presseau J, Brehaut J. "How Do Experts Determine Where to Intervene on Test Ordering? An Interview-Based Study." (In preparation)

How Do Experts Determine Where to Intervene on Test Ordering? An Interview-Based Study.

Eyal Podolsky, MSc (candidate)[1,2], Natasha Hudek, PhD[2], Kelly Carroll, MA[2], Christopher McCudden, PhD[2,3,4] Justin Presseau, PhD[1,2], Jamie Brehaut, PhD[1,2*]

[1]School of Epidemiology and Public Health, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, K1H 8M5, Canada

[2]Ottawa Hospital Research Institute, Clinical Epidemiology Program, The Ottawa Hospital, General Campus, 501 Smyth Road, Centre for Practice Changing Research, Box 201B, Ottawa, Ontario, K1H 8L6, Canada

[3]Division of Biochemistry, Department of Pathology and Laboratory Medicine, The Ottawa Hospital, 501 Smyth Road, Ottawa, ON K1H 8L6, Canada

[4]Eastern Ontario Regional Laboratory Association, 501 Smyth Road, Ottawa, ON K1H 8L6

*__Corresponding Author:__ Jamie Brehaut, C/O Kelly Carroll, The Ottawa Hospital, General Campus, 501 Smyth Road, Centre for Practice Changing Research, Box 201B, Ottawa, Ontario, K1H 8L6, Canada

Abstract

Background: It has been suggested that lab testing is a precursor for 70% of all medical decisions. Considering that 30% of tests may be inappropriately ordered, reduction of inappropriate use is a clear target for improving efficiency and care. Test-ordering intervention development experts often have large datasets on the test-ordering practices within their jurisdiction but have no clear guidance on how to use this data to optimize the appropriateness of test-ordering practice. The goal of this study was to interview guidance experts about which factors should guide decisions about what tests to target for intervention, how these factors can be measured, and what processes are most helpful in deciding what tests can most fruitfully be targeted for intervention.

Methods: We conducted semi-structured interviews with test ordering guidance experts to explore four broad themes 1) the process they followed to determine priority tests for intervention; 2) the factors considered when deciding which tests to target, and relative importance of these factors; 3) how the factors were measured; 4) input on a framework being designed to help decide what tests can most fruitfully be targeted for intervention. The participants were authors of studies identified in a previous scoping review or were recommended by experts that we had already interviewed. In total, 24 experts were contacted with requests for interviews. Of the 24 who were contacted, eight experts participated in our study, one expert responded that they were unable to make time to participate as a result of the COVID-19 pandemic and 15 experts did not respond to our original or follow-up emails.

Results: Eight intervention development experts from across North America were interviewed in-person, by phone or through video teleconference; interviews lasted in the range of 36-104 minutes. The most commonly cited and important factors for choosing certain tests for intervention were the prevalence of the test, test value and patient care. The processes most followed were the existence of an explicit clinical standard such as guideline and consultation with the local team.

Discussion: Intervention developers face difficult choices when deciding where to put their intervention resources. Our study provided insight on which factors and processes participants believed to be most important when deciding which tests to target for intervention.

Registration: N/A

Keywords: Laboratory Testing, Low-Value, Semi-structured Interview

Background

Laboratory testing is one of the highest volume activities in health care. Testing accounts for 3- 5% of all medical costs[1] and is thought to guide up to 70% of medical decisions determining subsequent, more costly care.[1-3] The demand for testing is also increasing.[4;5] Against this backdrop, estimates suggest that 20-30% of tests ordered are low-value, i.e. unnecessary, not indicated, or potentially harmful.[3;6] Inappropriate repeat ordering of six of the most common tests alone may result in $160 million per year wasted in Canada.[7]

Barriers to ameliorating testing practices may include the immense amount of administrative test ordering data available, the large number of tests to choose from, and the wide range of circumstances under which these tests are ordered. Intervention developers, i.e. those staff at individual institutions who are tasked with designing different approaches to improving test-ordering practice, can find it difficult to decide which among the hundreds of routinely ordered tests if targeted for intervention would result in improved and more efficient care.

Large-scale initiatives such as *Choosing Wisely* and NICE Guidelines have sought to improve many clinical practices including the efficiency of test ordering.[9,10] *Choosing Wisely* guidance statements have developed over 250 recommendations pertaining to reducing low-value care, including unnecessary testing, treatments and procedures. [9] In terms of reducing testing, these recommendations do not always give clear guidance on which tests to intervene on,[11] and often do not provide detailed guidance on how to implement such interventions.

In a previous study, we performed a scoping review of guidance statements and empirical studies focused on improving test-ordering practice, and explored factors reported as important in deciding which tests should be targeted for intervention.[12] Interventions were defined as any systematic effort to increase, decrease, or improve appropriateness of test-ordering practice. Results showed that different processes support these interventions, including literature reviews, consultation with local and national/international experts, and that many factors were considered when choosing target tests, including test value, cost and prevalence of the test.

As a further step towards designing guidance for intervention developers, this interview study sought the experience of those who had made these decisions in the past to describe which factors are most important, how they can be measured, and the processes most helpful in arriving at effective decisions.

Methods

Ethics

The study was approved by the Ottawa Health Science Network Research Ethics (OHSN-REB# 20180626-01H). To ensure comprehensive reporting, we used the consolidated criteria for reporting qualitative research (COREQ) guidelines.

Identification of Participants

Adult English-speaking individuals who had previously led initiatives to developing a guideline for improving test-ordering practice or led an empirical study for improving test-

ordering practice were eligible to participate in the study. Potential interview candidates were identified in waves. The first wave of candidates was obtained by contacting the primary author on ten selected articles from our previous scoping review;[12] the second and third waves of individuals were identified through snowball sampling. The authors for the first wave were chosen because the 10 articles that they had published had been the most comprehensive and descriptive as pertaining to factors and processes. We took this as an indication of careful thought about topics and thus that they would be most likely to have more in-depth information to add through interviews with them. This was achieved by asking the interviewees if they could recommend other experts who were doing similar work and could provide insight on this topic. Aside from recruitment emails no contact or relationship was developed between the interviewers and the individual study participants. We contacted eligible individuals through email and provided study information, and the consent form. If the participants did not respond to the initial email to request their participation in the study, they were sent two follow-up emails at intervals of 2-4 weeks.

Interview Guide

The interview guide was designed by the study team and piloted with two test-ordering experts (a critical care physician and head of a clinical testing lab) to ensure completeness and clarity. We followed a semi-structured interview approach which allowed the interviewers to ask specific questions about the prioritization of factors, measurement of factors and processes followed and yet provided the interviewers the flexibility to gather additional information on interesting themes or ideas that may arise naturally during the interview.

Our interview guide was composed of four general sections (Appendix D for study guide). First, after introductions, experts gave an overview of their professional background, how their role was related to test-ordering and making decisions related to testing, and whether they had intended for their intervention/guideline to be adopted locally or on a broader scale.

The second section covered the processes that the test-ordering intervention development experts followed to select tests for intervention. If they reported not using any formal processes (e.g. literature review or Delphi processes), they were asked to discuss what kind, if any, informal processes/conversations took place in order to choose certain tests over others for intervention. This section also included questions on which colleagues were involved in the decision-making process, and whether interviewees had any insight into processes to make the process more efficient.

The third section covered factors the interviewees thought important for choosing certain tests. Experts were first asked to freely recall any factors they had believed to be important. They were then provided with the list of factors obtained from a previous scoping review[12] and were asked to comment on any additional factors of note. Finally, we asked the experts how they prioritized among the factors and whether and how they measured those factors.

In the last section of the interview, we discussed the prioritization framework that we propose to develop with the intervention development experts. We asked the experts which elements they believed would make the framework most useful to them and others in their field moving forward. We concluded the interviews by asking the experts if they felt that they had

anything further to add about prioritizing tests that we had yet to discuss and us soliciting from them names of other potential interview candidates.

We decided to allow the participants an opportunity to discuss their own experiences with prioritizing (both factors and processes) before they were presented with the results of the scoping review. This provided the interviewees the opportunity to contribute their own answer prior to potentially having their answer biased by observing the result of our scoping review. This was more likely to be representative of their true experience with prioritization rather than their answer being clouded by a potential answer that they thought we were seeking.

Interview procedure

All interviews were conducted by two researchers, EP and NH. EP is a male who is currently in the process of completing his Masters (MSc) in Epidemiology and this was his first time conducting a qualitative study. NH is a female interviewer who holds a PhD in psychology and has had many experiences performing interviews on a broad array of topics and thus helped to ensure that the interviews were conducted as required by institutional standards. Both EP and NH work for the Ottawa Hospital Research Institute (OHRI) but neither had specific expertise pertaining to test ordering. Interviews were held either in-person, by phone or on the Zoom platform with both interviewers always present and no one else aside from the interviewee. At the start of the interview, confirmatory verbal consent was sought for participation (and written consent for any in-person interview) and recording of the interview, using a standard, secure audio tape recorder. Participants were informed that they could skip a question if they felt uncomfortable answering it and that they could stop the interview anytime if they were no longer willing to participate in the study.

Data analysis

Interviews were transcribed verbatim and anonymized. We did not return transcripts to interviewees for comment. Two coders (EP, NH) independently reviewed the interview transcripts and coded them using Excel 2011. The extraction sheet was created based on the interview guide with each separate question from the guide receiving its own separate section in the extraction sheet. The extraction was piloted using the transcript from the first interview and it became apparent that some sections could be amalgamated into broader categories as certain questions often received a repetitive answer from the participant. We employed a deductive thematic analytic approach, the final extraction sheet collecting information on four broad themes 1) the process(es) followed to prioritize certain tests for intervention; 2) the factors considered when deciding which tests to target and which factors among all were most important; 3) how factors were measured; 4) the essential components of a framework for helping to decide which tests to target. The themes identified by interviewees were coded based on the processes identified during a previous scoping review[12] and if they did not fall into any appropriate category, an additional theme was added. Additionally, the 18 factors which were previously identified in our scoping review,[12] were presented to the interviews and they commented on which were most important and added any others that they considered to be important. The coders met after coding the first transcript to compare their results and develop a

uniform scheme for coding the quotes. The approach consisted of coders reviewing the transcript and attempting to place the quotes into pre-existing themes (e.g. separating processes, factors, and other issues). If the interviewee discussed an idea that did not belong in any theme that already existed, the coders created a new theme to place the quote into. After both coders finished coding all the interviews, they met to discuss whether certain themes could be amalgamated into unifying themes. Any disagreements were discussed and resolved through consensus or decided upon by a third party (JCB).

Results

Participants

Interviews were conducted between February 2020 and April 2020 and lasted in the range of 36-104 minutes (median n=52 minutes). In total, 24 experts were contacted with requests for interviews. Of the 24 who were contacted, eight experts participated in our study, one expert responded that they were unable to make time to participate as a result of the COVID-19 pandemic and 15 experts did not respond to our original or follow-up emails. Eight intervention development experts (six males, two females) were interviewed in-person, by phone or through Zoom with no repeat interviews being conducted. All eights experts were from institutions across North America (6 Canada; 2 United States). Seven individuals were affiliated with a university research institution and one was affiliated with a private insurance company. The experts were asked to discuss their experiences in the context of developing a specific guideline for improved test-ordering practice or their experience with a specific empirical study in this same domain. Recruitment of interview participants was planned to occur until thematic saturation was achieved but due to the occurrence of the COVID-19 pandemic, further recruitment was stopped at eight interviews.

Themes

Processes used in decision-making of tests to target for intervention

Table 1 contains the 18 themes and 42 subthemes describing the processes that the intervention developers used to choose certain tests over others (n=8). Most participants (5) described the process of consulting with some form of **external clinical standard** to determine which tests to target, whether that was consulting a Choosing Wisely recommendation (5), or some other form of government mandated ordering criteria (1). Additionally, half of the participants (4) reported **consulting with their local team** about which tests to target. This could have meant inquiring with the local staff physicians (4) which tests they think need to be targeted or engaging clinical leadership (2) to ensure that they support pursuing certain tests over others.

Some participants also mentioned considering **local data** (3), **performing a literature review** (2), considering the **difference between academic and community hospitals** (2), **pursuing measurable outcomes** (2) and **assessing the ability to change test-ordering culture** (2) as important processes. **Local data** was usually accessed through an electronic medical record system (3); the decision makers would consult the data to determine which tests they felt warranted intervention based on certain criteria (3). **Literature reviews** were discussed in the

context of determining whether other intervention developers in the same clinical specialty had performed similar work. The interviewees also mentioned taking into consideration the **differences that exist between working in a large academic centre as compared to in the community** and how that may impact the success of their intervention. They mentioned that they often have fewer resources (2) at their disposal as well as have far more stakeholders in the community setting which significantly impacted their decision-making process about which tests to pursue for intervention (1). This was related to the process of **pursuing measurable outcomes** in that the participants commented that one element of difficulty of measuring the outcomes was that in the community it is a much more arduous process (2). Finally, the interviewees mentioned determining whether they believed they could **change the test-ordering culture** associated with a certain test or with a bundle of related tests within their institution.

There were also five themes that were mentioned by one interviewee each**: following a formal consensus process**, **consulting with a lab agency**, **gaining buy-in from end-users**, **considering the broadly important factors first** and **consulting with an internal clinical standard**. Following a consensus process involved consulting with national or international experts outside of the local institution and using a formalized method of reaching consensus on a test that was worth pursing for intervention. The interviewee mentioned using a survey and modified Delphi-process (1) to make the decision about which test should be targeted. The interviewee that mentioned **consulting with the lab** agency explained that they sought guidance from the agency about which tests they were permitted to target (1), and this helped them narrow down their list of which tests they considered to target for intervention. The interviewee that discussed gaining **buy-in from end-users** explained that first they tried to gain acceptance from their local staff who were using the tests that overall change needed to occur and once they were convinced, they solicited their input on which tests to pursue for intervention (1). Another process reported by an interviewee was **considering the broadly important factors.** which as they explained, meant that they decided which factors were most important to them in the specific instance (i.e cost, prevalence of test, patient care) and searched for tests that would allow them to best fulfill the factors that they felt were most important (1). Finally, the last factor mentioned was **consulting with an internal clinical standard** (1) which meant using or consulting a practice followed within the institution but not necessarily by the governing body.

