CHOOSING INDIRECT

EVIDENCE FOR CLINICAL PRACTICE GUIDELINE PANELS

OPTIMIZING THE PRESENTATION OF INDIRECT EVIDENCE FOR CLINICAL PRACTICE GUIDELINE PANELS THAT USE THE GRADE APPROACH FOR DECISION-MAKING

By JOHN JOSEPH RIVA, D.C., M.Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2020)

Hamilton, Ontario (Health Research Methodology Program).

TITLE: Optimizing The Presentation Of Indirect Evidence For Clinical Practice Guideline Panels That Use The GRADE Approach For Decision-Making

AUTHOR: John J. Riva, D.C., M.Sc. (McMaster University)

SUPERVISOR: Associate Professor Jan L. Brozek, M.D., Ph.D.

NUMBER OF PAGES: xviii, 286 pages.

LAY ABSTRACT

While many clinical practice guidelines have ample published evidence to support their recommendations, certain situations (e.g. rare, rapidly emerging, or understudied diseases) may be limiting the available literature. As a result, it is common for guideline developers to seek out indirect evidence from other, but related areas, to fill in these evidence gaps. Selection of available indirect evidence may be better than basing decisions on no evidence, in particular in situations of clinical equipoise. However, including all potentially relevant indirect evidence may represent an overuse of evidence. Indirect evidence refers to information sources with related populations, interventions, outcomes or comparisons, which could reasonably be extrapolated; but, are not entirely specific to the research topic at hand. This confluence of both indirect information considerations and desire by many for increasing literature sources to draw from weighs on the simplicity of an overall summary of literature presented during a guideline recommendation decision-making process. Herein, firstly, we described an example of explicit decision-rules for including indirect evidence that were specific and the implications of the rules for presenting results to decision-makers. Secondly, we provided a comprehensive overview of how guideline developers currently report economic information across guideline frameworks, in particular with respect to indirectness. Lastly, we described the most important study characteristics suggested by economists to consider as decision-rules when assessing economic evaluations for use as research evidence in a guideline.

This work presents important concepts for guideline developers to consider when choosing indirect evidence sources in their clinical practice guidelines. Our findings have the potential to simplify the presentation of indirect evidence for guideline panels and developers, as well as, to reduce decision-making confusion, time demands and guideline funder costs.

ABSTRACT

While many clinical practice guidelines that use the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach have ample published evidence to support their recommendations, certain situations (e.g. rare, rapidly emerging, or understudied diseases) may be limiting. As a result, it is common for guideline developers to seek out indirect evidence from other areas to fill in these evidence gaps. The GRADE evidence-to-decision (EtD) framework, which offers a structured and transparent development process for guidelines, includes additional research evidence domains (e.g. feasibility, acceptability, equity) for panels to consider in their decision-making process. This confluence of both considerations of indirect information and increasing literature domain sources to draw from when making decisions weighs on the simplicity of literature presentation. Herein, firstly, we described an example of specific decision-rules for including indirect evidence and the implications of the rules for presenting results to decision-makers. Secondly, we provided a comprehensive overview of how guideline developers currently report economic information across GRADE evidence-to-decision frameworks. Lastly, we ranked the most important study characteristics suggested in the literature by economists to consider as decision-rules when assessing indirectness (transferability) of economic evaluations chosen as research evidence in a GRADE guideline. We conclude that developers, with the help of their panels, should work to establish and report clear decision-rules and the rationale for indirect evidence that they select for their clinical practice guidelines. This has the potential to simplify the presentation of indirect evidence for panels and developers, as well as, to reduce decision-making confusion, time demands and guideline funder costs.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor, members of my PhD committee and mentors, Dr. Jan Brozek, Dr. Jason Busse, Dr. Feng Xie and Dr. Holger Schünemann. You have collectively and independently guided me to become a better researcher. My experience with each of you has been invaluable. Thank you for your patience and direction.

Thank you to Lorraine Carroll at the department of Health Research Methods, Evidence, and Impact (HEI) for your tireless advocacy and helping me through this process. Without you a lot would not be possible. Also, a special thanks to my colleagues and friends at HEI for the support and unforgettable moments.

A big thanks to my friends and family who have been understanding of the time demands that come along with this type of commitment and their offers of support. Someone close once said, "It's not what you say that is important; just what you do in the end." Building on this idea, it is also what you publish that matters.

John J. Riva

Hamilton, Ontario - September 2020

PREFACE

The work in this dissertation is presented as three manuscripts which have been published, accepted for publication or prepared for submission framed by an introduction (chapter 1) and summary and conclusion (chapter 5). The manuscript in Chapter 2, "Predictors of prolonged opioid use following initial prescription for acute musculoskeletal injuries in adults: a systematic review and meta-analysis of observational studies", was published August 18, 2020 in the journal *Annals of Internal Medicine*. The manuscript in Chapter 3, "Reporting of economic information in GRADE guidelines that use evidence-to-decision (EtD) frameworks: a systematic survey", is planned for submission to the *Journal of Clinical Epidemiology*. The manuscript in Chapter 4, "Selection of economic evaluations for use in GRADE clinical practice guidelines: a systematic review of transferability (indirectness) factors", is also planned for submission to the *Journal of Clinical Epidemiology*.

The systematic review and meta-analysis presented in Chapter 2 serves as part of the evidence base for a joint clinical practice guideline from the American College of Physicians and American Academy of Family Physicians; led by my committee member, Dr. Jason Busse. I developed the search strategy (with help from a resource librarian Rachel Coubin, MISt), developed the protocol and screened, abstracted data, analyzed data (with help of statistician Li Wang, PhD to create 4 forest plots in Stata) and drafted summaries of our findings to present to the guideline panel, with input of my committee member Dr. Jason Busse. I drafted the manuscript which was circulated to co-authors. I incorporated feedback from the co-authors, prepared and submitted the manuscript and responded to journal peer-review feedback for submission. Chapter 3 was a systematic survey that I conceived and coordinated under the supervision of my supervisor Dr. Jan Brozek. I developed the search approach, screened and

abstracted data from guideline evidence-to-decision frameworks, along with a research team. I analysed the data and prepared the manuscript for submission. Chapter 4 was a systematic review that I also conceived of and coordinated with my supervisor Dr. Jan Brozek. I developed the search, screened and abstracted data from articles along with a research team. I drafted the manuscript, incorporated feedback from the co-authors and prepared the manuscript for submission.

I received a PhD tuition award from the NCMIC Foundation (https://www.ncmicfoundation.org/), who had no part in the conception, development or manuscripts related to these projects. The systematic review and meta-analysis presented in Chapter 2 in this dissertation was funded by the National Safety Council to inform the American College of Physician and American Academy of Family Physicians; guidelines. The systematic survey and systematic review described in Chapters 3 and 4, respectively, were otherwise unfunded.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter Two

Chapter Three

Chapter Four

LIST OF TABLES

Chapter Two

Chapter Three

Chapter Four

LIST OF APPENDICES

Chapter Two

Chapter Three

Chapter Four

LIST OF ABBREVIATIONS AND SYMBOLS

ARDS: Acute Respiratory Distress Syndrome

ARI: Absolute Risk Increase

$AU: Australian Dollars

CI: Confidence Interval

CDC: Centers for Disease Control and Prevention

CERQual: Confidence in the Evidence from Reviews of Qualitative research

CoE: Certainty of Evidence

COVID-19: Coronavirus Disease of 2019

CPG: Clinical Practice Guideline

EP: Evidence Profile

EtD: Evidence-to-Decision

EU: European Union

GIN: Guidelines International Network

GRADE: Grading of Recommendations Assessment, Development and Evaluation

HR: Hazard Ratio

HIV: Human Immunodeficiency Virus

HTA: Health Technology Assessment

ICD-9: International Clinical Diagnosis Code version 9

ICD-10: International Clinical Diagnosis Code version 10

IQR: Interquartile Range

n: sample size

OR: Odds Ratio

PCT: Patent Cooperation Treaty

PICO: Population, Intervention, Comparison and Outcome

QUIPS: Quality in Prognosis Studies Tool

RCT: Randomized Controlled Trial

RoB: Risk of Bias

SoF: Summary of Findings

Stata: StataCorp statistical software

SR: Systematic Review

UK: United Kingdom

USA: United States of America

WHO: World Health Organization

DECLARATION OF ACADEMIC ACHIEVEMENT

I declare that I, jointly with my supervisor, Associate Professor Jan L. Brozek, and committee member Associate Professor, Jason W. Busse, played the primary role in the conception, design, and execution of the studies here included. We obtained feedback and advice from other PhD committee members Professors Xie and Schünemann, as well as from clinical and methodological experts that co-authored the work.

This work is original research that I conducted. I am the principle contributor and first author of all the manuscripts contained in this dissertation. I conducted all analyses, designed figures and tables, and organized meetings; I developed one search strategy in conjunction with a resource librarian (Rachel Coubin, MA, MISt). I performed the meta-analysis in conjunction with a statistician (Assistant Professor, Li Wang, PhD) for the 4 forest plots in one paper. I wrote the manuscripts with editorial advice and supervision of Professors Brozek and Busse, and received feedback and advice from Professors Xie and Schünemann. The co-authors on each paper contributed significantly with important comments and advice for the final manuscripts.

For all the three manuscripts composing this "sandwich" thesis, earlier drafts of parts of this research have been presented at academic conferences, including the Canadian Pain Society's 40th Annual Scientific Meeting in Toronto, April 2019 and the local GRADE working group meeting in Hamilton, June 2019. The first manuscript (chapter 1) has been published in the journal *Annals of Internal Medicine* on August 18, 2020. The second and third manuscripts are final drafts, which will be submitted to the *Journal of Clinical Epidemiology*.

CHAPTER 1. INTRODUCTION

Reporting indirect evidence to GRADE guideline panels

The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach is one of the most widely endorsed guideline development systems.[1] While some clinical practice guidelines that use the GRADE approach have ample published evidence to support their recommendations, certain situations (e.g. rare, rapidly emerging, or understudied diseases) may be limiting. As a result, it is common for guideline developers to seek out indirect evidence from different, but related areas to fill in these evidence gaps. Historically, the GRADE working group has suggested the consideration of these indirect information sources in four ways, through: 1) population indirectness, including demographic, setting or species differences; 2) intervention indirectness, including component, jurisdictional, or delivery; 3) outcome indirectness, including time-frame or use of surrogates; and 4) indirect comparisons.[2]

There has been a push by all research communities to be better reflect patient values and preferences literature.[3] At the same, the GRADE evidence-to-decision (EtD) frameworks have been developed to support the development process, which includes additional research evidence domains (e.g. feasibility, acceptability, equity) for panels to consider in their decision-making process.[4] This confluence of the above mentioned indirect information considerations and increasing literature domain sources to draw from weighs on the simplicity of literature presentation to panel members and the burden required by development teams to produce synthesized summaries that remain useful for recommendation decision-making.

For example, GRADE guideline developers and panel members may struggle with the balance between use of indirect information and simplicity. Too little use of indirect evidence may result in an inability to make a recommendation, although guideline consumers prefer that those developing recommendations make them even in the face of very low certainty evidence.[5] However, unclear decision-rules for selecting indirect evidence can lead to either overwhelming amounts or irrelevant information, which in turn complicate the presentation of information in the summary of findings for panel members that they process during their deliberations.[6] This lack of decision-rules hampers efforts to narrow literature summaries from including any or all information remotely relevant to a topic that guideline developers may find.[7] Setting clear inclusion and exclusion criteria for indirect evidence may help overcome this barrier to appropriate use of evidence, but for certain domains of evidence (e.g. economic evaluations) methodological approaches have not been developed in the context of GRADE.

Why is this research important?

Some pragmatic examples highlight why this research is important. In recent American Society of Hematology (ASH) guidelines, venous thromboembolism was rare; but, it had the potential for severe consequences in the context of pregnancy or in pediatric populations.[8,9] As a result, large proportions of indirect information were used to inform many research question GRADE summary of findings tables. In the context of pregnancy (indirect 'non-pregnant' literature) and with pediatric (indirect 'adult' literature) populations, both direct and indirect sources were presented alongside each other and essentially doubling the size of outcome tables that panel members reviewed.

With respect to rapidly emerging disease, the Canadian Medical Association Journal (CMAJ) recently published a GRADE guideline that considered corticosteroids in patients with severe corona virus disease of 2019 (COVID-19) and acute respiratory distress syndrome (ARDS), authors reported outcomes one direct study and seven indirect studies from populations in other viral, bacterial or non-infectious causes to support their recommendation.[10]

Beyond the information about benefits and harms, systematic reviews of patients values and preferences information were done in GRADE guidelines by the Kingdom of Saudi Arabia Ministry of Health to identify local information. After no eligible studies were identified, the guideline development group on migraine headaches decided to widen their inclusion to use information on values and preferences related migraine headache from elsewhere in the world.[4]

While these strategies were thorough, the examples suggest that, without a framework the information may be too indirect (e.g. global values and preferences) or too complicated for panel members to keep track and weigh multiple outcomes, both supported by direct and indirect evidence, to formulate a guideline recommendation. Therefore, improving the clarity around selection and presentation of indirect evidence has the potential to impact the simplicity of summary information presented to guidelines panels.

Objectives

The goals of this dissertation were threefold:

1. Describe an example of decision-rules for GRADE evidence profile outcome indirectness that were specific and the implications to the simplicity of the resulting absolute measures of combined (direct and indirect) prevalence and risk factor summary information that

panel members used in their GRADE guideline recommendation decision-making (Chapter 2);

2. Provide a comprehensive overview of how guideline developers currently report economic information, with widely varying levels of indirectness, across a convenience sample of recently published GRADE evidence-to-decision frameworks in the absence of well formulated guidance (Chapter 3);
3. Prioritize the most important study characteristics suggested in the literature by economists to consider as decision-rules when assessing indirectness of economic evaluations chosen as research evidence in a GRADE guideline (Chapter 4).

Thesis overview

This work comprises three main concepts described above with a final chapter that integrates the work done on the three chapters and raising the implications of these findings for future research.

As stated above, Chapter 2 presents a systematic review and meta-analysis of observational studies that reports predictors of prolonged opioid use following prescription for an acute musculoskeletal injury and provides clear decision rules for including indirect evidence and absolute measures that combine both direct and indirect evidence sources. Chapter 3 entails a systematic survey of GRADE evidence-to-decision frameworks from published guidelines that describes the wide variability of strategies developers use to report economic evidence from the literature, in the absence of clear selection decision-rules. Chapter 4 includes the systematic review summary of study characteristics that economists suggest are the most important decision-rules to assess the indirectness of economic evaluations when choosing studies to use as research

evidence. Lastly, Chapter 5 presents an overall summary of these findings and a discussion including implications for future research.

References

1. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ; GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008 Apr 26;336(7650):924-6.
2. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, et al; GRADE Working Group. GRADE guidelines: 8. Rating the quality of evidence--indirectness. J Clin Epidemiol. 2011 Dec;64(12):1303-10.
3. Zhang Y, Coello PA, Brożek J, Wiercioch W, Etxeandia-Ikobaltzeta I, et al. Using patient values and preferences to inform the importance of health outcomes in practice guideline development following the GRADE approach. Health Qual Life Outcomes. 2017 May 2;15(1):52.
4. Moberg J, Oxman AD, Rosenbaum S, Schünemann HJ, Guyatt G, Flottorp S, Glenton C, Lewin S, Morelli A, Rada G, Alonso-Coello P; GRADE Working Group. The GRADE Evidence to Decision (EtD) framework for health system and public health decisions. Health Res Policy Syst. 2018 May 29;16(1):45.
5. Neumann I, Alonso-Coello P, Vandvik PO, Agoritsas T, Mas G, et al. Do clinicians want recommendations? A multicenter study comparing evidence summaries with and without GRADE recommendations. J Clin Epidemiol. 2018 Jul;99:33-40.
6. Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G,. GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. J Clin Epidemiol. 2013 Feb;66(2):158-72.

7. Hill SR, Olson LG, Falck-Ytter Y, Cruz AA, Atkins D, Baumann M, Jaeschke R, Woitalla T, Schünemann HJ; ATS/ERS Ad Hoc Committee on Integrating and Coordinating Efforts in COPD Guideline Development. Incorporating considerations of cost-effectiveness, affordability, and resource implications in guideline development: article 6 in Integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report. Proc Am Thorac Soc. 2012 Dec;9(5):251-5.
8. Bates SM, Rajasekhar A, Middeldorp S, McLintock C, Rodger MA, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: venous thromboembolism in the context of pregnancy. Blood Adv. 2018 Nov 27;2(22):3317-3359.
9. Monagle P, Cuello CA, Augustine C, Bonduel M, Brandão LR, et al. American Society of Hematology 2018 Guidelines for management of venous thromboembolism: treatment of pediatric venous thromboembolism. Blood Adv. 2018 Nov 27;2(22):3292-3316.
10. Ye Z, Rochwerg B, Wang Y, Adhikari NK, Murthy S, et al. Treatment of patients with nonsevere and severe coronavirus disease 2019: an evidence-based guideline. CMAJ. 2020 Apr 29:cmaj.200648.

CHAPTER 2. PREDICTORS OF PROLONGED OPIOID USE FOLLOWING INITIAL PRESCRIPTION FOR ACUTE MUSCULOSKELETAL INJURIES IN ADULTS: A SYSTEMATIC REVIEW AND META-ANALYSIS OF OBSERVATIONAL STUDIES

Status: Published in the journal *Annals of Internal Medicine, August 18, 2020.*

Predictors of prolonged opioid use following initial prescription for acute musculoskeletal injuries in adults: a systematic review and meta-analysis of observational studies

Authors

John J. Riva[1,2], Salmi T. Noor[3], Li Wang[4], Vahid Ashoorion[4], Farid Foroutan[2], Behnam Sadeghirad[2,4], Rachel Couban[4], Jason W. Busse[2,4,5,6]

[1] Department of Family Medicine, McMaster University, Hamilton, Ontario, Canada.

[2] Department of Health Research Methods, Evidence & Impact, McMaster University, Hamilton, Ontario, Canada.

[3] Department of Medicine, University of Ottawa, Ottawa, Ontario, Canada.

[4] Department of Anesthesia, McMaster University, Hamilton, Ontario, Canada.

[5] Michael G. DeGroote National Pain Centre, McMaster University, Hamilton, Ontario, Canada.

[6] Chronic Pain Centre of Excellence for Canadian Veterans, Hamilton, Ontario, Canada.

Corresponding author: John J. Riva, Department of Family Medicine, David Braley Health Sciences Centre, 100 Main Street West, 6th Floor, Hamilton, ON L8P 1H6; email: rivaj@mcmaster.ca

Previous presentations: Presented as a poster presentation at Canadian Pain Society's 40th Annual Scientific Meeting, April 2-5, 2019 in Toronto, Ontario.

Key points

1. Prolonged use in those prescribed opioids for acute musculoskeletal injuries is common.
2. Our findings, from both direct and indirect evidence, suggest avoiding prescribing opioids to patients with past or current substance use disorder and, when prescribed, restricting duration to ≤ 7 days and lower doses.

Abstract

Background: Opioids are frequently prescribed for acute musculoskeletal injuries and may result in long-term use and consequent harms.

Purpose: A systematic review to explore factors associated with persistent opioid use after prescription for acute musculoskeletal injury.

Data sources: MEDLINE, EMBASE, Web of Science and Google Scholar, through January 2020, and reference lists of selected articles.

Study selection: Observational studies of adults with opioid prescriptions for outpatient acute musculoskeletal injuries, in an adjusted model, that explored risk factors for prolonged use.

Data extraction: Six reviewers, working in pairs, independently extracted data, rated the quality of studies, and evaluated the certainty of evidence.

Data synthesis: Fourteen cohorts with 13 263 393 participants were included. The overall prevalence of prolonged opioid use after musculoskeletal injury for high-risk populations (that is, patients receiving workers' compensation benefits, Veterans Affairs claimants, or patients with high rates of concurrent substance use disorder) was 27% (95% CI, 18% to 37%). The prevalence among low-risk populations was 6% (CI, 4% to 8%; P for interaction < 0.001). Moderate-certainty evidence showed increased odds of persistent opioid use with older age (absolute risk increase [ARI] for every 10-year increase, 1.1% [CI, 0.7% to 1.5%]) and physical comorbidity (ARI, 0.9% [CI, 0.1% to 1.7%]). Low-certainty evidence suggested increased risk for persistent opioid use with past or current substance use disorder (ARI, 10.5% [CI, 4.2% to 19.8%]),

prescriptions lasting more than 7 days (median ARI, 4.5%), and higher morphine milligram equivalents per day.

Limitation: Sparse, heterogeneous data with suboptimal adjustment for potential confounders.

Conclusion: Avoiding prescribing opioids for acute musculoskeletal injuries to patients with past or current substance use disorder and, when prescribed, restricting duration to ≤ 7 days and lower doses are potentially important targets to reduce rates of persistent opioid use.

Keywords: opioids, prolonged use, predictors, prognosis, GRADE, guidelines.

Background

From 1996 to 2011, approximately 9% of the U.S. population reported a musculoskeletal injury. This amounted to 23 to 25 million reports each year, with older Americans reporting higher rates of injury than younger ones.[1] Opioid analgesics are often prescribed to manage pain associated with acute musculoskeletal injuries and, from 2011 to 2015, 25% of U.S. patients presenting to an emergency department for a sprained ankle received a prescription for opioids.[2] Acute use of opioids may lead to persistent use, which can be associated with misuse, dependence, addiction, and overdose.

Purpose

We conducted a systematic review to explore factors associated with prolonged opioid use in patients with acute musculoskeletal injuries managed in an outpatient setting. This review serves as part of the evidence base for a joint clinical practice guideline from the American College of Physicians and American Academy of Family Physicians.

Methods

We reported our systematic review in accordance with the MOOSE (Meta-analysis of Observational Studies in Epidemiology) statement[3] and registered our protocol (PROSPERO: CRD42018104968) on 5 September 2018. We made the following changes to our protocol: We did not impute data for statically nonsignificant predictors for which no data were reported, we did not conduct meta-regression for the relationship between length of follow-up and prevalence of prolonged opioid use because of limited variability in length of follow-up across studies, and

we used the QUIPS (Quality In Prognosis Studies) checklist[4] to assess risk of bias for individual studies.

Data sources and searches

We searched MEDLINE, EMBASE, Web of Science, and Google Scholar from inception to 6 January 2020, without language restrictions, with terms related to prolonged opioid use, prognosis, and acute musculoskeletal injuries (see the summary of search strategy and results in the Supplement, available at Annals.org). We reviewed reference lists of eligible studies for additional articles.

Study selection

Six reviewers (J.J.R., S.T.N., V.A., F.F., B.S., R.C.) worked in pairs to screen, independently and in duplicate, the titles and abstracts of identified citations from an EndNote library (version 7.8 [Thomson Reuters]) and, subsequently, the full texts of potentially eligible studies. Reviewers resolved disagreements by discussion or with the help of an adjudicator when consensus could not be reached. We included observational studies that explored risk factors for prolonged opioid use—as defined by authors—after an initial prescription for an acute musculoskeletal injury (≤4 weeks) in an inception cohort of adults (prospective or retrospective) using an adjusted analysis. Studies were ineligible if they enrolled hospitalized patients, patients with injuries requiring surgery, or patients experiencing acute flare-ups of chronic conditions; their reference group was nonopioid users; more than 20% of enrolled patients had nonacute musculoskeletal injuries and the study reported only aggregate results; or all adjusted models contained statistically significant predictors collected more than 30 days after prescription. In such instances, the status of the predictor may be a result, rather than a cause, of prolonged opioid use. When we were able to

ascertain that a study included an opioid-naive, acute pain population but were not able to determine the proportion of patients presenting with musculoskeletal injuries, we included these studies and downgraded the certainty of evidence for indirectness.

Data extraction and risk of bias assessment

Using standardized, pilot-tested data extraction forms (MS-Excel 2011), pairs of reviewers (JJR, STN, JWB) extracted data and used the Quality In Prognosis Studies (QUIPS) tool (JJR, BS) to assess risk of bias from articles, independently and in duplicate.[4] Predictive models were considered at low risk of bias if adjusted for, at minimum, age, sex and injury severity. Disagreements were resolved by discussion to achieve consensus or, if consensus could not be reached, by an arbitrator. We extracted information for inception cohorts within study populations when required. For example, the cohort study by Berecki-Gisolf et al. reported on 54,931 injured workers; however, only 8,267 received an opioid prescription after their injury.[5]

Data synthesis and analysis

Among eligible studies, we pooled the prevalence of prolonged opioid use and used the Freeman–Tukey transformation to stabilize the variance.[6] Without this transformation, very high or very low prevalence estimates can produce confidence intervals (CIs) that contain values lower than 0% or greater than 100%. When studies reported prevalence of prolonged opioid use according to methods proposed by Deyo and colleagues[7] and Shah and colleagues,[8,9] we prioritized the latter approach (defined as discontinuation of opioid treatment with ≥180 continuous days without opioid use from the end date of the last prescription) on the basis of independent consultation with 2 clinical experts in addiction medicine. When possible, unless there was large heterogeneity present, we pooled all factors assessed for an association with

prolonged opioid use that were reported by more than 1 study using random-effects models and the DerSimonian–Laird method.[10] We presented pooled measures of association as odds ratios (ORs) and the absolute risk increase (ARI), both with associated 95% CIs, to facilitate interpretation. When the association for age was reported using categories, we assumed the association between age and the dependent variable (persistent opioid use) was linear in each age category and the associations across categories were independent of each other. We used Bucher's approach to calculate the OR and 95% CI for each age category[11] and pooled the ORs using the inverse variance method to produce a single OR for each study.[12] We used the pooled prevalence from studies that enrolled patients representative of the general population to derive a baseline risk of 6% for prolonged opioid use after prescription for an acute musculoskeletal injury.

We explored the consistency of association between our pooled results and studies reporting the same predictors that were not possible to pool. We used 3 criteria to identify predictors that were not amenable to pooling and that showed promise for future research: a sample size of more than 500 participants, a highly statistically significant association with prolonged opioid use ($P \leq 0.01$), and a large magnitude of association (OR of 2.0 or greater or 0.5 or less).

If more than 1 adjusted model exploring risk factors for prolonged use was reported in a single study, we used only the most adjusted model to avoid clustering. We evaluated heterogeneity for all pooled estimates through visual inspection of forest plots because statistical tests of heterogeneity can be misleading when sample sizes are large and CIs are therefore narrow.[13] We performed all statistical analyses using Stata, version 13.1 (StataCorp). All comparisons were 2-tailed, with a P value less than 0.05 considered statistically significant.

Subgroup analysis

We generated 4 a priori hypotheses to explain variability between studies, assuming larger associations with higher-risk populations, studies at greater risk of bias, longer duration of follow-up, and indirect populations. We defined high-risk populations as patients receiving wage replacement benefits[14] or defined as high risk by the study authors (that is, high prevalence of workers' compensation recipients, Veterans Affairs claimants, or patients with concurrent substance use disorder). We considered Veterans Affairs populations to be at higher risk for prolonged opioid use because of higher rates of substance use disorder and posttraumatic stress disorder among this population than the general public.[15] We did not conduct subgroup analyses if there was only 1 study in a given subgroup. We reported pooled associations with a combination of direct and potentially indirect study populations when there was no statistically significant subgroup effect between studies; otherwise, we reported only pooled estimates from direct populations.

Certainty of evidence

We used the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach to summarize the certainty of evidence for all meta-analyses.[16-18] Given a 6% baseline risk for prolonged opioid use after surgery, we estimated that a 2.5% increase in absolute risk would likely be sufficient to address modifiable risk factors in the context of a clinical encounter for the management of an acute musculoskeletal injury, and a 5% increase in risk for a nonmodifiable factor would be sufficient to identify high-risk candidates for intervention. We therefore downgraded for imprecision when the CI overlapped an absolute risk difference of

2.5% for modifiable factors or 5% for nonmodifiable factors. We assessed publication bias if there were at least 10 studies that contributed to a meta-analysis.[12]

Results

Of 11 747 unique records, we retrieved 134 full-text articles for review; 13 retrospective studies representing 14 cohorts (13 263 393 patients) proved eligible (Figure 1).[2,5,7-9,19-28] We successfully contacted 6 of 8 authors to confirm eligibility.[2,7,27,29-31] Eleven studies enrolled patients from the United States,[2,7-9,19-27] 1 enrolled patients from Australia,[5] and 1 enrolled patients from Malaysia,[28] and all included patients receiving workers' compensation benefits, Veterans Affairs claimants, and injuries in the general population. Ten studies considered multiple acute pain reports,[5,7-9,19,20,22,23,25-28] and 2 restricted their study population to low back pain[21,24] or ankle sprains.[2] The median length of follow-up was 12 months (range, 3 to 24 months). One study did not report a source of funding,[21] whereas the remaining 12 reported financial support from not-for-profit sources. The definition of prolonged opioid use varied across studies (Table 1).

Risk of bias

All studies were at risk of bias for at least 1 domain. Two studies could not confirm that patients were opioid naive at the time of enrollment[5,26] and, for all studies, despite matching the time of injury with an opioid prescription, there remained the possibility that opioids were prescribed for an indication aside from an acute musculoskeletal injury. Loss to follow-up was low among studies (range, <1% to 10%) in which this information was reported; 1 study[8,9] did not report the proportion of missing outcome data. Two studies reported that important confounders (for example, substance use disorder) may have been underestimated because of limitations of their

registry data,[26,27] and only 1 study[19,20] was able to confirm prolonged opioid use was related to initial musculoskeletal injury. Nine studies did not report adequately adjusted regression models,[5,7,19-22,24,26-28] and 3 studies used data-driven adjusted regression models in which not all selected factors were included in their final model (Supplement Table 1, available at Annals.org).[19-21,25]

Prevalence of prolonged opioid use

The overall pooled prevalence of prolonged opioid use across included studies was 10.6% (95% CI, 5.9% to 16.5%); however, substantial heterogeneity was associated with this estimate. Eight studies enrolled patients from the general public (low risk);[2,7-9,23-25,27,28] 3 studies (with 4 cohorts) enrolled Veterans Affairs claimants, patients receiving workers' compensation benefits, or high proportions of patients with substance use disorder (high risk);[5,21,22] and 2 enrolled a mixed population of patients, with at least some receiving wage replacement benefits (uncertain risk).[19,20,26] Subgroup analysis revealed no difference in rate of prolonged opioid use among studies enrolling low-risk and uncertain-risk patients (5.7% [CI, 3.6% to 8.3%] vs. 5.3% [CI, 5.1% to 5.5%]; P for interaction = 0.85), and we therefore included studies of uncertain risk in the low-risk category. Subgroup analysis found that high-risk patients were more likely to develop prolonged opioid use (26.9% [CI, 18.2% to 36.6%]) than low-risk patients (5.9% [CI, 4.0% to 8.2%]; P for interaction < 0.001) (Figure 2).

Predictors of prolonged opioid use

The 13 studies eligible for review reported the association of 47 independent variables with prolonged opioid use after prescription for an acute musculoskeletal injury, 3 of which were suitable for meta-analysis on the basis of our criteria.

Sociodemographic Factors

We found moderate-certainty evidence for small, but statistically significant, associations between prolonged opioid use and older age in adults (OR for every 10-year increase in age, 1.20 [CI, 1.12 to 1.27]; ARI, 1.1% [CI, 0.7% to 1.5%]) and greater physical comorbidity (OR, 1.16 [CI, 1.02 to 1.31]; ARI, 0.9% [CI, 0.1% to 1.7%]), as well as low-certainty evidence for a statistically significant association with past or present substance use disorder (OR, 3.14 [CI, 1.79 to 5.52]; ARI, 10.5% [CI, 4.2% to 19.8%]) (Table 2; Supplement Figure, available at Annals.org). Substance use disorder was typically defined using International Classification of Diseases, Ninth or 10th Revision codes, including general definitions, such as any drug abuse[2] or nonopioid use disorders[27] as well as more specific codes for alcohol, marijuana, methamphetamine, benzodiazepine, or cocaine use disorders.[21,25]

Among sociodemographic factors that were not amenable to pooling, 12 predictors were consistently associated with prolonged opioid use (Supplement Table 2, available at Annals.org), including sleep disorders, opioid use disorder, history of suicide attempt or self-injury, lower socioeconomic status, higher household income, rural residency, lower education level, early work disability lasting more than 2 weeks, permanent work disability, being injured in a motor vehicle accident, receipt of Medicaid, and incurring high hospital expenses. Medical claim–only costs were associated with a lower likelihood of prolonged opioid use (Supplement Table 2). Five of these factors (opioid use disorder, suicide attempt or self-injury history, early work disability lasting >2 weeks, receipt of Medicaid, and medical claim–only costs) met our criteria for promising predictors for future research. Six factors (sex, anxiety, depression, smoking status, occupation, and injury type) showed conflicting associations (Supplement Table 3, available at

Annals.org), and 6 factors (race, alcohol abuse, psychosis, episodic mood disorders, obesity, and non–full-time employment status) were consistently not associated with prolonged opioid use (Supplement Table 4, available at Annals.org).

Prescribing Factors

No prescribing factors were amenable to meta-analysis, but 4 factors were reported by several studies and showed a consistent association with increased risk for prolonged opioid use: prescribing opioids for more than 7 days (5 cohorts; 2 087 624 patients; median ARI, 4.5%; low-certainty evidence); higher morphine milligram equivalent dose (6 cohorts; 2 624 355 patients; ARI varied widely on the basis of dose and reference category; low-certainty evidence); long-acting versus short-acting opioids (3 cohorts; 1 924 421 patients; ARI range, 0.6% to 23.4%; very-low-certainty evidence); and more than 1 refill in the first month (3 cohorts; 1 230 243 patients; median ARI, 2.5%; very-low-certainty evidence) (Supplement Table 2).

Among factors reported by single studies, 5 were consistently associated with prolonged opioid use: primary care visit within 30 days of injury, non–emergency department prescriptions, hydrocodone versus oxycodone prescription, tramadol versus other opioids, and coprescription of benzodiazepine. Physical therapy within 30 days of injury was associated with a lower likelihood of prolonged opioid use (Supplement Table 2). Two of these factors (non–emergency department prescriptions and hydrocodone vs. oxycodone prescription) met our criteria as promising for future study. Eight predictors were not associated with prolonged opioid use: 1) year sampled from 2012 to 2015; 2) U.S. region; 3) early diagnostic imaging; 4) visiting an emergency department; 5) surgeon or other specialist consultation; or coprescription with 6) nonsteroidal anti-inflammatory drugs, 7) muscle relaxants, or 8) oral steroids (Supplement Table 4).

Discussion

The prevalence of prolonged opioid use after prescription for an acute musculoskeletal injury was 27% for high-risk populations (that is, workers' compensation patients receiving disability benefits, Veterans Affairs claimants, and patients with high prevalence of comorbid substance use disorder) and 6% among patients representative of the general population. We found moderate-certainty evidence that older age and greater physical comorbidity, and low-certainty evidence that past or present substance use disorder, are associated with prolonged opioid use after its prescription for acute musculoskeletal injury. The strongest of these associations was with past or present substance use disorder, with an absolute increase in prolonged opioid use of 11%. Among predictors that could not be pooled, prescribing opioids for more than 7 days, higher morphine milligram equivalent opioid doses, higher number of refills in the first month, non–emergency department prescriptions, hydrocodone versus oxycodone prescription, opioid use disorder, suicide attempt or self-injury history, early work disability lasting more than 2 weeks, receipt of Medicaid, and medical claim–only costs met our criteria for promising associations for future research.

Our finding of a relationship between older age in adults and prolonged use differs from a recent systematic review of children and adults on predictors of opioid misuse after prescription for acute or chronic pain; however, only 5 of the 64 included studies enrolled acute pain populations.[32] A clinical practice guideline by the Centers for Disease Control and Prevention, which informed a Health Quality Ontario standard, recommends avoiding prescribing more than 7 days of opioids at one time for acute pain because of increased risk for prolonged use.[33,34] Others have, as we found, reported that higher doses of opioids are associated with prolonged use,[2] and our findings are consistent with recent reviews of opioid-naive patients receiving

opioids for any pain condition that reported history of substance use disorder was significantly associated with the development of opioid use disorder[35] and opioid misuse.[32] Our finding from a single study[2] that alcohol abuse was not associated with prolonged opioid use is likely because of the small number of patients with alcohol use disorder (73 of 6463), leading to high imprecision in the estimate of association.