There were two themes that we found during the scoping review[13] that we expected would be discussed by our interviewees but were not discussed at all: **Vetting a Guideline** and **Case reviews**.

Staff involved in decision-making of tests to target for intervention

Table 2 reports on the different staff that were consulted with by intervention development experts and includes 13 groups and 13 sub-groups of people. The intervention development experts reported consulting with a wide array of different people when making the decision of which tests to target for intervention. **Clinical staff** (6) were consulted by almost all the experts which included nursing staff, respiratory therapists, occupational therapists and others. They gathered the opinions of all the different staff to see if they had ideas about which tests made the most sense to target for intervention. Three other groups that were consulted very commonly were **laboratory personnel** (5), **administration** (4), and **IT staff** (4). Finally, there

were a few other stakeholder groups that were consulted more occasionally including, **pharmacist** (3), **patient partners** (2), **KT Staff** (1), **clinical trainees** (1) **and Choosing Wisely Coordinator** (1).

Factors considered in decision-making of tests to target for intervention

Table 3 reports on the factors reported by the interviewees as having been considered when deciding which tests to target for intervention and includes 30 factors and 82 sub-themes. The two most common factors which were listed by every expert (8) were **prevalence/frequency of the test** and **test value**. Unsurprisingly, seven of the eight experts listed prevalence as one of their key factors to be considered but what came as a surprise was that only three of the experts reported test value to be a key factor in their decision-making process. Four other factors that were reported by most experts (7) were: **patient care**, **prior existence of a guideline** for the specific test**, strain on the healthcare system** and **evidence for/against using the test**. Patient care was also ranked to be extremely important as it was considered highly important by all seven of the experts who listed it as one of their determining factors.

A few other commonly considered factors were **cost associated** with the test (6), **evidence of inappropriate use** (6), **implications of a false positive test** (6) and **feasibility of changing the test-ordering behavior** (5). Only two of the experts reported the feasibility of changing the behavior as an important factor in their considerations. Finally, a few other factors were mentioned that did not appear on our list: choice of tests that are in **concordance with the institutions' strategic initiative** (3), whether the intervention on the specific test-ordering behavior **will have a large and measurable impact on both cost and patient care** (3), **sustainability of the project** (2), **concrete evidence of a performance gap** (2) and **medico-legal obligations** (1).

Measurement of factors considered in decision-making of tests to target for intervention

Table 4 contains the tools/techniques used by intervention development experts to measure the factors that they considered important; 18 themes and 9 sub-themes were identified. Aside from measuring the impact on financial cost, the use of measurement tools/techniques were rarely used or considered by intervention development experts. The most common methods used were the implementation of **IT Tools** (3) and **consultation with the** literature (3) to determine how others attempted to measure certain factors. Other concepts discussed were **internal tools used to assess the impact on patient care** (2) if a certain test-ordering behavior was intervened on. An interesting measurement technique mentioned by two separate experts was the use **of cost of testing solely to measure improvement rather than as a motivator for change**. Finally, a technique used by one of the experts was using the tools that they apply to **measure the quality of lab measurements of other labs** and apply them in-house to test whether their performance is up to standard.

Recommendations provided by intervention developers for a framework designed to change test-ordering practice

Table #5 reports on the suggestions of factors, processes or tools that the interviewees reported as being most helpful or important to be included in our proposed framework and

includes 28 distinct themes and 12 sub-themes. Some recommendations mentioned by two experts included a **built-in tool in the framework** that helps assess the ease of implementation of the intervention on the test-ordering practice. A method integrated into the framework that helps in **assessing the clinical significance of the project** and that **the most important factors** such as prevalence of test, risk of test, evidence for/against using the test are all **automatically integrated into the framework** and don't need to be inputted by the user. Other recommendations provided were tools that assist the user of the framework with **choosing appropriate staff for the project**, **assistance with financial resource allocation** to ensure success of the project. One of the participants suggested that the framework could have an **algorithm built-in which predicted the success of intervention** based on certain factors inputted into the system and the specific tests chosen. Another suggested that the framework considers **the level at which the intervention was taking place** and make test-ordering recommendations specific to whether it was being done at the hospital level or the unit level. Finally, one of our interviewees mentioned that the framework would be extremely useful if it was able to incorporate the **socio-political implications of making a test-ordering change** and how the change may cause certain problems to arise. In this situation, the clinician referred to a test that had stopped being funded by the provincial government because it wasn't used by typical clinicians. However, it was essential to the care of their specific population and this caused a big uproar from that clinician group, which could have been be avoided if due diligence had been taken in deciding which tests to select.

Discussion

We sought the experience of experts in intervening on test ordering practices, to understand how they went about deciding what tests to target for intervention. These interviews were important to help us understand what factors are thought to be most important, how they can be measured, and what processes are most helpful in making these decisions. Our interviewees were those who had previously led initiatives to developing a guideline for improving test-ordering practice or led an empirical study for improving test-ordering practice. In many instances, our interviewees were able to shed light on a factor or process that been mentioned by in the literature but not well elaborated on.

Most participants described the process of consulting with some form of explicitly stated clinical standard to determine which tests to target, whether that was consulting a Choosing Wisely recommendation, or some other form of government mandated ordering criteria. For example, in Ontario if a test is ordered to a public laboratory, even if the laboratory personnel do not believe it is a necessary test, they are obligated to run the test. However, many of these same people acknowledged that some of these same guidelines were not well supported by high quality evidence. This is a common theme throughout the literature that demonstrates that physicians will often turn to guidance statements for direction but will use the recommendations cautiously when applying them in clinical practice due to their inherent limitations.[13]

Half of the participants reported consulting with their local team about which tests to target. This could have meant inquiring with the local staff physicians which tests they think need to be targeted or engaging clinical leadership to ensure that they support pursuing certain tests over others. This approach has been widely supported within the health-care cost

management literature, with most experts agreeing that physician buy-in and local applicability is essential to make long lasting and meaningful change within a healthcare organization.[14-16] We found that the larger, academic centres often looked to internal experts, whereas the smaller, rural centres felt that they did not have the internal expertise to make these decisions and thus sought help from external experts. While it was commonly reported that expert experiences were gathered when making decisions, our interviewees did not mention following any formal type of consensus or Delphi process, which may have limited the effectiveness of the expert input.[17] Niederberger et el. argue that Delphi studies are especially useful because during the process experts rely on various sources of information to make their judgments. On the one hand, they can call on their personal expertise and, on the other hand, they can call on knowledge from other types of studies, e.g., randomized controlled trials or metanalysis. Thus, following a Delphi process to help inform which tests to target for intervention might be worth considering in many circumstances when resources allow.[18]

We explored the extent to which literature reviews informed test ordering intervention decisions. In these discussions, most interviewees mentioned literature reviews, but not necessarily comprehensive systematic reviews. This leaves open the possibility that they may have been missing out on key information when making their decisions on which tests to target for intervention. Though systematic reviews tend to be more time-consuming and expensive to perform, their ability to limit bias, and identify a comprehensive literature set allows intervention developers to be more certain if meaningful interventions could be applied to improve test-ordering practice.[19]

Interviewees reported consulting with a wide array of different people when making the decision of which tests to target for intervention, including non-physician clinical staff, laboratory personnel, and patient partners. This is consistent with the literature on health care implementation science, which argues that engaging relevant stakeholder groups involved in the change process can lead to much more effective and lasting change.[20-24] As a result, intervention development experts should be encouraged to identify and engage with relevant stakeholder groups when making test ordering intervention decisions.

The two most common factors, identified by every interviewee, were prevalence/frequency of the test and test value. Prevalence is naturally an important factor because the more frequently that a test is ordered the more it may contribute to undue burden to patients and the more it may contribute to the rising costs of test-ordering. Further, the more frequently that tests are ordered, the more likely it is that unnecessary testing due to repeat testing or standing orders is occurring with regards to that specific test. The interviewees were divided on whether absolute prevalence of tests ordered or the number of tests per patient was the more appropriate measure of prevalence, while the literature surrounding test reduction focuses primarily on a reduction in number of tests per patient.[1,4,7] Three other factors that were reported by most experts were: prior existence of a guideline for the specific test, strain on the healthcare system and evidence for/against using the test.

Patient care was also ranked to be extremely important as it was considered highly important by all seven of the experts who listed it as one of their determining factors. Interviewees differed in their definitions of the components of patient care most relevant to these decisions. The most typical reference to patient care was with regards to patient safety. However,

patient care was also more broadly applied to the increased length of stay in hospital that may arise as a result of unnecessary testing, whether it was for longer test processing times or to overcome anemia from unnecessary blood draws. Additionally, one interviewee included any downstream effects of misinterpretation of the test which may lead to increased cost to the patient or undue stress because of an incorrect diagnosis. As a result, we can conclude that more information is needed on patient care in terms of how to measure it and which elements of it deserve particular attention in this context.

Aside from the impact on financial cost, the use of measurement tools/techniques to assess the impact of test ordering interventions were rarely used or considered by the interviewees. The most common methods discussed included IT Tools and consultation with the literature to determine how others attempted to measure certain factors. Most commonly, intervention developers reported using a previously implemented electronic medical record and electronically available billing data. Several interviewees mentioned that development of tools to help measure improvement stemming from these interventions would be of use.

The final section of our semi-structured interviews asked our experts about recommendations for things they would hope to see in our framework. Indeed, much of the discussion for several interviews focused on directive tools that could help users of the framework with difficult aspects of the implementation process. Two experts recommended a tool as part of the framework that helps assess the ease of implementation of the change test-ordering practice. Other suggested tools to assist the user of the framework with choosing appropriate staff for the project and assistance with financial resource allocation to ensure success of the project.

Limitations of Study

Our study had several limitations that warrant consideration. One limitation stemmed from the emergence of the COVID-19 pandemic. The pandemic took hold globally just two weeks after we began recruitment for the semi-structured interviews. This impacted the interviews in two ways. We believe that the number of clinicians that were able to make time for an interview with us was more limited because of the acute nature of the pandemic. This issue created an imbalance in our interviewee sample, such that we had interviews with very few clinicians, and we feel that factors and process that clinicians may have reported as being important are insufficiently represented. Moreover, laboratory experts became so overwhelmingly busy in attempting to ensure that their facilities were effectively able to process all the testing needs of the current pandemic. Additionally, some interviews were cut short because of this extremely busy time and as such some parts of the interviews were hurried through. This led to a situation where the interviewers did not have the necessary time to elucidate more robust and detailed answers from the interviewees about how they made certain prioritization decisions. As a result of these limitations, we can't be entirely sure that the interviews that we performed allowed us to reach theme saturation nor can we be sure that certain important opinions were not missed as a result of the pressure placed on the interviews due to the pandemic. Finally, clinician framework is essential for our framework, because they are the ones actually ordering the tests and due to their position as essential stakeholders, we would hope to perform more interviews with practicing clinicians to ensure that their views were well represented in our framework.

Future Research and Conclusion

Intervention developers face difficult choices when deciding where to put their intervention resources. In triangulating the results from our previous scoping review[12] and these semi structured interviews, our plan is to use the combined results to help inform a framework designed to assist intervention developers in deciding which tests are worth pursing for intervention. Our interviews identified a wide range of factors, measurement techniques and processes used to make these choices. In order to ensure that our interviews were highly representative, more interviews are needed. After these additional interviews are performed, a consensus process will be pursued in order to ensure that there is agreement among experts as to which factors and processes are most important in deciding which tests to target for intervention.

RESULTS TABLES

Table #1: Processes used by intervention developers to decide which test-ordering practices to target for intervention.

Themes	Sub-theme	Sample Quote	Frequency (Out of 8)
Clinical Standard-External (Existence of a guideline)	Consulting with Choosing Wisely Recommendations	"And so in the case of utilization it could be for any of the specific tests or treatments that are listed on your some of the Choosing Wisely domains." (P10)	5
	Aim for concordance with officially published guidelines	"One would be community clinical practice guidelines that are endorsed by a Canadian national society that we feel we can support the implementation…" (P7)	3
	Government mandated ordering criteria	"So they've changed the ordering the government changed the ordering criteria around AST and one of the liver blood tests and around the order of red blood cell folate or folate serum folate." (P7)	2
	Guideline available for test of interest	"So what I did just in that case there was no there were no randomized trials. There was one study of before/after in another ICU that had introduced a guideline for blood gas measurement…" (P5)	2
	Guideline endorsed by governing body	"So, so in the, in the community setting there's two things we look at if we're gonna make try to target utilization change. One would be community clinical practice guidelines that are endorsed by a Canadian national society that we feel we can support the implementation of that guideline through how we manage testing within the community lab service." (P7)	1
Local Team	Local Team Buy-in	"The one he decided to do was chest x-rays for ribs and what was the other one and urine culture. Because these are were easier to get a consensus of the group." (P2)	5
	Need collaboration between stakeholders	"And that gets into another probably critical success factor which is you need good collaboration between those two groups." (P2)	4
	Physician Input	"…go and talk about with the clinical users so the Paediatric Gastroenterologists and the adult	4

		Gastroenterologists we got agreement from them." (P7)	
	Influence of engaged Clinical Leadership	"It takes a lot of leading by example. It takes a lot of, lot of leadership and meetings and getting the right people in the leadership positions." (P2)	3
	More likely to pursue projects that require fewer stakeholder groups	"So for, so for example some things are, are good ideas but you need a lot of people to have an impact on them if it crosses the lines into needing nursing and needing unions registration." (P2)	3
	Developed own guideline	"And so we actually then we came up with a guideline that took two forms. It turns out one profession the nurses prefer lists and the respiratory therapists prefer algorithms." (P5)	3
	Formal Committee	"It's nice to just say that in the end either a) you have a group to fall back on and to say yes you're doing the right thing or, you know, here's another consideration but then also when you're communicating it's not you, you're communicating on behalf of a committee on behalf on them, you know." (P8)	3
	Physician panel meeting	"And he asked by a show of hands how many physicians in the room had made a clinical decision based on a chloride lately? And there's probably five physicians who raised their hand." (P3)	2
	Incentivized participation in QI/test ordering projects	"You know now, now I'm saying to them if you want to get the stipend for your job there's an expectation that you'll be leading a choosing wisely effort and you'll be engaging in the following quality initiative." (P2)	2
	Administration-led projects	"So our approach was very much backward compared to a lot of other hospitals where most of them were led by like a [name] and those guys or clinicians and colleagues Choosing Wisely clinicians I hear of. [...] Ours was kind of a more admin-led initiative from the corner of the desk." (P3)	1
Consensus Process	Consensus by external experts/committee used to choose test	"…we went through a very broad process of first with the group who was working on Choosing Wisely with the Society of Hospital Medicine talking about what the, what we thought the tests and treatment that were overused	5

		in hospital medicine were." (P6)	
	Consensus can be more important than evidence	"And so relying on evidence sometimes works sometimes doesn't and I think getting the consensus is a lot easier. And that's when once you have the consensus they can't really disagree with, with what you're trying to write.." (P3)	1
	Survey of Experts	"Then we actually put a survey out to the membership of the Society of Hospital Medicine and asked them the same question." (P6)	1
	Modified Delphi-Process	"And then we, we did two rounds in a modified Delphi process of, of voting around what we thought the five tests or treatments should be." (P6)	1
Local Data	Use of Electronic Ordering system	"…so depending on if the let's say if we expect the problem to be tied to an issue with physician ordering that is related to say a computer physician order entry like CPOE….." (P10)	4
	Accessing data	"I think we're pretty data driven. …we're gonna pull the data out almost right away and then it's either gonna go or not…." (P6)	4
	Real-time data	"But one thing that helps us do that is we monitor every undefined test in real time…" (P8)	3
	IT tools	"Intelligence Tools and we had IT folks who wanted to, wanted to use those tools for clinical settings. […] And so we went again our IT guys went and started developing tools that was looking not only at lab utilization data from our EMR but also the coding database from digital support…" (P3)	1
	Comparison to other centers	"Like when we look at the Choosing Wisely ones I think they were at the 50% repeated CBCs at [hospital 4]. When we looked at ours we were at something like 15% already so there was not a lot of gain there to make." (P3)	1
	System changes provide opportunity to review order sets	"Yeah but as an example now we're about to go live in June with electronic ordering by physicians and last year or so there's been we've had physicians in each department as subject matter	1

		experts to review all the electronic order sets." (P3)	
	Chart Review	"And so requires a little bit of baseline data and chart reviews to understand whether there is a prevalence of the problem in our local setting." (P10)	1
Literature Review		"And that group we then did a literature review around to look at what were the, what the literature was on those topics." (P6)	4
Consideration of funding/resources available*		"But how we began, how we began was choosing criteria and our basic method is to choose criteria that is feasible for you given the resources and culture that your specific institution may have." (P8)	3
Process differs in large academic centre vs community hospital with regards to resource accessibility*	Less resources for test-ordering projects	"So I'm in a community hospital. I it, it's different than, you know, being in a large teaching centre where you potentially have a lot of resources. You have graduate students like yourself to help." (P2)	2
	More difficult to gain consensus in community setting	"So we instead of having a limited number of physicians you have to manage you're now managing 20,000 physicians across a province." (P7)	1
	Test user buy-in more important in small centers	"Yeah it depends on the like we're a small hospital so I think it, it's important maybe for a multi-site UHN type [hospital 3] maybe it's not as critical, but for us to get the buy-in I think it's important." (P3)	1
	Legislative barriers in some settings	"The other thing we can't do is in the community sector if someone's ordered a test we've actually got to provide service that's how the legislation is written at the moment." (P7)	1
Test-Ordering Culture*	Intend to change ordering culture of institution	"…our basic method is to choose criteria that is feasible for you given the resources and culture that your specific institution may have." (P8)	2
	Culture change can assist with buy-in	"I think that that's a critical piece is that you need to be able to show people those gains and those wins otherwise it's hard to get buy-in it's hard to get credit and it's hard to influence culture." (P2)	2
	Involve trainees in process to change	"So I think it's very important to involve in terms of the team. I may be	1

	culture	going off on a tangent here, but involving trainees because this is the future of our personnel in healthcare." (P5)	
Buy-In*	Pushback from test-orderers	"And then you get the, you'll get the people they're phoning and like having a rant on the phone to you and you'll then explain it and you'll debate it and after you bring them around. And if you don't bring them around you tell them they can't have it anyway." (P7)	2
	Bottom-up initiatives facilitate buy-in	"The message could be exactly the same it's just it comes across so much differently if it's bottom up rather than top down." (P6).	2
	Local Buy-In from End-Users	"And for a couple of them we thought we could just do it without telling anyone right on the server and for a couple of them we thought we needed physician sponsorship and leadership." (P2)	1
Consideration of Outcomes*	Measurable Intervention Outcomes	"….and then also be able to measure it." (P2)	2
	Impact of the intervention	"If you're gonna put in so much effort to identify the causes of the problem and to design an intervention and implement it and study it, it really needs to be something that's either fairly widespread or has such a compelling rationale or results in such great harms that it's really worth going after." (P10)	2
Clinical Standard-Internal (The internal standard of practice that is not formally established by the governing body)		"Again, it's guided by professional practice, you know, that would be the target would be what is the best, the best practice for that. And my team are able to sort of work out how they, how they do that in an effective way" (P7)	2
Evolving Approach to Choosing Tests*	Approach became more formalized/structured over time	"…this has been an evolving, an evolving process over the past eight years or so. And as part of some background we have seven, sevenish years ago because of these exact questions you're asking put together a group called…" (P8)	2
	SAP approach	"And one of the main things the, the Dr. [name 4] came up with was an approach. So the SAP approach so	1

		looking at the source so before tackling an initiative let's look at the science out there like in terms of so Choosing Wisely has recommendations backed by science." (P3)	
Start with easy wins/ small projects (low hanging fruit)*		"So where, you know, the fat's so thick (laugh) that you could skim off and, you know, that's sort of what we call the low hanging fruit. Tests that are being done all the time with absolutely no value those ones we targeted pretty early on and we were pretty aggressive about it with like what is called the switch approach where you just turn it off." (P10)	2
Consult with external groups/centers	Consult with Lab Agency	"And then the other approach is in British Columbia where they have a lab agency that has some ability to dictate what you can and can't do in a community setting." (P7)	1
	Consult other centers on their interventions	"And then also we use our member base to share their experiences as well. And that specific example is interesting because we learned early on there were like five different places that used the same intervention with different results." (P8)	1
Process differs each time*		"Mm it varies I mean, I mean I don't think it's there's one way. […] So I don't know that any way that I've done it or my colleagues have done it in a particularly useful fashion." (P9)	1
Consideration of organizational structure*		"So that system works where there is a centralized budget that is managed by a laboratory in conjunction with overall running of the hospital. So that piece that can work fairly well." (P7)	1
Vetting of a Guideline	N/A		
Case Review	N/A		

*Represents a process that was newly obtained from interviews and unrecorded during scoping review.

Table #2: People involved/consulted with by intervention developers when deciding which test-ordering practices to target for intervention.

Themes	Sub-theme	Sample Quote	Frequency (Out of 8)
Physicians	Physicians from Various Departments	"Like there's the grass roots like the physicians at the bedside coming up with ideas…" (P2)	6
	Local champion	"…the reality is that someone just gets an idea and if somebody else likes the idea they race forward with it." (P9)	5
	Internal committees/groups	"…there's also there's the hospital clinical managers, you know, the nursing committees the lab committees who frequently have good observations and good ideas as well." (P2)	2
	Department heads	"…but we also worked with the Head of Radiology and we developed and then we consulted with the attending physician." (P5)	2
	Influential voices	"...and the other thing is you always have certain people that carry a lot of sway. [...] And that's where a clever leader knows how to leverage that sort of what I refer to as the unofficial power mat of the organization." (P2)	2
	Diversity of physicians involved	"And then we whip in our internal experts as needed infectious disease for example or endocrinologists." (P8)	1
	Clinical-content expert	"And then we whip in our internal experts as needed infectious disease for example or endocrinologists." (P8)	1
Clinical Staff	Front-Line Staff (Non-physician) (e.g. nurses, genetic counselor)	"Anyway so I assembled the team of the people who were involved and so for blood gases it was the nurses the bedside nurses, respiratory therapist…" (P5) Clinical is interesting clinical is more on the nursing side and in our experience we've been lacking on that, on that one and because a lot of times it's not just physician behaviour but it's nursing behaviour change…" (P3)	6
	Clinical director	"If it's a therapeutic medication type of intervention then the question would	2

		first go to the Director of Pharmacy as an example." (P10)	
Laboratory Personnel		"You know they know like the lab people know, you know, the physicians aren't the best person to go to so for certain things…" (P1) "In this case we worked with people from the lab just so we could understand more about the tests…" (P5)	5
Administration	Administrative-Level Decision-Makers	"And so I think really getting that more mid to senior level administrator and leadership support or at least input to tell people whether it's gonna be or not be a success." (P10) "You know I, I moved us to having departmental leadership that was onboard with the idea that the goal is quality of care." (P2)	5
	Administration Level Buy-in	"…is that it's sponsored by the highest levels of the hospital… something get successful is that if you can get Board approval, you know, the, the higher the priority is of a project the more likely everybody is to help." (P1)	2
	Quality Director	"Quality director was the same thing again meeting standards." (P3)	1
Information Technology Staff		"And then this we had the solution architect again for I think getting the data was big actually that solutions architect now is at (Location), (location) hired him…" (P3)	4
Pharmacist		"If it's a therapeutic medication type of intervention then the question would first go to the Director of Pharmacy as an example." (P10)	4
Patient Partner/representative		"The patient partner was interesting and the patient and the patient feedback a lot of times tweaked our approach or because feedback that kept coming is the trust between the physician and the, and the patient. (P3)	3

End-users of interventions		"I think those who we think will be, you know, primarily involved as either the end receiver or participants of the intervention and/or in some way would be affected by this project…" (P10)	3
Team specific to each project		"...with our, our contact person at The Ottawa Hospital." (P2)	2
Choosing Wisely Coordinator		"The Choosing Wisely Coordinator we had one at the beginning who was an overarching like a 10,000 ft view of everything, but we didn't have any specific ones to tackle all the campaigns. And just last year we finally got somebody dedicated for this." (P3)	1
Resident/ Clinical Trainees	Trainee-led projects	"…that one was led by residents and so here's another opportunity and I think a very important piece of this work and I hope that you will build it in. This is work that we need to teach to our trainees and I mean like clinical trainee." (P5)	1
KT/QI Staff		"…you know, there are people from a performance improvement standpoint that are looking at it." (P6)	1
Researchers		"So a committee of different doctors and the VP of Academic Affairs and researchers to try to gather all the initiatives together and have a game plan." (P3)	1

Table #3 Factors considered by intervention developers when deciding which test-ordering practices to target for intervention.

Themes	Sub-theme	Frequency that the factor was discussed (*frequency that it was mentioned as a key factor in making decisions*) (n=8)	Sample Quote
Prevalence/frequency of the test	Absolute magnitude of tests ordered	8 (7)	"Frequency, you know, how many times but, you know, it's really how what volume? If you look at all the tests done which, which one is the #1 in volume that's the one I think people should be looking at." (P5)
	Frequency of inappropriate use of test	2 (2)	"So we would look at who had been ordering it, how frequently they meet patient criteria or not?" (P8)
	Testing out of habit/routine	2 (2)	"In fact sometimes things were done routinely in admission orders. In other words in preprinted routine admission orders these patients every patient would get one of these things. And so it became almost, you know, as they say a robotic habit."
	Test "volume" more important than frequency	2 (2)	"…these are tests that we do really frequently and so it's a very large volume of all the procedures that are done." (P10)
	Testing based on time of day vs need	1 (1)	"…you know, 7 am and 7 pm, you know, which makes no sense. And the idea was to try and, and that was one of the strategies was to try and order tests based on physiology not based on the clock." (P5)
	Repeat testing too frequently	1 (1)	"So I would focus on if there was a very high repeat order rate on something that had a clear guidance as to, you know, you needed to do it every a less frequent amount of time then that could be something you'd focus on. (P7)
Costs associated with Test	Overall cost of testing	8 (5)	"…we first started with any test that was over a $1,000 then we went back to every test over $500…. So the prevalence is

			important but the prevalence can be kind of trumped by the cost." (P6) "And it, it's if you focus on cost you're not gonna get consensus that's our experience." (P3)
	Fixed Vs Variable costs	1 (0)	"Like it's like fractions of a penny to actually conduct a lot of these tests, even though the charges may be listed as several dollars they come up with these sort of bogus. Because, you know, the fixed versus variable cost and stuff like that." (P9)
Ease of implementation of intervention for changing test-ordering behaviour	Modifiable in electronic ordering	8(2)	"So I, I looked for things that were order-based, you know, lab-based where I could leverage the technology…" (P2)
	Overall ability to implement intervention	5 (1)	"And like sometimes making a small change can actually be a lot of work. And other times you can really sit at your desk and just make it happen maybe with one or two other people." (P2)
	Fewer Stakeholders needed	1 (1)	"…that I needed a huge stakeholder group for. So for, so for example some things are, are good ideas but you need a lot of people to have an impact on them if it crosses the lines into needing nursing and needing unions…." (P2)
	Low hanging fruit	1 (0)	"But yeah so yeah if something was easy to implement you would you might pick up on that rather than so we did target things that one would call the low-hanging fruit. Things that we think people wouldn't worry about." (P7)
	Much more efficient with use of Computer Order Entry	1 (1)	"And system wide changes so and I talked a bit about it right earlier on in the sense that if we had an ordering system that linked into a proper database and it was like almost like an instant feed and you it's just like modern times and technology should work." (P7)
Test Value	Clinical Value of Test	7 (3)	"Tests that are being done all the time with absolutely no value those ones we targeted pretty early on and we were pretty aggressive..." (P10)