We found limited evidence that physical therapy early in care was associated with lower risk for prolonged opioid use.[21] A recent cross-sectional study of 88 985 opioid-naive patients with acute musculoskeletal pain found that early physical therapy was associated with lower risk for long-term opioid use and, among those prescribed opioids, a 10% reduction in the mean dose of opioids when compared with similar patients who did not receive early physical therapy.[36] The effect of health care provider attending to care is further supported by a study of 377 629 Medicare beneficiaries presenting to an emergency department in which those who saw high-intensity opioid prescribers were more likely to progress to long-term opioid use than those visiting low-intensity prescribers (OR, 1.3 [CI, 1.23 to 1.37]).[37]

Strengths and limitations

Strengths of our review include explicit eligibility criteria, a comprehensive search, and use of the GRADE approach to appraise the certainty of evidence. We have presented pooled measures of association as both relative and absolute risk increases, which we believe strengthens inferences about the importance of associations. Some authors have proposed inclusion of randomized controlled trials in prognostic reviews (38); however, we included only observational studies because of concerns that strict inclusion criteria used in many randomized trials would limit their generalizability.

Our study also has limitations, including imprecision for the prevalence of prolonged opioid use among high-risk populations and some risk of bias for most studies. One study that explored the same registry of Veterans Affairs claimants in 2011 and again in 2016 found that the prevalence of prolonged opioid use decreased from 29% to 17%,[22] suggesting that recent changes to Veterans Affairs policies to curb opioid use have been effective.[39-41] We were unable to pool predictors from 3 studies that reported only nonsignificant associations without accompanying data, and their inclusion would reduce the magnitude of associations for age, physical comorbidity, and past or present substance use disorder with prolonged opioid use (1 study for each predictor). However, the 1 study that was excluded from our pooled estimate for past or present substance use disorder acknowledged that its registry data underreported rates of comorbid mental illness[26] and identified only 60 of 9596 patients (0.6%) with substance use disorders. We found sparse information to inform the associations of some predictors, and data from our review came from patient or claim registries in which the reason for the original opioid prescription and subsequent prescriptions could typically not be definitively attributed to acute musculoskeletal injury.

Implications for future research

Future research would benefit from prospective studies in which both the initial prescription for, and continued use of, opioids was confirmed to be associated with an acute musculoskeletal injury. Regression models for prolonged opioid use should include, at a minimum, age, sex, injury severity, past and present substance use disorder, physical comorbidity, payer (for example, workers' compensation or Medicaid), and opioid prescribing factors (for example, duration, number of refills, dose, and type of opioid). Only 15% of the 13 studies eligible for our review included all of these factors in their adjusted regression model (Supplement Table 5,

available at Annals.org). Some regression models we reviewed included independent factors with few observations, resulting in highly imprecise measures of association. Future studies should set a threshold of a minimum number of observations per category for each independent factor (for example, ≥200) to provide some reassurance that each variable has sufficient discriminant power to detect an association with prolonged opioid use if an association exists. Studies should report multiple clinically meaningful categories for opioid duration and dose that reflect current legislative changes.[42] Further, prolonged opioid use is a surrogate for patient-important outcomes, such as addiction, overdose, and death, which should also be captured and reported.[43]

Conclusion

In conclusion, prolonged use is common among patients prescribed opioids for acute musculoskeletal injuries. Avoiding prescribing opioids for acute musculoskeletal injuries among patients with past or current substance use disorder, and restricting duration to 7 days or less and using lower doses when they are prescribed, are potentially important targets to reduce rates of persistent opioid use.

Acknowledgements

We thank James MacKillop, PhD, and Jennifer Brasch, MD, for guidance regarding selection of outcome measures for prolonged opioid use.

Authorship contributions

Conception and design: JJR, JWB

Data acquisition: JJR, STN, VA, FF, BS, RC

Data analysis: JJR, LW, JWB

Interpretation of results: JJR, LW, BS, JJR

Manuscript drafting: JJR, JWB

Critically revision of the manuscript and approval of the final version: JJR, STN, LW, VA, FF, BS, RC, JWB

Role of funding source

This study was funded by the National Safety Council. Dr. Riva is supported by a PhD training award from the National Chiropractic Mutual Insurance Company Foundation (www.ncmicfoundation.org).

References

1. Yelin E, Weinstein S, King T. The burden of musculoskeletal diseases in the United States [Editorial]. Semin Arthritis Rheum. 2016;46:259-260. [PMID: 27519477] doi:10.1016/j.semarthrit.2016.07.013

2. Delgado MK, Huang Y, Meisel Z, et al. National variation in opioid prescribing and risk of prolonged use for opioid-naive patients treated in the emergency department for ankle sprains. Ann Emerg Med. 2018;72:389-400.e1. [PMID: 30054152] doi:10.1016/j.annemergmed.2018.06.003

3. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA. 2000;283:2008-12. [PMID: 10789670]

4. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013;158:280-6. [PMID: 23420236] doi:10.7326/0003-4819-158-4-201302190-00009

5. Berecki-Gisolf J, Collie A, McClure RJ. Prescription opioids for occupational injury: results from workers' compensation claims records. Pain Med. 2014;15:1549-57. [PMID: 24641213] doi:10.1111/pme.12421

6. Freeman MF, Tukey JW. Transformations related to the angular and the square root. Ann Math Statist. 1950;21:607-11. doi:10.1214/aoms/1177729756

7. Deyo RA, Hallvik SE, Hildebran C, et al. Association between initial opioid prescribing patterns and subsequent long-term use among opioid-naïve patients: a statewide retrospective cohort study. J Gen Intern Med. 2017;32:21-27. [PMID: 27484682] doi:10.1007/s11606-016-3810-3.

8. Shah A, Hayes CJ, Martin BC. Factors influencing long-term opioid use among opioid naive patients: an examination of initial prescription characteristics and pain etiologies. J Pain. 2017;18:1374-1383. [PMID: 28711636] doi:10.1016/j.jpain.2017.06.010

9. Shah A, Hayes CJ, Martin BC. Characteristics of initial prescription episodes and likelihood of long-term opioid use—United States, 2006-2015. MMWR Morb Mortal Wkly Rep. 2017;66:265-269. [PMID: 28301454] doi:10.15585/mmwr.mm6610a1

10. Murad MH, Montori VM, Ioannidis JPA, et al. Fixed-effects and random-effects models. In: Guyatt G, Rennie D, Meade MO, et al, eds. Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice. 3rd ed. McGraw-Hill; 2015:507-14.

11. Bucher HC, Guyatt GH, Griffith LE, et al. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol. 1997;50:683-91. [PMID: 9250266]

12. Higgins JPT, Thomas J, Chandler J, et al, eds. Cochrane Handbook for Systematic Reviews of Interventions. 2nd ed. Wiley-Blackwell; 2019.

13. Rücker G, Schwarzer G, Carpenter JR, et al. Undue reliance on I(2) in assessing heterogeneity may mislead. BMC Med Res Methodol. 2008;8:79. [PMID: 19036172] doi:10.1186/1471-2288-8-79

14. Henschke N, Maher CG, Refshauge KM, et al. Prognosis in patients with recent onset low back pain in Australian primary care: inception cohort study. BMJ. 2008;337:a171. [PMID: 18614473] doi:10.1136/bmj.a171

15. Gallagher RM. Advancing the pain agenda in the veteran population. Anesthesiol Clin. 2016;34:357-78. [PMID: 27208715] doi:10.1016/j.anclin.2016.01.003

16. Iorio A, Spencer FA, Falavigna M, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. BMJ. 2015;350:h870. [PMID: 25775931] doi:10.1136/bmj.h870

17. Atkins D, Best D, Briss PA, et al; GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ. 2004;328:1490. [PMID: 15205295]

18. Foroutan F, Guyatt G, Zuk V, et al. GRADE guidelines 28: use of GRADE for the assessment of evidence about prognostic factors: rating certainty in identification of groups of patients with different absolute risks. J Clin Epidemiol. 2020;121:62-70. [PMID: 31982539] doi:10.1016/j.jclinepi.2019.12.023

19. Durand Z, Nechuta S, Krishnaswami S, et al. Prevalence and risk factors associated with long-term opioid use after injury among previously opioid-free workers. JAMA Netw Open. 2019;2:e197222. [PMID: 31314119] doi:10.1001/jamanetworkopen.2019.7222

20. Durand Z, Nechuta S, Krishnaswami S, et al. Prescription opioid use by injured workers in Tennessee: a descriptive study using linked statewide databases. Ann Epidemiol. 2019;32:7-13. [PMID: 30853149] doi:10.1016/j.annepidem.2019.02.001

21. Fritz JM, King JB, McAdams-Marx C. Associations between early care decisions and the risk for long-term opioid use for patients with low back pain with a new physician consultation and initiation of opioid therapy. Clin J Pain. 2018;34:552-558. [PMID: 29135698] doi:10.1097/AJP.0000000000000571

22. Hadlandsmyth K, Lund BC, Mosher HJ. Associations between initial opioid exposure and the likelihood for long-term use. J Am Pharm Assoc (2003). 2019;59:17-22. [PMID: 30409501] doi:10.1016/j.japh.2018.09.005

23. Halbert BT, Davis RB, Wee CC. Disproportionate longer-term opioid use among U.S. adults with mood disorders. Pain. 2016;157:2452-2457. [PMID: 27472400]

24. Harris RA, Kranzler HR, Chang KM, et al. Long-term use of hydrocodone vs. oxycodone in primary care. Drug Alcohol Depend. 2019;205:107524. [PMID: 31707268] doi:10.1016/j.drugalcdep.2019.06.026

25. Hooten WM, St Sauver JL, McGree ME, et al. Incidence and risk factors for progression from short-term to episodic or long-term opioid prescribing: a population-based study. Mayo Clin Proc. 2015;90:850-6. [PMID: 26141327] doi:10.1016/j.mayocp.2015.04.012

26. O'Hara NN, Pollak AN, Welsh CJ, et al. Factors associated with persistent opioid use among injured workers' compensation claimants. JAMA Netw Open. 2018;1:e184050. [PMID: 30646268] doi:10.1001/jamanetworkopen.2018.4050

27. Quinn PD, Hur K, Chang Z, et al. Incident and long-term opioid therapy among patients with psychiatric conditions and medications: a national study of commercial health care claims. Pain. 2017;158:140-148. [PMID: 27984526]

28. Zin CS, Nazar NI, Rahman NSA, et al. Patterns of initial opioid prescription and its association with short-term and long-term use among opioid-naïve patients in Malaysia: a retrospective cohort study. BMJ Open. 2019;9:e027203. [PMID: 31270113] doi:10.1136/bmjopen-2018-027203

29. Jeffery MM, Hooten WM, Hess EP, et al. Opioid prescribing for opioid-naive patients in emergency departments and other settings: characteristics of prescriptions and association with long-term use. Ann Emerg Med. 2018;71:326-336.e19. [PMID: 28967517] doi:10.1016/j.annemergmed.2017.08.042

30. Martin BC, Fan MY, Edlund MJ, et al. Long-term chronic opioid therapy discontinuation rates from the TROUP study. J Gen Intern Med. 2011;26:1450-7. [PMID: 21751058] doi:10.1007/s11606-011-1771-0

31. Quinn PD, Hur K, Chang Z, et al. Association of mental health conditions and treatments with long-term opioid analgesic receipt among adolescents. JAMA Pediatr. 2018;172:423-430. [PMID: 29532067] doi:10.1001/jamapediatrics.2017.5641

32. Cragg A, Hau JP, Woo SA, et al. Risk factors for misuse of prescribed opioids: a systematic review and meta-analysis. Ann Emerg Med. 2019;74:634-646. [PMID: 31229388] doi:10.1016/j.annemergmed.2019.04.019

33. Dowell D, Haegerich TM, Chou R. CDC guideline for prescribing opioids for chronic pain—United States, 2016. MMWR Recomm Rep. 2016;65:1-49. [PMID: 26987082] doi:10.15585/mmwr.rr6501e1

34. Health Quality Ontario. Quality standards: opioid prescribing for acute pain. 2018. Accessed at www.hqontario.ca/Evidence-to-Improve-Care/Quality-Standards/View-all-Quality-Standards/Opioid-Prescribing-for-Acute-Pain on 17 February 2020.

35. Klimas J, Gorfinkel L, Fairbairn N, et al. Strategies to identify patient risks of prescription opioid addiction when initiating opioids for pain: a systematic review. JAMA Netw Open. 2019;2:e193365. [PMID: 31050783] doi:10.1001/jamanetworkopen.2019.3365

36. Sun E, Moshfegh J, Rishel CA, et al. Association of early physical therapy with long-term opioid use among opioid-naive patients with musculoskeletal pain. JAMA Netw Open. 2018;1:e185909. [PMID: 30646297] doi:10.1001/jamanetworkopen.2018.5909

37. Barnett ML, Olenski AR, Jena AB. Opioid-prescribing patterns of emergency physicians and risk of long-term use. N Engl J Med. 2017;376:663-673. [PMID: 28199807] doi:10.1056/NEJMsa1610524

38. Riley RD, Moons KGM, Snell KIE, et al. A guide to systematic review and meta-analysis of prognostic factor studies. BMJ. 2019;364:k4597. [PMID: 30700442] doi:10.1136/bmj.k4597

39. Gellad WF, Good CB, Shulkin DJ. Addressing the opioid epidemic in the United States: lessons from the Department of Veterans Affairs. JAMA Intern Med. 2017;177:611-612. [PMID: 28288245] doi:10.1001/jamainternmed.2017.0147

40. Hadlandsmyth K, Mosher H, Vander Weg MW, et al. Decline in prescription opioids attributable to decreases in long-term use: a retrospective study in the Veterans Health Administration 2010-2016. J Gen Intern Med. 2018;33:818-824. [PMID: 29380212] doi:10.1007/s11606-017-4283-8

41. Lin LA, Bohnert ASB, Kerns RD, et al. Impact of the Opioid Safety Initiative on opioid-related prescribing in veterans. Pain. 2017;158:833-839. [PMID: 28240996] doi:10.1097/j.pain.0000000000000837

42. Davis CS, Lieberman AJ, Hernandez-Delgado H, et al. Laws limiting the prescribing or dispensing of opioids for acute pain in the United States: a national systematic legal review. Drug Alcohol Depend. 2019;194:166-172. [PMID: 30445274] doi:10.1016/j.drugalcdep.2018.09.022

43. Shaheed CA, McLachlan AJ, Maher CG. Rethinking “long term” opioid therapy. BMJ. 2019;367:l6691. [PMID: 31784472] doi:10.1136/bmj.l6691

Figures

Figure 1: Flow Diagram of Study Selection

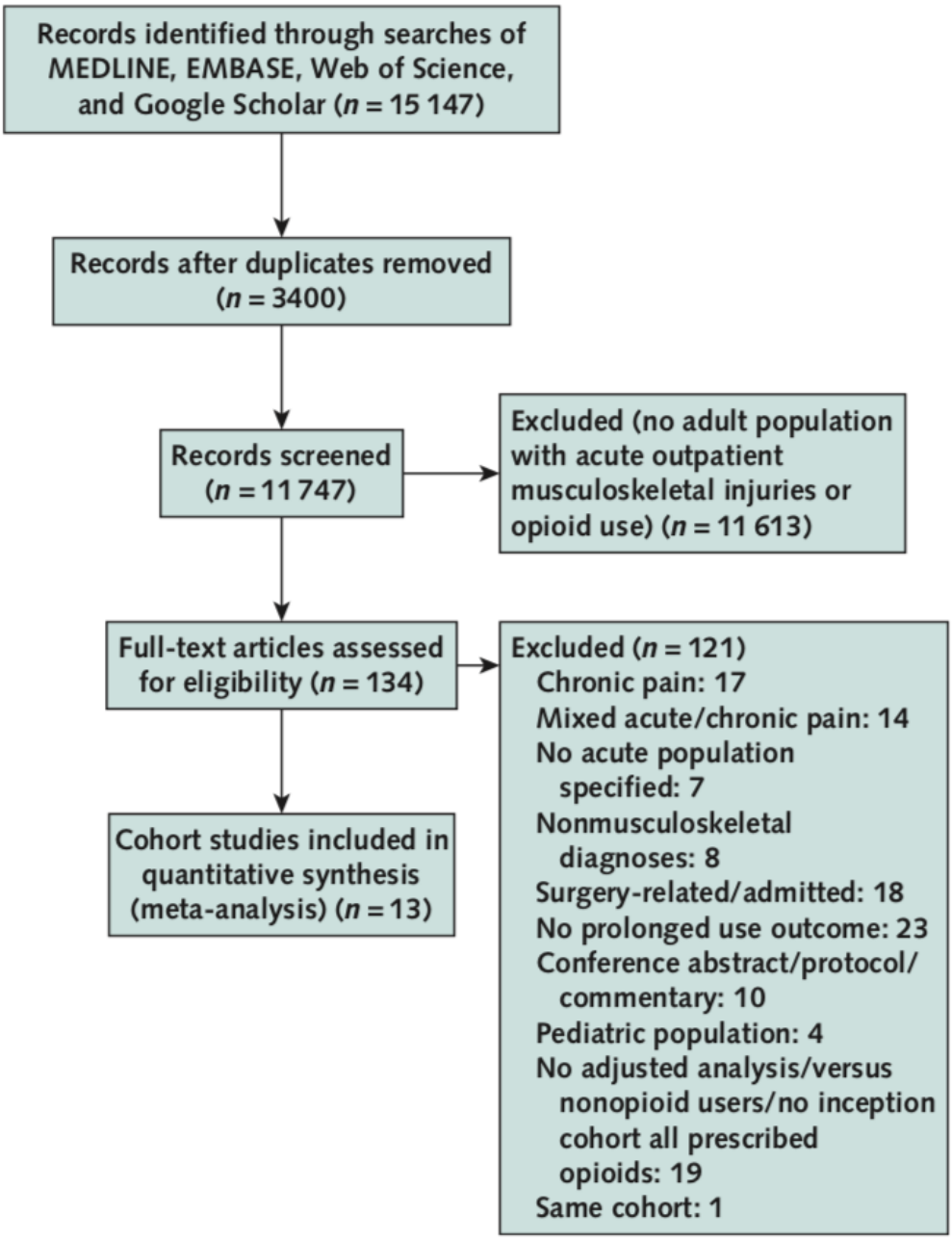
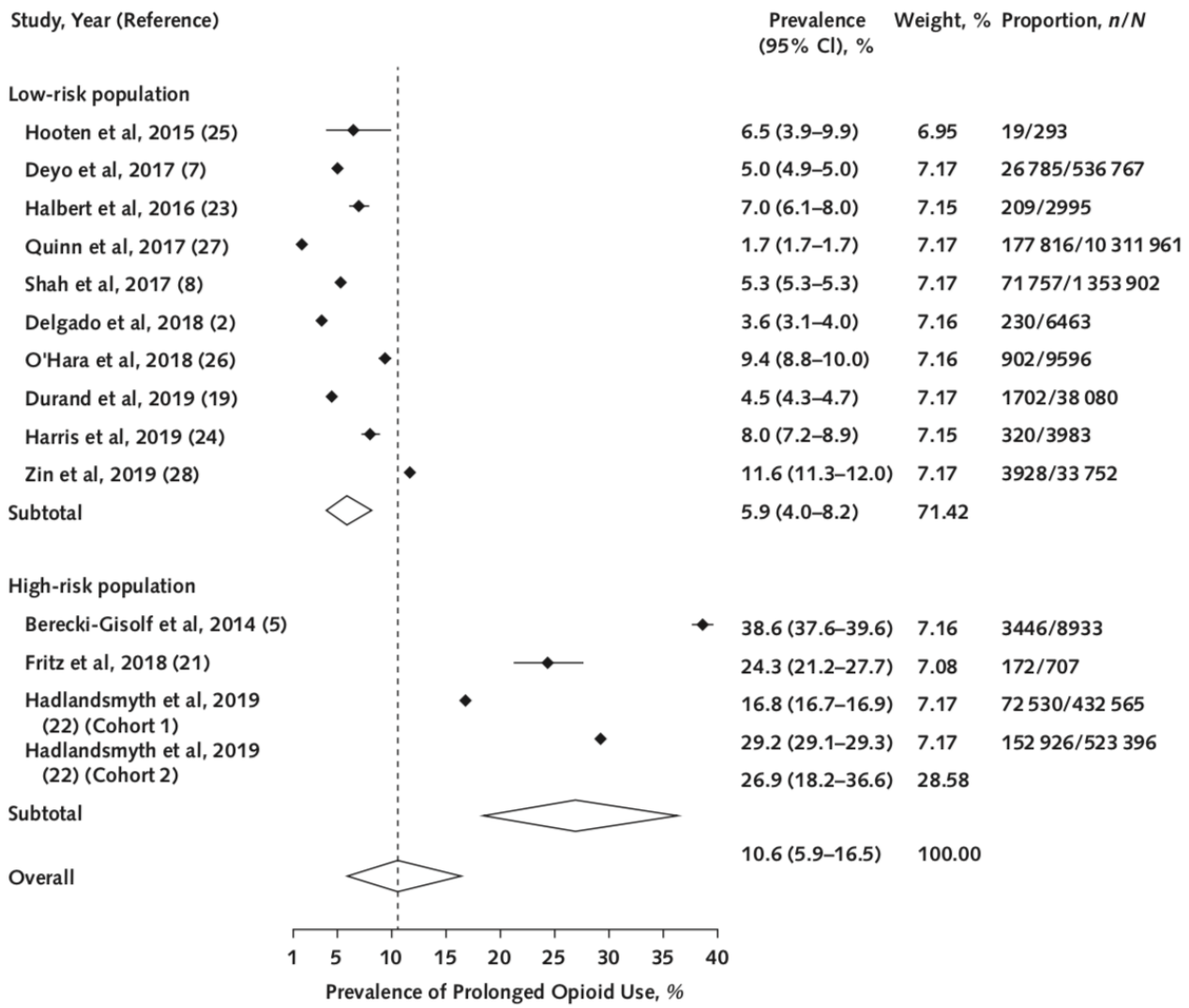


Figure 2: Pooled Prevalence of Prolonged Opioid Use



The figure shows a subgroup analysis of populations with high versus low risk for prolonged opioid use (P for interaction < 0.001). High-risk populations were injured workers receiving workers' compensation benefits,[5] Veterans Affairs claimants,[22] and patients with a high prevalence of substance use disorders.[21]

Tables

Table 1: Characteristics of Included Studies

Study, Year (Reference)	Population, Country	Definition of Prolonged Opioid Use	Follow-up	Risk of Bias	Funding Source and Competing Interests
Direct studies					
Berecki-Gisolf et al, 2014[5]	Musculoskeletal work injuries (*n* = 8933), Australia	≥1 opioid prescription in the second year after injury	24 mo	High	Nonprofit; Transport Accident Commission and Institute for Safety, Compensation and Recovery Research
Delgado et al, 2018[2]	Ankle sprains (*n* = 6463), United States	≥4 new opioid prescriptions 30–180 d after the initial prescription	6 mo	High	Nonprofit; National Institute on Drug Abuse, National Institutes of Health, and Leonard Davis Institute of Health Economics at the University of Pennsylvania PI reports honorarium for participating in an expert roundtable on innovative solutions for pain management convened by the UnitedHealth Group
Fritz et al, 2018[21]	Low back pain (*n* = 707), United States	≥120 d or >90 d with ≥10 fills during 1 year	12 mo	High	Not reported
O'Hara et al, 2018[26]	Musculoskeletal work injuries (*n* = 9596), United States	Filled an opioid prescription >90 d from the date of injury	12 mo	High	Nonprofit; Chesapeake Employers' Insurance Company
Durand et al, 2019[19]	Musculoskeletal work injuries (*n* = 38 080), United States	Receiving an opioid on most days for a 90-d period, measured as ≥45 prescription-days in 90 d after injury	3 mo	High	Nonprofit; Centers for Disease Control and Prevention's Prescription Drug Overdose: Prevention for States Program Coinvestigator served as a consultant and receiving personal fees from Western University of Health Sciences, Southern California University of Health Sciences, RAND Corporation, and EBSCO Information Services
Harris et al, 2019[24]	Low back pain (*n* = 3983), United States	Using the CONSORT (Consortium to Study Opioid Risks and Trends) criteria, prescription dates spanned ≥90 d from initial prescription to the	9 mo	High	Nonprofit; U.S. Department of Health and Human Services, Penn Center for AIDS Research, Penn Mental Health AIDS Research Center, and Veterans Integrated Service Network 4 Mental Illness Research, Education, and Clinical Center PI named inventor on PCT patent application: "Genotype-guided dosing of opioid agonists"

		run-out date of the last prescription, and included ≥120-d supply or ≥10 fills			
Indirect opioid-naive population studies					
Deyo et al, 2017[7]	Any acute pain condition (*n* = 536 767), United States	≥6 opioid fills in the 12 mo after the initiation month	12 mo	High	Nonprofit; National Institute on Drug Abuse and National Center for Advancing Translational Sciences PI reports receiving royalties from UpToDate for authoring topics on low back pain, and previous board membership at the nonprofit Informed Medical Decisions Foundation
Halbert et al, 2016[23]	Any acute noncancer pain condition (*n* = 2995), United States	≥3 opioid prescriptions during consecutive survey periods during 1-y follow-up	12 mo	High	Nonprofit; Institutional National Research Service Award, Ryoichi Sasakawa Fellowship Fund, Division of General Medicine and Primary Care at Beth Israel Deaconess Medical Center, Harvard Catalyst–The Harvard Clinical and Translational Science Center, National Center for Advancing Translational Sciences, and National Institutes of Health
Hooten et al, 2015[25]	Any acute pain condition (*n* = 293), United States	Episodes of prescribing lasting >90 d and ≥120 total days' supply	12 mo	High	Nonprofit; Rochester Epidemiology Project
Quinn et al, 2017[27]	Any new noncancer pain condition (*n* = 10 311 961), United States	Filled prescriptions for >90-d opioid supply during a 6-mo window and required 6-mo window had no gaps of >32 days' supply	12–18 mo	High	Nonprofit; National Institute on Drug Abuse and Indiana Clinical and Translational Sciences Institute
Shah et al, 2017[8]	Any acute noncancer pain condition (*n* = 1 353 902), United States	Opioid treatment discontinuation was defined as ≥180 continuous days without opioid use from the end date of the last opioid prescription	12 mo	High	Nonprofit; University of Arkansas for Medical Sciences Translational Research Institute, National Institutes of Health, and Translational Training in Addiction

Hadlandsmyth et al, 2019[22]	Any acute pain condition (cohort 1, *n* = 432 565; cohort 2, *n* = 523 396), United States	Deyo method: ≥6 opioid fills in the 12 mo after the initiation month Shah method: Opioid treatment discontinuation was defined as ≥180 continuous days without opioid use from the end date of the last opioid prescription	12 mo	High	Nonprofit; U.S. Department of Veterans Affairs, Veterans Health Administration, and the Health Services Research and Development Service
Zin et al, 2019[28]	Any acute noncancer pain condition (*n* = 33 752), Malaysia	Opioids were prescribed for ≥90 d per year after the index prescription over 12-mo follow-up	12 mo	High	Nonprofit; The Ministry of Education Malaysia (Fundamental Research Grant Scheme)

Abbreviations: (n) – study sample size; PI – principal investigator; PCT – Patient Cooperation Treaty.

Table 2: GRADE Evidence Profile of Pooled Predictors of Prolonged Opioid Use After Prescription for Acute Musculoskeletal Injuries

Sociodemographic Factor	Patients (Studies), Follow-up	Quality Assessment						Adjusted Relative Effect: OR (95% CI)	Anticipated Absolute Effect: Risk Difference† (95% CI)
		Risk of Bias	Inconsistency	Indirectness	Imprecision	Publication Bias	Overall Certainty of Evidence		
Age (every 10-y increase in adults)‡	29 016 patients (5 studies), 6–24 mo follow-up	Serious§	Not serious	Not serious	Not serious	Uncertain; only 6 studies‖	Moderate	1.20 (1.12–1.27)	1.1% more (0.7%–1.5%) patients per 10-y increase develop prolonged use
Past or present substance use disorder¶	10 319 424 patients (4 studies), 6–18 mo follow-up	Serious§	Not serious	Not serious	Serious**	Uncertain; only 5 studies††	Low	3.14 (1.79–5.52)	10.5% more (4.2%–19.8%) patients with substance use disorder‡‡ develop prolonged use
Comorbidity index§§	7170 patients (2 studies), 6–12 mo follow-up	Serious§	Not serious	Not serious	Not serious	Uncertain; only 3 studies‖	Moderate	1.16 (1.02–1.31)	0.9% more (0.1%–1.7%) patients with higher numbers of comorbidities‖‖ develop prolonged use

GRADE = Grading of Recommendations Assessment, Development and Evaluation; OR = odds ratio.

† Baseline risk of 6% based on pooled prevalence of 10 low-risk population studies (Figure 2).
‡ From references 2, 5, 21, 24, and 26.
§ Downgraded because of limitations reported in risk of bias summary (Supplement Table 1).
‖ The study by Hooten and colleagues[25] (n = 293) was not included in the pooled estimate because there were no data reported for age or physical comorbidity; both reported no statistically significant association with prolonged opioid use in an adjusted model.
¶ From references 2, 21, 25, and 27.
** Downgraded because 95% CI crossed 5% threshold for a nonmodifiable factor in the context of a clinical encounter for the management of an acute musculoskeletal injury.
†† The study by O'Hara and colleagues[26] (n = 9596) was not included in the pooled estimate because there were no data reported for substance use disorder, which was noted to show no statistically significant association with prolonged opioid use in an adjusted model.
‡‡ Substance use disorder was defined using International Classification of Diseases, Ninth or 10th Revision codes, including general codes, such as any drug abuse[2] or nonopioid use disorders,[27] as well as more specific codes, including alcohol, marijuana, methamphetamine, benzodiazepine, or cocaine use disorders.[21,25]
§§ From references 2 and 21.

|||| The study by Delgado and colleagues[2] measured comorbidity using the total number of Elixhauser comorbidities. The most common comorbidities represented in the sample were hypertension (23%), uncomplicated diabetes (9%), chronic pulmonary disease (9%), hypothyroidism (7%), and obesity (6%). The study by Fritz and colleagues[21] measured comorbidity using the Charlson Comorbidity Index and classified patients as having multiple comorbid conditions if the index score was ≥2. Obesity was present in 26% of patients with long-term opioid use; other comorbidities were not reported.

Appendices

Supplement Table 1. Risk of Bias Summary Using QUIPS Tool

Author, Year	1-Study Population	2-Study Attrition	3-Prognostic Factors	4-Outcome Measurement	5-Study Confounding*	6-Analysis & Reporting	Comments and Overall Determination
Direct Studies							
Berecki-Gisolf 2014	Moderate	Low	Low	Moderate	Moderate	Low	(1) High-risk population for prolonged opioid use by nature of workers' compensation claims; not restricted to opioid naïve. (4) Authors report that opioid use was limited to opioid reimbursement recorded in WorkSafe payments for medical and like expenses; an employer-paid excess of around $AU630 must be reached first. (5) Prior recent use not considered as confounder. Overall: High Risk of Bias
Delgado 2018	Moderate	Low	Low	Moderate	Low	Low	(1) Authors assumed that prescriptions were for ankle sprains and not for other indications that were not coded. (4) They also reported the opioid prescription rate was likely underreported as some patients may fill prescriptions by paying out of pocket instead of using health insurance. Overall: High Risk of Bias
Fritz 2018	Moderate	Low	Moderate	Moderate	Low	Moderate	(1) Authors reported that their study population was at higher risk for prolonged opioid use due to high prevalence of Medicaid coverage, mental health conditions, and substance use disorders. (3) Authors reported that individuals may have had more comorbidity than recorded as a result of seeking treatment for other conditions aside from low back pain.

							(4) To capture prolonged use authors only included those with a second low back pain claim within a year of analysis, to examine persistent symptoms. (6) Not all covariates selected were included in the final adjusted model. Overall: High Risk of Bias
O'Hara 2018	Moderate	Low	Low	Moderate	Moderate	Low	(1) Workers' compensations injuries with some including open wounds and poisoning; not restricted to opioid naïve and could not definitely confirm a new acute diagnosis. Authors also noted underreporting of mental illness. (4) Authors reported the potential for gaps in how prolonged use was measured between 90, 180 and 365 time windows, as dose and duration for the prescribed opioid were infrequently reported in the dataset. (5) Prior recent use not considered as confounder. Age and annual income data were reported as missing in < 10% of cases and multiple imputation was used to impute them. The final model was not adjusted for sex. Overall: High Risk of Bias
Durand 2019	Low	Low	Low	Low	High	Moderate	(4) No follow up on prolonged use beyond 90 days. (5) Authors reported that clinical data was not available for mental health status. Model not adjusted for age or sex. (6) Variables selected for adjusted model were data driven. Some statistically significant predictors in model were captured at different time points, past 30 days from injury. Overall: High Risk of Bias
Harris 2019	Moderate	Low	Low	Moderate	Moderate	Low	

							(1) Excluded non-monotherapy with specific oral opioids and patient who switched from one opioid to another. (4) Used prescription claims based on commercial insurance plans and Medicare advantage registry data and unable to confirm use of opioids. (5) Model not adjusted for injury severity. Overall: High Risk of Bias
Indirect Studies							
Deyo 2016	High	Low	Low	Moderate	Moderate	Low	(1) Unable to determine proportion of acute musculoskeletal pain patients. Authors were not able to confirm opioid naïve status in those with multiple clinicians or alternative sources of opioid medications. Intent to not use opioids long-term was not confirmed. Their population in their analysis excluded patients > 45 years old. (4) The authors noted that they were not able confirm that patients were not initiating opioids long-term for a chronic pain at the outset. (5) Model not adjusted for sex or injury severity. Overall: High Risk of Bias
Halbert 2016	High	Low	Low	Moderate	Moderate	Low	(1) Unable to determine proportion of acute musculoskeletal pain patients. Authors noted the potential for misclassification of acute and chronic pain cases. (4) The authors noted a limitation that they could not calculate morphine-milligram equivalents, total number days supplied per episode, or number of days for each distinct clinical episode. (5) Authors reported that residual confounding was possible due to broad definitions, when comparing those with or without mood disorders.

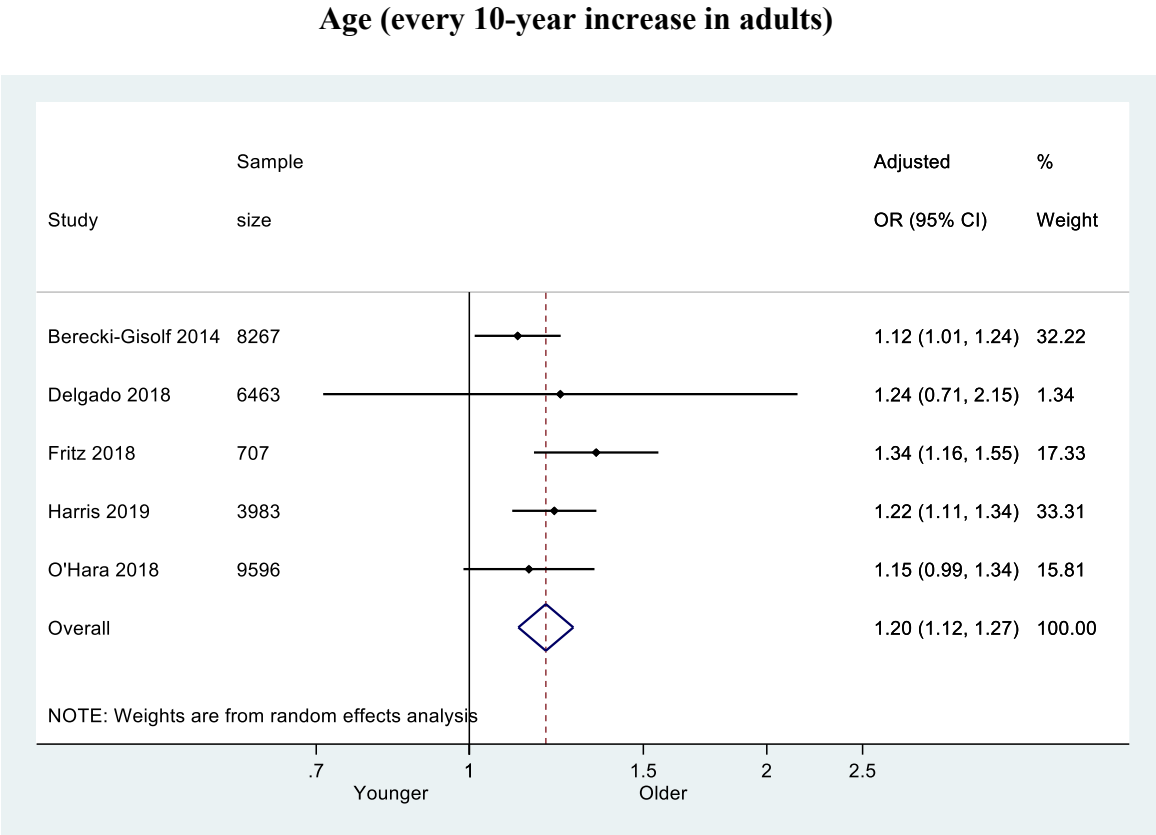
							Overall: High Risk of Bias
Hooten 2015	High	Low	Low	Moderate	Low	Moderate	(1) Unable to determine proportion of acute musculoskeletal pain patients. Authors reported that an unspecified group of patients from the sample was excluded because they used paper versus electronic charts. (4) Authors reported that prescriptions from a portion of the sample were not captured, as they did not utilize an electronic prescription system. (6) Authors reported their modest sample size (293 patients) limited the statistical power of their models. Also, values for statistically non-significant covariates where not reported. Overall: High Risk of Bias
Quinn 2017	High	Low	Moderate	Moderate	Moderate	Low	(1) Unable to determine proportion of acute musculoskeletal pain patients. The authors' patient sample excluded those > 65 years old, on Medicaid, or not using commercial insurance. (3) Authors reported that substance use disorder may have been underestimated, as only diagnoses linked to claims were captured. (4) Authors noted that they were not able to confirm if patients took prescriptions or if any other outside opioids were received. (5) Model not adjusted for injury severity. Overall: High Risk of Bias
Shah 2017	High	Moderate	Low	Moderate	Low	Low	(1) Unable to determine proportion of acute musculoskeletal pain patients. Authors were not able to confirm that patients were prescribed opioids for only acute conditions.

							(2) Authors noted loss to follow up but did not report the magnitude. (4) Authors noted that it was not possible to tell which long-term use cases stemmed directly from an intent for only acute use. Overall: High Risk of Bias
Hadlandsmyth 2019	High	Low	Low	Moderate	High	Low	(1) Unable to determine proportion of acute musculoskeletal pain patients. Authors were not able to confirm opioid naïve status in those with multiple clinicians or alternative sources of opioid medications. Also, intent to not use opioids for chronic long-term purposes at outset was not confirmed. (4) The authors noted that they were not able confirm that patients were not using opioids from other sources or for other purposes at the outset. (5) Model not adjusted for age, sex or injury severity. Overall: High Risk of Bias
Zin 2019	High	Low	Low	Moderate	Moderate	Low	(1) Unable to determine proportion of acute musculoskeletal pain patients. Authors noted they were unable to ascertain if patients were opioid-naïve, noting it was possible they had received opioid prescription elsewhere. (4) Authors noted that they were unable to confirm if opioids were dispensed or consumed. (5) Authors noted they were not able to adjust for diagnoses and felt this may confound their associations. Model not adjusted for injury severity. Overall: High Risk of Bias

$AU = Australian dollars

***** Predictive models were optimally adjusted: low risk of bias if adjusted for, at minimum, age, sex, and injury severity.

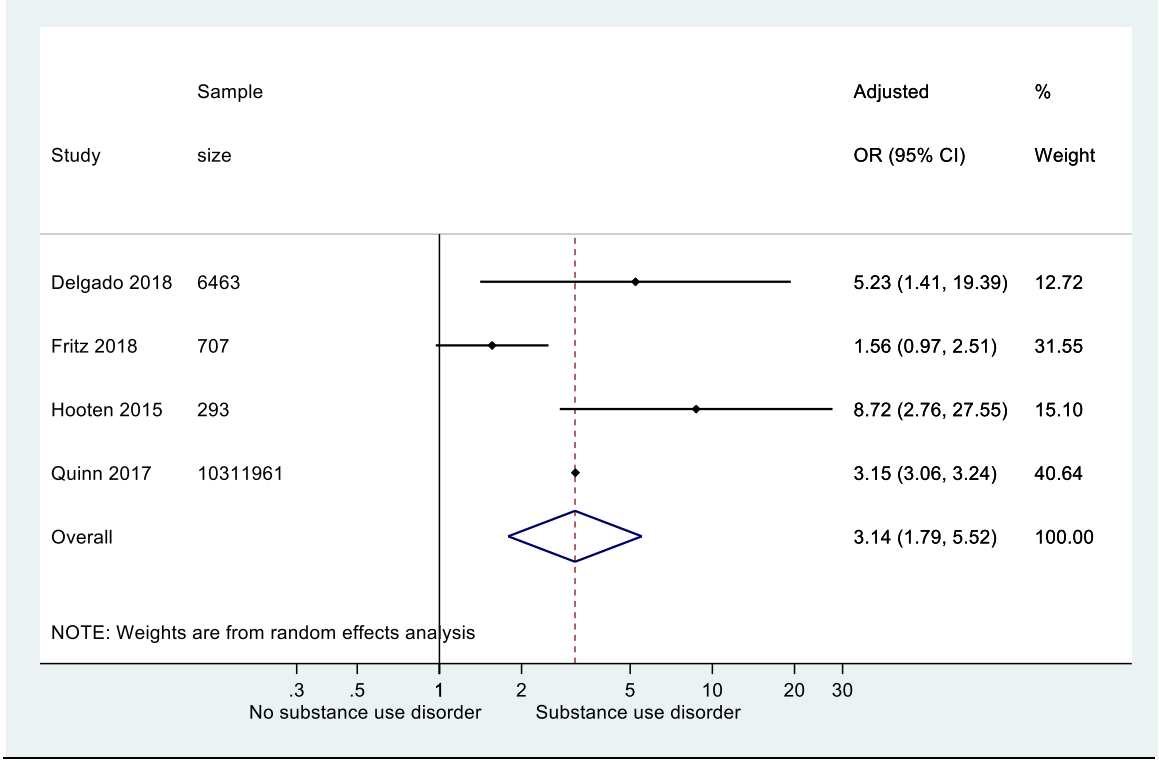
Supplement Figure. GRADE Evidence Profile Comparing Direct and Indirect Studies for Pooled Predictors of Prolonged Opioid Use After Prescription for Acute Musculoskeletal Injuries



Abbreviations: OR – odds ratio; 95% CI – 95% confidence interval

All studies were direct evidence. Delgado 2018, O'Hara 2018 and Harris 2019 were low risk populations for prolonged opioid use; Fritz 2018 and Berecki-Gisolf 2014 were high risk populations for prolonged use.
Subgroup analysis for age: high vs low-risk population, test of interaction p= 0.90

Past or Present Substance Use Disorder (SUD)

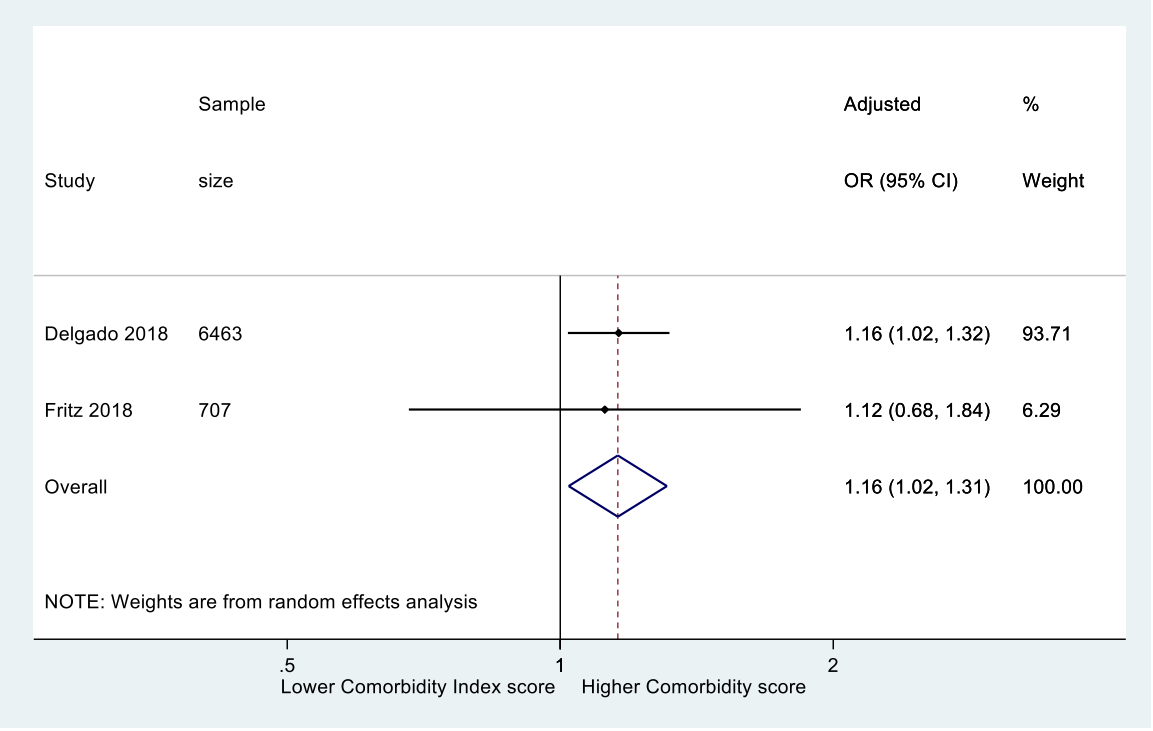


Abbreviations: OR – odds ratio; 95% CI – 95% confidence interval

Delgado 2018 and Fritz 2018 were direct evidence and Hooten 2015 and Quinn 2017 were indirect evidence. Delgado 2018, Hooten 2015 and Quinn 2017 were low risk populations and Fritz 2018 were high risk populations for prolonged opioid use. Subgroup analysis for substance use disorder: direct vs indirect evidence, interaction p=0.49

Substance use disorder was typically classified using ICD-9 or ICD-10 codes and included use disorders such as, alcohol, marijuana, or drugs (e.g. methamphetamine, benzodiazepine, cocaine) - one study (Quinn 2017) classified substance use as any non-opioid use disorder and one study (Delgado 2018) classified drug abuse and alcohol abuse as a separate predictors.

Comorbidity index



Abbreviatons: OR – odds ratio; 95% CI – 95% confidence interval

All studies were direct evidence. Delgado 2018 was a low risk population and Fritz 2018 was a high risk population for prolonged opioid use.

Delgado 2018 measured co-morbidity using the total number of Elixhauser co-morbidities. The most common co-morbidities represented in the sample were hypertension (23%), uncomplicated diabetes (9%), chronic pulmonary disease (9%), hypothyroidism (7%) and obesity (6%). Fritz 2018 measured co-morbidity using the Charlson co-morbidity index and classified patients as having multiple comorbid conditions if index score was ≥ 2. Obesity diagnosis was present in 26% of patients with long-term opioid use; other co-morbidities were not reported.

Supplement Table 2. GRADE Evidence Profile of Statistically Significant Unpooled Predictors of Prolonged Opioid Use After Prescription for Acute Musculoskeletal Injuries

Quality Assessment *					Adjusted Relative Effect (95%CI)	Anticipated Absolute Effect
Author, Year, Study Sample Size (population risk for prolonged use – low/high)		Indirectness	Imprecision	Overall Certainty of Evidence		Risk Difference (95%CI)[1]
Sociodemographic Factors – Predictors from Single Studies						
Quinn 2017 n=10,311,961 (low risk)	Sleep Disorders vs not	Serious[2]	Not Serious	Low	HR 1.78[3] (1.75-1.80)	4.1% more (4.0-4.2% more) patients with sleep disorders develop prolonged use[4]
Quinn 2017 n=10,311,961 (low risk)	Opioid Use Disorder vs not	Serious[2]	Not Serious	Low	HR 8.70[3] (8.20-9.24)	29.4% more (28.1-30.8% more) patients with opioid use disorder develop prolonged use[4]
Quinn 2017 n=10,311,961 (low risk)	Suicide attempt / self-injury vs not	Serious[2]	Not Serious	Low	HR 2.55[3] (2.21-2.94)	7.9% more (6.3-9.7% more) patients with history of suicide attempt or self-injury develop prolonged use[4]
Berecki-Gisolf 2014 n=8,267 (high risk)	1st-2nd SEIFA (socioeconomic status) deciles vs highest 9th-10th deciles	Not Serious	Not Serious	Moderate	OR 1.78 (1.51-2.10)	4.1% more (2.7-5.7% more) patients with lowest (more disadvantaged) SEIFA deciles of socioeconomic status develop prolonged use
	3rd-4th SEIFA deciles vs highest 9th-10th deciles	Not Serious	Not Serious		OR 1.44 (1.20-1.72)	2.4% more (1.1-3.8% more) patients with lower SEIFA deciles of socioeconomic status develop prolonged use
	5th-6th SEIFA deciles vs highest 9th-10th deciles	Not Serious	Not Serious		OR 1.48 (1.26-1.73)	2.6% more (1.4-3.9% more) patients with low SEIFA deciles of socioeconomic status develop prolonged use
	7th-8th SEIFA deciles vs highest 9th-10th deciles	Not Serious	Not Serious		OR 1.27 (1.10-1.47)	1.5% more (0.6-2.5% more) patients with low SEIFA deciles of socioeconomic status develop prolonged use
O'Hara 2018 n=9,596 (low risk)	Household Income 20,000-	Not Serious	Serious[5]	Low	OR 1.10 (0.85-1.41)	0.6% more (0.8% less to 2.2% more) patients with slightly higher

	39,999 USD vs < 20,000 USD					household incomes develop prolonged use
	Household Income 40,000-59,999 USD vs < 20,000 USD	Not Serious	Not Serious	Moderate	OR 1.40 (1.08-1.82)	2.2% more (0.4-4.3% more) patients with higher household incomes develop prolonged use
	Household Income ≥ 60,000 USD vs < 20,000 USD	Not Serious	Not Serious	Moderate	OR 1.48 (1.11-1.99)	2.6% more (0.6-5.2% more) patients with highest household incomes develop prolonged use
Deyo 2016 n=536,767 (low risk)	Rural residency vs Urban	Serious[2]	Not Serious	Low	OR 1.37 (1.34-1.41)	2.0% more (1.9-2.2% more) patients with rural residency develop prolonged use
Delgado 2018 n=6,463 (low risk)	High school diploma education vs Bachelor degree or higher	Not Serious	Serious[6]	Low	OR 2.30 (1.01-5.24)	6.7% more (0.1-18.8% more) patients with lower education levels, particularly high school, develop prolonged use
	Less than 12th grade education vs Bachelor degree or higher	Not Serious	Very Serious[5,7]	Very Low	OR <0.001 (<0.001->999.99) [5]	Not calculated; no statistically significant association.
	Less than Bachelor degree education vs Bachelor degree or higher	Not Serious	Serious[5]	Low	OR 2.02 (0.93-4.35)	5.3% more (0.4% less to 15.5% more) patients with less than a bachelor's degree education develop prolonged use
	Unknown education level vs Bachelor	Not Serious	Very Serious[5]	Very Low	OR 2.56 (0.37-17.8)	7.9% more (3.6% less to 46.8% more) patients with unknown education levels develop prolonged use

	degree or higher					
Berecki-Gisolf 2014 n=8,267 (high risk)	1-14 days of early work disability vs no early work disability	Not Serious	Serious[5]	Low	OR 0.63 (0.37-1.06)	2.1% less (3.6% less to 0.3% more) patients with 1-14 days of early work disability develop prolonged use
	> 14 days of early work disability vs no early work disability	Not Serious	Not Serious	Moderate	OR 2.17 (1.52-3.10)	6.1% more (2.8-10.4% more) patients with greater than 14 days of early work disability develop prolonged use
O'Hara 2018 n=9,596 (low risk)	Permanent total disability vs permanent partial disability	Not Serious	Not Serious	Moderate	OR 6.29 (1.68-23.6)	22.4% more (3.6-53.8% more) patients with permanent total disability develop prolonged use
	Temporary total disability vs permanent partial disability	Not Serious	Not Serious	Moderate	OR 0.66 (0.55-0.78)	1.9% less (1.2-2.6% less) patients with temporary total disability develop prolonged use
	Temporary partial disability vs permanent partial disability	Not Serious	Serious[5]	Low	OR 0.98 (0.58–1.64)	0.1% less (2.4% less to 3.4% more) patients with temporary partial disability develop prolonged use
	Medical only claims vs permanent partial disability	Not Serious	Not Serious	Moderate	OR 0.20 (0.16-0.26)	4.7% less (4.3-4.9% less) patients with medical only claims develop prolonged use
Quinn 2017 n=10,311,961 (low risk)	Motor Vehicle Crashes vs not	Serious[2]	Not Serious	Low	HR 1.99 [3] (1.86-2.14)	5.2% more (4.5-5.9% more) patients with motor vehicle crashes develop prolonged use

Fritz 2018 n=707 (high risk)	Medicaid Payer vs not	Not Serious	Not Serious	Moderate	OR 2.84 (1.62-5.00)	9.2% more (3.3-18.0% more) patients with Medicaid payers develop prolonged use
Berecki-Gisolf 2014 n=8,267 (high risk)	$AU 1-1,800 vs no hospital expenses	Not Serious	Not Serious	Moderate	OR 0.83 (0.72-0.96)	1.0% <u>less</u> (0.2-1.6% less) patients with lower expenses at hospital develop prolonged use
	$AU > 1,800 vs no hospital expenses	Not Serious	Not Serious	Moderate	OR 1.88 (1.66-2.14)	4.6% more (3.5-5.9% more) patients with higher expenses at hospital develop prolonged use
Prescribing Factors – Predictors from Multiple Studies						
Opioid Days Supplied						
Higher days supplied versus 2 days or less						
Shah 2017 n=1,353,902 (low risk)	3-4 days	Serious[2]	Not Serious	Low	HR 1.47 ** (1.47-1.49)	2.5% more (2.5-2.6% more) patients with 3-4 days supply develop prolonged use
	5-7 days		Not Serious	Low	HR 2.22 ** (2.17-2.22)	6.3% more (6.1-6.3% more) patients with 5-7 days supply develop prolonged use
	8-10 days		Not Serious	Low	HR 2.86 ** (2.78-2.86)	9.3% more (8.9-9.3% more) patients with 8-10 days supply develop prolonged use
	11-14 days		Not Serious	Low	HR 3.33 ** (3.33-3.45)	11.4% more (11.4-11.9% more) patients with 11-14 days supply develop prolonged use
	15-21 days		Not Serious	Low	HR 3.70 ** (3.70-3.85)	12.9% more (12.9-13.5% more) patients with 15-21 days supply develop prolonged use
	22 or more days		Not Serious	Low	HR 5.88 ** (5.56-5.88)	2.5% more (2.5%-2.6% more) patients with 22 or more days supply develop prolonged use
Higher days supplied versus 3 days or less						
Delgado 2018 n=6463 (low risk)	4-5 days	Not Serious	Serious[5]	Low	OR 0.59 (0.33-1.06)	2.3% <u>less</u> (3.9% less to 0.3% more) patients with 4-5 days supply develop prolonged use
	6-7 days		Serious[5]	Low	OR 1.92 (0.92-4.0)	4.8% more (0.4% less to 13.2% more) patients with 6-7 days supply develop prolonged use

	8 or more days		Serious[5]	Low	OR 1.86 (0.81-4.3)	4.5% more (1.1% less to 15.3% more) patients with 8 or more days supply develop prolonged use
Zin 2019 n=33,752 (low risk)	3-7 days	Serious[2]	Serious[5]	Very low	OR 1.06 (0.67-1.67)	0.3% more (1.9% less to 3.6% more) patients with 3-7 days supply develop prolonged use
	More than 7 days		Not Serious	Low	OR 16.47 (10.56-25.7)	44.9% more (33.9-55.8% more) patients with more than 7 days supply develop prolonged use
Higher days supplied versus 7 days or less						
Hadlandsmyth 2019 (model 1) n=317,367 (high risk)	8-14 days	Serious[2]	Not Serious	Low	OR 1.53 (1.42-1.64)	2.9% more (2.3-3.4% more) patients with 8-14 days supply develop prolonged use
	15-21 days		Not Serious	Low	OR 2.96 (2.73-3.21)	9.8% more (8.7-10.9% more) patients with 15-21 days supply develop prolonged use
	22-30 days		Not Serious	Low	OR 9.38 (8.90-9.88)	31.1% more (29.9-32.4% more) patients with 22-30 days supply develop prolonged use
	30 or more days		Not Serious	Low	OR 21.7 (20.1-23.5)	51.7% more (51.1-53.7% more) patients with 30 or more days supply develop prolonged use
Hadlandsmyth 2019 (model 2) n=376,140 (high risk)	8-14 days	Serious[2]	Serious[6]	Very Low	OR 1.44 (1.38-1.51)	2.4% more (2.1-2.7% more) patients with 8-14 days supply develop prolonged use
	15-21 days		Not Serious	Low	OR 2.43 (2.30-2.56)	7.3% more (6.7-7.9% more) patients with 15-21 days supply develop prolonged use
	22-30 days		Not Serious	Low	OR 7.35 (7.09-7.62)	25.6% more (24.9-26.4% more) patients with 22-30 days supply develop prolonged use
	30 or more days		Not Serious	Low	OR 15.5 (14.7-16.4)	43.4% more (42.1-44.8% more) patients with 30 or more days supply develop prolonged use
Opioid Dose - Morphine Milligram Equivalents (MME)						
Higher dose versus 1-75 MME						
Delgado 2018 n=6,463 (low risk)	76 -150 MME	Not Serious	Serious[5]	Low	OR 1.33 (0.82–2.15)	1.8% more (1.0% less to 6.0% more) patients with 76-150 MME develop prolonged use

	151-225 MME		Serious[5]	Low	OR 1.55 (0.65–3.73)	3.0% more (2.0% less to 13.1% more) patients with 151-225 MME develop prolonged use
	≥ 226 MME		Not Serious	Moderate	OR 4.15 (1.85–9.30)	14.7% more (4.5-30.9% more) patients with ≥ 226 MME develop prolonged use
Higher dose versus 1-119 MME[8]						
Deyo 2016 n=536,736 (low risk)	120 – 279 MME	Serious[2]	Serious[6]	Very Low	OR 1.42 (1.37-1.49)	2.3% more (2.0-2.6% more) patients with 120-279 MME develop prolonged use
	280 –399 MME		Not Serious	Low	OR 2.22 (2.10-2.34)	6.3% more (5.7-6.9% more) patients with 280–399 MME develop prolonged use
	400 –799 MME		Not Serious	Low	OR 2.96 (2.81-3.11	9.8% more (9.1-10.4% more) patients with 400-799 MME develop prolonged use
	800 –1599 MME		Not Serious	Low	OR 4.63 (4.37-4.92)	16.6% more (15.6-17.7% more) patients with 800-1599 MME develop prolonged use
	1600 –2399 MME		Not Serious	Low	OR 6.78 (6.21-7.40)	23.9% more (22.1-25.8% more) patients with 1600-2399 MME develop prolonged use
	2400 –3199 MME		Not Serious	Low	OR 11.27 (10.04-12.65)	35.5% more (32.7-38.3% more) patients with 2400-3199 MME develop prolonged use
	3200 –3999 MME		Not Serious	Low	OR 16.30 (13.71-19.37)	44.6% more (40.3-48.9% more) patients with 3200-3999 MME develop prolonged use
Higher dose versus 24 MME or less						
Shah 2017 n=1,353,902 (low risk)	25 -49 MME	Serious[2]	Not Serious	Low	HR 1.02 (1.02-1.03) **	0.1% more (0.1-0.2% more) patients with 25-49 MME develop prolonged use
	50 -89 MME		Not Serious	Low	HR 1.02 (1.01-1.03) **	0.1% more (0.1-0.2% more) patients with 50-89 MME develop prolonged use
	≥ 90 MME		Not Serious	Low	HR 1.04 (1.03-1.05) **	0.2% more (0.2-0.3% more) patients with ≥ 90 MME develop prolonged use
Higher dose versus 15 MME or less						
Hadlandsmyth 2019 (model 1) n=317,367 (high risk)	15.01 – 30 MME	Serious[2]	Not Serious	Low	OR 1.53 (1.47-1.59)	2.9% more (2.5-3.2% more) patients with 15.01-30 MME develop prolonged use
	30.01 – 45 MME		Not Serious	Low	OR 1.83 (1.74-1.92)	4.4% more (3.9-4.8% more) patients with 30.01-45 MME develop prolonged use

	> 45 MME		Not Serious	Low	OR 2.74 (2.59-2.90)	8.8% more (8.1-9.5% more) patients with greater than 45 MME develop prolonged use
Hadlandsmyth 2019 (model 2) n=376,140 (high risk)	15.01 – 30 MME	Serious[2]	Not Serious	Low	OR 1.22 (1.19-1.25)	1.2% more (1.0-1.4% more) patients with 15.01-30 MME develop prolonged use
	30.01 – 45 MME		Not Serious	Low	OR 1.37 (1.32-1.42)	2.0% more (1.7-2.3% more) patients with 30.01-45 MME develop prolonged use
	> 45 MME		Not Serious	Low	OR 1.85 (1.77-1.93)	4.5% more (4.1-4.9% more) patients with greater than 45 MME develop prolonged use
Higher dose versus 50 MME or less per day						
Zin 2019 n=33,752 (low risk)	50 – 100 MME	Serious[2]	Not Serious	Low	OR 3.79 (3.01-4.78)	13.3% more (10.0-17.2% more) patients with 50-100 MME per day develop prolonged use
	> 100 MME		Not Serious	Low	OR 7.12 (4.50-11.27)	25.0% more (16.1-35.5% more) patients with greater than 100 MME per day develop prolonged use
Long acting opioids[9]						
Shah 2017 n=1,353,902 (low risk)	Long acting versus short acting opioids	Serious[2]	Not Serious	Low	HR 1.10 (1.09-1.10) **	0.6% more (0.5-0.6% more) patients with long acting opioids develop prolonged use
Deyo 2016 n=536,767 (low risk)		Serious[2]	Not Serious	Low	Not Reported	Those prescribed long-acting opioids were more likely than those prescribed short-acting opioids to develop prolonged use (overall probability 24.5% with long-acting versus 3.5% with short-acting, p<0.001)
Zin 2019 n=33,752 (low risk)		Serious[2]	Not Serious	Low	OR 6.62 (4.90-8.94)	23.4% more (17.6-30.0% more) patients with long acting opioids develop prolonged use
Number of Prescription Refills in First Month[9]						
Deyo 2016, n=536,736 (low risk)	Two versus one prescription fills in first month	Serious[2]	Not Serious	Low	OR 2.25 (2.17-2.33)	6.5% more (6.1-6.8% more) patients filling two opioid prescriptions in first month develop prolonged use[8]
Hadlandsmyth 2019 (model 1) n=317,367 (high risk)		Serious[2]	Not Serious	Low	OR 1.24 (1.17-1.31)	1.3% more (0.9-1.7% more) patients filling two opioid prescriptions in first month develop prolonged use
Hadlandsmyth 2019 (model 2) n=376,140 (high risk)		Serious[2]	Serious	Very Low	OR 1.47 (1.41-1.53)	2.5% more (2.2-2.9% more) patients filling two opioid prescriptions in first month develop prolonged use

Deyo 2016 n=536,736 (low risk)	Three versus one prescription fills in first month	Serious[2]	Not Serious	Low	OR 2.62 (2.49-2.76)	8.2% more (7.6-8.9% more) patients filling three opioid prescriptions in first month develop prolonged use[8]
Hadlandsmyth 2019 (model 1) n=317,367 (high risk)		Serious[2]	Not Serious	Low	OR 0.91 (0.84-0.99)	0.5% less (0.1-0.9% less) patients filling three opioid prescriptions in first month develop prolonged use
Hadlandsmyth 2019 (model 2) n=376,140 (high risk)		Serious[2]	Not Serious	Low	OR 1.18 (1.12-1.24)	1.0% more (0.7-1.3% more) patients filling three opioid prescriptions in first month develop prolonged use
Deyo 2016 n=536,736 (low risk)	Four versus one prescription fills in first month	Serious[2]	Not Serious	Low	OR 3.32 (3.11-3.53)	11.3% more (10.4-12.2% more) patients filling four or more opioid prescriptions in first month develop prolonged use[8]
Hadlandsmyth 2019 (model 1) n=317,367 (high risk)		Serious[2]	Serious[6]	Very Low	OR 1.41 (1.28-1.55)	2.2% more (1.5-3.0% more) patients filling four opioid prescriptions in the first month develop prolonged use
Hadlandsmyth 2019 (model 2) n=376,140 (high risk)		Serious[2]	Serious[6]	Very Low	OR 1.49 (1.39-1.59)	2.6% more (2.1-3.2% more) patients filling four opioid prescriptions in the first month develop prolonged use
Prescribing Factors - Predictors from Single Studies						
Fritz 2018 n=707 (high risk)	Early primary care visit vs not	Not Serious	Serious[6]	Low	OR 1.66 (1.12-2.46)	3.5% more (0.7-7.5% more) patients with early primary care visits develop prolonged use
	Early physical therapy visit vs not	Not Serious	Serious[6]	Low	OR 0.44 (0.22-0.89)	3.2% less (0.6-4.5% less) patients with early physical therapy visits develop prolonged use
Zin 2019 n=33,752 (low risk)	Non-emergency (ER) hospital department vs ER prescription	Serious[2]	Not Serious	Low	OR 14.04 (12.48-15.78)	40.9% more (38.0-43.8% more) patients with non-ER department prescriptions develop prolonged use
Harris 2019 n=3,983 (low risk)	Hydrocodone vs Oxycodone	Not Serious	Not Serious	Moderate	OR 2.62 (1.77-3.88)	8.2% more (4.1-13.7% more) patients prescribed hydrocodone develop prolonged use
Shah 2017 n=1,353,902 (low risk)	Tramadol Use versus other Opioids	Serious[2]	Not Serious	Low	HR 1.12 (1.11-1.12)[3]	0.7% more (0.6-0.7% more) patients prescribed Tramadol develop prolonged use

Harris 2019 n=3,983 (low risk)	Average number of tablets per day	Not Serious	Not Serious	Moderate	OR 0.64 (0.59-0.69)	2.0% less (1.8-2.3% less) patients with higher average numbers of opioid tablets per day develop prolonged use
Fritz 2018 n=707 (high risk)	Benzodiazepine Co-prescription	Not Serious	Serious[6]	Low	OR 1.87 (1.01-3.48)	4.6% more (0.1-12.0% more) patients with early benzodiazepine co-prescription develop prolonged use

Abbreviations: OR = odds ratio; HR = hazard ratio; SEIFA = socioeconomic status, lower deciles indications more disadvantaged individuals; $AU = Australian dollars; USD = United States denomination; MME = morphine milligram equivalent

* GRADE domain of Inconsistency not applicable with individual studies; all studies were high Risk of Bias – see Supplement Table 1; Publication Bias uncertain as < 10 studies for any predictor
** Direction of Hazard Ratio was converted for presentation, as authors evaluated likelihood of opioid discontinuation. Hazard Ratio was considered consistent with an Odds Ratio as overall prolonged use rate in study (5.3%) was < 10%.

1 Baseline risk of 6% based on pooled prevalence of 10 low risk population studies (Figure 2).
2 Rated down as indirect study population.
3 Hazard Ratio was considered consistent with an Odds Ratio as the overall prolonged use rate in the study (1.7%) was < 10%.
4 Quinn 2017 reported event rates based on a random 40% sample of cohort.
5 Rated down for Imprecision as 95% confidence interval overlapped a risk difference of 0 (no effect) or estimate not reported.
6 Rated down for Imprecision as 95%CI crosses 2.5% threshold for a modifiable risk factor or 5% for a non-modifiable factor.
7 Rated down for Imprecision due to small number of observations, only 33 in this study population of 6,463.
8 Selected opioid naïve analysis, short acting opioids prescribed.
9 Overall body of evidence across 3 studies rated down for Imprecision (Very low across studies) due ARI range across studies crossing 2.5% more patients with prolonged use.

Supplement Table 3. GRADE Evidence Profile of Inconsistent Unpooled Predictors of Prolonged Opioid Use After Prescription for Acute Musculoskeletal Injuries

Author, Year, Study Sample Size (population risk for prolonged use – low/high)	Indirectness	Imprecision	Overall Certainty of Evidence *	Adjusted Relative Effect (95%CI)	Risk Difference (95%CI)[1]
Sex (Female versus Male)					
Delgado 2018 n=6,463 (low risk)	Not Serious	Serious[2]	Low	OR 1.33 (0.87-2.05)	1.8% more (0.7% less to 5.5% more) female patients develop prolonged use
Harris 2019 n=3,983 (low risk)	Not Serious	Serious[2]	Low	OR 1.27 (0.99-1.62)	1.5% more (0.1% less to 3.3% more) female patients develop prolonged use
Fritz 2018 n=707 (high risk)	Not Serious	Serious[2]	Low	OR 0.69 (0.46-1.04)	1.8% less (3.1% less to 0.2% more) female patients develop prolonged use
Berecki-Gisolf 2014 n=8,267 (high risk)	Not Serious	Not Serious	Moderate	OR 1.67 (1.48-1.88)	3.6% more (2.6-4.6% more) female patients develop prolonged use
Hooten 2015 n=293 (low risk)	Serious[3]	Serious[2]	Very Low	Sex reported as not statistically significantly associated with prolonged use	
Anxiety					
Fritz 2018 n=707 (high risk)	Not Serious	Serious[4]	Low	OR 1.69 (1.12-2.55)	3.7% more (0.7-7.9% more) patients with anxiety develop prolonged use
O'Hara 2018 n=9,596 (low risk)	Not Serious	Serious[2]	Low	Anxiety reported as not statistically significantly associated with prolonged use	
Quinn 2017 n=10,311,961 (low risk)	Serious[3]	Not Serious	Low	OR 1.92 (1.89-1.95)	4.8% more (4.7-5.0% more) patients with anxiety develop prolonged use
Depression					
Delgado 2018, n=6,463 (low risk)	Not Serious	Serious[2]	Low	OR 1.04 (0.54-2.02)	0.2% more (2.6% less to 5.3% more) patients with depression develop prolonged use
Fritz 2018 n=707 (high risk)	Not Serious	Serious[2]	Low	OR 0.98 (0.64-1.51)	0.1% less (2.0% less to 2.7% more) patients with depression develop prolonged use

O'Hara 2018 n=9,596 (low risk)		Not Serious	Serious[2]	Low	Mood disorder reported as not statistically significantly associated with prolonged use	
Halbert 2016 n=2,995 (low risk)		Serious[3]	Not Serious	Low	OR 1.77 (1.15-2.72)	4.1% more (0.8-8.7% more) patients with depression develop prolonged use
Quinn 2017 n=10,311,961 (low risk)		Serious[3]	Not Serious	Low	OR 1.94 (1.92-1.97)	4.9% more (4.8-5.1% more) patients with depression develop prolonged use
Smoking Status						
Fritz 2018 n=707 (high risk)		Not Serious	Serious[4]	Low	OR 1.53 (1.03-2.28)	2.9% more (0.2-6.6% more) patients with positive smoking status develop prolonged use
Hooten 2015 n=293 (low risk)		Serious[3]	Very Serious[4,5]	Very Low	OR 2.12 (0.66-6.80)	5.8% more (1.9% less to 24.0% more) patients with positive smoking status develop prolonged use
Occupation						
Occupations versus Professionals						
Berecki-Gisolf 2014 n=8,267 (high risk)	Trade persons / related workers	Not Serious	Not Serious	Moderate	OR 1.23 (1.01-1.49)	1.3% more (0.1-2.6%) trade persons develop prolonged use
	Managers / administrators		Serious[2]	Low	OR 0.85 (0.63-1.15)	0.8% less (2.1% less to 0.8% more) managers / administrators develop prolonged use
	Associate professionals		Serious[2]	Low	OR 1.22 (0.98-1.51)	1.2% more (0.1% less to 2.7% more) associate professionals develop prolonged use
	Advanced clerical and service workers		Serious[2,4]	Low	OR 1.61 (0.97-2.66)	3.3% more (0.2% less to 8.4% more) advanced clerical and service workers develop prolonged use
	Intermediate clerical workers		Serious[2]	Low	OR 1.17 (0.96-1.43)	0.9% more (0.2% less to 2.3% more) intermediate clerical workers develop prolonged use
	Intermediate production and		Not Serious	Moderate	OR 1.36 (1.13-1.64)	2.0% more (0.7-3.4%) intermediate production and transport workers develop prolonged use

	transport workers					
	Elementary clerical workers		Not Serious	Moderate	OR 1.41 (1.07-1.85)	2.2% more (0.4-4.5%) elementary clerical workers develop prolonged use
	Laborers and related workers		Not Serious	Moderate	OR 1.60 (1.34-1.92)	3.2% more (1.9-4.8%) laborers and related workers develop prolonged use
Occupations versus Operatives and Technicians						
O'Hara 2018 n=9,596 (low risk)	Service workers	Not Serious	Serious[2]	Low	OR 1.02 (0.85-1.24)	0.1% more (0.8% less to 1.3% more) service workers develop prolonged use
	Laborers and helpers		Serious[2]	Low	OR 1.17 (0.95-1.46)	0.9% more (0.3% less to 2.5% more) laborers and helpers develop prolonged use
	Professionals		Serious[2]	Low	OR 0.86 (0.63-1.16)	0.8% less (2.1% less to 0.9% more) professionals develop prolonged use
	Office workers		Serious[2]	Low	OR 1.13 (0.86-1.49)	0.7% more (0.8% less to 2.6% more) office workers develop prolonged use
	Sales workers		Serious[2]	Low	OR 0.78 (0.48-1.26)	1.2% less (3.0% less to 1.4% more) sales workers develop prolonged use
	Unrecorded		Serious[2]	Low	OR 0.32 (0.08-1.31)	3.9% less (5.4% less to 1.7% more) patients develop prolonged use
Injury Type						
Injury type versus other musculoskeletal injuries						
Berecki-Gisolf 2014 n=8,267 (high risk)[6]	Tendon trauma	Not Serious	Not Serious	Moderate	OR 0.54 (0.48-0.61)	2.6% <u>less</u> (2.2-3.0% less) patients with tendon trauma develop prolonged use
	Wounds	Serious[3]	Not Serious	Low	OR 0.37 (0.31-0.43)	3.6% <u>less</u> (3.3-4.0% less) patients with wounds develop prolonged use
	Fractures	Not Serious	Not Serious	Moderate	OR 0.36 (0.30-0.42)	3.7% <u>less</u> (3.3-4.1% less) patients with fractures develop prolonged use
Injury type versus soft tissue or contusion injuries						

O'Hara 2018 n=9,596 (low risk)	Strain or sprain	Not Serious	Not Serious	Moderate	OR 1.54 (1.27-1.86)	2.9% more (1.5-4.5%) patients with strain or sprain develop prolonged use
	Fracture	Not Serious	Not Serious	Moderate	OR 1.38 (1.05-1.81)	2.1% more (0.3-4.3%) patients with fracture develop prolonged use
	Crush injury	Not Serious	Not Serious	Moderate	OR 1.45 (1.08-1.94)	2.4% more (0.4-4.9%) patients with crush injury develop prolonged use
	Other or unspecified injury	Not Serious	Serious[4]	Low	OR 2.18 (1.23-3.85)	6.1% more (1.3-13.5%) patients with other or unspecified injuries develop prolonged use
	Open wound	Serious[3]	Serious[2]	Very Low	OR 1.24 (1.00-1.54)	1.3% more (0.0-2.9%) patients with open wound develop prolonged use
	Poisoning[7]	Serious[3]	Serious[2]	Very Low	OR 1.30 (0.90-1.86)	1.6% more (0.6% less to 4.5% more) patients with poisoning develop prolonged use

Abbreviation: OR = odds ratio

* GRADE domain of Inconsistency rated as serious, to very low, all across multiple combined predictors; all studies were rated as high Risk of Bias – see Supplement Table 1.