Patient Care	Avoid patient harm	7 (7)	I'm gonna start with my usual I think the patient's safety when there's patient safety implications will be first." (P8) "So I think patient care it's for me is important so you'd want to make sure that you've got a good eye on patient care whatever you're doing with this." (P7)
	Quality of care	4 (4)	"The challenge with patient care is that there's very few tests that have direct patient outcomes linked to them, but they are there's a number that affect patient care directly so you want to make sure that patient care isn't compromised by, by the testing change that you're gonna look at." (P7)
	Harm from misinterpretation/downstream effects	4 (4)	"So cost to patient in terms of preventable harm or a cascade of events that would result in additional testing and emotional harm and other things like that." (P10)

	Length of stay	1 (1)	"And then and one of the things, again it used to be about money and it infringes the 4Ps but looking at the return on the intervention. So we might not necessarily save money but it might have a huge impact on patients or length of stay or quality of care." (P3)
Prior existence of a guideline for the specific test(s)	Guidance statement available for test	7 (5)	"…but, you know, a lot of doctors make a very valid point which is that we're actually ahead of the guideline." (P2) "Prior existence of guidelines is really important and I think of the way I look at things now I'd say if there were good guidelines we would certainly try and follow them. And I would include at the moment Choosing Wisely has a fairly good set of some are focused on lab testing guidelines so or recommendations." (P7)
	Clear guideline recommendations	2 (0)	"Things like transfusing only above 80 not below 80 so there were very clear-cut things that we could just hang our hat on." (P2)
	Choosing wisely	2 (2)	"Yes definitely ideal to have something to point to like the Choosing Wisely guidelines or any other consensus guidelines that come out." (P8)
	Guidelines not available	1 (1)	"…except for the chest x-ray in which the American College of Radiology had changed their recommendations about indications for chest x-rays in the ICU based on that cluster randomized trial that supported us. There really weren't any other high grade guidelines for the other tests and I don't think I have any others." (P5)

Quality of supporting evidence for/against using the test	High quality or abundant literature that the test should not be performed	7 (4)	"And then to think about and then tests which have lots of literature behind them that might be less indicated." (P8)
	Quality includes age of data	1 (1)	"I mean things are constantly evolving,…..so it's, it's hard to keep on top of what's happening." (P2)
	Evidence does not have to be high grade	1 (1)	"Supporting evidence for/against, yes well we used, you know, again I told you it's not high grade it's not like we have randomized trials but except for the, the chest x-rays where we did. And that was a compelling argument having a cluster randomized trial showing no benefit was really helpful." (P5)
Implications of a false positive test	False Positive	7 (1)	"Implications of a false positive test so yeah so false positive tests are obviously a major problem particularly when they have downstream impact in the in other parts of healthcare." (P7)
	False Negative	1 (1)	"I'm thinking now not false positives but we're having all these false negatives in testing with Covid and what the antibody testing might mean for that too in terms of the false positives." (P8)
Strain on Health-Care System	Overall strain on various aspects of Health-care system	6 (2)	"I think strain on the healthcare system is sort of is the nice thing about that one is it sort of bundles up the cost, the laboratory, the downstream effect so I actually kind of like that one." (P9)
	Important to keep in mind but does not help choose between tests	1 (0)	"So that would be yeah I think you'd have to consider it, but it's I'm not sure it's a high priority in my list on this list when I'm trying to choose between tests. Because I think they all put a strain on the system because I think the system's strained as it is." (P7)
Test Quality	Overall test quality (i.e specificity & sensitivity)	6 (2)	"Okay so diagnostic test accuracy yeah if something has very poor parameters then I would target that as something to take out of service from all together." (P7) The other aspect of test quality that I would highlight and maybe you haven't asked this to other people, but we have accuracy and bias assessments from the proficiency

			testing perspective and there are certain tests run by certain labs that don't meet those specifications....” (P7)
	One test replaces another	2 (2)	“Or you just really don't need this test at all it's been replaced by a better, a better test.” (P8)
	Proficiency testing (i.e. lab not equipped to perform test well)	1 (1)	“The other aspect of test quality that I would highlight and maybe you haven't asked this to other people, but we have accuracy and bias assessments from the proficiency testing perspective and there are certain tests run by certain labs that don't meet those specifications…” (P7)
Relevance to Current Practice	Test useful for clinicians in their daily practice	5 (2)	“Relevant to current practice is important because, you know, as we develop new technology current practice sometimes falls behind where technology is taking you. And so making sure that we keep the testing relevant to current practice is important…” (P7)
	Keeping testing up to date with advances in science (e.g replacing old test for newer test)	3 (3)	“Relevant to current practice is important because, you know, as we develop new technology current practice sometimes falls behind where technology is taking you.” (P7)
Feasibility of changing test ordering behaviour	Project is within scope of authority/ having control over test ordering practice	5 (1)	“I think feasibility of changing ordering behaviour is very important and then what's the scope of changing the behaviour which fits within the feasibility.” (P6)
	Overall feasibility of changing test-ordering	5 (1)	“And, you know, what's been known about and what's been shown to be effective and can be rolled out or implemented in a feasible fashion within our setting?” (P10)
	Low-hanging fruit vs complex problem	4 (0)	Feasibility of changing test ordering behaviour that's a it's yeah it's a big, it's a big factor and it comes back to the low hanging fruit versus the tougher one.” (P3) “Sometimes some people think it's a low-hanging fruit, which it never is….” (P9)
	Smaller target group	1 (1)	“So are you just targeting one group of providers or is this changing the, you know, you like the urinary catheter thing, you know, is, is all kinds of clinicians plus nurses and everybody else you've got to change behaviour with.” (P6)
	Reduced ability to use clinical judgement	1 (1)	“And it's, it's very little room nowadays for us to say I used clinical judgement.” (P2)

	Consensus/buy-in from those ordering test	1 (0)	"So, so that's sort of how I chose that it had a combination of impact, measurability, stakeholder analysis and a desire to, to be able to get recognized and, and get people onboard." (P1)
	Computer Order Entry	2 (0)	"…feasibility yes I think they were because they were, they were things that were in our order entry system and they were well-known to everybody…" (P5)
	Changing behaviour difficult regardless of test type	1 (0)	"…heard that changing test ordering behaviour is quite challenging in any context… I wouldn't and I think I wouldn't consider that a major factor…And it's difficult to do so and I think that's the same regardless of what the test is." (P7)
	Infrastructure in-place to make change	1 (0)	"whether more so of whether there's infrastructure in either the microbial system where you're implementing an intervention that could accommodate an intervention that would change the behaviour." (P10)
	Benchmark/scorecard for changing behaviour	1 (1)	"Use of scorecards, benchmarking so yeah I've talked a bit about those things so those are the things that I so I think the benchmarking, scorecards and the payment structure piece and I've mentioned that already." (P7)
	Limited by resources	1 (1)	"But how we began, how we began was choosing criteria and our basic method is to choose criteria that is feasible for you given the resources…" (P8)
	Limited by culture of institution	1 (1)	"…and culture that your specific institution may have." (P8)
	Low vs high leverage for changing behaviour	1 (1)	"…and I don't know if you're familiar with this, but Choosing Wisely always talks about the implementation spectrum about how you can do either low leverage interventions or, or high level interventions and they always have examples for this." (P3)
	Managing expectations	2 (2)	"So it's, it's important in managing expectations like especially at higher admin is they know we're gonna tackle telemetry but we told them it's not gonna take 6 it's gonna take a year to two years to get the results you need so." (P3)
	Need incentives for lab to implement change	1 (1)	"And so as soon as we do anything that cuts the amount of service we do we, we cut them out of revenue regenerated. So there's no incentive for a community lab to, to actually do any utilization and it actually

			process money." (P7)
	More difficult in certain contexts	1 (1)	"Feasibility of changing test ordering behaviour I think you've heard that changing test ordering behaviour is quite challenging in any context and it has more challenges in some contexts than others, but, you know, do we consider that?" (P7)
	Change fatigue	1 (1)	"It's for sure rare that they consider a lot of other even general principles like how likely they are to succeed? What the interference is gonna be with other sort of like change fatigue I should say with other things going on? (P8)
Actionability of test results	Only perform testing if results affect care	4 (1)	If you don't have this test like how will it impact your management of either a specific patient or all patients? And so we try to capture that as well." (P8)
	Few tests have directly actionable results or difficult to determine actionability	2 (0)	"But, you know, it's really hard on a test by test basis because I don't think most people could tell you what actionability for half the things that are ordered." (P8)
Laboratory Workload	Overall lab-personnel workload	4 (2)	"So I would look at where I can run a test and how easy it is to run it and I'll target things that were more complicated if I was looking at tests to pick that is one I would look at from the lab workload perspective." (P7)
	Complexity of lab usage for test(s)	1 (1)	"So I would look at where I can run a test and how easy it is to run it and I'll target things that were more complicated if I was looking at tests to pick that is one I would look at from the lab workload perspective." (P7)
Test being bundled with other tests	Bundling is a reason to target to test	4 (1)	"…if something was bundled I would probably avoid trying to target it as something from a utilization perspective because it's to try and pull it out of all the bundles becomes a challenge unless you've got really good evidence to do it…" (P7)
	Order sets	4 (2)	"n fact sometimes things were done routinely in admission orders. In other words in preprinted routine admission orders these patients every patient would get one of these things. And so it became almost, you know, as they say a robotic habit." (P5)

	Necessary test bundled with unnecessary test	2 (2)	"So if you've got a test that has a lot of value but it's bundled into many other pieces then how do you manage, manage that? Those are quite difficult to deal with." (P7)
	No control over bundling	1 (0)	"Test being bundled with other tests well that happens already and in fact we didn't have much control over that because for example in the magnesium project we had no influence over titre magnesium that was bundled as part of nutritional assessments…" (P5)
	Bundling inflates costs	1 (1)	"Yeah correct because you're just you think you're paying for x and it's x is appropriate but you're paying for x, y, z and then the other thing happens if y and z," (P6)
	Assay bundling in lab	1 (1)	"One of the challenges there was one of the assay manufacturers brought out a multiplexed assay that does all of that on a single, on a single bead. So you actually if you're running that instrument you actually have to suppress some of the results…" (P7)
	Physicians may build own sets (barrier to intervention success)	1 (1)	"What happened to that one is it worked for about three months and then people started rebuilding their order sets through different channels into the system and they got the test anyway." (p7)
	Unbundling is difficult without good evidence	1 (1)	"…becomes a challenge unless you've got really good evidence to do it out of all of the bundles."
	Difficult to change behaviour for bundling vs order sets	1 (1)	"Yeah so that, that's a consideration for yeah I guess for me it would be how successful any of this is gonna be. Because if it's not in order sets that means they're doing it on purpose a lot and that's a harder thing to change." (p8)
	Bundling is a reason to avoid targeting tests	1 (1)	"So from a test menu perspective if something was bundled I would probably avoid trying to target it as something from a utilization perspective because it's to try and pull it out of all the bundles becomes a challenge unless you've got really good

			evidence to do it out of all of the bundles." (P7)
Expert Experience	Experience with tests based on personal practice	4 (1)	"How about the fancy word heuristic… based on experience…this wasn't just gut feeling that was unsubstantiated it's by seeing it everyday and seeing the and knowing the lack of utility." (P5)
	External experts/specialty groups	2 (2)	"The things we've done internally I'd say have been more, you know, they're more it's I guess somewhat similar but it's more grassroots where people feel like there are tests or treatment in their area which are overused." (P6) "Expert experience would be important if you, if you need to get people to buy into the change." (P7)
	Individual physician heuristics	2 (2)	"How about the fancy word heuristic, you know, it's sort of I mean heuristics is sort of the way we practice, you know, based on experience. And I mean this was just this wasn't just gut feeling that was unsubstantiated it's by seeing it everyday and seeing the and knowing the lack of utility." (P3)
	Helps to get buy-in	1 (1)	"But if you had something that was a bit more had less evidence less clarity behind the use of the test and the effect on patient care having expert experience would be helpful." (P7)
Hospital's strategic initiatives*		3 (2)	"So how you factor in which intervention you're gonna do will be most successful if you can align with those strategic initiatives. You don't want to be the one saying now we're gonna expand all this testing right after the hospital saying we're gonna, you know, furlough everyone." (P8) "I think about in general with quality improvement stuff is that it is sometimes worth thinking whether going after this test will lay the groundwork for anything else or is synergistic with anything else going on locally for instance." (P9)
Potential for large impact on care/costs		3 (1)	"… but is it a compelling problem meaning that we think that there is significant harm or some kind of negative effects If you're gonna put in so much effort to identify the

			causes of the problem and to design an intervention and implement it and study it, it really needs to be something that's either fairly widespread or has such a compelling rationale or results in such great harms that it's really worth going after." (P10) And so I mean we, we recognize that resource stewardship requires, you know, thinks about cost but we also think about non monetary costs. So cost to patient in terms of preventable harm or a cascade of events.." (P10)
Evidence of Inappropriate use	Unclear source of evidence	3 (3)	"…they're not just gonna take clinicians aren't just gonna buy hey I do it too much. And you need evidence that that incidence of unjustified variation between clinicians is also important." (P6)
	Local evidence	3 (3)	"Evidence of inappropriate use would be a really important one. So if you knew people were using a test badly or an inappropriate context, inappropriate settings that would be an important one to target. So I would put that higher up the list." (P7)
	Evidence that intervention is feasible	2 (1)	"…one other one is, you know, do we know enough about this quality problem or utilization problem if you will where there are some demonstrated evidence-based types of studies that could help us." (P10)
	Evidence that variation by practitioner is problematic	1 (1)	"So being able to show hey here's 10 docs at your practice in your system and, you know, these two are outliers. That actually helps a lot to get those two to say okay I realize I'm not doing what the rest of my colleagues are doing and I'm gonna change behaviour." (P6)
	Variation in ordering is evidence of inappropriate use	1 (1)	"No I think we've gone after things, you know, that in the new evidence situation that certainly we've gone after to say, but we've gone after even if everybody was doing it wrong I mean, you know, there was a time when we had all our Orthopaedic Surgeons that we felt were doing, you know, DVT prophylaxis wrong. And, you know, we went after that, you know, and there was—but even then there was variation because they were all doing it wrong (laugh) that's a whole other story." (P6)

Sustainability of project*		3 (0)	"What will be required to sustain it if it does succeed, you know, these types of things?" (P2)
Ensuring that there is truly a performance gap*		2 (2)	"having there actually be a performance gap because some of these things there's just not a performance gap so why would you go after it is a piece of it?" (P6) "We're thinking of an area of problem, a QI problem if you will, where there is a demonstrated gap in terms of how we're currently performing compared to where we should be." (P10)
Prevalence of Target Disease	High prevalence disease in population	2 (1)	"Prevalence of disease is important, you know, something that that goes along with the frequency of the test as well." (P6)
	Don't want to miss rare disease just because not prevalent	1 (1)	"We deal with in the labs we deal with so many diseases that are really uncommon and a lot of the reason people do a broader spectrum of tests is to try and make sure they don't miss the really uncommon stuff." (P7)
Potential for large impact on care/costs*		2 (0)	I think it's a couple things one is the scale of the potential improvement…when around using PPIs outside the ICU the reality is when we ran the data it wasn't, you know, it wasn't an issue…" (P6)
Target Repeat testing*		2 (0)	And there's certain testing and that is, you know, once you've had it done you don't really need to get it done again... And particularly now that we do a lot more DNA testing. Once you've had a DNA test done you don't really need to get it done again…" (P7)
Measurable Outcome*	Measurable outcomes for buy-in and culture change	1 (1)	"I think that that's a critical piece is that you need to be able to show people those gains and those wins otherwise it's hard to get buy-in it's hard to get credit and it's hard to influence culture." (P2)
	Baseline data helps decide if test should be targeted	1 (1)	"So, you know, the measurement was or getting the baseline data it served two foci; one was to say is this something you should go after or not?" (P6)
	Baseline data helps manage change	1 (1)	"But the second thing is is there are, were instances and I can't think of any off the top of my head, but we just couldn't measure it very well. And if you couldn't measure it

			very well it didn't behoove you necessarily to go after it because you weren't gonna be able to manage it very well if you couldn't measure it very well." (P6)
Test-Ordering Culture*		1 (1)	"I mean if you have a certain way things are always being done it's very hard to break away from that." (P2)
Having a strong team in place*	Team focused on area of concern	1 (1)	"I mean if I had to think of like the one thing that helps me move things forward is having that coalition of allies that are ready to work." (P2)
	Stakeholders at all levels	1 (1)	"We basically have a Choosing Wisely resource team already assembled called the Transfusion Committee which was made up of critical leadership, somebody from administration, somebody from the labs, somebody from Ottawa." (P2)
	Dedicated time of members	1 (1)	"…everybody's doing stuff so but it's getting, getting people together who have the time and that's the real challenge… we've become very LEAN, so I think finding people that can spend time doing things is hard." (P2)
Medico-legal responsibility*		1 (0)	"…physicians nowadays I think are very medical/legally defensive as well. I, I we have quite a few doctors that I think every single kid with a fever they see is getting testing because they just don't have confidence that they would be supported should the child return sicker…" (P2)
Patient-directed testing*		1 (0)	"Either regular things that I just want, you know, I have this additional piece of information a family history or whatever and I want my kid or themselves to have additional testing or they read in the news there's this new test. They want it or there's these kits filtering in themselves that you can order online and it still requires a doctor order." (P8)
Personnel time (other than lab)*		1 (1)	"And the last one was personnel, which is staff so we might not be making a huge effort in savings and so on but we might be saving an hour/day of nursing time at the

			end of the day." (P3)

*Represents a factor that was newly obtained from interviews and unrecorded during scoping review.

Table #4 Measurement tools and techniques for factors considered by intervention developers.

Themes	Sub-theme	Sample Quote	Frequency (Out of 8)
Accessing electronic data	Importance of electronic record for doing any kind of measurement	"Well I mean that's where I think it's very important to be able to do that and that's where you really want, you've really got to hope your hospital is on an electronic platform." (P2)	3
	Need IT assistance/tools to sort through abundance of data	"Yeah for us it was, it was getting our own data from our own system. It's the business intelligence tools is, is what would be the ones in Q6 that I said strain on healthcare system." (P3)	3
Literature review to guide measurement	Literature to measure quality of evidence	"If there wasn't a guideline you would then go back and look a the base literature to look and see if there was say for example any systematic reviews that had been performed." (P7)	3
	Literature to measure patient harm/care	"So, you know, when thinking about what tests to target we sort of start with a literature review to understand what's already been published and known about the problem. And so generally if there is evidence that quantifies what kind of harm and in what population" (P10)	1
Cost as a metric		"…you bring in cost but you can show that we spent half a million bucks on a specific test…. Did you realize that this, this many tests meant this much money? And, and one of our Medical Directors is really good at explaining. I mean saving $120,000/year is, is an extra nurse…" (P3)	3
Assessing impact on patient care	Measuring patient harm (e.g. death, long term outcomes, minor outcomes like temporary fear)	"…we have a method and I'm not gonna say that we do it a lot, but the hospital has like this ranking system that you could use to think through the elements of various types of harm. And it includes many things that are on this list… You would give a ranking for the relative like seriousness if, if the test was misinterpreted would that mean, you know, they would die? Would they would have some type of bad outcome that they wouldn't have had before?" (P8)	2
Manual measurement possible but may		"So I actually had to go to the pharmacy and do an inventory and count them. And every month I'd go in and say how many did you	1

not be feasible		start with? How many did you order? How many are missing? So unless you're electronic you're kind of screwed." (P2)	
Measurement/data used to assess intervention success		"So I, I think that when you make an intervention to improve quality it's very important to say well what is it we're trying to improve? How do we measure it and how do we, how do we see if we've moved the needle? So but it, it's that's data that we'll ask for." (P2)	2
Define current state of ordering		"And as I said an Ishikawa diagram that sort of thinks about like why is it that the test is being ordered commonly? That is I think there's no substitute for doing that before deciding how to move forward." (P9)	2
Having an organized system/checklist	Use factors together in a ranking system/checklist	"…but the hospital has like this ranking system that you could use to think through the elements of various types of harm. And it includes many things that are on this list so like you would give a ranking for this relative frequency. You would give a ranking for the relative like seriousness if, if the test was misinterpreted would that mean, you know, they would die? Would they would have some type of bad outcome that they wouldn't have had before? Or would that mean that they would just be scared for a little bit?" (P8)	2
	Assess levels of influence on ordering (e.g. patient factors, provider factors, infrastructure, staffing issues)	"Like I mean you need to sort of have a kind of a checklist or, or a framework like that and apply it to a given problem…" (P9)	1
Measurement based on what is in guideline		"Mm okay so let's start so if there's a guideline we would work with the guideline and then hopefully it covered off those points. So that would be the sort of top level of evidence that I would look at for that." (P7)	1
Consistent metrics measured for every test		"…I think every time we tackled one of these things we went, you know, we would figure out who was, you know, how many were ordered? Who was ordering them? Where they were ordering them and then what the cost was?" (P6)	1
Assessing quality of lab		"…how we think about external labs and how we judge their quality. And there's like a table	1

measurements		of different things that we look at… So we have things that we think about for quality of a lab and specific tests which you could take some of those and apply to in-house tests as well." (P8)	
Test quality measurement	Metrics specific to test (e.g. turnaround time, correction rate, presentation of results, normative billing)	"…how we think about external labs and how we judge their quality. And there's like a table of different things that we look at like the turnaround time; correction rate of the lab do we know if they make a lot of errors? Are the reports for how they provide the results reasonable or normative or easy to understand." (P8)	1
Blood volume as a metric	Difficult to get data	"No well no we tried it and it's tough to get the data. [...] So that's what [hospital 4] did we kind of but we can make the point without having specific milliliters like it's, people just know.[...] So yeah so that's one of the things where we tried to get the data we couldn't." (P3)	1
Nursing time as a metric		"And I forget the numbers they came up with but that's one of the metrics they're gonna use in the initiative is look upfront. We're gonna measure how much time it takes and then we're gonna say well we can save 30% of the time." (P3)	1
Readmission as a metric		"But what they were actually saying is just show to me that I'm not coming off more patients or my readmission rate is constant and a lot of times that's what they meant." (P3)	1
No measurement done for CHOOSING test		"…but the actual measurement of the numbers of tests was done as part of the project. We did it during a baseline period and we started our interventions and then we did a follow-up. So it's before/after these were all before/after studies in one ICU." (P5)	1
Measured WHO is ordering		"…we would figure out who was, you know, how many were ordered? Who was ordering them?" (P6)	1
Measured which environments had high test-ordering		"Where they were ordering them and then what the cost was?" (P6)	

Table #5 Recommendations provided by intervention developers for a framework designed to change test-ordering practice.

Themes	Sub-theme	Sample Quote	Frequency (Out of 8)

Assessing ease of implementation		"Well again, ease of implementation and signs of impact." (P2)	2
Most important factors built in		"I think we really need to the second piece is to include in the framework problems that are again I think area of high degree of overuse or frequency." (P10) "It's a similar question and so high volume, high cost, high risk, new technology, new evidence, you know, those are kind of criteria that you could be that you could use to choose a test." (P5)	2
Indication that methods may differ for large and small institutions		"…because an approach in an academic hospital versus a community hospital versus like the community lab setting you're gonna have different factors you're gonna need to consider." (P7)	2
Stepwise approach	Structured, sequential approach	"But forgetting about the framework was a key mistake is that's why having that that model we have and I think that step-by-step focuses you and, and makes you realize that this is not a two-weeks initiative it's gonna take 18 months to make it happen." (P3)	2
	Steps/model helps set expectations	"…the framework there from the get-go and	1

		making people realizing how much time it's gonna take and how much effort is, is probably—and it's a good thing not to discourage people but it's to make sure you have the right people around the table and you might need twice the amount of resource you think." (P3)	
	Ishikawa diagram	"And as I said an Ishikawa diagram that sort of thinks about like why is it that the test is being ordered commonly? That is I think there's no substitute for doing that before deciding how to move forward." (P9)	1
Measurable change		"You know I think that the lab and the admin have to tell us which changes are impactful…" (P2)	2
Updated evidence on test value/quality		"Or if there's new evidence that comes out about the lack of utility of a particular test for example sometimes there's evidence that comes out about, you know, how a test is not actually detecting what you think it's detecting stuff like that. So new evidence about the utility of a test…." (P5)	2
Importance of collaboration	Representation from key stakeholders	"…I think you need to have representation from all the key groups, right. So, so if you're I think you need well depending on the departments you need, you need somebody in	2

		critical leadership, somebody in administrative leadership. I think the danger is always underestimating how important somebody can be." (P1)	
	Consensus process for prioritizing test targets	"So if the people that are stewarding the resources can get together with the people using them they can probably come up with a pretty good focusing strategy." (P2)	1
Assess context and consider appropriate methods	Large vs small institutions	"Yeah so I think, I think the starting point would be, you know, what, what sort of environment are you working in because an approach in an academic hospital versus a community hospital versus like the community lab setting you're gonna have different factors you're gonna need to consider." (P7)	2
	Availability of experts	"And as I said in a, in a community hospital setting in a smaller hospital they just don't have enough experts to really support this sort of change. And that makes it much, much more difficult." (P7)	1
	Appropriate system in place for changing behaviour	"So do you have so like for that physician group do you have an annual review meeting where you could present them with say feedback data? And actually tie it you've got to meet certain change expectations in your	1

		practice so that you get, get your credentials back that you can still practice in the hospital. And you've got to demonstrate that you're working towards that on an ongoing basis." (P7)	
	Success may differ by institutions using similar interventions	"And that specific example is interesting because we learned early on there were like five different places that used the same intervention with different results." (P8)	1
Organize factors	Combine/consolidate factors	"Yeah probably I think, I think the list is a little bit too, too long. I think that you could probably put them into, into a few subcategories in terms of how to think about them for sure." (P2)	2
	Use a matrix	"Yeah I wonder if it's like a matrix." (P8)	1
Project with Clinical significance		"Yeah so one of my mentors always says, you know, don't go after projects that nobody cares about." (P10) "…..then the clinical piece needs to come from the physicians in terms of does this matter to us." (P2)	1
Help with Resource allocation		"And it helps, and then it helps also putting the right amount of resource. So if you have that that framework that step-by-step you know that you can't do it by	1

		yourself." (P3)	
Tools to choose Staff		"And it helps, and then it helps also putting the right amount of resource. So if you have that that framework that step-by-step you know that you can't do it by yourself." (P3)	1
Routine Surveillance system built-in		"…having a dashboard that shows volume of tests. You know if you do something really infrequently it's, it's, you know, it's not really worthwhile to go after." (P5)	1
Measure local variation in practice		"…then understanding what the, the variation is within the practice your local practice… But generally that doesn't happen usually there's if there's variation you're gonna have some people doing it, you know, by a guideline and some people doing unnecessary work…" (P6)	1
Method to ensure the testing menu is appropriate for setting		"So that would always be the starting place when your framework is actually menu review, making sure what, your menu's appropriate for your setting." (P7)	1
Tailored to address various goals of implementing project		"So I mean this is, is this something to address a clinical need or a budget restriction	1