1 Baseline risk of 6% based on pooled prevalence of 10 low risk population studies (Figure 2).
2 Rated down for Imprecision as 95% confidence interval at or overlapped a risk difference of 0 (no effect) or estimate not reported.
3 Rated down as indirect study population.
4 Rated down for Imprecision as 95%CI crosses 2.5% threshold for a modifiable risk factor of 5% for a non-modifiable factor.
5 Rated down for Imprecision due to small study sample size.
6 The study sample of Berecki-Gisolf 2014 consisted of workers with an approved workers' compensation claim with an injury onset in 2008–2009. Those with a primarily mental health claim were excluded. Those with injury to nerves and spinal cord (N = 21) were also excluded, as opioid use was extremely high in this very small group. In total, 54,931 claims were included. In this workers' compensation population with a mix of work-related injury and disease, opioid use was determined and described. Further analysis and modeling was done using a more homogeneous sample of injured workers with any of the four most common injury categories: musculoskeletal, fractures, wounds, and tendon trauma.
7 In the study sample of O'Hara 2018, non-opioid poisoning represented 4.5% of study population.

Supplement Table 4. Statistically Nonsignificant Associations of Unpooled Predictors With Prolonged Opioid Use

Author, Year	Predictor	Comparison	Study Sample Size	Adjusted OR (95%CI)	Overall Certainty of Evidence *	Interpretation
Sociodemographic Factors						
Delgado 2018	Race	Asian vs White Black vs White Hispanic vs White Unknown vs White	6,463	OR 0.67 (0.09-4.97) OR 0.98 (0.49-1.95) OR 0.78 (0.36-1.70) OR 0.74 (0.20-2.78)	Low; Rated down for Risk of Bias and Imprecision.	No statistically significant association was found between race and prolonged opioid use.
Delgado 2018	Alcohol abuse	No alcohol abuse	6,463	OR <0.001 (<0.001->999.99)[1]	Very Low; Rated down for Risk of Bias and Imprecision.	No statistically significant association was found between alcohol abuse, psychosis, episodic mood disorder or obesity and prolonged opioid use.
Delgado 2018	Psychosis	No psychosis	6,463	OR 1.28 (0.25-6.65)	Low; Rated down for Risk of Bias and Imprecision.	
Fritz 2018	Episodic mood disorder[2]	No episodic mood disorder	707	OR 0.98 (0.61-1.56)		
Fritz 2018	Obesity	No obesity	707	OR 1.32 (0.84-2.05)		
O'Hara 2018	Employment status - non-full time employment; vs full-time[3]		9,596	Not reported; excluded from multivariable analysis	Low; Rated down for Risk of Bias and Imprecision.	No statistically significant association was found for non-full time employment status and prolonged opioid use.
Prescribing Factors						
Delgado 2018	Year sampled	Year 2012 vs 2011 Year 2013 vs 2011 Year 2014 vs 2011 Year 2015 vs 2011	6,463	OR 1.54 (0.90-2.65) OR 0.75 (0.38-1.44) OR 0.63 (0.30-1.31) OR 0.80 (0.40-1.60)	Low; Rated down for Risk of Bias and Imprecision.	No statistically significant association was found between later years 2012-2015 (vs 2011 year) and prolonged opioid use.
Harris 2019	Region (Central / Western USA)	Eastern USA	3,983	OR 1.22 (0.94-1.59)	Low;	No statistically significant

				Rated down for Risk of Bias and Imprecision.	association was found between region of USA and prolonged opioid use.
Fritz 2018	Early diagnostic imaging; emergency room, surgeon or other specialist consultations; co-prescription with NSAIDs, muscle relaxers or oral steroids; vs not	707	Not reported; excluded from multivariable analysis	Low; Rated down for Risk of Bias and Imprecision.	No statistically significant association was found for early diagnostic imaging (radiographs or advanced imaging); emergency room, surgeon or other specialist consultations, or co-prescription with NSAIDs, muscle relaxers or oral steroids.

Abbreviations: NSAIDs = nonsteroidal anti-inflammatory drugs; OR = odds ratio

* Domain of Inconsistency not applicable.

1 Only 73 observations of 6463 in this category.
2 Episodic mood disorders were classified as bipolar, manic affective and major depressive disorders.
3 Employment status was not entered into models with 180-day or 365-day follow-up based stepwise technique with a minimum Akaike information criterion.

Supplement Table 5. Consideration of Important Independent Factors Across Included Studies

Author, Year	Important Independent Factors Across Studies						
	Age	**Sex**	**Injury Severity**	**Physical Co-morbidities**	**Past or Present Substance Use Disorder**	**Payer (e.g. Disability, Medicaid)**	**Opioid Prescribing Factors**
Berecki-Gisolf 2014	Included	Included	Included			Included	
Delgado 2018	Included	Included	Included	Included	Included	Included	Included
Fritz 2018	Included	Included	Included	Included	Included	Included	Included
O'Hara 2018	Included	Included	Included		Included	Included	
Durand 2019			Included				Included
Harris 2019	Included	Included					Included
Deyo 2016							Included
Halbert 2016			Included				
Hooten 2015	Included	Included	Included	Included	Included		
Quinn 2017					Included		
Shah 2017	Included	Included	Included		Exclusion Criteria	Included	Included
Handlandsmyth 2019							Included
Zin 2019	Included	Included					Included

Summary of Search Strategy and Results

Database	Total
MEDLINE	4777
EMBASE	7001
Web of Science	2379
Google Scholar	990
Subtotal	15147
- Duplicates	- 3400
Total	11747

MEDLINE

Database: OVID Medline Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present
Search Strategy:
---
1 (opioid adj2 "use*").mp. (9521)
2 "opioid use".ti,ab. (7000)
3 "use of opioid*".ti,ab. (3462)
4 (Opioid* or opiate* or narcotic* or analges* or oxycodone* or hydrocodone*).ti. (89391)
5 (prescription* or prescrib* or long-term or longer-term or early or late or sustain* or prolong* or persistent or repeat or recurrent or problematic or user* or usage).ti. (848330)
6 4 and 5 (6344)
7 1 or 2 or 3 or 6 (15518)
8 ((duration or length) adj3 (therapy or treatment)).ab. (45676)
9 4 and 8 (332)
10 7 or 9 (15730)
11 ((Opioid* or opiate* or narcotic* or analges*or morphine* or MED or MEQ or oxycodone* or hydrocodone* or hydromorphone* or fentanyl* or codeine*) adj3 (Early or late or naive or initial* or initiat* or prolong* or sustain* or long-term or longer-term or fill* or repeat* or pharmacovigil* or recurrent or problematic or user* or usage)).ti,ab. (7358)
12 10 or 11 (19371)
Annotation: Opioid use block free-text
13 exp *Analgesics, Opioid/ad, ae, tu [Administration & Dosage, Adverse Effects, Therapeutic Use] (36547)
14 exp Opioid-Related Disorders/ (25507)
15 13 or 14 (53936)
Annotation: opioid use MeSH
16 12 or 15 (64642)
Annotation: opioid use block
17 prognosis/ (491540)
18 ep.fs. and (opioid* or opiate* or narcotic*).tw. (9452)

19 exp risk/ (1170987)
20 exp PROBABILITY/ (1341700)
21 exp Regression Analysis/ (414546)
22 "analysis of variance"/ or multivariate analysis/ (333737)
23 exp Epidemiologic Studies/ (2419109)
24 (prognosis or prognostic or predict* or risk*).tw. (3647808)
25 ((univariate or covariate or variance or covariance or multivariate or regression or adjusted or unadjusted or logistic or diagnostic) adj2 (analys* or model*)).tw. (720527)
26 (logistic adj2 regress*).tw. (268856)
27 ((cohort or observational) adj3 (study or studies or analy*)).tw. (341937)
28 (longitudinal or retrospective or cross sectional or prospective).tw. (1494855)
29 (Follow up adj (study or studies)).tw. (48134)
30 or/17-29 (6361717)
Annotation: prognosis
31 16 and 30 (25031)
Annotation: opioid use AND prognosis
32 exp arm injuries/ or athletic injuries/ or exp joint dislocations/ or exp fractures, bone/ or fractures, cartilage/ or exp hand injuries/ or exp hip injuries/ or exp leg injuries/ or exp neck injuries/ or occupational injuries/ or exp shoulder injuries/ or exp soft tissue injuries/ or exp "sprains and strains"/ or exp tendon injuries/ or exp Compartment Syndromes/ or exp Bone Malalignment/ (326498)
33 (exp Musculoskeletal system/ or musculoskeletal diseases/ or osteitis/ or exp cartilage diseases/ or exp fasciitis/ or exp bursitis/ or exp metatarsalgia/ or exp synovitis/ or muscle cramp/ or myalgia/ or exp tendinopathy/) and pain*.ti,ab. (99258)
34 Musculoskeletal Pain/ or Neck Pain/ or Acute Pain/ or exp Arthralgia/ (23367)
35 (Arthralgi* or bursitis or capsulit* or epicondyalgia* or epicondylit* or fasciopath* or fasciitis or fascitis or metatarsalgi* or myalgi* or myelitis or myopath* or myosit* or osteitis or osteochondritis or osteomyelitis or polymyosit* or radiculopath* or radiculit* or synovit* or tend?nopath* or tendinit* or tenosynovit* or whiplash or WAD).tw. (122302)
36 ((ligament or tendon or supraspinatus or infraspinatus or subscapularis or teres minor or teres major or trapezius or deltoid or bicep* or bicipital or coracobrachialis or deltoid or fibularis or talofibular or calcaneofibular or calcaneotibial or tibio* or rotator cuff) adj3 (injur* or impair* or imping* or sprain* or strain* or tear or torn)).tw. (16718)
37 ((myofascial or neck* or cervical* or musculoskeletal* or MSK or elbow* or arm* or finger* or hand* or wrist* or forearm* or leg or ankle* or knee* or hip* or foot* or toe* or femur* or radius or radii or tibia* or ulna* or humerus or humeri or metatarsal* or metacarpal* or fibula* or patella* or patellofemoral or carpal* or tarsal* or phalange* or clavicle* or scapula* or bone* or joint* or muscle* or shoulder*) adj3 (sprain* or strain* or injur* or impair* or fractur* or break* or broken or disorder* or pain*)).tw. (266903)
38 (sciatica or backache or dorsalgia or lumbago or toothache or migraine*).mp. [mp=title, abstract, original title, name of substance word, subject heading word,

floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (52922)
39 exp Back Pain/ or exp Toothache/ or headache/ or exp Headache disorders/ or exp dentistry/ (495659)
40 or/32-39 (1133274)
Annotation: acute msk disorders
41 exp Emergency Service, Hospital/ (74797)
42 exp Emergency Medicine/ (13340)
43 emergency.ti,jw. (158353)
44 accident.ti,jw. (19492)
45 ((emergency or trauma or triage) adj3 (care or healthcare or department* or unit or units or room* or treatment* or centre or centres or center or centers or ward*)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (162541)
46 or/41-45 (283598)
Annotation: emergency care population
47 Primary Health Care/ or Outpatients/ or Medicine/ or Specialization/ (149068)
48 ((primary adj3 (care or healthcare)) or specialt* or outpatient* or communit*).ti,ab. (838522)
49 ((health or healthcare) and (data or database* or record or records)).ti,ab. (590959)
50 or/47-49 (1387232)
Annotation: primary care hedge
51 sn.fs. and (opioid* or opiate* or narcotic*).mp. (9298)
52 Drug Utilization/ or Economics, Pharmaceutical/ or Pharmacoepidemiology/ (24441)
53 (pharmacoeconomic* or pharmacoepidemiol*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, floating sub-heading word, keyword heading word, organism supplementary concept word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms] (8333)
54 prescrip*.ab. /freq=2 (31208)
55 prescrib*.ab. /freq=2 (36599)
56 or/51-55 (89796)
Annotation: prescribing trends hedge
57 Pain Management/ (32344)
58 pain*.jw,ti. (224721)
59 pain*.ab. /freq=2 (279509)
60 exp *pain/dt (48154)
61 or/57-60 (401553)
Annotation: pain hedge
62 40 and 61 (135904)
Annotation: acute msk pain population
63 50 and 56 (30591)

64 50 and 61 (34248)
65 56 and 61 (6604)
Annotation: population hedge
66 46 or 62 or 65 (420544)
Annotation: 3 population concepts
67 31 and 66 (4777)
68 31 and (46 or 62 or 65) (4777)

EMBASE

Database: Embase <1974 to 2020 January 03>
Search Strategy:
---
1 prescription/ and (dt.fs. and opioid*.mp.) (3682)
2 *opiate/ and (dt.fs. and opioid*.mp.) (6460)
3 *opiate/ (29955)
4 narcotic analgesic agent/dt [Drug Therapy] (10188)
5 *opiate derivative/dt [Drug Therapy] (428)
6 or/3-5 (40334)
7 prescription/ (193537)
8 6 and 7 (6838)
9 1 or 2 or 8 (12996)
Annotation: emtree opioid use terms from validation set
10 (opioid adj2 "use*").mp. (14759)
11 "opioid use".ti,ab. (10929)
12 "use of opioid*".ti,ab. (5472)
13 (Opioid* or opiate* or narcotic* or analges* or oxycodone* or hydrocodone*).ti. (110507)
14 (prescription* or prescrib* or long-term or longer-term or early or late or sustain* or prolong* or persistent or repeat or recurrent or problematic or user* or usage).ti. (1056974)
15 13 and 14 (8614)
16 10 or 11 or 12 or 15 (23016)
17 ((duration or length) adj3 (therapy or treatment)).ab. (79762)
18 13 and 17 (521)
19 16 or 18 (23347)
20 ((Opioid* or opiate* or narcotic* or analges*or morphine* or MED or MEQ or oxycodone* or hydrocodone* or hydromorphone* or fentanyl* or codeine*) adj3 (Early or late or naive or initial* or initiat* or prolong* or sustain* or long-term or longer-term or fill* or repeat* or pharmacovigil* or recurrent or problematic or user* or usage)).ti,ab. (11449)
21 19 or 20 (29022)
Annotation: Opioid use block free-text
22 9 or 21 (35269)
Annotation: opioid use block (Emtree OR freetext)
23 cohort analysis/ or trend study/ or pharmacoepidemiology/ or sensitivity analysis/ or prognosis/ (1220076)

24 risk/ or risk assessment/ or risk factor/ or exp regression analysis/ or "analysis of variance"/ or multivariate analysis/ or propensity score/ or treatment duration/ (2531708)
25 (prognosis or prognostic or predict* or risk*).tw. (5081499)
26 ((univariate or covariate* or covariance or multivaria* or regression or adjusted or unadjusted or logistic) adj2 (analys* or model*)).tw. (1015116)
27 (logistic adj2 regress*).tw. (396916)
28 ((cohort or observational) adj3 (study or studies or analy*)).tw. (527518)
29 ((association* or associated) and "opioid use").ti,ab. (5012)
30 ((duration or long-term or longer-term or sustain* or prolong* or persist*) and "opioid use").ti,ab. (3057)
31 or/23-30 (6654703)
Annotation: prognosis concept
32 22 and 31 (17243)
Annotation: Use of opioids AND prognosis
33 limb injury/ or exp arm injury/ or exp leg injury/ or exp limb fracture/ or sport injury/ or exp joint injury/ or musculoskeletal injury/ or exp cartilage injury/ or exp "ligament and tendon injury"/ or medial tibial stress syndrome/ or muscle injury/ or overexertion/ or exp sprain/ or fracture/ or avulsion fracture/ or clavicle fracture/ or comminuted fracture/ or fracture dislocation/ or exp fracture healing/ or intraarticular fracture/ or exp joint fracture/ or exp limb fracture/ or exp multiple fracture/ or exp scapula fracture/ or stress fracture/ or neck injury/ or whiplash injury/ or occupational accident/ or soft tissue injury/ or compartment syndrome/ or musculoskeletal pain/ or neck pain/ or shoulder pain/ or arthralgia/ (494349)
34 (Arthralgi* or bursitis or capsulit* or epicondyalgia* or epicondylit* or fasciopath* or fasciitis or fascitis or metatarsalgi* or myalgi* or myelitis or myopath* or myosit* or osteitis or osteochondritis or osteomyelitis or polymyosit* or radiculopath* or radiculit* or synovit* or tend?nopath* or tendinit* or tenosynovit* or whiplash or WAD).tw. (162559)
35 ((ligament or tendon or supraspinatus or infraspinatus or subscapularis or teres minor or teres major or trapezius or deltoid or bicep* or bicipital or coracobrachialis or deltoid or fibularis or talofibular or calcaneofibular or calcaneotibial or tibio* or rotator cuff) adj3 (injur* or impair* or imping* or sprain* or strain* or tear or torn)).tw. (19570)
36 ((myofascial or neck* or cervical* or musculoskeletal* or MSK or elbow* or arm* or finger* or hand* or wrist* or forearm* or leg or ankle* or knee* or hip* or foot* or toe* or femur* or radius or radii or tibia* or ulna* or humerus or humeri or metatarsal* or metacarpal* or fibula* or patella* or patellofemoral or carpal* or tarsal* or phalange* or clavicle* or scapula* or bone* or joint* or muscle* or shoulder*) adj3 (sprain* or strain* or injur* or impair* or fractur* or break* or broken or disorder* or pain*)).tw. (348095)
37 or/33-36 (800228)
Annotation: acute MSK pain
38 32 and 37 (1373)
Annotation: opioid use AND prognosis AND acute MSK pain

39 emergency health service/ or hospital emergency service/ or emergency care/ or emergency ward/ (255808)
40 (accident or emergency).ti,jw. (212619)
41 ((emergency or trauma or triage) adj3 (care or healthcare or department* or unit or units or room* or treatment* or centre or centres or center or centers or ward*)).mp. (292646)
42 or/39-41 (436752)
Annotation: emergency care
43 32 and 42 (1486)
Annotation: opioid use AND prognosis AND emergency care
44 primary medical care/ or outpatient/ or outpatient care/ or specialization/ (253727)
45 ((primary adj3 (care or healthcare)) or specialt* or outpatient* or communit*).ti,ab. (1103359)
46 ((health or healthcare) and (data or database* or record or records)).mp. (1200173)
47 44 or 45 or 46 (2138930)
Annotation: primary care
48 *"statistics and numerical data"/ (48641)
49 prescrip*.ab. /freq=2 (55692)
50 prescrib*.ab. /freq=2 (66002)
51 pharmacoeconomics/ or drug utilization/ (27008)
52 or/48-51 (175879)
Annotation: analysis of prescription trends hedge
53 analgesia/ (120617)
54 pain*.jw,ti. (301492)
55 pain*.ab. /freq=2 (425125)
56 exp pain/dt (164895)
57 or/53-56 (669313)
Annotation: pain hedge
58 47 and 52 (81662)
Annotation: primary care with prescription trends hedge
59 47 and 57 (73463)
Annotation: primary care with pain hedge
60 52 and 57 (10573)
Annotation: set of two hedges: pain AND analysis of prescription trends
61 58 or 59 or 60 (156516)
Annotation: set of three pairs of hedged results for primary care, pain and prescription trends
62 32 and 61 (5495)
Annotation: opioid use AND prognosis AND triple concept (primary care, pain, prescription trends)
63 38 or 43 or 62 (7001)
64 37 or 42 or 61 (1331120)
Annotation: acute MSK pain OR Emergency care OR primary care-pain-prescribing trends (population concept)
65 32 and 64 (7001)

Annotation: population concept AND opioid use AND prognosis

Web of Science

# 9	**2,379**	#8 AND #5 *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI* *Timespan=1976-2020*
# 8	**7,381**	#7 AND #6 *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI* *Timespan=1976-2020*
# 7	**1,603,620**	TI=(prescription* or prescrib* or long-term or longer-term or early or late or sustain* or prolong* or persistent or repeat or recurrent or problematic or user* or usage) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI* *Timespan=1976-2020*
# 6	**94,352**	TI=(Opioid* or opiate* or narcotic* or analges* or oxycodone* or hydrocodone*) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI* *Timespan=1976-2020*
# 5	**6,668,704**	#4 OR #3 OR #2 OR #1 *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI* *Timespan=1976-2020*
# 4	**381,515**	TS=((cohort or observational) NEAR/3 (study or studies or analy*)) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI* *Timespan=1976-2020*
# 3	**278,269**	TS= (logistic NEAR/2 regression) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI* *Timespan=1976-2020*
# 2	**957,088**	TS=((univariate or covariate or variance or covariance or multivariate or regression or adjusted or unadjusted or logistic or diagnostic) NEAR/2 (analysis or analyses or model or models or modelling)) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI* *Timespan=1976-2020*
# 1	**6,090,172**	TS=(prognosis or prognostic or predict* or risk*) *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI* *Timespan=1976-2020*

GoogleScholar (via Harzing's Publish or Perish)
https://harzing.com/resources/publish-or-perish

predictor, prolonged duration sustained, opioid use, NOT surgery
Publish or Perish 7.15.2643.7260
Windows (x64) edition, running on Windows 6.1.7601 (x64)
Search terms
Keywords: predictor, prolonged duration sustained, opioid use, NOT surgery
Years: all

Data retrieval
Data source: Google Scholar

Search date: 2020-01-06 12:51:37 -0500
Cache date: 2020-01-06 13:43:38 -0500
Search result: [1027]

Metrics
Reference date: 2020-01-06 13:43:38 -0500
Publication years: 1971-2020
Citation years: 49 (1971-2020)
Papers: 990

CHAPTER 3. REPORTING OF ECONOMIC INFORMATION IN GRADE GUIDELINES THAT USE EVIDENCE-TO-DECISION (ETD) FRAMEWORKS: A SYSTEMATIC SURVEY

Status: Final draft manuscript, to be submitted to the *Journal of Clinical Epidemiology*.

Reporting of Economic Information in GRADE Guidelines that use Evidence-to-decision (EtD) Frameworks: a Systematic Survey

Authors

John J. Riva[1,2,3], Meha Bhatt[3], David J. Brunarski[4], Jason W. Busse[5,6,7], Carolina C. Martins[8], Feng Xie[3,9], Holger J. Schünemann[2,3,10], Jan L. Brozek[2,3,10]

[1] Department of Family Medicine, McMaster University, Hamilton, Ontario, Canada.

[2] Michael G. DeGroote Cochrane Canada and McMaster GRADE Centres, Hamilton, Ontario, Canada.

[3] Department of Health Research Methods, Evidence & Impact, McMaster University, Hamilton, Ontario, Canada.

[4] Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands.

[5] Department of Anesthesia, McMaster University, Hamilton, Ontario, Canada.

[6] Michael G. DeGroote National Pain Centre, McMaster University, Hamilton, Ontario, Canada.

[7] Chronic Pain Centre of Excellence for Canadian Veterans, Hamilton, Ontario, Canada.

[8] Department of Pediatric Dentistry, Federal University of Minas Gerais, Belo Horizonte, Brazil.

[9] Centre for Health Economics and Policy Analysis (CHEPA), McMaster University, Hamilton, Ontario, Canada.

[10] Department of Medicine, Hamilton, Ontario, Canada.

Corresponding author: John J. Riva, Department of Family Medicine, David Braley Health Sciences Centre, 100 Main Street West, 6th Floor, Hamilton, ON L8P 1H6; email: rivaj@mcmaster.ca

Previous presentations: Presented as an oral presentation at the full GRADE working group meeting, June 13, 2019 in Hamilton, Ontario.

Key points

1. The majority of GRADE guideline developers report economic evidence and many panels justify their clinical practice recommendations with this information.
2. Reporting of economic information is quite variable with regards to detail and directness.
3. Linking economic information to recommendation justifications serves to improve adoption, adaptation of recommendations, and transparency of GRADE EtDs.

Abstract

Background: The GRADE guideline Evidence-to-Decision (EtD) framework was developed to improve structure, transparency and reusability for end users. Little is known about how developers and panel members report cost and cost effectiveness considerations in these EtD frameworks.

Purpose: A systematic survey to explore approaches and factors contributing to variability in economic information reporting in GRADE EtD frameworks.

Data sources: Guideline organization websites, suggested by GRADE working group members, were systematically searched to create a convenience sample of guidelines, from 2012 when EtDs experienced wider use, up to 2018.

Study selection: Reviewers screened websites to identify published EtD frameworks from GRADE guidelines, without language restrictions.

Data extraction: One author extracted EtD data verbatim from websites and generated frequencies of reporting approaches, with 3 authors checking work for consistency. We used thematic analysis, independently and duplicate, to summarize themes of factors related to variability of economic information reporting if they were used at least once in a guideline.

Data synthesis: We included 142 guidelines, with 1625 EtD frameworks. The overall rate of reporting at least some economic information was high (91%); but, there was variability across completion of pre-defined EtD Likert-type judgments (70%), noting information as not identified across EtD framework domains (57%), and providing remarks used to justify recommendations (38%). Six themes emerged related to: intervention, population, payor, provider, healthcare resource use and economic model building factors that contributed to economic information variability. Only two guidelines performed a GRADE certainty appraisal of economic outcomes.

Conclusion: Completing pre-defined EtD Likert-type judgments, specifically reporting the literature review approach, study selection criteria and economic model building limitations, as well as linking these to recommendation justification remarks are potential areas for improved use, adoption and adaptation of recommendation, and transparency of GRADE EtD frameworks.

Keywords: economics, resource use, costs, evidence-to-decision, GRADE, guidelines.

Background

Innovations in the GRADE guideline development approach led to the advancement of the evidence-to-decision (EtD) framework through its European Union funded DECIDE project in 2011. The four key GRADE criteria considered initially were: 1) balance of effects; 2); certainty of evidence; 3) values and preferences; and 4) resource allocation.[1,2] In the DECIDE project, the GRADE Working Group developed a more comprehensive framework intended to allow for improved structure for developers with varying level of competencies and transparency for readers across a wide variety of practice contexts (e.g. treatment, diagnosis, public health) for decision-making.[3-8] This new EtD framework included the four GRADE factors, as well as additional domains from other fields that addressed other important decision-making factors, such as cost effectiveness,[9] feasibility[10] and equity.[11] With regards to information about resources, the EtD framework contains pre-defined Likert-type questions (weighing magnitudes of costs and cost-effectiveness, certainty of evidence, as well as whether they favour the intervention or comparison) indicating a judgment, research evidence and blank free text fields allowing for entry of other cost, cost-effectiveness or additional considerations that the guideline team or panel decides to include.

While placed into widespread use around 2012, further testing and development work will focus on how EtD framework use may affect panel member recommendations, transparency of reporting for readers, and reusability for end users making new decisions.[12] Many GRADE guidance papers exist on domains of balance of effects, certainty of evidence and patient values and preferences, but there have only been a few reports with respect to how guideline developers

may wish to report economic information (resource allocation) within this new EtD framework over time.[13]

Purpose

To review a convenience sample of GRADE guideline EtD frameworks published online from first use in 2012 until 2018, identified by GRADE working group members, to assess reporting approaches used by clinical practice guideline developers for considering economic information.

Methods

Data sources

As EtD frameworks are often not published as part of indexed journals manuscripts, we sent an e-mail in October 2017 to the entire GRADE working group (over 600 members, personal communication H.J.S.) (members@lists.gradeworkinggroup.org) asking members for online links to specific guideline development agency websites or any links to other published or unpublished EtD frameworks they had access to on guideline development platforms, such as GRADEpro-GDT (https://gdt.gradepro.org/app/) and MAGICapp (https://app.magicapp.org/). From links provided, we systematically searched these websites for any additional GRADE guideline EtD frameworks.

Guideline selection

We included all guidelines with available EtD frameworks that were identified from organization website links provided by GRADE working group members, in any language, up to the end of

2018. We systematically reviewed websites for any new GRADE guidelines by manually searching up and down the link hierarchy from each website. Frameworks that did not allow the ability to capture free text fields, due to access restrictions, were excluded.

Data extraction

We piloted an Excel spreadsheet abstraction form on a sample of 22 known Kingdom of Saudi Arabia Ministry of Health GRADE guidelines.[117-138] From this, we refined the form to allow for extraction of demographic information for each guideline (i.e. guideline agency, country, publication year, title), whether EtD frameworks from guidelines completed the pre-defined Likert-type judgments, and any EtD framework free text information provided in the research evidence and additional consideration sections for the domains of cost, cost-effectiveness, acceptability, feasibility, equity, or related to the recommendation section remarks used as justification.

One author (JJR) extracted all information per guideline and 3 others (MB, DB, CM) checked this work for consistency, with disagreements resolved by consensus. Non-English content was converted to English using Google Translate (https://translate.google.ca/). For guidelines with 8 or more EtD frameworks, we used an online random number generator (https://www.random.org/) and selected 10% of EtD frameworks and at least 2 if there were less than 20 EtD frameworks. If saturation (where no new information was identified) was not apparent, further EtD frameworks were abstracted. To avoid clustering, for agencies with more than 25 guidelines, we selected 20% or least 10 of most recently published guidelines. Again, if saturation was not apparent, further guidelines were abstracted until no new information was apparent. We created a per-guideline summary, making the assumptions that items presented in

at least one EtD framework of a guideline would: (1) have had equal consideration within a guideline; and (2) that agencies with multiple guidelines may potentially use different summary approaches between guidelines year-to-year.

Data synthesis

We generated frequencies for the economic information (pre-defined Likert-type judgments completed and free text content), reported at least once, in an EtD framework per GRADE guideline. For free text information we additionally performed thematic analysis, with two reviewers (JJR, MB) developing a preliminary coding system to categorize themes and sub-themes after a discussion of a small sample of 15 comments, using a previously established approach.[14] We then applied this system, independently and in duplicate, to written comments in free text fields until coding became stable, as evidenced by no new codes and disagreements among reviewers being minimal. Each block of free text information could contribute to more than one theme or sub-theme per criterion on the EtD framework domain; however, each theme or sub-theme was only counted once across a single guideline to address the issue of clustering for total counts across the entire sample of guidelines.

Results

We received 18 replies to our e-mail request from GRADE working group members and after systematic searching of associated websites, we included 142 guidelines[15-156] that reported on 1625 sets of EtD framework information (Appendix 1). We excluded guidelines from one organization[157] because their document security features did not allow cut and paste of free text.

Of the one organization that published 42 guidelines, we included their 10 most recent[89-98] in our systematic survey.

Guideline characteristics

The 142 guidelines were from 29 different settings worldwide, with a mean of 6 (range: 1 to 42) guidelines per organization (Appendix 2). The vast majority were non-government (e.g. association, professional society) organizations (48.3%) or government (e.g. health ministry) groups (31.0%), with guidelines published since 2017. Most organizations were from North America (24.0%), Europe (21.1%) or part of an international organization (26.8%), such as the World Health Organization (WHO) and World Allergy Organization. The clinical area of guidelines represented a diverse spectrum of decision-making situations; but, most commonly they related to communicable disease management (e.g. vaccination, infection control) (21.1%) or hospital and injury management (e.g. emergency medical services, fracture repair) (19.0%). The large majority (74.0%) of guidelines had 10 or less EtD frameworks; although, there was a very wide variation with a median of 6 (range: 1 to 264) (Table 1).

Frequencies of reported economic information

We considered three avenues or locations in the EtD framework allow for reporting economic information: (1) completing pre-defined Likert-type judgments in the resources required (judging the magnitude of costs), certainty of evidence of resources required (judging from very low to high), and cost-effectiveness (judging from favouring intervention to favouring comparison) domains; (2) reporting economic free text information in the research evidence and additional considerations sections for the domains of costs, cost-effectiveness, acceptability, feasibility and

equity; and (3) reporting economic justification information in the remarks sections below a recommendation statement.

Developers reported these forms of information with varying levels of directness, with 99 of 142 (69.7%) guidelines completing pre-defined EtD Likert-type judgments in the resources required (judging magnitude of costs), certainty of evidence of resources required (judging from very low to high), or cost-effectiveness (judging from favouring intervention or favouring comparison) domains. Of these, only 13 of 142 (9.2%) guidelines left the Likert-type judgments and all other sections of the EtD frameworks completely blank from economic information.

Research evidence and additional considerations

With respect to research evidence and additional considerations in the EtD framework (Table 2), 57 of 142 (40.1%) guidelines supplied descriptions of economic information and 22 of 142 (15.5%) explicitly reported that no economic information was identified. Additionally, of those guidelines that reported free text economic information, 9 of 57 (15.8%) specifically noted doing a new systematic review, while 23 of 57 (40.4%) presented economic study information; but, did not clarify if this information was from a systematic review or not. Others reported adopting existing systematic reviews in 10 of 57 (17.5%) guidelines, one reported developing a specific economic evaluation[100] and another noted that they adapted an existing evaluation.[59] Of the 17 of 59 (28.8%) guidelines that reported economic information in a qualitative format in the research evidence and additional considerations free text boxes, all reported performing a systematic review. Although, overall, 61 of 142 (43.0%) guidelines left these free text boxes in the EtD framework completely blank, which is not typical of the intended use.[158]

Some guideline groups reported information on unit costs; but not cost-effectiveness in 18 of 59 (30.5%) guideline EtD frameworks; similarly, others reported cost-effectiveness information, but not information on unit costs in 18 of 59 (30.5%) guidelines. Most commonly, unit cost boxes contained specific information related to interventions in 35 of 59 (59.3%) guidelines, which were itemized by procedure, testing, equipment, overhead, and provider factors or summarized by clumps of factors (e.g. cost per procedure, facility, patient, or GRADE outcome). For example, with respect to outcomes selected by the panel as critical to decision-making, in some venous thromboembolism (VTE) guidelines[76,77,79,81] they reported cost per major bleed or cost per VTE (i.e. deep vein thrombosis, pulmonary embolism) or as cost per false positive in breast cancer screening.[87] Only two guidelines[87,129] reported unit costs per GRADE outcome and performed a GRADE certainty appraisal.

Costs Effectiveness was most often reported using incremental cost effectiveness ratio (ICER), per quality adjusted life year (QALY), effects in 47 of 59 (79.7%) guidelines (e.g. life years gained or recovered, disability adjusted). Other guidelines reported these cost-effectiveness results as coming from specifically cost-utility,[50,116,133,137] cost-benefit,[103,154,155] or cost analysis[59,80,100,128,] evaluations. No guideline EtD framework reported on what inclusion or exclusion criteria they used for selecting economic evidence.

With respect to certainty of cost effectiveness outcomes, one guideline[87] reported a GRADE certainty appraisal, two guidelines[87,150] reported ICER certainty as a factor of best and worst case scenarios and another three guidelines described this through sensitivity.[50,118,128] With respect to length of follow up for economic evaluations, reported by nine guidelines, short,[118] long,[76] both

short and long-term,[50] and lifetime[59,81,87,120,128,150] time horizons for consequences were considered.

Remarks and justifications

In the remarks sections of EtD framework, used to justify recommendation statements (Table 2), 54 of 142 (38.0%) guidelines used economic factors as part of their rationale. Of these recommendation rationales, 22 of 54 (40.7%) guidelines described economic factors without any economic information reported in the EtD framework, aside from 15 of 22 (68.2%) having completed the cost and cost-effectiveness Likert-type judgments – 8 of 22 (36.4%) of these justifications usually related to phrasing about the low cost of an intervention or 14 of 22 (63.6%) related to a recommendation being of conditional strength based on cost-implication decisions.

Lastly, with respect to cost and cost effectiveness information used in the remarks section of the EtD framework to justify recommendation statements in 35 of 142 (24.6%) guidelines, there were five common economic justifications across our sample, which were: (1) 11 of 35 (31.4%) guidelines reported the decision for an option favours one cost-effective option over another; (2) 8 of 35 (22.9%) reported the decision against an option was based on high cost versus low effectiveness (i.e. not cost-effective); (3) 7 of 35 (20.0%) reported the incremental cost was very small and did not play into the decision; (4) 5 of 35 (14.3%) noted there was uncertainty about incremental costs (e.g. no research identified); or (5) 4 of 35 (11.4%) described that the incremental cost balance was similar between interventions. (Figure 1)

Themes of reported economic information

Across the guideline EtD framework domains of cost, cost-effectiveness, acceptability, feasibility, and equity, along with the remarks section used to justify recommendations, we identified 6 themes (Table 3) and 61 sub-themes: (Appendix 4)

Theme I – Intervention factors were most commonly reported across 98 of 142 (69.0%) guidelines and had 14 sub-themes, the most frequent of which were: price, access, frequency and effectiveness.

Theme II – Population factors were reported in 53 of 142 (37.3%) guidelines and had 15 sub-themes, the most frequent of which were: sub-group risk, country, setting, and adverse events. Particularly, in the domain of equity a few new population sub-themes of marginalized populations and caregiver burden were distinct.

Theme III – Payor factors were reported in 32 of 142 (22.5%) guidelines and had 13 sub-themes with coverage, program costs, budget considerations being most common. In the domain of acceptability a few new payor sub-themes of risk management, raising awareness and advocacy group interests were distinct.

Theme IV – Provider factors were also reported in 32 of 142 (22.5%) guidelines and had 7 sub-themes most often described as fees, training and standards of practice. The domain of equity, which usually applies to patients, would not have been expected to contain any sub-themes related to providers.

Theme V – Healthcare resource use factors were again reported in 35 of 142 (24.7%) guidelines and had 7 sub-themes most commonly being hospitalization costs, length of stay and equipment costs.

Theme VI – Economic model building factors were reported exclusively in the domains of cost and cost-effectiveness. The domains of acceptability, feasibility and equity, as well as the remarks section would not have been expected to contain any sub-themes related to model design considerations and limitations. A minority, with 22 of 142 (15.5%) guidelines, reported on these considerations and had 5 sub-themes of model perspective, time horizon, sensitivity, transferability and reference year.

Discussion

Brief Summary

The overall rate of reporting of at least some form of economic information, through Likert-type judgments or free text, was high (90.8%) across our sample of 142 GRADE guidelines; but, there was very large variability across Likert-type judgment completion (69.7%), economic free text summary information reported across EtD framework domains (57.1%) and remarks used to justify recommendations (38.0%). A recent 2019 review, considering 67 organization developer handbooks, which included some GRADE groups using the EtD framework, found a similar high rate (88.1%) of reporting, as well as variability in economic information use throughout their guideline development process.[159]

As the Likert-type judgments were only being used 69.7% of the time, there is room for improvement since it involves a very small amount of panel time and no guideline development cost to complete. However, completing these GRADE EtD framework judgments without supporting research evidence or additional considerations alone has been raised as a concern in WHO guidelines due to it potentially becoming a "check box exercise" rather than a fulsome consideration of domains beyond benefits and harms.[158] In our sample of WHO guidelines, 15 of 18 (83.3%) fully reported this information in the EtD framework, but, only 5 of 18 (27.7%) made reference to economic factors in their justification of their recommendation. Similar to the WHO, 24.6% of guidelines across our entire sample made at least one remark related to economic factors in their recommendation justifications.

Key Findings

It is not often possible to perform a full economic evaluation as part of clinical practice guidelines, primarily due to time and funding constraints - ideally this information would use the effect estimates, values and other evidence that is specific for the actual guideline, as opposed to using existing research summaries that often include different assumptions. Indeed, some guidelines included a full new systematic review, performing a 'de novo' specific economic evaluation, or did a re-analysis of previous work to adapt it to a guideline. For example, in our sample of GRADE guidelines, following the principles of GRADE-'adolopment',[160] 9 of 57 (15.8%) guidelines specifically reported doing a new systematic review, whereas 23 of 57 (40.4%) used adoption through a summary of previously published individual economic evaluations - many of these did not explicitly report that presented information was obtained

from a systematic review. Others fully adopted existing systematic reviews in 10 of 57 (17.5%) guidelines, one guideline reported developing a specific 'de novo' economic evaluation[100] and another reported that they adapted an existing evaluation by re-running the analysis using local context inputs.[59] This variability may be explained by a lack of clear economic evaluation systematic review methodology guidance – a review of over 200 recently published systematic reviews of health economic evaluations reported challenges with systematic review methodological quality, assessment of transferability and synthesis of quantitative economic information.[161]

With only two guidelines[87,129] performing a formal GRADE certainty appraisal, this dearth of reporting may be due to a lack of clear guidance in applying the traditional five GRADE domains to an economic outcome or comparing more than two alternatives. To bridge this gap, some have proposed modified assessment criteria, simplified into three domains of economic information certainty: (1) transferability to the decision context (similar to directness in GRADE); (2) model limitations (e.g. model inputs, model design, time horizon, reference year, perspective, sensitivity) of the economic evaluation, and, if applicable, (3) consistency of findings across multiple cost-effectiveness models.[162] In our sample, economic model design was considered and reported in 23 of 142 (16.2%) guidelines and was concordant with these assessment criteria. The sub-theme of transferability was only considered in one guideline[61] and improved reporting of these details appears important for GRADE guideline developers and users to consider.

In recent years there has been a push for simpler language to better reflect the spectrum of developers, readers and users of GRADE guidelines, who may not have extensive epidemiology training. While often considered a minimum standard of reporting when information may be

lacking,[163] in our sample of guidelines that reported economic information, 17 of 59 (28.8%) used purely qualitative summaries of their systematic review. The remainder used either quantitative or a combination of both to report identified economic information. It is not known if guideline panels and users find qualitative statements more useful to decision-making over quantitative ones; but, recent GRADE guidance suggests a preference for simpler statements that still retain aspects of both magnitude and certainty of estimates.[164]

We identified economic information variability themes across EtD framework domains and the remarks section used to justify recommendations that related to interventions, populations, payors, providers, healthcare resource use and model building. A recent overview of 36 systematic reviews of decision-making criteria reported most commonly cited elements across reviews that were concordant with the themes in our sample of GRADE guidelines – those being aspects of unit cost, relation of cost to benefits, budget impact, burden of disease, and affordability of intervention population sub-group risk, payor coverage, human resources and infrastructure; two additional decision factors from their overview of reviews, cost-minimization and innovation were less represented in our sample of guidelines.[165] Cost-minimization, considered, as high value by a guideline panel, was reported in six guidelines[28,29,46,60,78,141] to justify recommendations. Considering these additional decision factors more explicitly moving forward could potentially be useful for GRADE developers and panel members to foster proactive advances in clinical practice. As an example, from the American Society of Hematology guideline on Diagnosis of Venousthromboembolism,[78] the panel used the following remarks:

> <u>Panel</u>: *"The panel considered a strategy with D-dimer testing first to reduce cost, ensure feasibility, and reduce radiation exposure."*

A 2018 systematic review of 19 coverage decision frameworks, used when potentially paying for new expensive healthcare interventions, has suggested modifications to the GRADE EtD framework by adding the consideration of limitations of the alternative strategies in use (as an elaboration of benefits and harms), impact of efficiency (opportunity cost of moving from one efficient strategy combination to another) and mitigating inappropriate use.[166] In our sample, only three guidelines[52,59,67] used opportunity cost to justify their recommendation and one considered limitations of alternative strategies already in use - the Canadian National Pain Centre, Guideline for Opioid Therapy and Chronic Non-Cancer Pain,[74] reported the following remarks:

> Panel: *"In recognition of the cost of formal multidisciplinary opioid reduction programs and their current limited availability/capacity, an alternative is a coordinated multidisciplinary collaboration that includes several health professionals whom physicians can access according to their availability."*

In summary, the degree to which economic factors should be weighed into a particular recommendation decision remains unclear, but it should be a requirement to have research evidence or additional consideration information to make an informed judgment. Therefore, either linking the Likert-type and free text justification decisions together with reported economic information to culminate into a recommendation justification remark (Figure 1), or prioritizing guideline questions that most warrant fulsome economic information for decision-making[167] and otherwise providing an explicit statement that economic factors were not considered in decision-making, may be a useful guideline developer and panelist target to consider moving forward when completing EtD frameworks.

Strengths and limitations

Strengths of our review include broad eligibility criteria, an additional scan of related websites search, and the resulting large sample of GRADE guidelines and associated EtD frameworks. We also considered guidelines in all languages.

We used thematic analysis to summarize important themes related to free text information reported, along with duplicate checking of all abstracted information.

Our study also has a number of limitations. While we identified a large number of guidelines representing over 1600 frameworks, our convenience sample may not be representative because it was identified by GRADE working group members who may or may not have had an interest in economic evaluations. However, we systematically searched websites for additional guidelines to overcome this possible identification bias. Also, a very large of number of GRADE guidelines have been developed in recent years that lack access to published EtD frameworks. A recent review[168] identified 98 guidelines reporting use of GRADE methodology and we included only 9 in our convenience sample that had published their EtD frameworks. Also, we used Google Translate to convert 35 of 142 (24.7%) guidelines from various languages - Danish (12), Norwegian (12), Spanish (9), Italian (1), and Japanese (1) using Google Translate; this could have introduced some language translation errors, but none were obvious throughout our review.

While we did a comprehensive review of the published EtD frameworks we identified, we did not review other associated published guideline manuscripts or documents to identify if any new economic information was added or changed. For instance, the panel may have considered

economic evidence but did not report it as part of recommendation justification in the EtD framework, or elsewhere, and later added more context in the published manuscript. As well, no guideline in our sample reported on what inclusion and exclusion criteria they used for selecting economic evidence, although this may been reported elsewhere.

We made the assumption that, if in a particular guideline, they reported economic information in at least one EtD framework that there would have been equal opportunity for this reporting to occur across the entire guideline. It is possible there was variation in the developers and panelists across EtD frameworks and this was not the case; although, we also only counted this reporting instance once per guideline, if it occurred across multiple EtD frameworks in a guideline.

Implications for practice

Guideline developers that use the EtD framework should make efforts always complete the pre-defined Likert-type judgments and populate the research evidence and additional considerations free text sections related to resource use, certainty of evidence of resource use, and cost-effectiveness. As parts of the guideline development process can become fragmented between manuscripts, EtD frameworks and panel meeting deliberation transcripts over the time span of a project, it is advisable to offer some level of clarity in the EtD framework with respect how economic evidence was identified (e.g. inclusion/exclusion criteria) and any particular assumptions the guideline development team may have made with respect to choosing economic models or cost inputs, as well as a brief description of the systematic review methodology employed.

Adding these features will improve the transparency of the reporting within the EtD framework and better facilitate their use be other guideline groups attempting to development similar guideline recommendations based on these research questions.

Implications for research

Future research in this area could address panel preferences towards the best phrasing and wording format to present economic summary information and reflect the associated estimates. As well, understanding the threshold of directness (transferability) of available published economic information to a particular guideline question and informing a subsequent recommendation is still largely unknown.

Little is known about the weight that economic information plays in particular decision-making contexts; identifying when or when not to consider this source of evidence would reduce both necessary guideline developer costs and panelist time.

For GRADE guideline methodology, a next research step would be to further refine and identify best reporting model limitation factors to consider, including those from economics, that would be useful for clinical practice guideline panels to assess certainty.

Conclusion

The large majority of GRADE guideline developers report economic evidence and many panels justify their recommendation with this information. Although, reporting of economic information was quite variable in their level of detail and directness. Completing the EtD pre-defined Likert-

type judgments, specifically reporting the literature review approach, study selection criteria and economic model building limitations, as well as linking this information to recommendation justification remarks are potential areas for improved reusability and transparency of GRADE Evidence-to-Decision frameworks.

Acknowledgements

We would like to thank GRADE working members Ainsley Moore, Bram Rochwerg, Claire Glenton, David Rigai Comas, Eddy Lang, Ein-Soon Shin, Gihad Nesrallah, Itziar Etxeandia, Juan Pablo Peña-Rosas, Lorenzo Moja, Luciane Lopes, Maria Ximena Rojas Reyes, Melanie Marti, Morio Aihara, Paul Elias Alexander and Valerie Lavergny for responding to our e-mail request with suggestions on websites with published Evidence-to-Decision frameworks.

Authorship contributions

Conception and design: JJR, JLB

Data acquisition: JJR, MB, DJB, CCM

Data analysis: JJR, FX, JWB, JLB

Interpretation of results: JJR, MB, JWB, FX, HJS, JLB

Manuscript drafting: JJR, JWB, JLB

Critical revision of the manuscript and approval of the final version: JJR, MB, DJB, JWB, CCM, FX, HJS, JLB

Disclosure of Conflicts of Interest

This study was unfunded. JJR is supported by a PhD training award from the NCMIC foundation (https://www.ncmicfoundation.org/). The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Disclosure of Competing Interests

JJR is a Member of the Canadian Task Force on Preventive Health Care (https://canadiantaskforce.ca/) and the GRADE working group (https://www.gradeworkinggroup.org/). HJS is co-chair of the GRADE Working Group. These groups had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

References

1. Schünemann HJ, Hill SR, Kakad M, Vist GE, Bellamy R, Stockman L, Wisløff TF, Del Mar C, Hayden F, Uyeki TM, Farrar J, Yazdanpanah Y, Zucker H, Beigel J, Chotpitayasunondh T, Hien TT, Ozbay B, Sugaya N, Oxman AD. Transparent development of the WHO rapid advice guidelines. PLoS Med. 2007 May;4(5):e119. doi: 10.1371/journal.pmed.0040119.
2. Guyatt GH, Oxman AD, Kunz R, Falck-Ytter Y, Vist GE, Liberati A, Schünemann HJ; GRADE Working Group. Going from evidence to recommendations. BMJ. 2008 May 10;336(7652): 1049-51.
3. Moberg J, Oxman AD, Rosenbaum S, Schunemann HJ, Guyatt G, Flottorp S, et al. The GRADE Evidence to Decision (EtD) framework for health system and public health decisions. Health Res Policy Syst. 2018;16(1):45.
4. Parmelli E, Amato L, Oxman AD, Alonso-Coello P, Brunetti M, Moberg J, et al. GRADE EVIDENCE TO DECISION (EtD) FRAMEWORK FOR COVERAGE DECISIONS. Int J Technol Assess Health Care. 2017:1-7.
5. Schunemann HJ, Mustafa R, Brozek J, Santesso N, Alonso-Coello P, Guyatt G, et al. GRADE Guidelines: 16. GRADE evidence to decision frameworks for tests in clinical practice and public health. J Clin Epidemiol. 2016;76:89-98.
6. Neumann I, Brignardello-Petersen R, Wiercioch W, Carrasco-Labra A, Cuello C, Akl E, et al. The GRADE evidence-to-decision framework: a report of its testing and application in 15 international guideline panels. Implement Sci. 2016;11(1):93.
7. Alonso-Coello P, Schunemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent

approach to making well informed healthcare choices. 1: Introduction. BMJ. 2016;353:i2016.

8. Alonso-Coello P, Oxman AD, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: Clinical practice guidelines. BMJ. 2016;353:i2089.
9. Fischer KE. A systematic review of coverage decision-making on health technologies-evidence from the real world. Health Policy. 2012 Oct;107(2-3):218-30.
10. Tromp N, Baltussen R. Mapping of multiple criteria for priority setting of health interventions: an aid for decision makers. BMC Health Serv Res. 2012 Dec 13;12:454.
11. Guindo LA, Wagner M, Baltussen R, Rindress D, van Til J, Kind P, Goetghebeur MM. From efficacy to equity: Literature review of decision criteria for resource allocation and healthcare decision making. Cost Eff Resour Alloc. 2012 Jul 18;10(1):9.
12. Rosenbaum SE, Moberg J, Glenton C, Schünemann HJ, Lewin S, Akl E, Mustafa RA, Morelli A, Vogel JP, Alonso-Coello P, Rada G, Vásquez J, Parmelli E, Gülmezoglu AM, Flottorp SA, Oxman AD. Developing Evidence to Decision Frameworks and an Interactive Evidence to Decision Tool for Making and Using Decisions and Recommendations in Health Care. Glob Chall. 2018 Jan 10;2(9):1700081.
13. Brunetti M, Shemilt I, Pregno S, Vale L, Oxman AD, Lord J, Sisk J, Ruiz F, Hill S, Guyatt GH, Jaeschke R, Helfand M, Harbour R, Davoli M, Amato L, Liberati A, Schünemann HJ. GRADE guidelines: 10. Considering resource use and rating the quality of economic evidence. J Clin Epidemiol. 2013 Feb;66(2):140-50.

14. Green J, Willis K, Hughes E, Small R, Welch N, Gibbs L, Daly J. Generating best evidence from qualitative research: the role of data analysis. Aust N Z J Public Health. 2007;31:545-50.
15. World Health Organization (WHO). Latent Tuberculosis Infection: Updated and Consolidated Guidelines for Programmatic Management. 2018. Available at: < https://www.who.int/tb/publications/2018/latent-tuberculosis-infection/en/ >
16. World Health Organization (WHO). Bacille Calmette-Guérin (BCG) Vaccination. 2018. Available at: < https://www.who.int/immunization/documents/positionpapers/en/ >
17. World Health Organization (WHO). Cholera Vaccination. 2018. Available at: < https://www.who.int/immunization/documents/positionpapers/en/ >
18. World Health Organization (WHO). Rabies Vaccination. 2018. Available at: < https://www.who.int/immunization/documents/positionpapers/en/ >
19. World Health Organization (WHO). Dengue Vaccination. 2017. Available: < https://www.who.int/immunization/documents/positionpapers/en/ >
20. World Health Organization (WHO). Diphtheria Vaccination. 2017. Available at: < https://www.who.int/immunization/documents/positionpapers/en/ >
21. World Health Organization (WHO). Hepatitis B Vaccination. 2017. Available at: < https://www.who.int/immunization/documents/positionpapers/en/ >
22. World Health Organization (WHO). Human Papillomavirus (HPV) Vaccination. 2017. Available at: < https://www.who.int/immunization/documents/positionpapers/en/ >
23. World Health Organization (WHO). Treatment of Drug-susceptible Tuberculosis and Patient Care (2017 update). 2017. Available at: < https://www.who.int/tb/publications/2017/dstb_guidance_2017/en/ >

24. World Health Organization (WHO). Measles Vaccination. 2017. Available at: < https://www.who.int/immunization/documents/positionpapers/en/ >

25. World Health Organization (WHO). Tetanus Vaccination. 2017. Available at: < https://www.who.int/immunization/documents/positionpapers/en/ >

26. World Health Organization (WHO). Fractional Dose Yellow Fever Vaccination. 2017. Available at: < https://www.who.int/immunization/documents/positionpapers/en/ >

27. World Health Organization (WHO). Malaria Vaccination. 2016. Available at: < https://www.who.int/immunization/documents/positionpapers/en/ >

28. World Health Organization (WHO). Sexually Transmitted Infection: Chlamydia Trachomatis. 2016. Available at: < https://www.who.int/reproductivehealth/publications/rtis/chlamydia-treatment-guidelines/en/>

29. World Health Organization (WHO). Sexually Transmitted Infection: Syphilis. 2016. Available at: < https://www.who.int/reproductivehealth/publications/rtis/policy/en/ >

30. World Health Organization (WHO). Treatment of Drug-resistant Tuberculosis. 2016. Available at: < https://apps.who.int/iris/bitstream/handle/10665/250125/9789241549639-eng.pdf >

31. World Health Organization (WHO). Pertussis Vaccination. 2015. Available at: < https://www.who.int/immunization/documents/positionpapers/en/ >

32. World Health Organization (WHO). Systematic Screening for Active Tuberculosis. 2013. Available at: < https://www.who.int/tb/tbscreening/en/ >

33. BMJ WikiRecs. QiukuiClopidogrel plus aspirin versus aspirin or clopidogrel alone for the treatment of acute minor ischemic stroke or high-risk TIA. 2018. Available at: < https://app.magicapp.org/ >

34. BMJ WikiRecs. Primary Care Rapid Rec Alpha blockers for Treatment of Ureteric Stones in Primary Care. 2018. Available at: < https://app.magicapp.org/ >
35. BMJ WikiRecs. Oxygen for Acutely ill Patients. 2018. Available at: < https://app.magicapp.org/ >
36. BMJ WikiRecs. Prostate-Specific Antigen (PSA) Screening in men Without Symptoms of Prostate Cancer. 2018. Available at: < https://app.magicapp.org/ >
37. BMJ WikiRecs. Corticosteroids for Sepsis. 2018. Available at: < https://app.magicapp.org/ >
38. BMJ WikiRecs. Patent foramen ovale Closure, Antiplatelet or Anticoagulation Therapy for Management of Cryptogenic Stroke. 2018. Available at: < https://app.magicapp.org/ >
39. BMJ WikiRecs. Atraumatic versus Conventional Lumbar Puncture Needles. 2018. Available at: < https://app.magicapp.org/ >
40. BMJ WikiRecs. Antibiotics for Uncomplicated Skin Abscesses. 2018. Available at: < https://app.magicapp.org/ >
41. BMJ WikiRecs. Corticosteroids for Sore Throat. 2018. Available at: < https://app.magicapp.org/ >
42. BMJ WikiRecs. Antiretroviral therapy for pregnant women living with HIV. 2017. Available at: < https://app.magicapp.org/ >
43. BMJ WikiRecs. Transcatheter aortic valve implantation versus surgical aortic valve replacement for severe aortic stenosis in low-intermediate risk patients. 2017. Available at: < https://app.magicapp.org/ >
44. BMJ WikiRecs. Vasopressors - Blood pressure Targets. 2017. Available at: < https://app.magicapp.org/ >

45. BMJ WikiRecs. Arthroscopic Surgery for Degenerative Knee Disease. 2017. Available at: < https://app.magicapp.org/ >

46. BMJ WikiRecs. Low Intensity Pulsed Ultrasound for Bone Healing. 2017. Available at: < https://app.magicapp.org/ >

47. BMJ WikiRecs. Probiotics for Children Receiving Antibiotics. 2016. Available at: < https://app.magicapp.org/ >

48. BMJ WikiRecs. Adjunctive Corticosteroid Therapy for Adults Hospitalized with Community-acquired Pneumonia. 2015. Available at: < https://app.magicapp.org/ >

49. European Forum for Research and Education in Allergy and Airway Diseases. Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines-2016 revision. 2017. Available at: < https://www.euforea.eu/aria >

50. World Allergy Organization (WAO). World Allergy Organization-McMaster University Guidelines for Allergic Disease Prevention (GLAD-P): Prebiotics. 2016. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/26962387 >

51. World Allergy Organization (WAO). World Allergy Organization-McMaster University Guidelines for Allergic Disease Prevention (GLAD-P): Vitamin D. 2016. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/27274360 >

52. World Allergy Organization (WAO). World Allergy Organization-McMaster University Guidelines for Allergic Disease Prevention (GLAD-P): Probiotics. 2015. Available at: < https://www.ncbi.nlm.nih.gov/pubmed/26998191 >

53. Canadian Task Force on Preventive Health Care (CTFPHC). Screening for Asymptomatic Bacteriuria in Pregnancy. 2018, Canada. Available at: < https://canadiantaskforce.ca/ >

54. Canadian Task Force on Preventive Health Care (CTFPHC). Screening for Impaired Vision. 2018, Canada. Available at: < https://canadiantaskforce.ca/ >

55. Canadian Task Force on Preventive Health Care (CTFPHC). Screening for Breast Cancer – update. 2018, Canada. Available at: < https://canadiantaskforce.ca/ >
56. Canadian Task Force on Preventive Health Care (CTFPHC). Screening for Abdominal Aortic Aneurysm. 2017, Canada. Available at: < https://canadiantaskforce.ca/ >
57. Canadian Task Force on Preventive Health Care (CTFPHC). Prevention of Tobacco Smoking in Children and Adults. 2017, Canada. Available at: < https://canadiantaskforce.ca/ >
58. Canadian Task Force on Preventive Health Care (CTFPHC). Screening for Hepatitis C. 2017, Canada. Available at: < https://canadiantaskforce.ca/ >
59. Canadian Task Force on Preventive Health Care (CTFPHC). Screening for Colorectal Cancer. 2016, Canada. Available at: < https://canadiantaskforce.ca/ >
60. Canadian Task Force on Preventive Health Care (CTFPHC). Screening for Developmental Delay. 2016, Canada. Available at: < https://canadiantaskforce.ca/ >
61. Canadian Task Force on Preventive Health Care (CTFPHC). Screening for Lung Cancer. 2016, Canada. Available at: < https://canadiantaskforce.ca/ >
62. Canadian Task Force on Preventive Health Care (CTFPHC). Pelvic Exam. 2016, Canada. Available at: < https://canadiantaskforce.ca/ >
63. Canadian Task Force on Preventive Health Care (CTFPHC). Screening for Cognitive Impairment. 2015, Canada. Available at: < https://canadiantaskforce.ca/ >
64. Canadian Task Force on Preventive Health Care (CTFPHC). Prevention of Obesity in Adults. 2015, Canada. Available at: < https://canadiantaskforce.ca/ >
65. Canadian Task Force on Preventive Health Care (CTFPHC). Prevention of Obesity in Children. 2015, Canada. Available at: < https://canadiantaskforce.ca/ >

66. Canadian Task Force on Preventive Health Care (CTFPHC). Screening for Prostate Cancer. 2014, Canada. Available at: < https://canadiantaskforce.ca/ >

67. Canadian Task Force on Preventive Health Care (CTFPHC). Screening for Depression. 2013, Canada. Available at: < https://canadiantaskforce.ca/ >

68. Canadian Task Force on Preventive Health Care (CTFPHC). Screening for Cervical Cancer. 2013, Canada. Available at: < https://canadiantaskforce.ca/ >

69. Canadian Task Force on Preventive Health Care (CTFPHC). Screening for Type 2 Diabetes. 2012, Canada. Available at: < https://canadiantaskforce.ca/ >

70. Canadian Task Force on Preventive Health Care (CTFPHC). Screening for Hypertension. 2012, Canada. < https://canadiantaskforce.ca/ >

71. Canadian Chiropractic Guideline Initiative (CCGI). Spinal Manipulative Therapy and Other Conservative Treatments for Low Back Pain. 2017, Canada. Available at: < https://www.chiropractic.ca/guidelines-best-practice/ >

72. Canadian Chiropractic Guideline Initiative (CCGI). Management of Neck Pain Associated Disorders. 2016, Canada. Available at: < https://www.chiropractic.ca/guidelines-best-practice/ >

73. Canadian Chiropractic Guideline Initiative (CCGI). Management of Whiplash Associated Disorder. 2016, Canada. Available at: < https://www.chiropractic.ca/guidelines-best-practice/>

74. Busse JW, Craigie S, Juurlink DN, Buckley DN, Wang L, Couban RJ, Agoritsas T, Akl EA, Carrasco-Labra A, Cooper L, Cull C, da Costa BR, Frank JW, Grant G, Iorio A, Persaud N, Stern S, Tugwell P, Vandvik PO, Guyatt GH. Guideline for opioid therapy and chronic noncancer pain. CMAJ. 2017 May 8;189(18):E659-E666.

75. Canadian Blood Services. Controlled Pediatric Donation after Circulatory Determination of Death (pDCD). 2017, Canada. Available at: < https://www.blood.ca/en , http://gdt.gradepro.org >

76. Schünemann HJ, Cushman M, Burnett AE, Kahn SR, Beyer-Westendorf J, Spencer FA, Rezende SM, Zakai NA, Bauer KA, Dentali F, Lansing J, Balduzzi S, Darzi A, Morgano GP, Neumann I, Nieuwlaat R, Yepes-Nuñez JJ, Zhang Y, Wiercioch W. American Society of Hematology 2018 guidelines for management of venous thromboembolism: prophylaxis for hospitalized and nonhospitalized medical patients. Blood Adv 2018, 2(22): 3198-3225.

77. Witt DM, Nieuwlaat R, Clark NP, Ansell J, Holbrook A, Skov J, Shehab N, Mock J, Myers T, Dentali F, Crowther MA, Agarwal A, Bhatt M, Khatib R, Riva JJ, Zhang Y, Guyatt G. American Society of Hematology 2018 guidelines for management of venous thromboembolism: optimal management of anticoagulation therapy. Blood Adv 2018, 2(22):3257-3291.

78. Lim W, Le Gal G, Bates SM, Righini M, Haramati LB, Lang E, Kline JA, Chasteen S, Snyder M, Patel P, Bhatt M, Patel P, Braun C, Begum H, Wiercioch W, Schünemann HJ, Mustafa RA. American Society of Hematology 2018 guidelines for management of venous thromboembolism: diagnosis of venous thromboembolism. Blood Adv 2018, 27(22): 3226-3256.

79. Monagle P, Cuello CA, Augustine C, Bonduel M, Brandão LR, Capman T, Chan AKC, Hanson S, Male C, Meerpohl J, Newall F, O'Brien SH, Raffini L, van Ommen H, Wiernikowski J, Williams S, Bhatt M, Riva JJ, Roldan Y, Schwab N, Mustafa RA, Vesely SK. American Society of Hematology 2018 Guidelines for management of venous thromboembolism: treatment of pediatric venous thromboembolism. Blood Adv 2018, 2(22):3292-3316.

80. Cuker A, Arepally GM, Chong BH, Cines DB, Greinacher A, Gruel Y, Linkins LA, Rodner SB, Selleng S, Warkentin TE, Wex A, Mustafa RA, Morgan RL, Santesso N. American Society of Hematology 2018 guidelines for management of venous thromboembolism: heparin-induced thrombocytopenia. Blood Adv 2018, 2(22): 3360-3392.
81. Bates SM, Rajasekhar A, Middeldorp S, McLintock C, Rodger MA, James AH, Vazquez SR, Greer IA, Riva JJ, Bhatt M, Schwab N, Barrett D, LaHaye A, Rochwerg B. American Society of Hematology 2018 guidelines for management of venous thromboembolism: venous thromboembolism in the context of pregnancy. Blood Adv 2018, 2(22): 3317-3359.
82. American Thoracic Society (ATS). Treatment of Idiopathic Pulmonary Fibrosis. 2018, United States. Available at: < http://gdt.gradepro.org >
83. American Thoracic Society (ATS). Diagnosis of Idiopathic Pulmonary Fibrosis. 2018, United States. Available at: < https://www.thoracic.org >
84. American Thoracic Society (ATS). Diagnosis of Primary Ciliary Dyskinesia: An Official ATS Clinical Practice Guideline. 2018, United States. Available at: < https://www.thoracic.org >
85. Society of Critical Care Medicine (SCCM); ICU Liberation. Management of Pain, Agitation and Delirium in Adults in the Intensive Care Unit (PAD-ES). 2018, United States. Available at: < http://gdt.gradepro.org >
86. National Highway Traffic Safety Administration (NHTSA). Evidence-Based Guidelines For Fatigue Risk Management In Emergency Medical Services. 2018, United States. Available at: < https://www.nhtsa.gov/ >

87. European Commission Initiative on Breast Cancer. Recommendations from European Breast Guidelines. 2018, European Union (EU). Available at: < https://ecibc.jrc.ec.europa.eu/recommendations/ >
88. Nationalt akutkirurgisk tvaerfagligt forum (National emergency surgical interdisciplinary forum). National Clinical Guideline (NCR) for Spinal Stabilization of Adult Trauma Patients in Denmark, 2018, Denmark. Available at: < https://app.magicapp.org/ >
89. Sundhedsstyrelsen (Health). National Clinical Guideline for Dementia and Medicine. 2018, Denmark. Available at: < https://app.magicapp.org/ >
90. Sundhedsstyrelsen (Health). National Clinical Guideline for the Prevention and Treatment of Behavioral and Mental Symptoms in People with Dementia. 2018, Denmark. Available at: < https://app.magicapp.org/ >
91. Sundhedsstyrelsen (Health). National Clinical Guideline for Treatment of Alcohol Addiction. 2018, Denmark. Available at: < https://app.magicapp.org/ >
92. Sundhedsstyrelsen (Health). National Clinical Guideline for Opioid Treatment of Chronic Non-Malignant Pain. 2018, Denmark. Available at: < https://app.magicapp.org/ >
93. Sundhedsstyrelsen (Health). National clinical Guideline for the study and Treatment of ADHD in Children and Adolescents. 2018, Denmark. Available at: < https://app.magicapp.org/ >
94. Sundhedsstyrelsen (Health). National clinical Guideline for the Diagnosis of mild Cognitive Impairment and Dementia. 2018, Denmark. Available at: < https://app.magicapp.org/ >
95. Sundhedsstyrelsen (Health). National Clinical Guideline for the Treatment of Emotionally Unstable Personality Structure, Borderline Type. 2017, Denmark. Available at: < https://app.magicapp.org/ >

96. Sundhedsstyrelsen (Health). National clinical Guideline for Post-operative effects of early Breast Cancer. 2017, Denmark. Available at: < https://app.magicapp.org/ >
97. Sundhedsstyrelsen (Health). National Clinical Guideline for Treatment of Brain Metastases. 2017, Denmark. Available at: < https://app.magicapp.org/ >
98. Sundhedsstyrelsen (Health). National clinical Guideline for Non-pharmacological Treatment in Children and Adolescents with Asthma. Denmark, 2017. Available at: < https://app.magicapp.org/ >
99. Aarhus Universitetshospital (Aarhus University Hospital). The Effect of Treatment Methods Within Gambling Addiction. 2016, Denmark. Available at: < https://app.magicapp.org/ >
100. Helsedirektoratet (Directorate of Health). National Academic Guideline for Dental Health Services for Children and Adolescents 0–20 years. 2018, Norway. Available at: < https://app.magicapp.org/ >
101. Helsedirektoratet (Directorate of Health). National Professional Guideline for Diabetes. 2018, Norway. Available at: < https://app.magicapp.org/ >
102. Helsedirektoratet (Directorate of Health). National Academic Guideline for Gestational Diabetes. 2018, Norway. Available at: < https://app.magicapp.org/ >
103. Helsedirektoratet (Directorate of Health). National Academic Guideline for Prevention of Cardiovascular Disease. 2018, Norway. Available at: < https://app.magicapp.org/ >
104. Helsedirektoratet (Directorate of Health). National Academic Guideline for Treatment and Rehabilitation by Stroke. 2017, Norway. Available at: < https://app.magicapp.org/ >
105. Norsk Selskap for Trombose og Hemostase (Norwegian Company for Thrombosis and Hemostasis). Guidelines for Antithrombotic Treatment and Prophylaxis. 2016, Norway. Available at: < https://app.magicapp.org/ >