		or a general sort of change in process that you're trying to implement? And each of them would come with a different thing." (P7)	
Tiering of different tests		"But, you know, proposing frameworks like have tiering. So there's Tier 1 tests which anyone can order because they're common and low cost. There's Tier 2 tests that might be limited to certain provider specialties…" (P8)	1
Algorithm predicting success of intervention based on certain factors and tests chosen		Yeah I'm thinking like, like an online algorithm… and then like questions are asked, you answer questions and then at the end you get a score or some type of interpretation… not even a yes or no, but like your test example will be difficult to implement but the value could be these things. Or this will be easy to implement with relatively little value or, you know, whatever combination of how, how you want to word it." (P8)	1
Consideration of level/method needed for intervention		"I guess the only missing thing for the framework I guess I would say is, is thinking about what level or in what way	1

		you're intervening? Like are you intervening at the macro level at the meso level at the micro level…" (P9)	
Table structure/organization		"Yeah, yeah. Yeah I guess I don't know I think I'm envisioning kind of like a table with the different questions that maybe like an example. I think an example use case would be helpful." (P8)	1
Examples to illustrate factors and cases that did/did not pass to intervention		"I think an example use case would be helpful. [...] Right yeah and maybe there's a few examples like this is one that made it through and why it did. How it passed the threshold and one that didn't make it through or reasons why you could interpret the same results differently (laugh) you know, like I don't know how." (P8)	1
Framework to help with resource allocation		"And it helps, and then it helps also putting the right amount of resource. So if you have that that framework that step-by-step you know that you can't do it by yourself. You know that you might need five different people you might need some expertise that you don't have." (P3)	1
Focus on changing test-ordering culture		"So that's a big, a big thing yeah and I would say also on the framework focus on the	1

		culture. And that has to do with going like to the Medical Advisory Committee the MAC and getting rubberstamping on from the leader saying yeah this is the way we're gonna do things now. And at so yeah" (P3)	
Low hanging fruit	Begin by reviewing/ cleaning up ordering menu	"Then how do you the framework so the framework needs to start off with are they things that are—what's on your menu and are there things that you are already aware of that shouldn't be there. And there's a whole lot of stuff that people don't clean up their menu so when did you last review your menu or testing is it all appropriate?" (P7)	1
Clearly identify goals/motivation for project to guide methods (e.g. cost, improving care, saving time)		"And I think after that it would then be how do you pick something if you wanted to address something so what is your, your purpose of doing it as well? So I mean this is, is this something to address a clinical need or a budget restriction or a general sort of change in process that you're trying to implement?" (P7)	1
Providing resources to help guide intervention for specific test		"And so but we've stayed out of we are basically an operational what I would sell [stewardship group] as is an operational house. So like we have all these tools but we're not gonna tell you how	1

		to order a specific test."	
Socio-political implications of making a test-ordering change		"I guess I mean the nice thing the, the resource I mean I guess like the resources required for macro things are often deceptive like not actually that high. But it's more the risks or the potential backlash if you get if you do something that some of us loud people like complain about." (P9)	1
Importance of buy in at all levels		"So that's one thing for people to consider and well maybe just one other thing under the umbrella when we say people care about but also be important that the organization in which the improver is going to do this work is working in they would have to care about it or be convinced that it's of value. So I think that's an important piece." (P10)	1

Declarations

Ethics Approval and Consent to Participation
The study was approved by the Ottawa Health Science Network Research Ethics (OHSN-REB# 20180626-01H).

Consent for Publication
Not Applicable.

Availability of Data and material
The authors may be contacted for data and material requests. However, the data and materials pertaining to individual participants will not be shared to protect privacy.

Competing Interests
The authors declare that they have no competing interests.

Funding
EP received a University of Ottawa Graduate Studies Scholarship and held a graduate studentship with the Ottawa Hospital Research Institute. Funding bodies had no role in the design of the study, collection, analysis, interpretation of data, or in the writing of the manuscript.

Authors' Contributions
JCB was responsible for the conception of this project and provided guidance and expertise throughout the entire project. EP drafted the manuscript, and JCB and NH, provided critical input and aided in the revision of the manuscript. EP and NH completed the interviews. Data extraction and quality assessment were completed by EP and confirmed by NH. The guarantor of this review is EP.

Appendix A- Interview Guide

Understanding how priorities for practice improvement are developed: Draft interview guide

Pre-amble: Thank you for taking the time to meet with us. We're interested in your views and experience in identifying priorities for practice improvement in test utilization.

As you know, laboratory testing is a high-volume activity in health care that has huge implications for costs and patient outcomes, and evidence suggests that many tests are sometimes used inappropriately.

When thinking about where to focus efforts to change practice around test ordering, (for example, deciding to focus on one test instead of others), there are potentially a lot of factors to consider (cost of the test, prevalence, etc). We are talking to experts who have experience in this prioritization process, in order to develop a framework to aid others with these types of decisions.

We've provided a consent form by email. Any questions before we start? Ok if I record our conversation so that I can review it later?

Q1: First, I wonder if you can tell me about the process you and your colleagues went through to decide on which test ordering practices needed to be changed?
Prompts:

A. If the answer to above, is "no formal process" can you tell me what you did do?

B. I notice you decided to focus a recommendation on (test X). Were there discussions about which tests to choose? **Among who?** What other options were on the table? How were the decisions made?

Q2: Could you talk a little bit about your current position and how it ties into test ordering practice?

Q3: Could you give us a sense of whether your guideline/intervention was intended for local use, in your center, or was it meant for a broader audience, such as a particular clinical specialty?

Q4: **Having gone through the process to develop these recommendations, is there anything you would change about it for next time?**

Q5: Which factors did you consider when deciding which test(s) to focus on?

Q6: I want to show you a list of factors the literature identifies as potentially being relevant to the quality and value of diagnostic tests in general. We'd like you to go through each item on this list, and tell me whether you considered these factors when deciding which tests your own recommendations would focus on, ("looking at this list, which factors did you actually consider") (list is at the End)

Q7: Are there any factors that aren't on the list that were important in your process?

Q8: Considering the factors that you mentioned at first and then after having looked at the list, are there factors that you felt were more important than others?
Prompt: how would you organize these factors in terms of importance?

Q9: Now after having gone through the list, are there any factors you would have included or think might they be important for others going through a process like yours, and why or why not?"

Q10: For each of the factors that you mentioned to be most important in Q6, can you tell me a little about how it was measured (if it was); that is, what sort of data were used to measure it?
Prompts: SR from the literature? An ad hoc literature review? Local data? Expert opinion? General impressions of the group?

Q11: We are hoping to develop a framework that will help other units decide which tests they should focus their attention on when seeking to improve test utilisation. This framework will include the factors that were generally considered most important in deciding which tests to target for intervention and the hope is that it will make it easier in the future to choose tests that will be most amenable to change.

A. What would you find useful in a framework like this?

B. How could we make a framework like this most useful if you were doing the project again?

Q12: Is there anything else about prioritizing tests to target that we haven't considered?

Q13: Could you recommend anyone else we should be speaking to about this?

List of factors for consideration when choosing among tests

A. Prevalence/frequency of the test
B. Patient Care
C. Relevance to Current Practice
D. Costs associated with Test
E. Test Value
F. Feasibility of changing test ordering behaviour
G. Expert Experience
H. Strain on Health-Care System
I. Evidence of Inappropriate use
J. Prevalence of Target Disease
K. Laboratory Workload
L. Test Quality
M. Implications of a false positive test
N. Actionability of test results
O. Test being bundled with other test
P. Quality of supporting evidence for/against using the test
Q. Ease of implementation of intervention for changing test-ordering behaviour
R. Risk/harm of administering test
S. Prior existence of a guideline for the specific test(s)
T. Other

References:

1. Thomas RE, Vaska M, Naugler C, Turin TC. Interventions at the laboratory level to reduce laboratory test ordering by family physicians: Systematic review. *Clinical Biochemistry* 2015;48(18):1358-65.

2. Busby J, Schroeder K, Woltersdorf W, et al. Temporal growth and geographic variation in the use of laboratory tests by NHS general practices: using routine data to identify research priorities. *Br J Gen Pract* 2013;63(609):e256-66.

3. Australian Association of Pathology Practices Inc. An Analysis of Pathology Test use in Australia. Family Medicine Research Centre, University of Sydney. 2011.

4. Cadogan SL, Browne JP, Bradley CP, Cahill MR. The effectiveness of interventions to improve laboratory requesting patterns among primary care physicians: a systematic review. *Implement Sci* 2015;10(1):167

5. Canadian Agency for Drugs and Technologies in Health (CADTH). Initiatives to Optimize the Utilization of Laboratory Tests: Environmental Scan. October 2014.

6. Zhi M, Ding EL, Theisen-Toupal J, Whelan J, Arnaout R. The landscape of inappropriate laboratory testing: a 15-year meta-analysis. *PLoS One* 2013;8(11):e78962.

7. Naugler C. Misutilization of laboratory tests: pathways to correction. *CIHR Foundation Scheme: 2014 1st Live Pilot*.

8. O'Cathain A, Croot L, Duncan E, et al. Guidance on how to develop complex interventions to improve health and healthcare. BMJ Open. 2019;9(8):e029954. Published 2019 Aug 15. doi:10.1136/bmjopen-2019-029954

9. Choosing Wisely Canada. Clinician Recommendations. Available at http://www.choosingwiselycanada.org/recommendations/. Accessed May 29, 2019.

10. National Institute for Health-Care Excellence. Available at https://www.nice.org.uk/search?q=TEST+ORDERING. Accessed May 29, 2019

11. Choosing Wisely Canada. Critical Care: Recommendations. Available at https://choosingwiselycanada.org/critical-care/. Accessed May 29, 2019.

12. Podolsky E, Hudek H, Carroll K, McCudden C, Presseau J, Brehaut J. “Factors Reported in Determining Laboratory Tests to Target for Intervention: A Scoping Review.” (In preparation)

13. Woolf, S. H., Grol, R., Hutchinson, A., Eccles, M., & Grimshaw, J. (1999). Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. BMJ (Clinical research ed.), 318(7182), 527–530. https://doi.org/10.1136/bmj.318.7182.527.

14. Milliken A. D. (2014). Physician engagement: a necessary but reciprocal process. *CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne*, *186*(4), 244–245. https://doi.org/10.1503/cmaj.131178.

15. Taitz, J. M., Lee, T. H., & Sequist, T. D. (2012). A framework for engaging physicians in quality and safety. *BMJ quality & safety*, *21*(9), 722–728. https://doi.org/10.1136/bmjqs-2011-000167

16. Scott, C. G., Thériault, A., McGuire, S., Samson, A., Clement, C., & Worthington, J. R. (2012). Developing a physician engagement agreement at The Ottawa Hospital: a collaborative approach. *Healthcare quarterly (Toronto, Ont.)*, *15*(3), 50–53. https://doi.org/10.12927/hcq.2013.23020

17. Jorm A. F. (2015). Using the Delphi expert consensus method in mental health research. *The Australian and New Zealand journal of psychiatry*, *49*(10), 887–897. https://doi.org/10.1177/0004867415600891

18. Niederberger, M., & Spranger, J. (2020). Delphi Technique in Health Sciences: A Map. *Frontiers in public health*, *8*, 457. https://doi.org/10.3389/fpubh.2020.00457

19. Gopalakrishnan, S., & Ganeshkumar, P. (2013). Systematic Reviews and Meta-analysis: Understanding the Best Evidence in Primary Healthcare. *Journal of family medicine and primary care*, *2*(1), 9–14. https://doi.org/10.4103/2249-4863.109934

20. Bosch, B., & Mansell, H. (2015). Interprofessional collaboration in health care: Lessons to be learned from competitive sports. *Canadian pharmacists journal : CPJ = Revue des pharmaciens du Canada : RPC*, *148*(4), 176–179. https://doi.org/10.1177/1715163515588106

21. Romanow RJ. *Building on values: the future of health care in Canada*. Ottawa (ON): Canada, Commission on the Future of Health Care in Canada; 2002. Available: http://publications.gc.ca/collections/Collection/CP32-85-2002E.pdf (accessed March 4, 2015).

22. Wheelan, S. A., Burchill, C. N., & Tilin, F. (2003). The link between teamwork and patients' outcomes in intensive care units. *American journal of critical care : an official publication, American Association of Critical-Care Nurses*, *12*(6), 527–534.

23. Martinez, A. S., Saef, J., Paszczuk, A., & Bhatt-Chugani, H. (2013). Implementation of a pharmacist-managed heart failure medication titration clinic. *American journal of health-system pharmacy : AJHP : official journal of the American Society of Health-System Pharmacists*, *70*(12), 1070–1076. https://doi.org/10.2146/ajhp120267

24. Chang, W. Y., Ma, J. C., Chiu, H. T., Lin, K. C., & Lee, P. H. (2009). Job satisfaction and perceptions of quality of patient care, collaboration and teamwork in acute care hospitals. *Journal of advanced nursing*, *65*(9), 1946–1955. https://doi.org/10.1111/j.1365-2648.2009.05085.x

25. Fernandez, M., ten Hoor, G., van Lieshout, S., Roiguez, S., Beidas, R., Parcel, G., Ruiter, R., Markham, C., & Kok, G. (2019). Implementation Mapping: Using Intervention Mapping to Develop Implementation Strategies. *Frontiers in Public Health*, *7*, 158–158. https://doi.org/10.3389/fpubh.2019.00158

Chapter 4: Integration and Discussion of Results

Discussion of Results

There have been many initiatives and efforts in the last 20 years to educate health care practitioners on the wastefulness and potential harms of unnecessary care. Of the total expenditures on health services in the USA, as much as 30% is thought to be unnecessary.[1-4] This value is congruent with a more recent analysis of several Canadian tests and interventions which observed that 30% may have been avoidable.[5] While financial losses are among the most discussed,[6] other reasons for reducing and eliminating unnecessary medical care include adverse events from over-testing, further unnecessary work stemming from false positive tests, and opportunity costs that can affect the care of other patients by reducing the availability of resources.[3,7-10]

The choice around which tests to focus organizational effort on is a challenging one, as the results of the intervention must be worth the time, effort and resources necessary to make the change on an organizational level. Intervention developers, i.e. those staff at individual institutions who are tasked with designing different approaches to improving test-ordering practice, can find it difficult to decide which among the hundreds of tests routinely ordered would most benefit care if targeted for intervention. Intervention developers often have immense amounts of data on the test-ordering practices of the clinicians that they oversee due to the increase in the availability of administrative data. As a result, a framework is needed to help test-ordering intervention developers decide which tests to pursue for improvement using their limited resources.