106. Norsk Selskap for Trombose og Hemostase (Norwegian Company for Thrombosis and Hemostasis). Prevention of VTE in Orthopedic Surgery Patients: A Norwegian adaptation of the 9th ed. of the ACCP Antithrombotic Therapy and Prevention of Thrombosis. 2015, Norway. Available at: < https://app.magicapp.org/ >
107. Norsk Selskap for Trombose og Hemostase (Norwegian Company for Thrombosis and Hemostasis). VTE, Thrombophilia, Antithrombotic Therapy and Pregnancy: A Norwegian adaptation of the 9th ed. of the ACCP Antithrombotic Therapy and Prevention. 2015, Norway. Available at: < https://app.magicapp.org/ >
108. Sykehuset Innlandet (Hospital Inland). Ataxia Telangiectasia. 2017, Norway. Available at: < https://app.magicapp.org/ >
109. Sykehuset Innlandet HF, Avdeling for Fysikalsk medisin og rehabilitering, Ottestad (Hospital Innlandet HF, Department of Physical Medicine and Rehabilitation, Ottestad). Knowledge-based Professional Guideline for Rehabilitation After Acquired Upper Limb Amputation in Norway. 2016, Norway. Available at: < https://app.magicapp.org/ >
110. Norsk Ortopdisk Foren (Norwegian Orthopaedic Association). Treatment of Distal Radius Fractures in Adults. 2015, Norway. Available at: < https://app.magicapp.org/ >
111. Norsk Ortopdisk Foren (Norwegian Orthopaedic Association). Adult Wrist Fracture Treatment. 2015, Norway. Available at: < https://app.magicapp.org/ >
112. Duodecim Medical Publications Ltd. Topical NSAIDs for Chronic Musculoskeletal Pain in Adults. 2017, Finland. Available at: < https://app.magicapp.org/ >
113. EBMeDS® Clinical Decision Support. Antiviral Treatment for Preventing Postherpetic Neuralgia. 2015, Finland. Available at: < https://app.magicapp.org/ >

114. The Scandinavian Society of Anaesthesiology and Intensive Care Medicine (SSAI). Fluid in Resuscitation of Critically ill Patients with Acute Circulatory Failure. 2016, Scandinavia. Available at: < https://app.magicapp.org/ >

115. The Scandinavian Society of Anaesthesiology and Intensive Care Medicine (SSAI). Fluid and Drug Therapy in Adults with Acute Respiratory Syndrome. 2016, Scandinavia. Available at: < https://app.magicapp.org/ >

116. National Prevention Plan Vaccinal (PNPV) 2012-2014. Decision-Making Model For the Adoption of Antivaricella Vaccine. 2016, Italy. Available at: < http://www.recentiprogressi.it/allegati/02152_2016_02/fulltext/06_Articolo%20originale%20-%20Gonzalez.pdf >

117. Kingdom of Saudi Arabia Ministry of Health. Diagnosis of First Suspected Deep Vein Thrombosis of Lower Extremity. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >

118. Kingdom of Saudi Arabia Ministry of Health. Venous Thromboembolism Treatment. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >

119. Kingdom of Saudi Arabia Ministry of Health. Prevention of Venous Thromboembolism with Stroke. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >

120. Kingdom of Saudi Arabia Ministry of Health. Use of Thrombolytic Therapy in Acute Stroke. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >

121. Kingdom of Saudi Arabia Ministry of Health. Antithrombotic Treatment for Non-valvular Atrial Fibrillation. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >

122. Kingdom of Saudi Arabia Ministry of Health. Use of Screening Strategies for Detection of Breast Cancer. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >
123. Kingdom of Saudi Arabia Ministry of Health. Screening and Treatment of Precancerous Lesions for Cervical Cancer Prevention. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >
124. Kingdom of Saudi Arabia Ministry of Health. Role of Vitamin D, Calcium and Exercise in Fracture Prevention in Elderly. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >
125. Kingdom of Saudi Arabia Ministry of Health. Timing of Initiation of Dialysis. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >
126. Kingdom of Saudi Arabia Ministry of Health. Allergic Rhinitis in Asthma. 2015, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >
127. Kingdom of Saudi Arabia Ministry of Health. Management of ST-elevation myocardial infarction (STEMI). 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >
128. Kingdom of Saudi Arabia Ministry of Health. Management of Breast Lump and Primary Breast Cancer. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >
129. Kingdom of Saudi Arabia Ministry of Health. Management of Eclampsia. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >
130. Kingdom of Saudi Arabia Ministry of Health. Management of Overweight and Obese Adults. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >

131. Kingdom of Saudi Arabia Ministry of Health. Management of Pre-Eclampsia. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >

132. Kingdom of Saudi Arabia Ministry of Health. Management of Sickle Cell Disease. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >

133. Kingdom of Saudi Arabia Ministry of Health. Management of Thalassemia – Iron chelation therapy, Bisphosphonates and Zinc supplementation. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >

134. Kingdom of Saudi Arabia Ministry of Health. Migraine Headache: Diagnosis & Management. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >

135. Kingdom of Saudi Arabia Ministry of Health. Prevention of VTE in Surgical Patients. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >

136. Kingdom of Saudi Arabia Ministry of Health. Prophylaxis of VTE in Medical Patients and Long Distance Travelers. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >

137. Kingdom of Saudi Arabia Ministry of Health. Screening for Colorectal Cancer. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >

138. Kingdom of Saudi Arabia Ministry of Health. Screening for Hypertension. 2014, Saudi Arabia. Available at: < https://www.moh.gov.sa/en/Pages/default.aspx >

139. Ministerio de Salud, Chile (Chile Ministry of Health). Conservative Non-Dialytic Treatment of Chronic Kidney Disease. 2018, Chile. Available at: < https://diprece.minsal.cl/temas-de-salud/temas-de-salud/guias-clinicas-no-ges/ >

140. Ministerio de Salud, Chile (Chile Ministry of Health). Prevention and Treatment of Dental Caries in boys and girls with Primary Dentition. 2018, Chile. Available at: < https://diprece.minsal.cl/temas-de-salud/temas-de-salud/guias-clinicas-no-ges/ >

141. Ministerio de Salud, Chile (Chile Ministry of Health). Anxiety Disorder Clinic Guide. 2018, Chile. Available at: < https://diprece.minsal.cl/temas-de-salud/temas-de-salud/guias-clinicas-no-ges/ >

142. Ministerio de Salud, Chile (Chile Ministry of Health). Technical aids for Mobility and Tissue Preservation for People Aged 15 to 64. 2017, Chile. Available at: < https://diprece.minsal.cl/temas-de-salud/temas-de-salud/guias-clinicas-no-ges/ >

143. Ministerio de Salud, Chile (Chile Ministry of Health). Oral Health in Adolescents 10-19 years: Prevention, Diagnosis and Treatment of Periodontal Disease. 2017, Chile. Available at: < https://diprece.minsal.cl/temas-de-salud/temas-de-salud/guias-clinicas-no-ges/ >

144. Ministerio de Salud, Chile (Chile Ministry of Health). Meningococcal Disease. 2017, Chile. Available at: < https://diprece.minsal.cl/temas-de-salud/temas-de-salud/guias-clinicas-no-ges/ >

145. Ministerio de Salud, Chile (Chile Ministry of Health). Influenza. 2017, Chile. Available at: < https://diprece.minsal.cl/temas-de-salud/temas-de-salud/guias-clinicas-no-ges/ >

146. Ministerio de Salud, Chile (Chile Ministry of Health). Outpatient use of Oseltamivir for people with Suspected or Diagnosed influenza. 2015, Chile. Available at: < https://diprece.minsal.cl/temas-de-salud/temas-de-salud/guias-clinicas-no-ges/ >

147. Ministerio de Salud, Chile (Chile Ministry of Health). Prevention of Oral Mucositis in People with Cancer. 2014, Chile. Available at: < https://diprece.minsal.cl/temas-de-salud/temas-de-salud/guias-clinicas-no-ges/ >

148. Australian Stroke Foundation. Stroke Management - Prehospital Care. 2018, Australia. Available at: < https://app.magicapp.org/ >

149. Australian Stroke Foundation. Stroke Management - Early Assessment and Diagnosis. 2018, Australia. Available at: < https://app.magicapp.org/ >

150. Australian Stroke Foundation. Stroke Management - Acute Medical and Surgical Management. 2018, Australia. Available at: < https://app.magicapp.org/ >

151. Australian Stroke Foundation. Stroke Management - Secondary Prevention. 2018, Australia. Available at: < https://app.magicapp.org/ >

152. Australian Stroke Foundation. Stroke Management – Rehabilitation. 2018, Australia. Available at: < https://app.magicapp.org/ >

153. Australian Stroke Foundation. Stroke Management - Managing Complications. 2018, Australia. Available at: < https://app.magicapp.org/ >

154. Australian Stroke Foundation. Stroke Management - Discharge Planning and Transfer of Care. 2018, Australia. Available at: < https://app.magicapp.org/ >

155. Australian Stroke Foundation. Stroke Management - Community Participation and Long-term Care. 2018, Australia. Available at: < https://app.magicapp.org/ >

156. Japan Society of Respiratory Therapy & Japan Society of Intensive Therapeutics Medicine. Acute Respiratory Distress Syndrome. 2016, Japan. Available at: < https://www.jrs.or.jp/modules/guidelines/index.php?content_id=88 >

157. Ministry of Health, Peru. Clinical Practice Guideline for the prevention and management of pre-eclampsia and eclampsia. 2017, Peru. Available at: < http://bvs.minsa.gob.pe/local/MINSA/4220.pdf >

158. Rehfuess EA, Stratil JM, Scheel IB, Portela A, Norris SL, Baltussen R. The WHO-INTEGRATE evidence to decision framework version 1.0: integrating WHO norms and values and a complexity perspective. BMJ Glob Health. 2019 Jan 25;4(Suppl 1):e000844.
159. Sanabria AJ, Kotzeva A, Selva Olid A, Pequeño S, Vernooij RWM, Martínez García L, Zhang Y, Solà I, Thornton J, Alonso-Coello P. Most guideline organizations lack explicit guidance in how to incorporate cost considerations. J Clin Epidemiol. 2019 Dec;116:72-83.
160. Schünemann HJ, Wiercioch W, Brozek J, Etxeandia-Ikobaltzeta I, Mustafa RA, Manja V, Brignardello-Petersen R, Neumann I, Falavigna M, Alhazzani W, Santesso N, Zhang Y, Meerpohl JJ, Morgan RL, Rochwerg B, Darzi A, Rojas MX, Carrasco-Labra A, Adi Y, AlRayees Z, Riva J, Bollig C, Moore A, Yepes-Nuñez JJ, Cuello C, Waziry R, Akl EA. GRADE Evidence to Decision (EtD) frameworks for adoption, adaptation, and de novo development of trustworthy recommendations: GRADE-ADOLOPMENT. J Clin Epidemiol. 2017 Jan;81:101-110.
161. Luhnen M, Prediger B, Neugebauer EAM, Mathes T. Systematic reviews of health economic evaluations: A structured analysis of characteristics and methods applied. Res Synth Methods. 2019 Jun;10(2):195-206.
162. Thornton J, Alderson P, Tan T, Turner C, Latchem S, Shaw E, Ruiz F, Reken S, Mugglestone MA, Hill J, Neilson J, Westby M, Francis K, Whittington C, Siddiqui F, Sharma T, Kelly V, Ayiku L, Chamberlain K. Introducing GRADE across the NICE clinical guideline program. J Clin Epidemiol. 2013 Feb;66(2):124-31.
163. Edejer TT. Improving the use of research evidence in guideline development: 11. Incorporating considerations of cost-effectiveness, affordability and resource implications. Health Res Policy Syst. 2006 Dec 5;4:23.

164. Santesso N, Glenton C, Dahm P, Garner P, Akl EA, Alper B, Brignardello-Petersen R, Carrasco-Labra A, De Beer H, Hultcrantz M, Kuijpers T, Meerpohl J, Morgan R, Mustafa R, Skoetz N, Sultan S, Wiysonge C, Guyatt G, Schünemann HJ; GRADE Working Group. GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. J Clin Epidemiol. 2020 Mar;119:126-135.
165. Stratil JM, Baltussen R, Scheel I, Nacken A, Rehfuess EA. Development of the WHO-INTEGRATE evidence-to-decision framework: an overview of systematic reviews of decision criteria for health decision-making. Cost Eff Resour Alloc. 2020 Feb 11;18:8.
166. Morgan RL, Kelley L, Guyatt GH, Johnson A, Lavis JN. Decision-making frameworks and considerations for informing coverage decisions for healthcare interventions: a critical interpretive synthesis. J Clin Epidemiol. 2018 Feb;94:143-150.
167. Guyatt G, Baumann M, Pauker S, Halperin J, Maurer J, Owens DK, Tosteson AN, Carlin B, Gutterman D, Prins M, Lewis SZ, Schünemann H. Addressing resource allocation issues in recommendations from clinical practice guideline panels: suggestions from an American College of Chest Physicians task force. Chest. 2006 Jan;129(1):182-7.
168. Cabrera PA, Pardo R. Review of evidence based clinical practice guidelines developed in Latin America and Caribbean during the last decade: an analysis of the methods for grading quality of evidence and topic prioritization. Global Health. 2019 Feb 19;15(1):14.

Figure

Figure 1: Common Economic Justifications Used in Remark Sections of Recommendations

- The decision favours one cost-effective option over another.
- The decision is against an option based on high cost versus low effectiveness (i.e. not cost-effective).
- The incremental cost was very small and did not play into the decision.
- There was uncertainty about incremental costs (i.e. no research identified).
- The incremental cost balance was similar between options.

Tables

Table 1: Characteristics of Included GRADE Guidelines

Organization (n=29)	n	%
Non-Government	14	48.3
Government	9	31.0
Government Group	2	6.9
University	2	6.9
Commercial	2	6.9
Publication Year (n=142)		
2018	50	35.2
2017	31	21.8
2016	20	14.1
2015	12	8.5
2014	24	16.9
2012-13	5	3.5
Setting (n=142)		
International	38	26.8
North America	34	24.0
Europe	30	21.1
Middle East	22	15.5
South America	9	6.3
Australia	8	5.6
Asia	1	0.7
Clinical Topic (n=142)		
Communicable Disease	30	21.1
Hospital, Injury, Pain Management	27	19.0
Cardiovascular Disease	23	16.2
Primary Care	22	15.5
Cancer	14	9.8
Autoimmune, Allergic Disease	13	9.2
Venousthromboembolism	13	9.2
EtD Frameworks (n=142)		
< 5	60	42.3
5 - 10	45	31.7
11 - 20	20	14.1
> 20	17	11.9

Abbreviations: n – sample size; (%) - percentage

Table 2: Frequency of Economic Information Reported across Pre-defined Likert-type Justifications, Cost, Cost Effectiveness and Remarks Section Boxes of 142 Guideline EtD Frameworks

Economic Literature Reported (n=142)	n	%
Reported Economic Information	57	40.2
Reported No Research Identified	22	15.5
Reported Adaptation of Economic Evaluation	1	0.7
Reported De novo Economic Evaluation	1	0.7
Not Reported (Blank)	61	42.9
Pre-defined EtD Likert-type Justifications Completed (n=142)		
Yes	99	69.7
No	43	30.3
Units Reported from Economic Research Identified (n=59)		
Both Unit Costs and Cost Effectiveness	23	39.0
Unit Costs only	18	30.5
Cost Effectiveness only	18	30.5
Research Identified Presented as Qualitative Information (n=59)		
Yes	17	28.8
No	42	71.2
Recommendation Remarks Mention Economic Information (n=142)		
Yes	54	38.0
No	88	62.0
Recommendation Remarks Mention Economic Factors without Literature Reported/Identified (n=54)		
Yes	22	40.7
No	32	59.3

Abbreviations: n – sample size; (%) - percentage

Table 3: Themes of Economic Information Across All Text Fields of EtD Frameworks from 142 Guidelines

Themes of Factors Affecting Economic Information Variability	n (%)	Evidence-to-Decision Framework Domain					
	Total Guidelines*	Remarks Section	Unit Costs	Cost Effectiveness	Acceptability	Feasibility	Equity
I - Intervention *(e.g. price, type, frequency, effectiveness, access)*	98 (69)	34 (24)	85 (60)	43 (30)	11 (8)	8 (6)	14 (10)
II - Population *(e.g. sub-group risk, country, setting, compliance, burden)*	53 (37)	8 (6)	29 (20)	39 (28)	10 (7)	5 (4)	7 (5)
III - Payor *(e.g. coverage, socioeconomics, programs, policy, budget)*	32 (23)	6 (4)	24 (17)	9 (6)	15 (11)	5 (4)	18 (13)
IV - Provider *(e.g. training, competency, standards of care, volunteers)*	32 (23)	1 (1)	28 (20)	6 (4)	3 (2)	3 (2)	N/A
V – Healthcare Resource Use *(e.g. hospitals, equipment, telemedicine, monitoring)*	35 (25)	2 (1)	32 (23)	11 (8)	1 (1)	1 (1)	1 (1)
VI - Economic Model Design *(e.g. perspective, time horizon, sensitivity, transferability)*	22 (16)	N/A	2 (1)	20 (14)	N/A	N/A	N/A

Abbreviations: n – sample size; (%) – percentage; N/A – not applicable.
* Theme contributing to a guideline at least once.

Appendices

Appendix 1: Summary of 142 GRADE Guideline EtD Frameworks (Published online between 2012-2018)

Organization / Source	Country / Year	Title	Number of Evidence-to-Decision (EtD) Frameworks	Economic Review Performed, Unit of Measure	Economic Information Quotes Reported in Recommendation Remarks Section in EtD Frameworks	Likert-type Justifications completed in EtD Framework	Quantitative Economic Information Presented	Variability Sub-Theme Factors				
								Unit Costs	Cost Effectiveness	Acceptability	Feasibility	Equity
INTERNATIONAL												
World Health Organization (WHO) https://www.who.int/tb/publications/2018/latent-tuberculosis-infection/en/	International 2018	Latent Tuberculosis Infection: Updated and Consolidated Guidelines for Programmatic Management	7	Yes, Unit Costs, Cost per Life-Years Gained	Remarks: The GDG agreed that cost–effectiveness favours 3RH because of the higher completion rate, safer profile and fewer resources required. The GDG also noted that, although direct evidence for the cost–effectiveness of 3RH in children is limited, the cost–effectiveness of	Yes	Prevention; Transmission; Country; Program Costs	Transmission	Providers, Laboratory Access	Blank	Blank	More Options Increases Equity; Patient Ability to Pay

					shorter preventive treatment including 3RH is supported by a body of evidence in adult populations. … There was consensus in the GDG that the cost–effectiveness of 3HP depends mainly on the cost of the drug and mode of administration, which would affect the costs to patients and health systems							
World Health Organization (WHO) https://www.who.int/immunization/documents/positionpapers/en/	International 2018	Bacille Calmette-Guérin (BCG) Vaccination	6	Yes, Cost per Year of Health Life Recovered	None	Yes	Yes	Program Costs; Endemicity of Disease; Access; Opportunistic Assessments; Setting	Subgroup Risk; Repeat Intervention; Transmission	Payors; Limited benefits	Blank	Blank

World Health Organization (WHO) https://www.who.int/immunization/documents/positionpapers/en/	International 2018	Cholera Vaccination	1	Yes, ICER	None	Yes	Qualitatively only	Program Costs; Drug Stockpiles; Payor; Setting	Setting	Blank	Blank	Blank
World Health Organization (WHO) https://www.who.int/immunization/documents/positionpapers/en/	International 2018	Rabies Vaccination	4	Yes, Unit Costs, ICER	Remarks: Previously WHO-recommended IM PEP regimens are still considered valid options, but may not be as cost-, dose-, or time sparing. … Countries opting for other PEP regimens should consider the regimen's (a) feasibility (i.e. cost and number of doses), …	Yes	Qualitatively only	Program Costs; Provider Costs; Country; Payor	Frequency of Intervention	Low Resource Settings; Price of Intervention; Travel Time; Subgroup Risk; Remote locations	Blank	Affordability; Accessibility; Subgroup Risk; Marginalized Populations
World Health Organization (WHO) https://www.who.int/immunization/doc	International 2017	Dengue Vaccination	2	Yes, No Research Identified	None	Yes	No Research Identified	Price of Intervention; Country; Screeni	Price of Intervention; Country, Region	Burden; Risk of Disease; Risk Management,	Blank	Blank

uments/positionpapers/en/								ng tests and Surveillance; Budget Availability; Program Costs	al Programs; Hospitalization Rates; Budget Impact; Lab Costs	Risk Communication; Access to Intervention; Program Costs		
World Health Organization (WHO) https://www.who.int/immunization/documents/positionpapers/en/	International 2017	Diphtheria Vaccination	1	Yes, No Research Identified	None	Yes	No Research Identified	Blank	Price of Intervention; Frequency of Intervention	Blank	Blank	Blank
World Health Organization (WHO) https://www.who.int/immunization/documents/positionpapers/en/	International 2017	Hepatitis B Vaccination	2	Yes, ICER	None	Yes	Qualitatively only	Provider fees; Outreach fees; Setting	Medication Dosing	Repeat Intervention; Effectiveness Payor	Blank	Blank
World Health Organization (WHO) https://www.who.int/immunization/doc	International 2017	Human Papillomavirus (HPV) Vaccination	3	Yes, ICER	Remarks: It should also offer opportunities for economies of scale in delivery …	Yes	Qualitatively only	Procurement Costs; Price of Intervention	Coverage; Sex Differences	Country; Payor	Donor Support; Large population, Hardwa	Blank

uments/positionpapers/en/					Immunization of multiple cohorts of girls is cost-effective in the age range 9–14 years, in particular when the recommended extended 2-dose schedule is used. The incremental cost effectiveness for each additional age cohort of girls and women aged ≥15 years depends on country context						re, Software, Space for Equipment	
World Health Organization (WHO) https://www.who.int/tb/publications/2017/dstb_guidance_2017/en/	International 2017	Treatment of Drug-susceptible Tuberculosis and Patient Care (2017 update)	22	Yes, No Research Identified	None	Yes	No Research Identified	Blank	Blank	Payors; Agency belief that cost should not be best driver of recommendation; Provide	Blank	Agency belief that cost should not be best driver of recommendation

										rs, Laboratory Access		
World Health Organization (WHO) https://www.who.int/immunization/documents/positionpapers/en/	International 2017	Measles Vaccination	2	Yes, No Research Identified	None	Yes	No Research Identified	Price of Intervention; Healthcare System Costs; Opportunistic Visits	Type of Intervention; Wastage rate per Intervention	Blank	Blank	Blank
World Health Organization (WHO) https://www.who.int/immunization/documents/positionpapers/en/	International 2017	Tetanus Vaccination	1	No	None	Yes	None	Frequency of Intervention	Price of Intervention; Frequency of Intervention	Blank	Blank	Blank
World Health Organization (WHO) https://www.who.int/immunization/documents/positionpapers/en/	International 2017	Fractional Dose Yellow Fever Vaccination	1	Yes, No Research Identified	None	Yes	No Research Identified	Program Costs; Social Mobilization Costs	Price of Intervention; Frequency of Intervention	Blank	Blank	Blank
World Health Organization (WHO)	International 2016	Malaria Vaccination	1	Yes,	None	Yes	Yes	Procurement Costs;	Subgroup Risk;	Country; Payor	Blank	Blank

https://www.who.int/immunization/documents/positionpapers/en/				ICER per DALY				Price of Intervention; Repeat visit costs	Hospital Usage; Setting			
World Health Organization (WHO) http://gdt.gradepro.org	International 2016	Sexually Transmitted Infection: Chlamydia	4	Yes, Unit Cost	Remarks: …When high value is placed on reducing costs, doxycycline in a standard dose may be the best choice; … Doxycycline delayed release (ER) may be an alternative to twice daily dosing of doxycycline, but the high cost of the delayed release formulation may prohibit its use. …	Yes	Yes	Price of Intervention; Full Course of Treatment, Procurement	No Research Identified; Type and Frequency of Intervention	Blank	False Beliefs; Infrastructure (phone, connection)	Blank
World Health Organization (WHO) http://gdt.gradepro.org	International 2016	Sexually Transmitted Infection: Syphilis	10	Yes, Unit Costs, Cost per Life-	Remarks: Data for drug prices and procurement indicate that doxycycline is cheaper than azithromycin and	Yes	Yes	Price of Intervention; Course of Treatment	Lives Saved	Blank	Blank	Blank

				Years Gained	erythromycin, although the latter drugs are still inexpensive. … Doxycycline is preferred over ceftriaxone due to its lower cost and oral administration.							
World Health Organization (WHO) https://www.who.int/tb/areas-of-work/drug-resistant-tb/treatment/resources/en/	International 2016	Treatment of Drug-resistant Tuberculosis	2	Yes, No Research Identified	None	Yes	No Research Identified	Blank	Blank	Blank	Price of Intervention; Generic Manufacturers	Price of Intervention; Price of Programs
World Health Organization (WHO) https://www.who.int/immunization/documents/positionpapers/en/	International 2015	Pertussis Vaccination	1	Yes, No Research Identified	None	Yes	No Research Identified	Price of Intervention; Country; Implementation Costs; Coverage	No Research Identified	Blank	Blank	Blank
World Health Organization (WHO)	International 2013	Systematic Screening for	9	Yes, No Researc	None	Yes	No Research	Blank	Blank	Blank	Blank	Blank

https://www.who.int/tb/tbscreening/en/		Active Tuberculosis		h Identified			Identified					
BMJ WikiRecs https://app.magicapp.org/	International 2018	QiukuiClopidogrel plus aspirin versus aspirin or clopidogrel alone for the treatment of acute minor ischemic stroke or high-risk TIA	2	No	Remarks: There is probably not any benefit to continuing clopidogrel beyond 21 days, but a longer duration increases the risk of bleeding, inconvenience, and costs.	Yes	None	Price of Intervention	Intervention Clearly Cost-Effective	Blank	Blank	Blank
BMJ WikiRecs https://app.magicapp.org/	International 2018	Primary Care Rapid Rec Alpha blockers for Treatment of Ureteric Stones in Primary Care	2	No	None	Yes	None	Price of Intervention; Payor	Blank	Blank	Blank	Blank
BMJ WikiRecs https://app.magicapp.org/	International 2018	Oxygen for Acutely ill Patients	3	No	None	Yes	None	Price of Providers; Something Providers Already Doing	Blank	Blank	Blank	Blank

BMJ WikiRecs https://app.magicapp.org/	International 2018	Prostate-Specific Antigen (PSA) Screening in men Without Symptoms of Prostate Cancer	1	Yes. ICER	Cost-Effectiveness: Although the panel focused on the patient-perspective rather than that of society, its recommendation is compatible with these findings	No	Yes	Blank	Patient and Society Perspective Similar	Blank	Blank	Blank
BMJ WikiRecs https://app.magicapp.org/	International 2018	Corticosteroids for Sepsis	1	No	None	Yes	None	Price of Hospital Length of Stay	Blank	Blank	Blank	Blank
BMJ WikiRecs https://app.magicapp.org/	International 2018	Patent foramen ovale Closure, Antiplatelet or Anticoagulation Therapy for Management of Cryptogenic Stroke	3	No	Cost Effectiveness: Implementation of this recommendation is likely to have an important impact on the costs for health funders which warrants cost-effectiveness data.	Yes	None	Price of Intervention; Future Care Costs	Patient and Society Perspective; Payors	Blank	Blank	Blank
BMJ WikiRecs	International 2018	Atraumatic versus Conventional	1	Yes, Unit Costs	Remarks: In the absence of conclusive cost-	Yes	Qualitatively only	Price of Equipment;	Intervention Costs;	Blank	Blank	Blank

https://app.magicapp.org/		Lumbar Puncture Needles			effectiveness analyses, potential cost implications are offset by the lack of effect modification with needle type and the fact that the cost of certain atraumatic needles is equivalent to conventional needles.			Manufacturer of Equipment	Payors; Work Lost Time			
BMJ WikiRecs https://app.magicapp.org/	International 2018	Antibiotics for Uncomplicated Skin Abscesses	3	No	Remarks: There is a close balance between the expected benefits from antibiotics (a modest reduction in treatment failure, abscess recurrence, and pain) and the expected harms (gastrointestinal side effects), burdens of treatment, and costs.	Yes	None	Setting, Price of Intervention; Price of Treatment Failure	Blank	Blank	Blank	Blank

BMJ WikiRecs https://app.magicapp.org/	International 2018	Corticosteroids for Sore Throat	1	No	Remarks: Due to their low cost, resources did not play an important role when formulating this recommendation.	Yes	None	Price of Intervention	Blank	Blank	Blank	Blank
BMJ WikiRecs https://app.magicapp.org/	International 2017	Antiretroviral therapy for pregnant women living with HIV	2	No	None	Yes	None	Setting, Payor	Blank	Blank	Blank	Blank
BMJ WikiRecs https://app.magicapp.org/	International 2017	Transcatheter aortic valve implantation versus surgical aortic valve replacement for severe aortic stenosis in low-intermediate risk patients	5	Yes, No Research Identified	None	Yes	No Research Identified	Blank	Blank	Blank	Blank	Blank
BMJ WikiRecs https://app.magicapp.org/	International 2017	Vasopressors - Blood pressure Targets	1	No	None	No	None	Subgroup Risk	Blank	Blank	Blank	Blank
BMJ WikiRecs https://app.magicapp.org/	International 2017	Arthroscopic Surgery for Degenerative Knee Disease	1	Yes, ICER	None	Yes	Qualitatively only	Price of Intervention	Payor or Societal	Blank	Blank	Blank

									Perspective			
BMJ WikiRecs https://app.magicapp.org/	International 2017	Low Intensity Pulsed Ultrasound for Bone Healing	1	Yes, Unit cost and Burden	Remarks: …combined with the potential burden and high costs of treatment represent waste of health care resources where we believe all or nearly all well informed patients would elect not to apply LIPUS for healing their fractures.	Yes	Qualitatively only	Patient and Society Perspective; Payor	No Research Identified	Blank	Blank	Blank
BMJ WikiRecs https://app.magicapp.org/	International 2016	Probiotics for Children Receiving Antibiotics	2	No	Remarks: Probiotics are generally inexpensive and widely available.	No	None	Payor, Private Insurance	Blank	Blank	Blank	Blank
BMJ WikiRecs https://app.magicapp.org/	International 2015	Adjunctive Corticosteroid Therapy for Adults Hospitalized with Community-acquired Pneumonia	1	No	Remarks: Corticosteroids are generally inexpensive and almost certainly cost-effective in this setting.	No	None	Price of Intervention	Intervention Clearly Cost-Effective	Blank	Blank	Blank

European Forum for Research and Education in Allergy and Airway Diseases https://www.euforea.eu/aria	International 2017	Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines-2016 revision	6	Yes, cost per patient per year	Remarks: The panel members acknowledged that the choice of treatment will mostly depend on patient preferences and local availability and cost of treatment.	Yes	Yes	Country	Lack of Effectiveness	Blank	Blank	Coverage, Socioeconomic status
World Allergy Organization (WAO) https://waojournal.biomedcentral.com/articles/	International 2016	World Allergy Organization-McMaster University Guidelines for Allergic Disease Prevention (GLAD-P): Prebiotics	2	Yes, ICER	None	Yes	Yes	Subgroup risk	Time Horizon	Blank	Blank	Blank
World Allergy Organization (WAO) https://waojournal.biomedcentral.com/articles/	International 2016	World Allergy Organization-McMaster University Guidelines for Allergic Disease Prevention (GLAD-P): Vitamin D	3	Yes, drug price	Remarks: This recommendation is based on very low certainty of evidence, as well as this strategy is not cost-effective and probably increasing health inequities.	Yes	Yes	Country, Region, Drug formulation	Lack of Effectiveness	Blank	Blank	Socioeconomic status, Policy-maker coverage, Effectiveness

World Allergy Organization (WAO) https://waojournal.biomedcentral.com/articles/	International 2015	World Allergy Organization-McMaster University Guidelines for Allergic Disease Prevention (GLAD-P): Probiotics	3	No	Remarks: We agreed that the considerations of values and preferences, resource implications and equity are likely similar to those in pregnant women. We also noted that the cost of probiotics is much lower than cost of a formula which may have an impact on the assessment of opportunity cost	Yes	Price of Intervention; Setting; Private Insurance	Setting, Country, Insurance access	Blank	Blank	Blank	Socioeconomic status, Policy-maker coverage
NORTH AMERICA												
Canadian Task Force on Preventive Health Care (CTFPHC) https://canadiantaskforce.ca/	Canada 2018	Screening for Asymptomatic Bacteriuria in Pregnancy	1	Yes, No Research Identified	None	No	No Research Identified	Blank	Blank	Blank	Blank	Blank
Canadian Task Force on Preventive Health Care (CTFPHC)	Canada 2018	Screening for Impaired Vision	1	No	None	No	None	Blank	Blank	Blank	Blank	Blank

https://canadiantaskforce.ca/												
Canadian Task Force on Preventive Health Care (CTFPHC) https://canadiantaskforce.ca/	Canada 2018	Screening for Breast Cancer – update	6	No	Remarks: In the judgment of the task force, current recommendations are both feasible and acceptable to women and clinicians and are not expected to have an increased negative effect on health equity or to pose additional costs to the health care system.	No	None	Blank	Blank	Blank	Blank	Blank
Canadian Task Force on Preventive Health Care (CTFPHC) https://canadiantaskforce.ca/	Canada 2017	Screening for Abdominal Aortic Aneurysm	3	Yes, ICER	None	No	Yes	Blank	Disease Prevalence; Price of Intervention	Blank	Blank	Blank
Canadian Task Force on Preventive Health Care (CTFPHC)	Canada 2017	Prevention of Tobacco Smoking in Children and Adults	2	No	None	No	None	Price of Intervention	Blank	Blank	Blank	Blank

https://canadiantaskforce.ca/												
Canadian Task Force on Preventive Health Care (CTFPHC) https://canadiantaskforce.ca/	Canada 2017	Screening for Hepatitis C	1	Yes, treatment price and budget impact	Remarks: Therefore, a recommendation in favour of screening would increase the number of people with known HCV (and who are potentially susceptible to harms of stigma and anxiety) who could not access treatment, thus deriving no clear benefit despite the potential for harm from a diagnosis combined with treatment ineligibility.	No	Yes	Price of Intervention; Socioeconomic status	Budget Impact	Blank	Blank	Coverage
Canadian Task Force on Preventive Health Care (CTFPHC) https://canadiantaskforce.ca/	Canada 2016	Screening for Colorectal Cancer	4	Yes Re-analysis - Modelling study	Remarks: It reflects a relatively higher value placed on the lack of direct evidence from RCTs of incremental	No	Qualitatively only	Blank	Intervention Frequency	Blank	Intervention type, provider type, setting (hospital vs	Blank

					benefit for colonoscopy and on the opportunity costs of using colonoscopy for population screening.						outpatient)	
Canadian Task Force on Preventive Health Care (CTFPHC) https://canadiantaskforce.ca/	Canada 2016	Screening for Developmental Delay	1	No	Remarks: …following the recommendation should allow clinicians to focus on more effective and cost-effective services, for example, attending to children at risk for or identified with development delay.	No	None	Blank	Blank	Blank	Blank	Blank
Canadian Task Force on Preventive Health Care (CTFPHC) https://canadiantaskforce.ca/	Canada 2016	Screening for Lung Cancer	3	Yes, ICER	None	No	Yes	Blank	Context; Transferability	Blank	Blank	Blank
Canadian Task Force on Preventive	Canada 2016	Pelvic Exam	1	No	None	No	None	Blank	Blank	Blank	Blank	Blank

Health Care (CTFPHC) https://canadiantaskforce.ca/												
Canadian Task Force on Preventive Health Care (CTFPHC) https://canadiantaskforce.ca/	Canada 2015	Screening for Cognitive Impairment	1	No	None	No	None	Time for Assessment, False positive rate	Blank	Blank	Blank	Blank
Canadian Task Force on Preventive Health Care (CTFPHC) https://canadiantaskforce.ca/	Canada 2015	Prevention of Obesity in Adults	5	Yes, Rate of Service use	Remarks: Therefore, the task force recommends calculation of BMI, placing a relatively high value on a low-cost, easily calculated measure with widely accepted cut-off points to base guidance for weight-gain prevention or management.	No	Qualitatively only	Rate of Service use	Inconsistent Findings	Blank	Blank	Blank
Canadian Task Force on Preventive	Canada 2015	Prevention of Obesity in Children	2	No	Remarks: In the judgment of the task force, growth	No	None	Blank	Blank	Blank	Blank	Blank

Health Care (CTFPHC) https://canadiantaskforce.ca/					monitoring is a long-standing, feasible, low-cost intervention that is unlikely to result in harms…							
Canadian Task Force on Preventive Health Care (CTFPHC) https://canadiantaskforce.ca/	Canada 2014	Screening for Prostate Cancer	3	No	None	No	None	Blank	Blank	Blank	Blank	Blank
Canadian Task Force on Preventive Health Care (CTFPHC) https://canadiantaskforce.ca/	Canada 2013	Screening for Depression	2	Yes	Remarks: The time clinicians take to screen for depression reduces their availability to deliver other services of known clinical benefit (opportunity cost).	No	Qualitatively only	Blank	Setting; Remission of Disease	Blank	Blank	Blank
Canadian Task Force on Preventive Health Care (CTFPHC) https://canadiantaskforce.ca/	Canada 2013	Screening for Cervical Cancer	5	Yes ICER	None	No	Qualitatively only	Blank	Intervention Frequency	Blank	Blank	Blank