There is a clear need to develop a prioritization framework to assist healthcare organizations and intervention developers in deciding which tests are worth targeting for intervention. This thesis combines the results of a scoping review[11] and semi-structured interviews,[12] to contribute directly to informing the content of such a framework. The framework should provide intervention developers with the range of factors (i.e. cost, relevance to practice, actionability) that have guided other test prioritization decisions, as well as those considered most important in the experience of experts in this area. In the long term, such a framework could result in increased efficiency of resource utilization and improved clinical decision making as a result of reduced use of low-value testing.[13-15]

In this thesis, we identified a wide array of factors that were reported to be considered in other test targeting exercises, and the processes that led to the decisions about which tests to target for intervention. By exploring all relevant guideline statements and related empirical studies, we gathered the full range of factors to consider when deciding which test(s) to target for intervention. Overall, we identified 18 factors in our review and 30 in our interviews, highlighting the complexity of these decisions. While our studies were not the first to demonstrate that targeting tests is complex and that many factors must go into the decision of deciding which tests to target[16-23], our work has identified the most comprehensive range of factors available. We also explored processes reported by the studies that led to these decisions. Individual processes (e.g. literature review, consultation with the local team and consensus processes) were rarely mentioned alone, meaning that articles and interviewees that did report on their process usually discussed several processes that were completed either sequentially or simultaneously. Of the nine processes identified in our review, some were identified in a majority of cases (literature review, followed by clinical standard-implicit and consensus process), and

other processes far more infrequently. Some processes were not clearly reported meaning that the relative frequency of the different processes may be somewhat different that that what is reported. We believed that interviews of relevant stakeholders would help prioritize among those processes that are most effective in helping decide which tests should be pursued for intervention and in asking our interviewees about the processes that they used in deciding which tests to target for intervention, they mentioned 18 unique processes.

The three most frequently identified factors were unsurprising: test value, cost, and prevalence of test; these were also the most commonly cited factors in our scoping review.[11] The fact that these factors were the most commonly identified reflects that ultimately financial considerations are top of mind within most test-ordering intervention projects. In the past, many initiatives have been pursued due to budgetary pressures.[3,8,24,25] Fully seven of the eight experts listed prevalence and cost among their key factors to be considered.

Interestingly, only three of the experts reported test value (i.e. does it directly impact decisions made around patient care) to be a key factor in their decision-making process. In our scoping review, test value was mentioned by 80% of articles which was concordant with the frequency with which we saw it mentioned by our interviewees. This may be a result of the allopathic medical training in which medical trainees are taught to exhaust all diagnostic avenues prior to making medical decisions. As a result, a test would not necessarily become a target for improved practice even though it is not essential for decision making; potentially because physicians are inclined to order tests in order to be comprehensive, with less regard for the utility of the individual test.

Other commonly mentioned factors could be classified into two groups: patient considerations and test-specific factors. The patient factors included: the impact of false positives, the actionability of test results and patient care. The factors that were test-specific included test quality, quality of evidence for/against using the test and the existence of a guideline for appropriate use of test. Our clinician interviewees tended to focus on the patient specific factors, while the lab-based interviews tended to be more concerned with the test-specific factors. This indicates that there is more to be learned about how a person's role may influence the factors that they are more likely to consider when deciding which test to target for intervention.

Two interesting factors mentioned by our interviewees but rarely seen in our scoping review were: the feasibility of changing test-ordering behavior and the ease with which the intervention could be implemented to improve testing practice. These two factors are clearly related to the likelihood of success of an intervention and should likely be considered more often than they seemed to be in the literature. This is consistent with a review that found that the perceived difficulty of barriers associated with implementing an intervention was extremely relevant to the success of its implementation.[26] These factors were uncommonly seen in the literature but certainly seemed to be guiding decision making on which tests to pursue for intervention. As a result, these are important factors that will be included in our future framework.

Finally, a few other factors not identified in the scoping review but we found to be both interesting and potentially important: choice of tests that are in concordance with the institution's strategic initiative, whether the intervention on the specific test-ordering behavior will have a large and measurable impact on both cost and patient care,

sustainability of the project and concrete evidence of a performance gap. All of these factors pertain to obtaining project support from higher level management which is ultimately likely to contribute to the on-going success of the intervention and we would certainly include them as recommendations in our framework.

One important difference between our scoping review and interview data with regards to the processes followed to decide which tests to target for intervention drew our attention. Most participants (63%) described the process of consulting with some form of explicitly stated clinical standard to determine which tests to target, whether that was consulting a Choosing Wisely recommendation, or some other form of government mandated ordering criteria. This is in keeping with the overall literature surrounding the use of clinical practice guidelines which show that over the last 30 years their use by physicians has significantly increased.[27] This was an interesting finding because in our scoping review,[13] authors referred to an explicit standard (33%) relatively infrequently at nearly the same rate as unofficial and/or implicit clinical standards (39%). This might stem from a sample bias of our interviewees, who may be more likely to adopt best clinical practices, or alternatively may because our interviewees chose tests for which there were already existing explicit standards.

Consultation with the local team was also frequently discussed. About half of the interviewees reported consulting with their local team about which tests to target, which stands in contrast to our scoping review, where 10% of authors reported consulting with their local team. This could have meant inquiring with the local staff physicians which tests they think need to be targeted or engaging clinical leadership to ensure that they support pursuing certain tests over others. Much of the literature surrounding sustained change in health care environments discuss the need for physician buy-in [28-30] and it was encouraging and useful to hear that the interviewees echoed this sentiment. According to interviewees, consultation with the local team was a logical step on the way to deciding which test to target for intervention; whether this is not as obvious in the public literature, or whether it is simply not commonly mentioned in written reports is unclear but warrants considerations for improved reporting.

Literature review was reported commonly in the literature but, was seldom mentioned by the participants in our interviews. There is much evidence to show that systematic reviews and meta-analysis are an excellent place to start when undertaking a new project[31] and we would certainly seek to elucidate if or why they are not being reported as primary starting points during test-ordering intervention projects. Systematic literature reviews are standard of practice, and it may be assumed that they were likely performed as part of the standard process and hence not perceived as worth mention. Additionally, there were two processes that were reported by the authors in our review that were not even mentioned by one participant in our interviews: vetting of a guideline and case review. We are tempted to attribute it to the lack of saturation in our interviews and would hope to resolve this difference as more interviews are performed.

Finally, several processes that were not reported in our scoping review, were reported by a significant number of our interviewees. Some participants mentioned considering local data, considering the difference between academic and community hospitals, pursuing measurable outcomes, and assessing the ability to change test-ordering culture (2) as important processes. Local data was usually accessed through an electronic medical record system or through a computer provider order entry (CPOE); the

decision makers would consult the data to determine which tests they felt warranted intervention based on certain criteria. We believe that some of these additional processes may have been incorporated into the more general processes that were reported throughout the scoping review but would have been difficult to isolate without more details about each one from the authors themselves. Many of these more specific tasks can be bundled under more general terms such as using IT tools in order to conserve space during the publication process.

Upon consideration of the data from our scoping review one of the members of the Thesis advisory committee, prior to the start of interviews, recommended gathering data on the stakeholders that were engaged in by our interviewees. In terms of relevant stakeholder groups to engage as part of this decision process, interviewees reported consulting with a wide array of different people when making the decision of which tests to target for intervention. Clinical staff were consulted by almost all the experts, including nursing staff, respiratory therapists, occupational therapists and others. Three other groups that were consulted very commonly were laboratory personnel, administration, and IT staff. Finally, there were a few other stakeholder groups that were consulted more occasionally including, pharmacist, patient partners, KT Staff, clinical trainees and Choosing Wisely Coordinators. The importance of including the correct people test-ordering projects was later echoed by many of our interviewees. Many studies have shown the importance of engaging a wide array of stakeholders when trying to make meaningful change within the health care setting and this should be no exception.[32-36] As a result, this would certainly be an important component of our framework, an element that assists in ensuring all of the correct people are being engaged in the decision of which tests should be targeted for intervention.

In addition to asking interviewees about which factors guided their decisions on which tests to target, they were also asked about if and how they went about quantifying these factors. We explored whether experts used techniques (beyond financial cost) to standardize the quantification of their test-reduction efforts. Aside from tangibly measuring the impact on financial cost, the use of measurement tools/techniques such as an electronic medical record or CPOE system were rarely used or considered by intervention development experts. The most common methods used were the implementation of IT Tools and consultation with the literature to determine how others attempted to measure certain factors. Other concepts discussed were internal tools used to assess the impact on patient care if a certain test-ordering behavior was intervened on. An interesting measurement technique mentioned by two separate experts was the use of cost of testing solely to measure (based on the premise that reduced cost is directly a result of improved test-ordering practice) improvement rather than as a motivator for change. We would recommend the further development of more objective and standardized measurement techniques to be used in future interventions and would integrate these tools into our framework.

The final section of our semi-structured interviews asked our experts about recommendations for things they would hope to see in our framework, and this might be a starting point for our consensus process which would seek to synthesize all of our previous steps in to the functional framework that we hope to create. Some recommendations mentioned by two experts included creating a built-in tool in the framework that helps assess the ease of implementation of the intervention on the test-

ordering practice i.e. a method integrated into the framework that helps assess the clinical significance of the project. Other recommendations included creating tools that assist the user of the framework with choosing appropriate staff for the project and suggesting ways to obtain assistance with financial resource allocation to ensure success of the project. Another recommendation of interest was that the framework consider the level at which the intervention was taking place and make test-ordering recommendations specific to whether it was being done at the hospital level or the unit level. Finally, one of our interviewees mentioned that the framework would be extremely useful if it was able to incorporate the socio-political implications of making a test-ordering change and how the change may cause unintended consequences to arise. For example, removing the funding for a specific test that only a select number of very specialized clinicians use but is essential to the care of their specific population causes a big uproar from that clinician group, which can be avoided if due diligence is taken in deciding which tests to select.

Figure 1. Preliminary framework of the essential elements involved in implementing an intervention designed to improve test-ordering practice.

Considerations for Future Research

Following the analysis on the results of both our scoping review and semi-structured interviews we have set the stage for developing a prioritization framework for intervention development experts to use to choose which tests to pursue for intervention. After having assessed our combined results, we feel that we are not quite at the stage to propose an entire framework. We believe that the scoping review and interviews set an excellent base for our desired framework, but that additional work is needed to create a standalone framework that is ready to be used by other institutions to improve testing at their respective facility.

We believe that more interviews are needed to elicit a greater breadth of knowledge from the intervention development experts. While we believe we were

nearing saturation with themes, we were not able to conduct enough interviews to be sure, particularly given that different themes were coming from different kinds of stakeholder. We hope to perform approximately 5-7 more interviews, which would bring us to a level where we feel saturation would be achieved. Among the interviewees, we noticed certain trends in their answers which may start to be an indication of saturation on certain themes.[37] Intervention development experts who were in similar roles seemed to have similar answers to the factors and processes that they found to be important. As a result of this trend, we believe that performing further interviews would assist us in solidifying our knowledge on how each participant's answer relates to their specific role. This information could be used to create more tailored approaches for use of our framework, depending on the background of the intervention development expert that would be using it. It also suggests our framework may need to take a clearer and more specific perspective as to who it would be targeted than we had originally envisioned. Again, further interviews would likely assist in determining whether the need and benefit would be greatest if the framework took an approach centered on clinical intervention developers or lab-centered developers. Each would come with their own unique advantages and challenges and would likely lead to the framework being framed slightly differently.

In addition, we believe that a consensus process or a Modified Delphi might be helpful after the remaining interviews were completed in pursuing the greater overall endeavor of creating our framework.[38-40] A wide variety of factors and processes were proposed by our interviewees such that it isn't clear which are most effective or important in deciding which tests to target for intervention. The future work on this project will likely present the processes and factors reported by our interviewees to another panel of experts and have them rate and prioritize certain processes and factors, the most important of which would be included in our framework.[40] Additionally, our interviewees provided us with specific recommendations on important elements to be included in our framework, from which attempts to prioritize ourselves and then gain input from the same panel of experts would be done.

Information pertaining to which personnel, factors and processes are essential during every test-ordering intervention project and those that depend on the context is still outstanding and could be addressed using a consensus process. Among our interviewees, there was variation among who was consulted (stakeholder groups) to help make the decision of which tests to target for intervention but there were also some common trends on certain staff that seemed to be consistently approached for advice. As a result, we believe that it would be important for our framework to include guidance on who should be involved in the decision to target certain tests. An example of someone that would only be consulted in specific circumstances would be consulting with the pharmacist on the improved test-ordering practice of a test that is used for therapeutic drug monitoring. In this instance the pharmacist's expertise would be invaluable for determining whether a specific test is good target for intervention.

In a similar vein, we found a wide array of factors between our scoping review and our interviews. Based on the information provided by our interviewees, we believe

that there are certain factors (e.g. cost, test value, prevalence of test) that must be considered every time a decision is being made on which tests to target for intervention. On the other hand, there are certain factors that need only be considered in certain circumstances and we would expect to use our consensus process to gather further information on this and provide guidance on the clinical instances in which certain factors merit greater consideration than others. Some example of factors that may not apply in all situations would be *test being bundled with other tests* or *laboratory workload of the test*. These factors wouldn't necessarily always apply because some tests are never bundled with others and other tests are run by automatic machines that do not require human resources for the operation of the test nor the interpretation of the test. Thus, the test would have a negligible impact on laboratory workload and this factor would not need to be considered.