Canadian Task Force on Preventive Health Care (CTFPHC) https://canadiantaskforce.ca/	Canada 2012	Screening for Type 2 Diabetes	3	Yes, testing price	None	No	Qualitatively only	Price of Intervention; Price of Testing	Time to Start Intervention; Frequency of Intervention	Blank	Blank	Blank
Canadian Task Force on Preventive Health Care (CTFPHC) https://canadiantaskforce.ca/	Canada 2012	Screening for Hypertension	3	Yes, No Research Identified	None	No	No Research Identified	Blank	Blank	Blank	Blank	Blank
Canadian Chiropractic Guideline Initiative (CCGI) https://www.chiropractic.ca/guidelines-best-practice/	Canada 2017	Spinal Manipulative Therapy and Other Conservative Treatments for Low Back Pain	5	No	None	Yes	None	Price of Intervention, Equipment, Provider Training	Blank	Blank	Blank	Blank
Canadian Chiropractic Guideline Initiative (CCGI) https://www.chiropractic.c	Canada 2016	Management of Neck Pain Associated Disorders	11	No	Remarks: The relative small cost of providing the option would make it more acceptable to stakeholders and	No	None	Price of Intervention	Blank	Blank	Blank	Blank

a/guidelines-best-practice/					feasible to implement							
Canadian Chiropractic Guideline Initiative (CCGI) https://www.chiropractic.ca/guidelines-best-practice/	Canada 2016	Management of Whiplash Associated Disorder	2	No	None	No	None	Price of Intervention	Blank	Blank	Blank	Blank
National Pain Centre - McMaster University https://app.magicapp.org/	Canada 2017	The 2017 Canadian Guideline for Opioids for Chronic Non-Cancer Pain	10	Yes, Medical Costs, Willingness to Pay per QALY gained	Remarks: Recognizing the cost of formal multidisciplinary opioid reduction programs and their current limited availability / capacity, an alternative is a coordinated multidisciplinary collaboration…	Yes	Yes	Price of Intervention, Providers; Coverage, Cultural and Geographical Barriers; Societal Burden	Blank	Blank	Blank	Blank
Canadian Blood Services https://www.blood.ca/en http://gdt.gradepro.org	Canada 2017	Controlled Pediatric Donation after Circulatory Determinatio	5	No	None	No	None	Blank	Blank	Blank	Blank	Blank

		n of Death (pDCD)										
American Society of Hematology (ASH) http://www.hematology.org/VTE	United States 2018	Prophylaxis for hospitalized and Non-hospitalized Medical Patients	18	Yes, Unit Costs	Remarks: Pharmacological thromboprophylaxis was deemed to be of high cost and probably not acceptable to stakeholders.	Yes	Yes	Cost per GRADE Outcome; Outpatient, Hospital, Pharmacy	Treatment vs No Treatment; Time Horizon	Blank	Blank	Setting; Illness Severity; Payor
American Society of Hematology (ASH) http://www.hematology.org/VTE	United States 2018	Optimal Management of Anticoagulation Therapy	21	Yes, Unit Costs	Remarks: …given the high cost of prescription oral vitamin K tablets and the variable vitamin K content of available over-the-counter products the panel conditionally recommends against administering oral vitamin K.	Yes	Yes	Cost per GRADE outcome; Testing Costs; Hospitalization Costs	No Research Identified	Payor; Test Commonly done in Practice	Blank	Price of Intervention; Access to Intervention
American Society of Hematology (ASH)	United States 2018	Diagnosis of Venous Thromboembolism	18	Yes, Per Facility Costs; ICER	Remarks: The panel considered a strategy with D-dimer testing first to reduce cost,	Yes	Yes	Per Facility Testing Approa	Setting; Time-frame; Type of	Blank	Blank	Blank

http://www.hematology.org/VTE					ensure feasibility, and reduce radiation exposure.			ch Costs	Intervention			
American Society of Hematology (ASH) http://www.hematology.org/VTE	United States 2018	Treatment of Pediatric Venous Thromboembolism	29	Yes, Unit Costs	None	Yes	Yes	Costs per GRADE outcome; wholesale costs	No Research Identified	Blank	Blank	Blank
American Society of Hematology (ASH) http://www.hematology.org/VTE	United States 2018	Management of Heparin Induced Thrombocytopenia (HIT)	33	Yes, Procedure Costs, Cost Analysis	None	Yes	Yes	Hospital Charges; Procedure Costs	Subgroup Risks; Complications	Blank	Blank	Blank
American Society of Hematology (ASH) http://www.hematology.org/VTE	United States 2018	Venous Thromboembolism Management in Context of Pregnancy	18	Yes, Unit Cost, ICER	None	Yes	Yes	Costs per GRADE outcome; Price of Interventions, Setting and Providers	Time Horizon; Subgroup Risk; Complications	Blank	Patient Compliance	Frequency of Intervention; Caregiver Burden

American Thoracic Society (ATS) http://gdt.gradepro.org	United States 2018	Treatment of Idiopathic Pulmonary Fibrosis	9	No	None	No	None	Blank	Blank	Blank	Blank	Blank
American Thoracic Society (ATS) https://www.thoracic.org	United States 2018	Diagnosis of Idiopathic Pulmonary Fibrosis	9	No	None	No	None	Price of Intervention; Access	Blank	Blank	Blank	Blank
American Thoracic Society (ATS) https://www.thoracic.org	United States 2018	Diagnosis of Primary Ciliary Dyskinesia: An Official ATS Clinical Practice Guideline	4	Yes, Unit Costs	Remarks: The overall impact of avoiding direct costs, complications and burden of repeat testing justified using this extended panel genetic testing as a replacement to reference standards. Furthermore, extended genetic panel testing was probably cost-effective, …Centers must also routinely train laboratory	Yes	Yes	Testing and Equipment Costs; Provider Training Costs; Payor	Wait times for Results; Testing Burden; Complications	Price of Hardware, Software, Equipment, Space for Equipment	Blank	Price of Intervention; Reimbursement for Intervention

					personnel in standard operating procedures for nNO measurement, which may add additional costs to implementing nNO testing.							
Society of Critical Care Medicine (SCCM); ICU Liberation http://gdt.gradepro.org	United States 2018	Management of Pain, Agitation and Delirium in Adults in the Intensive Care Unit (PAD-ES)	37	No	Remarks: Apparent benefit in self-reported sleep quality, not costly.	Yes	None	Price of Intervention; Access to Intervention; Ongoing Costs; Program Costs; Country; Setting; Provider Costs; Practice Change Resources; Space	Complications; Provider Costs; Access to Intervention; Equipment	Payor; Type of Intervention; Country; Standard of Practice; Cost of Equipment; Space and Access for Equipment	Price of Intervention; Access to Volunteers; Space and Access for Equipment	Blank

								and Access for Equipment; Infrastructure				
National Highway Traffic Safety Administration (NHTSA) https://www.nhtsa.gov/	United States 2018	Evidence-Based Guidelines For Fatigue Risk Management In Emergency Medical Services	5	Yes, Indirect Costs	Remarks: …Some instruments may require payment prior to use. In general, the panel perceives the costs and burden associated with implementation as limited….Considerations for program implementation include, but are not limited to: program focus, content, length, delivery method, costs and resource needs.	Yes	Qualitatively only	Tool Licensing Fees; Provider Shift Duration; Provider Education Costs; Infrastructure Costs	Setting; Patient and Provider Safety; Program Costs	Blank	Price of Intervention; Payors; Budget Limitations	Blank
EUROPE												
European Commission Initiative on Breast Cancer	European Union (EU) 2018	Recommendations from European Breast Guidelines	23	Yes, Units Costs, ICER, Willing	Remarks: Cost-effectiveness probably favours the intervention in different countries	Yes	Yes	Price of Intervention; Frequency of	Age Subgroup Risk; Countr	Payor; Setting; Access; Advocacy	Country, Subgroup Risk;	Policy Decisions, Country,

https://ecibc.jrc.ec.europa.eu/recommendations/				ness-to-Pay (Life-Year Gained and QALY)	or settings but varies across them. Differences in the cost-effectiveness results could be explained by the differences in setting, outcomes and type of technology used. … Finally, there is a lack of evidence on the cost effectiveness of DBT compared to DM. Conditions that the GDG felt are favourable towards the use of a particular screening test. DM: where costs or patient preference are in favour of DM, the GDG recommends considering continued use of DM. … There are additional factors			Intervention Cost per GRADE outcome; Hospital or Outpatient Setting; Cost per Diagnosis; Cost per False-Positives and False Negatives; GRADE Quality of Study Outcomes	y; Setting; Type of Equipment; Policy; Access to Programs	Group Interests	Price of Equipment; Provider competency, Access	Access, Payor

					of increased costs for DBT plus DM over DM alone that must be considered on a country-by-country basis, depending on resources available, for breast cancer screening programmes.							
Nationalt akutkirurgisk tvaerfagligt forum (National emergency surgical interdisciplinary forum) https://app.magicapp.org/	Denmark 2018	National Clinical Guideline (NCR) for Spinal Stabilization of Adult Trauma Patients in Denmark	4	No	None	Yes	None	Blank	Blank	Blank	Blank	Blank
Sundhedsstyrelsen * (Health) https://app.magicapp.org/	Denmark 2018	National Clinical Guideline for Dementia and Medicine	4	No	None	No	None	Blank	Blank	Blank	Blank	Blank
Sundhedsstyrelsen * (Health)	Denmark 2018	National Clinical Guideline for the	10	No	None	No	None	Blank	Blank	Blank	Blank	Blank

https://app.magicapp.org/		Prevention and Treatment of Behavioral and Mental Symptoms in People with Dementia										
Sundhedsstyrelsen * (Health) https://app.magicapp.org/	Denmark 2018	National Clinical Guideline for Treatment of Alcohol Addiction	15	No	None	Yes	None	Blank	Blank	Blank	Blank	Blank
Sundhedsstyrelsen * (Health) https://app.magicapp.org/	Denmark 2018	National Clinical Guideline for Opioid Treatment of Chronic Non-Malignant Pain	10	No	None	No	None	Blank	Blank	Blank	Blank	Blank
Sundhedsstyrelsen * (Health) https://app.magicapp.org/	Denmark 2018	National clinical Guideline for the study and Treatment of ADHD in Children and Adolescents	15	No	None	No	None	Blank	Blank	Blank	Blank	Blank
Sundhedsstyrelsen * (Health)	Denmark 2018	National clinical Guideline for	8	No	None	No	None	Blank	Blank	Blank	Blank	Blank

https://app.magicapp.org/		the Diagnosis of mild Cognitive Impairment and Dementia										
Sundhedsstyrelsen * (Health) https://app.magicapp.org/	Denmark 2017	National Clinical Guideline for the Treatment of Emotionally Unstable Personality Structure, Borderline Type	6	No	None	No	None	Blank	Blank	Blank	Blank	Blank
Sundhedsstyrelsen * (Health) https://app.magicapp.org/	Denmark 2017	National clinical Guideline for Post-operative effects of early Breast Cancer	5	No	None	Yes	None	Time for Training Providers	Blank	Blank	Blank	Blank
Sundhedsstyrelsen * (Health) https://app.magicapp.org/	Denmark 2017	National Clinical Guideline for Treatment of Brain Metastases	5	No	None	Yes	None	Both options available regardless of cost	Blank	Blank	Blank	Blank
Sundhedsstyrelsen *	Denmark	National clinical	8	No	None	Yes	None	Distance	Blank	Blank	Blank	Blank

(Health) https://app.magicapp.org/	2017	Guideline for Non-pharmacological Treatment in Children and Adolescents with Asthma						Travelled; Participation costs				
Aarhus Universitetshospital (Aarhus University Hospital) https://app.magicapp.org/	Denmark 2016	The Effect of Treatment Methods Within Gambling Addiction	8	No	None	Yes	None	Telemedicine Prices; Provider Competence; Access to Service	Blank	Blank	Blank	Blank
Helsedirektoratet (Directorate of Health) https://app.magicapp.org/	Norway 2018	National Academic Guideline for Dental Health Services for Children and Adolescents 0–20 years	24	Yes, de novo - Cost Analysis	None	Yes	Yes	Provider time; Country; Scope and Organization of Current Practice	Blank	Blank	Blank	Blank
Helsedirektoratet (Directorate of Health)	Norway 2018	National Professional Guideline for Diabetes	36	Yes, Unit Costs	None	No	Yes	Price of Intervention, Equipment,	Blank	Blank	Blank	Blank

https://app.magicapp.org/								Coverage; Provider Training; Lab Costs				
Helsedirektoratet (Directorate of Health) https://app.magicapp.org/	Norway 2018	National Academic Guideline for Gestational Diabetes	9	Yes, Unit Costs	None	Yes	Yes	Cost to Effectiveness Ratio; Price of Intervention and Follow ups; Price of Adverse Events	Blank	Blank	Blank	Blank
Helsedirektoratet (Directorate of Health) https://app.magicapp.org/	Norway 2018	National Academic Guideline for Prevention of Cardiovascular Disease	19	Yes, Cost Benefit Analysis; ICER	None	No	Qualitatively only	Price of Testing, Treatment and Monitoring	Guideline concordance; Subgroup risk	Blank	Blank	Blank
Helsedirektoratet (Directorate of Health)	Norway 2017	National Academic Guideline for Treatment and	26	No	Remarks: The measure is also relatively easy to implement at low cost, and there are	No	None	Blank	Blank	Blank	Blank	Blank

https://app.magicapp.org/		Rehabilitation by Stroke			positive user experiences.							
Norsk Selskap for Trombose og Hemostase (Norwegian Company for Thrombosis and Hemostasis) https://app.magicapp.org/	Norway 2016	Guidelines for Antithrombotic Treatment and Prophylaxis	264	No	None	Yes	None	Price of Intervention	Monitoring Costs	Blank	Blank	Blank
Norsk Selskap for Trombose og Hemostase (Norwegian Company for Thrombosis and Hemostasis) https://app.magicapp.org/	Norway 2015	Prevention of VTE in Orthopedic Surgery Patients: A Norwegian adaptation of the 9th ed. of the ACCP Antithrombotic Therapy and Prevention of Thrombosis	9	No	None	Yes	None	No additional resource demands	Blank	Blank	Blank	Blank
Norsk Selskap for Trombose og Hemostase (Norwegian Company for	Norway 2015	VTE, Thrombophilia, Antithrombotic Therapy and	38	No	None	Yes	None	Price of Intervention	Blank	Blank	Blank	Blank

Thrombosis and Hemostasis) https://app.magicapp.org/		Pregnancy: A Norwegian adaptation of the 9th ed. of the ACCP Antithrombotic Therapy and Prevention										
Sykehuset Innlandet (Hospital Inland) https://app.magicapp.org/	Norway 2017	Ataxia Telangiectasia	18	No	None	No	None	Blank	Blank	Blank	Blank	Blank
Sykehuset Innlandet HF, Avdeling for Fysikalsk medisin og rehabilitering, Ottestad (Hospital Innlandet HF, Department of Physical Medicine and Rehabilitation, Ottestad) https://app.magicapp.org/	Norway 2016	Knowledge-based Professional Guideline for Rehabilitation After Acquired Upper Limb Amputation in Norway	136	No	None	Yes	None	Price of Intervention; Equipment Costs; Coverage	Intervention Clearly Cost-Effective; Sickness Disability Costs	Blank	Blank	Blank
Norsk Ortopdisk	Norway 2015	Treatment of Distal Radius	6	Yes,	None	No	No Researc	Too close to	Blank	Blank	Blank	Blank

Foren (Norwegian Orthopaedic Association) https://app.magicapp.org/		Fractures in Adults		No Research Identified			h Identified	tell; Price of Intervention				
Norsk Ortopdisk Foren (Norwegian Orthopaedic Association) https://app.magicapp.org/	Norway 2015	Adult Wrist Fracture Treatment	10	Yes, No Research Identified	None	No	No Research Identified	Too close to tell; Price of Intervention, Follow up visits	Blank	Blank	Blank	Blank
Duodecim Medical Publications Ltd. https://app.magicapp.org/	Finland 2017	Topical NSAIDs for Chronic Musculoskeletal Pain in Adults	1	No	None	No	None	Price of Intervention	Blank	Blank	Blank	Blank
EBMeDS® Clinical Decision Support https://app.magicapp.org/	Finland 2015	Antiviral Treatment for Preventing Postherpetic Neuralgia	1	No	None	No	None	Price of Intervention	Blank	Blank	Blank	Blank
The Scandinavian Society of Anaesthesiology and Intensive	Scandinavia 2016	Fluid in Resuscitation of Critically ill Patients with Acute	7	No	None	No	None	Limited Treatment Supply	Blank	Blank	Blank	Blank

Care Medicine (SSAI) https://app.magicapp.org/		Circulatory Failure										
The Scandinavian Society of Anaesthesiology and Intensive Care Medicine (SSAI) https://app.magicapp.org/	Scandinavia 2016	Fluid and Drug Therapy in Adults with Acute Respiratory Syndrome	9	No	None	No	None	Price of Intervention	Blank	Blank	Blank	Blank
National Prevention Plan Vaccinal (PNPV) 2012-2014 http://www.recentiprogressi.it/allegati/02152_2016_02/fulltext/06_Articolo%20originale%20-%20Gonzalez.pdf	Italy 2016	Decision-Making Model For the Adoption of Antivaricella Vaccine	1	Yes, Unit Costs, ICER, Cost-Utility	None	Yes	Qualitatively only	Price of Intervention; Price to Administer Intervention; Direct and Indirect Disease Costs; Adverse Event Costs	Economic Study Heterogeneity, Time Horizon; Healthcare System Perspective	Blank	Blank	Blank
MIDDLE EAST												

Kingdom of Saudi Arabia Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx	Saudi Arabia 2014	Diagnosis of First Suspected Deep Vein Thrombosis of Lower Extremity	24	Yes, Unit Costs	None	Yes	Yes	Price of Intervention	Blank	Blank	Blank	Blank
Kingdom of Saudi Arabia Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx	Saudi Arabia 2014	Venous Thromboembolism Treatment	8	Yes, Unit Costs	None	Yes	Yes	Price of Intervention	Outpatient Costs; Length of Stay	Blank	Blank	Blank
Kingdom of Saudi Arabia Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx	Saudi Arabia 2014	Prevention of Venous Thromboembolism with Stroke	8	Yes, No Research Identified	None	Yes	No Research Identified	Access to Intervention	Subgroup Risk	Blank	Blank	Blank
Kingdom of Saudi Arabia Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx	Saudi Arabia 2014	Use of Thrombolytic Therapy in Acute Stroke	6	Yes, ICER	None	Yes	Yes	Access to Intervention	Subgroup Risk	Blank	Blank	Blank

Kingdom of Saudi Arabia Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx	Saudi Arabia 2014	Antithrombotic Treatment for Non-valvular Atrial Fibrillation	9	Yes, ICER	Remarks: The Saudi Expert Panel considered that in patients at intermediate risk of stroke, the desirable consequences of using oral anti-coagulation rather than aspirin plus clopidogrel (i.e. stroke reduction) probably outweigh the undesirable consequences (i.e. burden of treatment and costs).	Yes	Qualitatively only	Blank	Subgroup Risk; Outdated Research	Blank	Blank	Blank
Kingdom of Saudi Arabia Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx	Saudi Arabia 2014	Use of Screening Strategies for Detection of Breast Cancer	5	Yes, ICER	None	Yes	Yes	Price of Intervention, Providers, and Training of Providers	Settings, Individual and Population Perspectives	Blank	Blank	Blank
Kingdom of Saudi Arabia Ministry of Health	Saudi Arabia 2014	Screening and Treatment of Precancerous	6	Yes, No Research	None	Yes	No Research	Price of Intervention	Training, Quality Control	Blank	Blank	Blank

https://www.moh.gov.sa/en/Pages/default.aspx		Lesions for Cervical Cancer Prevention		Identified			Identified		, Wait times			
Kingdom of Saudi Arabia Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx	Saudi Arabia 2014	Role of Vitamin D, Calcium and Exercise in Fracture Prevention in Elderly	4	Yes, Unit Costs	None	Yes	Yes	Price of Intervention; Price of Hospital Stay; Annual Costs	No Research Identified	Blank	Blank	Blank
Kingdom of Saudi Arabia Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx	Saudi Arabia 2014	Timing of Initiation of Dialysis	1	Yes, Unit Costs with Quality Appraisal	None	Yes	Yes	Price of Intervention; Hospitalizations, Outpatient Visits; Transportation	No Research Identified	Blank	Blank	Blank
Kingdom of Saudi Arabia Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx	Saudi Arabia 2014	Allergic Rhinitis in Asthma	8	Yes, Unit Cost, Cost per Patient	Remarks: The incremental cost is probably small relative to the net benefits, and the use of INSC rather than INAH would be acceptable and feasible.	Yes	Yes	Price of Intervention	No Research Identified	High costs affect patient and healthcare system acceptance	Blank	Blank

Kingdom of Saudi Arabia Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx	Saudi Arabia 2014	Management of ST-elevation myocardial infarction (STEMI)	11	Yes, No Research Identified	None	Yes	No Research Identified	Price of Intervention; Lab Costs	No Research Identified	Blank	Blank	Either Intervention Equally Available; Socioeconomic status
Kingdom of Saudi Arabia Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx	Saudi Arabia 2014	Management of Breast Lump and Primary Breast Cancer	12	Yes, Unit Costs, ICER, 5-year QALY	Remarks: The incremental cost of implementing ultrasonography was considered probably low relative to the net benefit...	Yes	Yes	Price of Intervention; Hospital Admissions; Ambulatory Costs; Payor	Price of Intervention; Payor; Recurrence Rates	Blank	Blank	Blank
Kingdom of Saudi Arabia Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx	Saudi Arabia 2014	Management of Eclampsia	7	Yes, Cost per patient, Incremental Cost, Willingness to Pay	Remarks: For pregnant women with pre-eclampsia without severe features, the undesirable consequences of high resource use due to admission to hospital and the risk for future interventions probably	Yes	Yes	Price of Intervention, Monitoring Costs, Provider Fees	Adverse Event Costs; GRADE Quality of Study Outcomes; Hospital Care Cost;	Blank	Blank	Low cost for Intervention; widely available

					outweigh the small benefits of reducing the risk of eclampsia				Cost to Prevent one case			
Kingdom of Saudi Arabia Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx	Saudi Arabia 2014	Management of Overweight and Obese Adults	11	No	None	Yes	None	Price of Intervention; Provider Fees	No Research Identified	Blank	Blank	Blank
Kingdom of Saudi Arabia Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx	Saudi Arabia 2014	Management of Pre-Eclampsia	7	No	Remarks: All benefits outweigh any costs, inconvenience, or side effects	Yes	None	Price of Intervention	Clearly Cost-Effective	Blank	Blank	Setting; Coverage
Kingdom of Saudi Arabia Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx	Saudi Arabia 2014	Management of Sickle Cell Disease	10	Yes, ICER, mean improvement in QALY per Patient	Remarks: In addition pre-operative transfusion was judged to be of low cost, probably cost effective, and probably feasible.	Yes	Yes	Price of Intervention	Subgroup Risk	Blank	Blank	Blank
Kingdom of Saudi Arabia Ministry of Health	Saudi Arabia 2014	Management of Thalassemia – Iron	6	Yes, Unit Cost	Remarks: Due to lower cost and assumed lower acceptability of	Yes	Yes	Price of Intervention	No Research	Payor and Provider	Blank	Coverage; Access to

https://www.moh.gov.sa/en/Pages/default.aspx		chelation therapy, Bisphosphonates and Zinc supplementation			combination therapy and possibly higher compliance and better safety profile with deferoxamine monotherapy, the panel suggests treatment with deferoxamine alone				Identified	acceptance; no side-effects; low cost		Intervention; Monitoring
Kingdom of Saudi Arabia Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx	Saudi Arabia 2014	Migraine Headache: Diagnosis & Management	18	Yes, No Research Identified	Remarks: There was clinical benefit with small additional cost.	Yes	No Research Identified	Price of Intervention	No Research Identified	Blank	Blank	Blank
Kingdom of Saudi Arabia Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx	Saudi Arabia 2014	Prevention of VTE in Surgical Patients	18	Yes, No Research Identified	None	Yes	No Research Identified	Price of Intervention; Provider Fees; Indirect Costs	Subgroup Risk; Not Effective so Not Cost Effective	Blank	Blank	Low cost for Intervention
Kingdom of Saudi Arabia	Saudi Arabia 2014	Prophylaxis of VTE in Medical	20	Yes, No Researc	Remarks: The panel was uncertain about	Yes	No Research	Price of Intervention	No Research	Blank	Blank	Blank

Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx		Patients and Long Distance Travelers		h Identified	cost-effectiveness, and judged the intervention to be probably feasible and acceptable.		Identified		Identified			
Kingdom of Saudi Arabia Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx	Saudi Arabia 2014	Screening for Colorectal Cancer	7	Yes, Cost-Utility; ICER	Remarks: That being said, the benefits of screening using FS were still felt to clearly outweigh any undesirable consequences such as procedure complications, cost and implementation barriers.	Yes	Yes	Government vs Private; Price of Intervention, Equipment; Mailing Costs; Frequency of Repeat Testing; Infrastructure	Screening Annually	Blank	Blank	Blank
Kingdom of Saudi Arabia Ministry of Health https://www.moh.gov.sa/en/Pages/default.aspx	Saudi Arabia 2014	Screening for Hypertension	12	Yes, ICER	Remarks: The incremental cost relative to the net benefits is small, and this is an intervention acceptable to key stakeholders and feasible to implement.	Yes	Qualitatively only	Equipment, Provider time; Population Age Subgroups	Diagnostic Accuracy	Blank	Access to Intervention	Blank

SOUTH AMERICA												
Ministerio de Salud, Chile (Chile Ministry of Health) https://diprece.minsal.cl/temas-de-salud/temas-de-salud/guias-clinicas-no-ges/	Chile 2018	Conservative Non-Dialytic Treatment of Chronic Kidney Disease	11	No	Remarks: The expert panel considers that the use of statins in people with stage 5 chronic kidney disease has an uncertain benefit and probably increases the potential risks and costs… Finally, since dietary restrictions require a change in lifestyle and could result in greater expense, it is important to involve the patient in the decision, informing them of the potential benefits of diet (delay in starting dialysis) versus the risks (malnutrition) and associated costs.	Yes	None	Blank	Blank	Blank	Blank	Blank

Ministerio de Salud, Chile (Chile Ministry of Health) https://diprece.minsal.cl/temas-de-salud/temas-de-salud/guias-clinicas-no-ges/	Chile 2018	Prevention and Treatment of Dental Caries in boys and girls with Primary Dentition	8	No	None	Yes	None	Blank	Blank	Blank	Blank	Blank
Ministerio de Salud, Chile (Chile Ministry of Health) https://diprece.minsal.cl/temas-de-salud/temas-de-salud/guias-clinicas-no-ges/	Chile 2018	Anxiety Disorder Clinic Guide	10	No	Remarks: In lower resource settings, the use of group therapy could be favorable to increase the intervention capacity of the teams and reduce costs.	Yes	None	Blank	Blank	Blank	Blank	Blank
Ministerio de Salud, Chile (Chile Ministry of Health) https://diprece.minsal.cl/te	Chile 2017	Technical aids for Mobility and Tissue Preservation for People	8	No	Remarks: The panel of experts estimated that the potential benefits of performing the interdisciplinary evaluation for the	Yes	None	Blank	Blank	Blank	Blank	Blank

mas-de-salud/temas-de-salud/guias-clinicas-no-ges/		Aged 15 to 64			indication of technical aids possibly outweigh the use of resources and adverse effects in most cases							
Ministerio de Salud, Chile (Chile Ministry of Health) https://diprece.minsal.cl/temas-de-salud/temas-de-salud/guias-clinicas-no-ges/	Chile 2017	Oral Health in Adolescents 10-19 years: Prevention, Diagnosis and Treatment of Periodontal Disease	14	No	None	Yes	None	Blank	Blank	Blank	Blank	Blank
Ministerio de Salud, Chile (Chile Ministry of Health) https://diprece.minsal.cl/temas-de-salud/temas-de-salud/guias-clinicas-no-ges/	Chile 2017	Meningococcal Disease	12	No	Remarks: This recommendation is conditional on patient preferences, cost of using oral cephalosporin, and potential adherence to treatment.	Yes	None	Blank	Blank	Blank	Blank	Blank

Ministerio de Salud, Chile (Chile Ministry of Health) https://diprece.minsal.cl/temas-de-salud/temas-de-salud/guias-clinicas-no-ges/	Chile 2017	Influenza	5	No	Remarks: It is also important to consider that some people may want an accurate diagnosis, so patients and clinicians should balance the potential benefits versus the costs of performing the test together to make a decision	Yes	None	Blank	Blank	Blank	Blank	Blank
Ministerio de Salud, Chile (Chile Ministry of Health) https://diprece.minsal.cl/temas-de-salud/temas-de-salud/guias-clinicas-no-ges/	Chile 2015	Outpatient use of Oseltamivir for people with Suspected or Diagnosed influenza	8	No	Remarks: Considering that in low-risk patients the potential benefits seem to be small magnitude, the panel estimated that the cost and adverse effects of oseltamivir are likely to exceed to the potential benefits.	Yes	None	Blank	Blank	Blank	Blank	Blank
Ministerio de Salud, Chile (Chile Ministry of Health)	Chile 2014	Prevention of Oral Mucositis in People with Cancer	3	No	None	Yes	None	Blank	Blank	Blank	Blank	Blank

https://diprece.minsal.cl/temas-de-salud/temas-de-salud/guias-clinicas-no-ges/												
AUSTRALIA												
Australian Stroke Foundation https://app.magicapp.org/	Australia 2018	Stroke Management - Prehospital Care	2	Yes, No Research Identified	None	Yes	No Research Identified	Blank	Blank	Blank	Blank	Blank
Australian Stroke Foundation https://app.magicapp.org/	Australia 2018	Stroke Management - Early Assessment and Diagnosis	17	Yes, Willingness to Pay per QALY gained, ICER	None	Yes	Yes	Blank	Health Utilities, Procedure and Imaging Prices	Blank	Blank	Blank
Australian Stroke Foundation https://app.magicapp.org/	Australia 2018	Stroke Management - Acute Medical and Surgical Management	37	Yes, ICER	None	Yes	Yes	Blank	Region of Country; Lifetime Time Horizon	Blank	Blank	Blank
Australian Stroke Foundation	Australia 2018	Stroke Management	24	Yes, ICER, life	Remarks: There is no clear evidence that statins	Yes	Yes	Blank	Reference Year;	Blank	Blank	Blank

https://app.magicapp.org/		- Secondary Prevention		year gained	provide any benefit to patients presenting with haemorrhagic stroke and there are concerns about cost and side effects.				Setting; Perspective			
Australian Stroke Foundation https://app.magicapp.org/	Australia 2018	Stroke Management - Rehabilitation	48	Yes, No Research Identified	None	Yes	No Research Identified	Blank	Blank	Blank	Blank	Blank
Australian Stroke Foundation https://app.magicapp.org/	Australia 2018	Stroke Management - Managing Complications	36	Yes, No Research Identified	Remarks: However routine screening for malnutrition is resource intensive.	Yes	No Research Identified	Blank	Blank	Blank	Blank	Blank
Australian Stroke Foundation https://app.magicapp.org/	Australia 2018	Stroke Management - Discharge Planning and Transfer of Care	3	Yes, Total Costs, Cost per patient; Cost Consequence Analysis; ICER	None	Yes	Yes	Provider time; Provider training costs; Length of Stay; Readmission Rate	Provider training; Information interventions	Blank	Blank	Blank

Australian Stroke Foundation https://app.magicapp.org/	Australia 2018	Stroke Management - Community Participation and Long-term Care	7	Yes, Cost Consequence Analysis; ICER	Remarks: There is a lack of evaluation of the cost-effectiveness and prevalence of simulator sickness in the post-stroke population.	Yes	Yes	Blank	Transportation costs; ER visits; Readmissions; Rate of Institutionalization	Blank	Blank	Blank
ASIA												
Japan Society of Respiratory Therapy & Japan Society of Intensive Therapeutics Medicine https://www.jrs.or.jp/modules/guidelines/index.php?content_id=88	Japan 2016	Acute Respiratory Distress Syndrome	15	Yes, Unit Costs	None	Yes	Yes	Price of Intervention; Access; Coverage; Hospital Costs	Not Effective so Not Cost Effective; No New Resources Needed; Provider Time	Blank	Blank	Blank

Abbreviations: EtD - Evidence-to-decision; GDG – guideline development group; ICER – incremental cost effectiveness ratio, QALY – quality adjust life year; DALY – disability adjusted life year

* Sundhedsstyrelsen (Health) – there were 42 guidelines in series; we included the most recent 10 published guidelines to reduce clustering.

Appendix 2: Organization Frequency of Guidelines and Type of Institution

	Organization	Number of Guidelines	Type of Institution
1.	World Health Organization (WHO)	18	United Nations Division
2.	BMJ WikiRecs	16	Commercial Company
3.	European Forum for Research and Education in Allergy and Airway Diseases	1	Non-Government Organization
4.	World Allergy Organization (WAO)	3	Non-Government Organization
5.	Canadian Task Force on Preventive Health Care (CTFPHC)	18	Non-Government Organization
6.	Canadian Chiropractic Guideline Initiative (CCGI)	3	Non-Government Organization
7.	National Pain Centre	1	University Organization
8.	Canadian Blood Services	1	Non-Government Organization
9.	American Society of Hematology (ASH)	6	Non-Government Organization
10.	American Thoracic Society (ATS)	3	Non-Government Organization
11.	Society of Critical Care Medicine (SCCM)	1	Non-Government Organization
12.	National Highway Traffic Safety Administration (NHTSA)	1	Government Agency
13.	European Union (EU)	1	Group of Government Agencies
14.	Nationalt akutkirurgisk tvaerfagligt forum	1	Government Agency
15.	Sundhedsstyrelsen	42	Government Agency
16.	Aarhus Universitetshospital	1	University Organization
17.	Helsedirektoratet	5	Government Agency

18.	Norsk Selskap for Trombose og Hemostase	3	Non-Government Organization
19.	Sykehuset Innlandet	1	Government Agency
20.	Sykehuset Innlandet HF, Avdeling for Fysikalsk medisin og rehabilitering, Ottestad	1	Government Agency
21.	Norsk Ortopdisk Foren	2	Non-Government Organization
22.	Duodecim Medical Publications Ltd	1	Non-Government Organization
23.	EBMeDS® Clinical Decision Support	1	Commercial Company
24.	The Scandinavian Society of Anaesthesiology and Intensive Care Medicine (SSAI)	2	Non-Government Organization
25.	National Prevention Plan Vaccinal (PNPV) 2012-2014	1	Government Agency
26.	Kingdom of Saudi Arabia Ministry of Health	22	Government Agency
27.	Ministerio de Salud, Chile	9	Government Agency
28.	Australian Stroke Foundation	8	Non-Government Organization
29.	Japan Society of Respiratory Therapy & Japan Society of Intensive Therapeutics Medicine	1	Non-Government Organization
		174	
There were a total of 174 Guidelines; The one agency with 42 guidelines (#15), we only took a sample of 10 guidelines to saturation. There were an average of 6 guidelines per guideline Organization			

Appendix 3: Themes Related to Remarks Section Used to Justify Recommendations in EtD Frameworks of Guidelines

Recommendation Remarks Section		
Themes of Factors Affecting Recommendation Being Justified by Economic Information	**Number of New Sub-Themes**	**Sub-Theme Endorsements**
Intervention	3	Price (low/high) (22), Mode of Delivery (4), Frequency (3), Time to Administer (2), Effectiveness (2), Dose (1)
Population	0	Sub-group Risks (3), Access (2), Country (2), Setting (1)
Payor	3	Payor (2), Perspective (1), Program Implementation (1), Socioeconomics (1), Economies of Scale (1)
Provider	1	Training (1)
Healthcare Resource Use	1	Equipment Available (2)
Recommendation Justification Considered Cost Effectiveness	5	Favours one cost-effective option over another (11), High cost versus low effectiveness (not cost-effective) (8), Incremental cost very small and did not play into decision (7), Uncertainty about incremental costs (5); the incremental cost balance was similar between interventions (4)

Appendix 4: Themes of Economic Information Variability Across Domains from EtD Frameworks of Guidelines

Units Costs		
Themes of Factors Affecting Economic Information Variability	**Number of Sub-Themes**	**Sub-Theme Endorsements**
Intervention	7	Price (57), Access/Travel (11), Frequency (8), Procurement (4), Stockpiles (2), Repeat Visits (2), Generics (1)
Population	6	Country (9), Setting (9), Adverse Events/ False Positives (5), Sub-group Risks (3), Endemicity/transmission (2), Burden (1)
Payor	5	Coverage (12), Program Costs (8), Socioeconomics (2), Budget (1), Private Insurance (1)
Provider	4	Fees/Training/Competency (20), Standard of Practice (4), Outreach (2), Opportunistic Encounters (2)
Healthcare Resource use	7	Hospital/Length of Stay (10), Equipment/Labs (7), Space/Overhead (7), Surveillance (3), Implementation (2), Telemedicine/mailing (2), Tool Licensing Fees (1)
Economic Model	1	Perspective (2)
Cost Effectiveness		

Themes of Factors Affecting Economic Information Variability	Number of Sub-Themes	Sub-Theme Endorsements
Intervention	3	Frequency (11), Price (9), Effectiveness (6), Type (8), Access (5), Waste/False Positives (4)
Population	3	Sub-group Risks (14), Setting (9), Adverse Events (6), Country (4), Transmission (2), Work Lost (2), Lives Saved (2)
Payor	0	Coverage (4), Regional Programs (3); Budget (2)
Provider	0	Fees/Training/Competency (6)
Healthcare Resource Use	0	Hospitalization Rate (6); Monitoring/lab Costs (5)
Economic Model	4	Perspective (9), Time Horizon (6), Sensitivity (3), Transferability (1), Reference Year (1)
Acceptability		
Themes of Factors Affecting Economic Information Variability	**Number of Sub-Themes**	**Sub-Theme Endorsements**
Intervention	1	Price (4), Access (2), Effectiveness (2), Burden (2), Frequency (1)
Population	1	Country (3), Setting (3), Sub-group Risks (2), Travel to Remote Locations (2)
Payor	4	Payor (8), Program Costs (3), Policy (1); Risk Management (1), Raising Awareness (1), Advocacy Groups Interests (1)
Provider	1	Standard of Care (2); Provider Acceptance (1)
Healthcare Resource Use	1	Monitoring (1)
Feasibility		
Themes of Factors Affecting Economic Information Variability	**Number of Sub-Themes**	**Sub-Theme Endorsements**
Intervention	0	Price (3), Access (3), Type (1), Generics (1)
Population	1	Setting (2), Country (1), Sub-group Risks (1), Compliance (1)
Payor	1	Payor (2), Coverage (1), Donor Support (1), Budget (1)
Provider	2	Provider Type (2); Volunteers (1)
Healthcare Resource Use	1	Infrastructure (1)
Equity		
Themes of Factors Affecting Economic Information Variability	**Number of Sub-Themes**	**Sub-Theme Endorsements**
Intervention	0	Price (5), Access (5), Effectiveness (3), Frequency (1)
Population	4	Country/Setting (3), Marginalized Populations (1), Sub-group Risks (1), Illness Severity (1), Caregiver Burden (1)
Payor	0	Coverage (8), Socioeconomic Status (6), Policy (2), Payor (1), Programs (1)
Healthcare Resource Use	0	Monitoring (1)

CHAPTER 4. SELECTION OF ECONOMIC EVALUATIONS FOR USE IN GRADE CLINICAL PRACTICE GUIDELINES: A SYSTEMATIC REVIEW OF TRANSFERABILITY (INDIRECTNESS) FACTORS

Status: Final draft manuscript, to be submitted for publication in *Journal of Clinical Epidemiology*.