We believe that a framework should also provide guidance on the most effective decision-making processes to use when making the decision which test to target for intervention. We understand that resources allocation (i.e. financial, resource, staff) is not equally distributed throughout all institutions and we hope to make our framework as useful for as many differently sized institutions as possible. As a result, our recommendations for certain processes would consider this disparity. For example, an ideal starting process to decide which test to target would be a systematic review of the literature.[31,40] For a large academic institution this is an entirely manageable step but for a smaller community institution it becomes much more difficult and therefore we might suggest a review of previous guidelines published addressing the test of interest rather than the more comprehensive systematic review. Overall, we would create process recommendations that allow a wide variety of institutions to use our framework.

Limitations and Considerations

Our studies had several limitations that warrant consideration. Our search strategy was implemented in only one database (Medline) as a result of limited resources. Future work should expand the search to include other relevant databases (Embase, CINHAL, Psycinfo). We sought to reduce the effects of this limitation by enhancing our review with snowball sampling to ensure that any relevant articles that may not have been in the initial search strategy were included. While we used the review to identify a wide range of factors potentially relevant to the decision of choosing which test to target for intervention, we cannot be sure which ones are most important overall, or in which circumstance some will become more important. Our study could have been limited by publication bias as it pertains to empirical studies because it is possible that only those that had succeeded in showing some kind of improvement of test ordering appropriateness would have been published.[41] One final limitation is that when categorizing whether a factor was mentioned or not, we used a rather inclusive approach such that almost any indication that an intervention developer referred to that factor was included. In some cases that factor may not have been instrumental in the planning stage of the study, but rather only given limited consideration or even considered post-hoc. While we tried to limit this by only including factors that were mentioned in the

introduction or methods sections of the study, the limitation remains and part of the rational for performing semi-structured interview was to address this limitation.

We believe that the unfortunate emergence of the COVID-19 pandemic led to limitations in our interview study. The pandemic took hold globally just two weeks after we began recruitment for the semi-structured interviews. This impacted the interviews in two ways. We believe that the number of experts that were able to make time for an interview with us was more limited because of the acute nature of the pandemic. Indeed, one potential clinician respondent, when asked for an interview, responded that the pandemic had already increased their clinical hours well beyond normal and thus they couldn't make time for an interview. Though only one clinician responded this way, we hypothesize that many of those potential interviewees who did not send us a reply, were unable to perform an interview for the same reason. This issue seemed to significantly slow down interviewee recruitment as interviews with clinicians, became exceedingly difficult to come by. We feel that factors and process that clinicians may have reported as being important are insufficiently represented. Moreover, laboratory experts became so overwhelmingly busy in attempting to ensure that their facilities were effectively able to process all the testing needs of the current pandemic. Thus, though we do not as feel as though our laboratory experts were inadequately represented, the number of participants that we were able to recruit was still more limited than we had hoped for our framework.

Conclusions

The current work prepares the way to develop a framework designed to help intervention developers choose which tests can most efficiently result in improved test-ordering processes. By identifying all relevant factors, processes, and stakeholders thought to be relevant to designing such interventions, we have laid the groundwork for such a framework. After additional interviews to ensure saturation of themes, we will be able to proceed with framework development, perhaps involving a consensus process of all relevant stakeholders. We hope to widely distribute our framework to assist intervention development experts working in a wide variety of milieus to help them decide which tests are worth targeting for intervention such that their respective institutions can provide the highest quality of care to patients.

References

1. Institute of Medicine (US) Roundtable on Evidence-Based Medicine. *The Healthcare Imperative: Lowering Costs and Improving Outcomes. Workshop Series Summary*. (Yong P, Saunders R, Olsen L, eds.). Washington, DC: National Academies Press; 2010. doi:10.17226/12750.

2. Lenzer J. Unnecessary care: are doctors in denial and is profit driven healthcare to blame? *Bmj*. 2012;345(oct02 3):e6230-e6230. doi:10.1136/bmj.e6230.

3. Wintemute K, McDonald K, Huynh T, Pendrith C, Wilson L. Addressing overuse starts with physicians. Choosing Wisely Canada. *Can Fam Physician*. 2016;62(3):199-200. doi:62/3/199 [pii].

4. Cho HJ. Countering Medicine's Culture of More. *Am Med Assoc J Ethics*. 2015;17(11):1079-1081.

5. Canadian Institute for Health Information. Canadians have more than 1 million potentially unnecessary medical tests and treatments every year. Canadian Institute for Health Information Web Page. https://www.cihi.ca/en/canadians-have-more-than-1-million-potentially-unnecessarymedical-tests-and-treatments-every-year. Published 2017. Accessed May, 2021.

6. Berwick DM, Hackbarth AD. Eliminating Waste in US Health Care. *JAMA*. 2012;307(14):1513- 1516. doi:10.1001/jama.2012.362.

7. Canadian Institute for Health Information. Unnecessary Care in Canada: Infographic. Canadian Institute for Health Information Web Page. https://www.cihi.ca/en/unnecessary-care-in-canadainfographic. Published 2017. Accessed May, 2021.

8. Thomas RE, Vaska M, Naugler C, Turin TC. Interventions at the laboratory level to reduce laboratory test ordering by family physicians: Systematic review. *Clin Biochem*. 2015;48(18):1358-1365. doi:10.1016/j.clinbiochem.2015.09.014.

9. Lewandrowski K, Baron J, Dighe A. Utilization Management in the Clinical Laboratory: An Introduction and Overview. In: Lewandrowski K, Sluss P, eds. *Utilization Management in the Clinical Laboratory and Other Ancillary Services*. Springer, Cham; 2017.

10. Sluss P. Calculating Costs and Savings in Utilization Management. In: Lewandrowski K, Sluss P, eds. *Utilization Management in the Clinical Laboratory and Other Ancillary Services.* Springer, Cham; 2017.

11. Podolsky E, Hudek H, Carroll K, McCudden C, Presseau J, Brehaut J. "Factors Reported in Determining Laboratory Tests to Target for Intervention: A Scoping Review." (In preparation)

12. Podolsky E, Hudek H, Carroll K, McCudden C, Presseau J, Brehaut J. “How Do Experts Determine Where to Intervene on Test Ordering? An Interview-Based Study.” (In preparation)

13. Naugler C. Laboratory Utilization Management in Canada. In: Lewandrowski K, Sluss P, eds. *Utilization Management in the Clinical Laboratory and Other Ancillary Services*. Springer, Cham; 2017.

14. Naugler C. A Perspective on Laboratory Utilization Management from Canada. *Clin Chim Acta*. 2014;427:142-144. doi:10.1016/j.cca.2013.09.022.

15. van Walraven C, Naylor CD. Do We Know What Inappropriate Laboratory Utilization Is? A Systematic Review of Laboratory Clinical Audits. *JAMA*. 1998;280(6):550-558. doi:10.1309/AJCPX1HIEM4KLGNU.

16. Yeh DD. Laboratory Utilization. In: Lewandrowski K, Sluss PM, eds. *Utilization Management in the Clinical Laboratory and Other Ancillary Services.* Cham: Springer International Publishing; 2017:203-210. doi:10.1007/978-3-319-34199-6_18.

17. Busby J, Schroeder K, Woltersdorf W, et al. Temporal growth and geographic variation in the use of laboratory tests by NHS general practices: Using routine data to identify research priorities. *Br J Gen Pract*. 2013;63(609):256-266. doi:10.3399/bjgp13X665224.

18. Ezzie ME, Aberegg SK, O’Brien JM. Laboratory Testing in the Intensive Care Unit. *Crit Care Clin*. 2007;23(3):435-465. doi:10.1016/j.ccc.2007.07.005.

19. Sedrak MS, Patel MS, Ziemba JB, et al. Residents’ Self-Report on Why They Order Perceived Unnecessary Inpatient Laboratory Tests. J Hosp Med. 2016;11(12):869-872. doi:10.1002/jhm.2645.

20. Sood R, Sood A, Ghosh AK. Non-Evidence-Based Variables Affecting Physicians’ Test-Ordering 130 Tendencies: A Systematic Review. *Neth J Med*. 2007;65(5):167-177. http://www.ncbi.nlm.nih.gov/pubmed/17519512.

21. Cadogan SL, Browne JP, Bradley CP, Cahill MR. The effectiveness of interventions to improve laboratory requesting patterns among primary care physicians: a systematic review. *Implement Sci*. 2015;10:167. doi:10.1186/s13012-015-0356-4.

22. Hindmarsh JT, Lyon AW. Strategies to Promote Rational Clinical Chemistry Test Utilization. *Clin Biochem*. 1996;29(4):291-299.

23. Brateanu A, Rothberg MB. Why do clinicians continue to order “routine preoperative tests” despite the evidence? *Cleve Clin J Med*. 2015;82(10):667-670. doi:10.3949/ccjm.82a.15118.

24. Schroeder SA, Kenders K, Cooper J, Piemme TE. Use of Laboratory Tests and Pharmaceuticals: Variation Among Physicians and Effect of Cost Audit on Subsequent Use. *JAMA*. 1973;225(8):969-973. doi:10.1001/jama.1973.03220360029007.

25. Morgen EK, Naugler C. Inappropriate Repeats of Six Common Tests in a Canadian City: A Population Cohort Study within a Laboratory Informatics Framework. *Am J Clin Pathol*. 2015;144(5):704-712. doi:10.1309/AJCPYXDAUS2F8XJY.

26. Geerligs, L., Rankin, N.M., Shepherd, H.L. *et al.* Hospital-based interventions: a systematic review of staff-reported barriers and facilitators to implementation processes. *Implementation Sci* **13,** 36 (2018). https://doi.org/10.1186/s13012-018-0726-9

27. Kredo, T., Bernhardsson, S., Machingaidze, S., Young, T., Louw, Q., Ochodo, E., & Grimmer, K. (2016). Guide to clinical practice guidelines: the current state of play. *International journal for quality in health care : journal of the International Society for Quality in Health Care*, *28*(1), 122–128. https://doi.org/10.1093/intqhc/mzv115

28. Milliken A. D. (2014). Physician engagement: a necessary but reciprocal process. *CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne*, *186*(4), 244–245. https://doi.org/10.1503/cmaj.131178.

29. Taitz, J. M., Lee, T. H., & Sequist, T. D. (2012). A framework for engaging physicians in quality and safety. *BMJ quality & safety*, *21*(9), 722–728. https://doi.org/10.1136/bmjqs-2011-000167

30. Scott, C. G., Thériault, A., McGuire, S., Samson, A., Clement, C., & Worthington, J. R. (2012). Developing a physician engagement agreement at The Ottawa Hospital: a collaborative approach. *Healthcare quarterly (Toronto, Ont.)*, *15*(3), 50–53. https://doi.org/10.12927/hcq.2013.23020

31. Gopalakrishnan, S., & Ganeshkumar, P. (2013). Systematic Reviews and Meta-analysis: Understanding the Best Evidence in Primary Healthcare. *Journal of family medicine and primary care*, *2*(1), 9–14. https://doi.org/10.4103/2249-4863.109934

32. Bosch, B., & Mansell, H. (2015). Interprofessional collaboration in health care: Lessons to be learned from competitive sports. *Canadian pharmacists journal : CPJ = Revue des pharmaciens du Canada : RPC*, *148*(4), 176–179. https://doi.org/10.1177/1715163515588106

33. Romanow RJ. *Building on values: the future of health care in Canada*. Ottawa (ON): Canada, Commission on the Future of Health Care in Canada; 2002. Available: http://publications.gc.ca/collections/Collection/CP32-85-2002E.pdf (accessed June, 2021).

34. Wheelan, S. A., Burchill, C. N., & Tilin, F. (2003). The link between teamwork and patients' outcomes in intensive care units. *American journal of critical care : an official publication, American Association of Critical-Care Nurses*, *12*(6), 527–534.

35. Martinez, A. S., Saef, J., Paszczuk, A., & Bhatt-Chugani, H. (2013). Implementation of a pharmacist-managed heart failure medication titration clinic. *American journal of health-system pharmacy : AJHP : official journal of the American Society of Health-System Pharmacists*, *70*(12), 1070–1076. https://doi.org/10.2146/ajhp120267

36. Chang, W. Y., Ma, J. C., Chiu, H. T., Lin, K. C., & Lee, P. H. (2009). Job satisfaction and perceptions of quality of patient care, collaboration and teamwork in acute care hospitals. *Journal of advanced nursing*, *65*(9), 1946–1955. https://doi.org/10.1111/j.1365-2648.2009.05085.x

37. Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2018). Saturation in qualitative research: exploring its conceptualization and operationalization. *Quality & quantity*, *52*(4), 1893–1907. https://doi.org/10.1007/s11135-017-0574-8

38. Jorm A. F. (2015). Using the Delphi expert consensus method in mental health research. *The Australian and New Zealand journal of psychiatry*, *49*(10), 887–897. https://doi.org/10.1177/0004867415600891

39. Niederberger, M., & Spranger, J. (2020). Delphi Technique in Health Sciences: A Map. *Frontiers in public health*, *8*, 457. https://doi.org/10.3389/fpubh.2020.00457

40. Woolf, S. H., Grol, R., Hutchinson, A., Eccles, M., & Grimshaw, J. (1999). Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. BMJ (Clinical research ed.), 318(7182), 527–530. https://doi.org/10.1136/bmj.318.7182.527.

41. Sterne J, Egger M, Moher D (editors). Chapter 10: Addressing reporting biases. In: Higgins J,Green S, eds. *Cochrane Handbook for Systematic Reviews of Intervention. Version 5.1.0 (Updated March 2011)*. The Cochrane Collaboration; 2011