Selection of Economic Evaluations for use in GRADE Clinical Practice Guidelines: a Systematic Review of Transferability (Indirectness) Factors

Authors

John J. Riva[1,2,3], Meha Bhatt[3], David J. Brunarski[4], Jason W. Busse[5,6,7], Carolina C. Martins[8], Feng Xie[3,9], Holger J. Schünemann[2,3,10], Jan L. Brozek[2,3,10]

[1] Department of Family Medicine, McMaster University, Hamilton, Ontario, Canada.

[2] Michael G. DeGroote Cochrane Canada and McMaster GRADE Centres, Hamilton, Ontario, Canada.

[3] Department of Health Research Methods, Evidence & Impact, McMaster University, Hamilton, Ontario, Canada.

[4] Department of Pediatric Dentistry, Federal University of Minas Gerais, Belo Horizonte, Brazil.

[6] Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands.

[6] Department of Anesthesia, McMaster University, Hamilton, Ontario, Canada.

[7] Michael G. DeGroote National Pain Centre, McMaster University, Hamilton, Ontario, Canada.

[8] Chronic Pain Centre of Excellence for Canadian Veterans, Hamilton, Ontario, Canada.

[9] Centre for Health Economics and Policy Analysis (CHEPA), McMaster University, Hamilton, Ontario, Canada.

[10] Department of Medicine, Hamilton, Ontario, Canada.

Corresponding author: John J. Riva, Department of Family Medicine, David Braley Health Sciences Centre, 100 Main Street West, 6th Floor, Hamilton, ON L8P 1H6; email: rivaj@mcmaster.ca

Previous presentations: Presented as an oral presentation at the full GRADE working group meeting, June 13, 2019 in Hamilton, Ontario.

Key points

1. Population, intervention and comparison elements of a research question, resource use estimation and methodology, and acceptability are the most important indirectness (transferability) study characteristics to consider when choosing to economic evidence for recommendation decision-making.
2. Guideline development teams would benefit from an economics competency when assessing indirectness of economic literature.

Abstract

Background: GRADE practice guideline developers often perform a systematic review of potential economic evaluations to inform recommendation decision-making. Little is known about what study characteristics related to indirectness of economic evidence influence the selection of these articles.

Purpose: A systematic review to identify indirectness characteristics regarding economic evaluations related to GRADE evidence-to-decision (EtD) theoretical frameworks.

Data sources: MEDLINE, EMBASE, CINAHL and EconLit, through May 2020.

Article selection: Four reviewers screened citations to identify articles of any type that explored study characteristics most important or relevant to economic evaluation transferability, restricted to English language.

Data extraction: Pairs of reviewers independently extracted data. We generated frequencies of article features and used thematic analysis to summarize study characteristics related to transferability. We assessed the certainty in the evidence with GRADE-CERQual.

Data synthesis: We included 57 articles in total, with 21 systematic reviews, 13 commentaries, 10 checklist developments, 7 literature summaries, 3 overviews using a data set, and 3 survey approaches. We identified 8 general themes and 28 sub-themes most important to transferability from 41% of articles. Moderate-to-high confidence in the evidence suggested that the GRADE EtD domains of population, intervention and comparison elements of a research question, resource use estimation and methodology, and provider and decision-maker acceptability are the most important indirectness study characteristics that economists consider when choosing an economic evaluation for use in recommendation decision-making.

Conclusion: We have identified factors important for guideline developers to consider when

selecting economic evaluations for use as research evidence. An economic competency on the development team facilitates these endeavors. This supports the GRADE Working Group's tenant of transparent reporting or availability of sufficient information elsewhere to assess indirectness.

Keywords: economics, transferability, cost-effectiveness, GRADE, indirectness, guidelines.

Background

When producing GRADE guidelines, the development process often involves gathering and reporting economic information for each research question to support the completion of Evidence-to-Decision frameworks by a guideline development group.[1] In considering economic evidence, a laudable goal would be to develop new or adapt existing economic evaluations for each research question, but these approaches require significant competency, time and funding.[2] A review of completed and published GRADE guideline Evidence-to-Decision frameworks (median of 6 research questions per guideline) noted that only 2 of 142 (1%) guidelines completed a new or adapted economic evaluation as part of their development process.[3] Therefore, rather than developing a new economic evaluation or re-running an existing one using local inputs, the most common strategy GRADE development teams use is to perform a systematic review of economic evaluations for each research question, which they use as research evidence.

In an absolute sense, some economists have suggested that an economic evaluation can never be transferred from one jurisdiction to another; yet, it is common practice to use effectiveness data from other settings and countries.[4] The GRADE approach handles this uncertainty primarily through assessment of the indirectness domain;[5,6] but, which study characteristics weigh into selection of economic evaluations for adaption from different contexts is not well understood. As well, deciding on whether to use an economic model or outputs from multiple models remains uncertain.[7] Thus, guideline development groups using GRADE may be provided with economic information that may be too indirect for panel decision-making. In economic literature this spectrum of indirectness is most commonly referred to as transferability and this gap is

highlighted by an overview of published systematic reviews of economic evaluations since 2015 that found only 3 of 202 (1.5%) such evaluations assessed the transferability of their review findings.[8]

Some economists have suggested that the term 'generalizability' (extrapolating to the whole population) best reflects this ability to directly apply study findings to other settings and the term 'transferability' is more specific to either adaptation[9] or the degree to which the study holds true in a different setting;[10] although, most often these terms are used synonymously.[11] The GRADE Work Group, on the other hand, groups concepts of transferability and generalizability together with similar concepts, such as external validity and translatability under the domain of indirectness.[12]

Purpose

To 1) perform a systematic review of economic publications that report on indirectness considerations when selecting economic evaluations for use in clinical practice guidelines and 2) categorize these characteristics into common domains of the GRADE Evidence-to-Decision framework.

Methods

Data sources and searches

We searched Medline, EMBASE, CINAHL and EconLit from inception to May 27, 2020, without language restrictions for any article reporting transferability study characteristics of

economic evaluations, with specific search terms related to economics ('economics' or 'cost-effectiveness') and transferability ('transferability', 'generalizability', 'generalisability' or 'external validity') and "exclude Medline" filters when possible. Two authors (JJR, DJB) reviewed reference lists of eligible articles and most common journal sources with targeted terms ('transferability', 'generalizability', or 'generalisability') for additional articles, in a 'snowballing' approach,[13] similar to others who considered the concept of transferability in a prior systematic review.[14]

Article selection

Three reviewers, acting in pairs, (JJR, MB, CCM) screened the titles and abstracts of identified and, subsequently, the full texts of potentially eligible studies, with disagreements resolved by consensus. We included articles that explored general study characteristics of transferability related to use of economic evaluations in other contexts. Articles were ineligible if: 1) they were non-English language, 2) they were conference abstracts, 3) they only considered indirectness study characteristics specific to a clinical situation or setting that would not be relevant outside of this context or 4) they considered methodologies specifically for adaptation. To avoid clustering (i.e. double counting) of reporting study characteristics, we also excluded articles that used a pre-existing transferability checklists of study characteristics verbatim, we already identified, without providing any additional information.

Data extraction and confidence in the evidence

We used a MS-Word table form, which was piloted and finalized with a sample of 21 articles to extract demographic information for each article (i.e. publication year, country of principle

investigator, journal source, type of article), publication characteristics (i.e. authorship team, how they defined importance of study characteristics), and free text quotations from articles.

We used the domains (methodological limitations, coherence, adequacy of data, relevance) of the GRADE Confidence in the Evidence from Reviews of Qualitative research (GRADE-CERQual) approach to assess the certainty in the evidence for most important themes and sub-themes of transferability study characteristics reported across included articles.[15] We considered: 1) methodological quality via limitations of underpinning publication approach (articles using only example references, using a literature summary or using a full systematic review) or clarity of interview methods; 2) coherence via fit between article findings and underpinning approach through a clear rationale from data to reporting in the article; 3) quantity of data through number of articles, participants or authors informing transferability study characteristics and richness of data through clarity of transferability study characteristic descriptions; and 4) relevance via assessing indirect, partial or unclear relevance of the transferability study characteristic from an article to the author definition of 'most important'. Two reviewers (JJR, CCM) performed these steps, independently and in duplicate, with discrepancies resolved through consensus.

Data synthesis

We generated frequencies for all abstracted demographic information content. To summarize transferability study characteristics reported by articles as most important, we performed thematic analysis for free text information, with one reviewer (JJR) using some of the domains of the GRADE evidence-to-decision theoretical framework (Research question – Population, Intervention, Comparison, Outcomes [PICO], Patient Values and Preferences, Resource Use, Acceptability, Feasibility, Equity)[16] as initial codes. We added additional themes and sub-themes

to the start codes that emerged and ranked them using a previously established approach.[17] One reviewer (MB) checked these for consistency, with any discrepancies resolved through consensus.

To summarize transferability study characteristics across articles, we performed a thematic analysis to categorize and rank information, but we used different initial codes. Due to the very large number of potential codes, we used the summary framework from a previously published systematic review by economists of 86 articles as initial codes and added any new sub-themes that emerged. The authors of that review noted they were unable to infer which transferability study characteristic themes and sub-themes were most important to consider based on the published literature available to them at that time.[18] Finally, two reviewers (JJR, JLB) recategorized the final summary of ranked economists' themes and sub-themes to the match the GRADE Evidence-to-Decision theoretical framework, allowing for cross-comparison with themes considered "most important". To avoid clustering of themes, within included articles, we only coded once: 1) from overlapping articles using the same references to support their arguments or 2) from systematic reviews, which used similar references and presented aggregate summaries.

Results

Of 3,198 unique records, we retrieved 180 articles in full text for review of which 57 proved eligible (Figure 1).[2,9-11,18-70] We found 9 of the included articles[26,36,38,48,51,63,64,69] by checking reference lists. For 6 of our 57 included articles, we only coded general transferability themes once from these overlapping articles to avoid clustering of themes.[20,41,42,66-68]

Article characteristics

With respect to country of origin, 28 of 57 (49.1%) article lead authors were from Europe and North America, particularly the United Kingdom), 11 of 57 (19.3%) from The Netherlands and 8 of 57 (14.0%) from Canada (Table 1). Of articles, 21 of 57 (36.8%) were systematic reviews of economic evaluation methodologies,[22,42] checklists,[19,26] factor variability,[18,58] guideline transferability,[43,48] and overviews of economic evaluations; as well, 13 of 57 (22.8%) were commentaries. Authorship of 36 of 57 (63.2%) articles had small or individual authorship teams and reflected an older literature base, as they were either published or used references in their reviews prior to the year 2005. The three most common journals sources were with 11 of 57 (19.3%) articles from PharmacoEconomics, 7 of 57 (12.3%) from Value in Health and 5 of 57 (8.7%) from the International Journal of Technology Assessment in Health Care (Appendix 1: Included Articles).

Confidence in the evidence for "most important" transferability themes

We assessed the CERQual domains per study (Appendix 2: GRADE-CERQual Evidence Profile) and a summary of findings across studies (Appendix 3: GRADE-CERQual Summary of Findings & Coding of Most Important Factors). Generally, there were methodological limitations across 12 of 23 (52.2%) articles due to a lack of empirical data from review articles and use of commentary approaches to explain themes and sub-themes. With respect to coherence,[71] the fit between important themes and the underpinning evidence, there were some minimal-to-moderate concerns across 9 of 23 (39.1%) articles about how authors progressed from the supportive evidence they cited and their themes reported as most important. Adequacy of the data was limited particularly around the richness of descriptions of methodological transferability factors

and small quantities of observations due to either small or single authorship teams in 26.1% (6 of 23) articles. Also, many final sub-themes identified were only reported by a single article. Lastly, relevance of the definition,[72] across 11 of 23 (47.8%) articles appeared only partially relevant to the specific concept of a "most important" theme to transferability.

Themes of transferability

Overall, we identified 8 general themes from the GRADE Evidence-to-Decision framework (population and disease, intervention and funding approach, usual care and funding approach, outcomes, patient values and preferences, resource use estimation and methodology, acceptability, feasibility) and 28 sub-themes in 23 of 56 (41.1%) articles that considered study characteristics of indirectness as most important (Table 2). Outside of the GRADE EtD framework, but a central tenant to the GRADE approach,[16] we identified a key theme of transparency of reporting in 12 of 23 (52.2%) articles with high certainty in the evidence and three sub-themes describing: 1) country specified; 2) generalisability discussed; and 3) language translatable, with low-to-very low certainty in the evidence. This was further supported by 23 of 131 (17.6%) other articles endorsing transparency of reporting as a general theme (Appendices 4 & 5: Coding of General Transferability Factors).

Table 2 shows the remaining themes and sub-themes as they relate to the GRADE evidence-to-decision framework, and they are described below as comparable research question (PICO) elements, values and preferences, resource use, acceptability, feasibility and equity: (Table 2)

Comparable research question (PICO) elements

High certainty in the evidence suggested themes of general population from 11 of 23 (47.8%) articles, intervention and funding approach from 8 of 23 (34.7%) articles and comparison (usual care) and funding approach from 7 of 23 (30.4%) articles, as well as, low certainty in the evidence suggested outcomes from 6 of 23 (26.1%) articles, were most important factors to consider when assessing indirectness. Common sub-themes for most important population factors related to demographics (e.g. age, sex, socioeconomic status), disease epidemiology (e.g. incidence, prevalence, progression) and the setting of the interventions, with moderate-to-high certainty in the evidence; other sub-themes such as disease severity, case-mix, mortality rate, life expectancy, co-morbidities and lifestyle were reported in articles, although not specified as most important. The sub-theme of comparable efficacy was the most important intervention factor, with moderate certainty in the evidence. Common sub-themes for most important comparison factors and funding approach related to availability of treatment options, substitutes and clinical practice norms and guidelines, with moderate certainty in the evidence. Comparable end points and valid outcome measures were the most important outcome sub-themes, but were associated with very low certainty in the evidence.

Comparable values and preferences, resource use, acceptability, feasibility, equity

Low certainty in the evidence from 2 of 23 (8.7%) articles suggested that general patient values and preferences were a most important theme when assessing indirectness. Common sub-themes for values and preferences factors related to health states (utilities), as well as adherence and compliance, with very low certainty in the evidence. Also, patients' intervention attitudes, religion, culture, hygiene, nutrition were sub-themes reported in articles, although not specified as most important.

High certainty in the evidence from 11 of 23 (47.8%) articles suggested that general resource use estimation and methodology was a most important theme when assessing indirectness. Common sub-themes for resource use estimation and methodology related to absolute and relatives prices, sensitivity analysis, estimation procedures, study perspective and model representativeness, with high-to-moderate confidence in the evidence. Also, low-to-very low certainty in the evidence suggested sub-themes of study conditions, conflicts of interest, exchange rate, time horizon, opportunity costs, and input mixes of equipment and personnel were most important.

Moderate certainty in the evidence from 2 of 23 (8.7%) articles suggested that general provider and decision-making acceptability were most important themes. Common sub-themes for acceptability related to incentives for providers and institutions, as well as decision-maker discount rates, willingness-to-pay thresholds and methods of remuneration, with moderate-to-very low certainty in the evidence.

While feasibility, such as comparable availability of staff, equipment, programs, delivery system, training, competency, technology advancement, regulation and potential for economies of scales were sub-themes reported in articles, they were not specified as most important. Lastly, no transferability themes or sub-themes were identified related to the GRADE EtD domain of equity.

Discussion

Brief Summary

Overall, our systematic review of 57 publications on concepts related to indirectness that are common in the economic scientific literature found that population, intervention and comparison elements of a research question PICO, resource use estimation and methodology, and provider and decision-maker acceptability are the most important study characteristics to consider when choosing an economic evaluation for use as research evidence for recommendation decision-making. Also, this information must be transparently reported by the economic evaluation or be available from other sources in sufficient detail (e.g. unit cost and resource use reported separately) to assess their level of indirectness, rather than being data that "happen to be available".[73]

As it is often the case that detailed information about economic evidence may not go into long explicit descriptions in scientific publications,[10,73] some effort may be required to find specific information from other sources. Therefore, a fulsome consideration of the research question, as it applies to the potential to adapt an existing economic evaluation is advisable.

Key Findings

Economists suggest that population, intervention and comparison elements of a research question PICO, resource use estimation and methodology, and provider and decision-maker acceptability are the most important indirectness study characteristics to consider when choosing an economic evaluation to inform the development of a guideline recommendation. As example, population-based screening may weigh the cost implications of high false-positives, such as in breast cancer screening.[74] Also, for questions on highly infectious disease populations like COVID-19, spread may vary from country-to-country, thereby affecting economic outcome estimates. Alternatively,

population disease severity may have cost consequences for questions based on frail older adults or for questions in HIV patients with co-morbidities.

Intervention characteristics may weigh substantially by differing long-term costs of medications like warfarin versus direct oral anticoagulants in the prevention of venous thromboembolism.[75] Even the conduct of an associated effectiveness trial, used to inform inputs of an economic evaluation, could generate unrealistic protocol-driven resource use estimates compared to use of a more pragmatic intervention.[27,44,69] Research question comparison characteristics may also have substantially differing cost implications when choosing between an inexpensive medication formulation or expensive non-pharmacological program alternative for chronic pain management.[76]

Usual care comparisons performed in low- and high-income countries may too influence results dramatically; so much so, that the World Health Organization have considered an international "reference case", by developing a null comparator - a hypothetical situation where all interventions were halted to allow for cross-country comparison.[77]

With respect to resource use estimation and methodology, various instruments have been developed to assess the quality of a cost-effectiveness study.[50,53,78,79] Some have also considered the more pragmatic use of reporting checklists; however, it has been argued that a poorly done economic evaluation that is reported thoroughly will still score highly and relying on reporting scores limits the ability to rank multiple economic evaluations against each other.[50] Others have gone further to suggest reporting is only the first step in a more explicit evaluation process, which therefore may act as an educational process for development teams.[80]

Provider and decision-maker acceptability in the decision-context also weighs into the ability of an economic evaluation to remain useful in another context – gathering of such information may rely largely on information gathered from outside of the economic evaluation itself through panel members, literature queries or external stakeholder consultation. It has long been understood that health care providers carry inherent and variable assumptions of disease and values they place on alternative interventions.[81] As well, willingness-to-pay thresholds would need to be comparable as an intervention may be cost-effective in one setting and not in another solely based on the differences in the decision-maker threshold.[39,57]

When reviewing transferability as a whole, two observations become apparent. First, similar to suggestions by members of the GRADE working group,[82] World Health Organization,[83] and others,[84] the addition of a competency in economics to a guideline development team would be useful, as assessing indirectness may go beyond the skillset of typical guideline development groups.[73] Secondly, some of the transferability characteristics are easier than others to find and assess for indirectness. Welte 2004 et al.[20] started this process and suggested that transferability characteristics, such as: perspective, discount rate, cost approach would require very little effort; establishing similar life expectancy would require a small amounts of effort; and establishing comparable practice norms, health status (utilities), absolute and relative prices, technology availability, and acceptability may require moderate-high amounts of effort to acquire information and assess indirectness.

Decision-making bodies may increasingly be looking at adapting cost-effectiveness evidence generated from other jurisdictions,[85] especially in situations when local data or funding to

conduct an economic evaluation is limited.[2,43] Even though it comes with pitfalls about applicability, recruitment of guideline group members for integrated knowledge translation has recently been suggested by modellers as means to improve the reliability of their products.[86] While the summary of such economic evidence should be clearer and simpler for panel members, the process for assessment of transferability (indirectness) remains complicated and similarly may act as basis for improved discourse between clinician panels making practice recommendations and economists traditionally making coverage decisions.

Strengths and limitations

A strength of our review is that we applied a formalized systematic process, with duplication of steps, to identify, summarize and assess the certainty in the evidence for themes and sub-themes suggested by economists that are important to transferability of economic evaluations. Also, to our knowledge this review, is the first to assign a degree of quality to inferences related to study characteristics of economic evaluations that are most important to indirectness.

However, our review is not without limitations. There was generally a dearth of empirically-based information to support various sub-themes related to transferability. While we did assess our certainty in the evidence through the GRADE-CERQual domain of methodological limitations,[87] it could be argued that our metric of article type may have been too forgiving – the transparency of the GRADE approach allows for others though to reapply more stringent confidence thresholds to fit their needs. As well, while many authors of included articles have described "most important" transferability themes and sub-themes, they may not have necessarily had this ranking intent.

Even though we did approach our literature search intentionally using a snowballing search strategy[13] based on experiences of others who have previously looking for articles related to transferability,[14] 9 of our 57 (16%) included articles were identified outside of our formal search strategy.

We relied on the coding by Goeree et al. 2007 systematic review,[18] as start codes and frequency counts for sub-themes,[88] which was based on 86 articles from previous older literature, identified often by hand searching entire paper journals, when library index terms about transferability was quite variable or absent in older year publications. These authors also reported they were unable to establish which factors were most important at the time based on the literature available to them up to 2005;[18] it is possible there were articles they included that made such suggestions. Although, we reviewed the original article titles from their reference list to check for and include any "important factor" themes or sub-themes that were reported in these individual articles.[20,40-42,67,68,89,90]

We considered all themes and sub-themes in the context of the GRADE domain of indirectness, as this was felt most important by GRADE working group members who have previously considered the concept of resource use.[5] Some of these indirectness factors may appear to cross over with other GRADE domains, such as risk of bias.[91] Likewise, individual transferability themes and sub-themes we identified may not have always been mutually exclusive. Indeed, both risk of bias and indirectness have in common that they express a systematic deviation from the truth. However, risk of bias typically relates to issues in the study design and execution, whereas, indirectness relates to the use of study findings in the interpretation of the results with a focus on the PICO domains.

Implications for practice

Guideline development groups using the GRADE EtD framework should aim to always have a competency in economics on their development teams, if they wish to consider the indirectness of published cost and cost-effectiveness literature for use in their recommendation decision-making. Also, it may also be helpful for developers to consider having an explicit process step that allows for a prioritization by the guideline panel on research questions most benefited by fulsome economic literature and a presentation by the development team, with an economics competency, on the indirectness limitations of existing published literature.

It could be argued that all published economic literature will have a degree of indirectness to it, and some economists have gone so far to note that economic models can never be adapted outside of their originally intended context. While it is commonplace to extrapolate the benefits and harms of effectiveness literature and outcomes (i.e. results from randomized controlled trials) to other contexts, making efforts with cost-effectiveness literature to differentiate more familiar effectiveness indirectness from the funding approach of an intervention indirectness may offer panel members and end-users of GRADE EtD frameworks a means to apply a level of magnitude of indirectness to available literature in the short term, until such time as a more fulsome GRADE economic model indirectness strategy is available.

Implications for research

The formalized assessment of indirectness of economic evaluations may serve to improve discourse between economists, guideline panels, systematic review teams and GRADE methodologists in guideline development. This also has the potential to save guideline developers' time, as well as a funders' costs; future research should assess the degree, budget implications and timing at which individuals with an economics competency should be brought into such situations.

Future research should also consider both the effort and the type of guideline team member required to find and assess these economic evaluation study characteristics of transferability in the context of guideline development process and budget. Preliminary work by Sanabria et al.[92] described the potential timing of such resource use gathering steps. Certain transferability study characteristics could be found by an evidence review team (e.g. perspective, discount rates, utilities), determined through input from the panel (e.g. practice norms, technology availability), clarified by decision-makers (willingness-to-pay thresholds), or judged by input from an individuals with economics and/or GRADE methodology competencies (e.g. absolute, relative costs, model representativeness).

Lastly, a next important step would be the development of a GRADE economic evaluation indirectness instrument, which balances ease and rigour, for use by guideline development groups in selecting either the most appropriate one or group of economics evaluations having a level of indirectness deemed useable for research evidence. These judgments should be a result of a cumulative assessment of indirectness of the input variables into an economic model as it relates to the indirectness of values (utilities)[93] and economic models,[7] but also related to the input variables that are more directly related to cost estimates.

Conclusion

Economists suggest that population, intervention and comparison elements of a research question PICO, resource use estimation and methodology, and provider and decision-maker acceptability are the most important study characteristics of transferability to consider when choosing an economic evaluation to inform the development of a guideline recommendation. Guideline developers should always strive to include a competency in economics on their teams to assess the indirectness of these study characteristics in published literature. This strategy relies on the GRADE tenant of transparent reporting or availability of sufficient information elsewhere to assess indirectness.

Acknowledgements

We would like to thank Laure Tessier, MSc for reviewing the manuscript and providing guidance on the theme summary approach.

Authorship contributions

Conception and design: JJR, JLB

Data acquisition: JJR, MB, CCM, DJB

Data analysis: JJR, MB, CCM

Interpretation of results: JJR, MB, CCM, JWB, FX, HJS, JLB

Manuscript drafting: JJR, JLB

Critical revision of the manuscript and approval of the final version: JJR, MB, CCM, DJB, JWB, FX, HJS, JLB

Disclosure of Conflicts of Interest

This study was unfunded. JJR is supported by a PhD training award from the NCMIC foundation (https://www.ncmicfoundation.org/). The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Disclosure of Competing Interests

JJR is a Member of the Canadian Task Force on Preventive Health Care (https://canadiantaskforce.ca/) and the GRADE working group (https://www.gradeworkinggroup.org/). HJS is the co-chair of the GRADE working group. These groups had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

References

1. Moberg J, Oxman AD, Rosenbaum S, Schünemann HJ, Guyatt G, Flottorp S, Glenton C, Lewin S, Morelli A, Rada G, Alonso-Coello P; GRADE Working Group. The GRADE Evidence to Decision (EtD) framework for health system and public health decisions. Health Res Policy Syst. 2018 May 29;16(1):45.
2. Hutton G, Baltussen R. Cost valuation in resource-poor settings. Health Policy Plan. 2005 Jul;20(4):252-9.
3. Riva JJ, Bhatt M, Brunarski DJ, Busse JW, Martins CC, Xie F, Schunemann HJ, Brozek JL. Reporting of Economic Information in GRADE Guidelines that use Evidence-to-decision (EtD) Frameworks: a Systematic Survey. J Clin Epi (pending submission) 2020.
4. Gulácsi L. The time for cost-effectiveness in the new European Union member states: the development and role of health economics and technology assessment in the mirror of the Hungarian experience. Eur J Health Econ. 2007 Jun;8(2):83-8.
5. Brunetti M, Shemilt I, Pregno S, Vale L, Oxman AD, Lord J, Sisk J, Ruiz F, Hill S, Guyatt GH, Jaeschke R, Helfand M, Harbour R, Davoli M, Amato L, Liberati A, Schünemann HJ. GRADE guidelines: 10. Considering resource use and rating the quality of economic evidence. J Clin Epidemiol. 2013 Feb;66(2):140-50.
6. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Falck-Ytter Y, Jaeschke R, Vist G, Akl EA, Post PN, Norris S, Meerpohl J, Shukla VK, Nasser M, Schünemann HJ; GRADE Working Group. GRADE guidelines: 8. Rating the quality of evidence--indirectness. J Clin Epidemiol. 2011 Dec;64(12):1303-10.
7. Brozek JL, Canelo-Aybar C, Akl EA, Bowen JM, Bucher J, et. al.; GRADE Working Group. GRADE Guidelines 30: The GRADE Approach to Assessing the Certainty of Modelled

Evidence - an Overview in the Context of Health Decision-making. J Clin Epidemiol. 2020 Sep 24:S0895-4356(20)31103-3. doi: 10.1016/j.jclinepi.2020.09.018. Epub ahead of print.

8. Luhnen M, Prediger B, Neugebauer EAM, Mathes T. Systematic reviews of health economic evaluations: A structured analysis of characteristics and methods applied. Res Synth Methods. 2019 Jun;10(2):195-206.
9. Drummond M, Barbieri M, Cook J, Glick HA, Lis J, Farzana M. Transferability of economic evaluations across jurisdictions: ISPOR Good Research Practices Task Force report. Value Health. 2009; 12: 409-18.
10. Knies S, Ament AJ, Evers SM, Severens JL. The transferability of economic evaluations: testing the model of Welte. Value Health. 2009;12(5):730-38.
11. Mason JM, Mason AR. The generalisability of pharmacoeconomic studies: issues and challenges ahead. Pharmacoeconomics. 2006;24(10):937-45.
12. Schünemann HJ, Tugwell P, Reeves BC, Akl EA, Santesso N, Spencer FA, et al. Non-randomized studies as a source of complementary, sequential or replacement evidence for randomized controlled trials in systematic reviews on the effects of interventions. Research Synthesis Methods. 2013;4(1):49-62.
13. Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. BMJ. 2005 Nov 5;331(7524):1064-5.
14. Schloemer T, Schröder-Bäck P. Criteria for evaluating transferability of health interventions: a systematic review and thematic synthesis. Implement Sci. 2018 Jun 26;13(1):88.
15. Lewin S, Booth A, Glenton C, Munthe-Kaas H, Rashidian A, Wainwright M, Bohren MA, Tunçalp Ö, Colvin CJ, Garside R, Carlsen B, Langlois EV, Noyes J. Applying GRADE-CERQual to qualitative evidence synthesis findings: introduction to the series. Implement Sci. 2018 Jan 25;13(Suppl 1):2.

16. Alonso-Coello P, Schünemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, Treweek S, Mustafa RA, Rada G, Rosenbaum S, Morelli A, Guyatt GH, Oxman AD; GRADE Working Group. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. BMJ. 2016 Jun 28;353:i2016.
17. Green J, Willis K, Hughes E, Small R, Welch N, Gibbs L, Daly J. Generating best evidence from qualitative research: the role of data analysis. Aust N Z J Public Health. 2007;31:545-50.
18. Goeree R, Burke N, O'Reilly D, Manca A, Blackhouse G, Tarride JE. Transferability of economic evaluations: approaches and factors to consider when using results from one geographic area for another. Curr Med Res Opin. 2007 Apr;23(4):671-82.
19. Goeree R, He J, O'Reilly D, Tarride JE, Xie F, Lim M, Burke N. Transferability of health technology assessments and economic evaluations: a systematic review of approaches for assessment and application. Clinicoecon Outcomes Res. 2011; 3:89-104.
20. Welte R, Feenstra T, Jager H, Leidl R. A decision chart for assessing and improving the transferability of economic evaluation results between countries. Pharmacoeconomics. 2004; 22(1):857-76.
21. Heyland DK, Kernerman P, Gafni A, Cook DJ. Economic evaluations in the critical care literature: do they help us improve the efficiency of our unit? Crit Care Med. 1996; 24(9):1591-8.
22. Späth HM, Carrère MO, Fervers B, Philip T. Analysis of the eligibility of published economic evaluations for transfer to a given health care system. Methodological approach and application to the French health care system. Health Policy. 1999; 49(3):161-77.

23. European Network for Health Technology Assessment (EUnetHTA), HTA Adaptation Toolkit, Work Package 5 (October 2011). Available at: https://eunethta.eu/wp-content/uploads/2011/01/EUnetHTA_adptation_toolkit_2011_version_5.pdf
24. Antonanzas F, Rodríguez-Ibeas R, Juárez C, Hutter F, Lorente R, Pinillos M. Transferability indices for health economic evaluations: methods and applications. Health Econ. 2009; 18(6):629-43.
25. Boulenger S, Nixon J, DruMmond M, Ulmann F, Rice S, de Pouvourville G. Can economic evaluations be made more transferable? Eur J Health Econom. 2005; 6:334-46.
26. Dyrvig AK, Kidholm K, Gerke O, Vondeling H. Checklists for external validity: a systematic review. J Eval Clin Pract. 2014; 20(6):857-64.
27. Burchett HE, Mayhew SH, Lavis JN, Dobrow MJ. When can research from one setting be useful in another? Understanding perceptions of the applicability and transferability of research. Health Promot Int. 2013; 28(3):418-30.
28. Ademi Z, Tomonaga Y, van Stiphout J, Glinz D, Gloy V, Raatz H, Bucher HC, Schwenkglenks M. Adaptation of cost-effectiveness analyses to a single country: the case of bariatric surgery for obesity and overweight. Swiss Med Wkly. 2018; 12;148:w14626.
29. Anderson R. Systematic reviews of economic evaluations: utility or futility? Health Econ. 2010; 19(3):350-64.
30. van Haalen HG, Severens JL, Tran-Duy A, Boonen A. How to select the right cost-effectiveness model? : A systematic review and stepwise approach for selecting a transferable health economic evaluation model for rheumatoid arthritis. Pharmacoeconomics. 2014; 32(5):429-42.
31. Gonzalez-Perez JG. Developing a scoring system to quality assess economic evaluations. Eur J Health Econ. 2002; 3(2):131-6.

32. Fukuda H, Lee J, Imanaka Y. Costs of hospital-acquired infection and transferability of the estimates: a systematic review. Infection. 2011; 39(3):185-99.
33. Wolfenstetter SB, Wenig CM. Economic evaluation and transferability of physical activity programmes in primary prevention: a systematic review. Int J Environ Res Public Health. 2010;7(4):1622-48.
34. Cook J, Drummond M, Heyse JF. Economic endpoints in clinical trials. Stat Methods Med Res. 2004; 13:157-76.
35. Li SC. How to improve pharmacoeconomic data generalisability between different countries. Expert Opin Pharmacother. 2007; 8(10):1409-13.
36. Jaime Caro J, Eddy DM, Kan H, Kaltz C, Patel B, Eldessouki R, Briggs AH; ISPOR-AMCP-NPC Modeling CER Task Forces. Questionnaire to assess relevance and credibility of modeling studies for informing health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. Value Health. 2014; 17(2):174-82.
37. Rutten F. Economic evaluation and health care decision-making. Health Policy. 1996; 36(3)215-29.
38. Bryan S, Brown J. Extrapolation of cost-effectiveness information to local settings. J Health Serv Res Policy. 1998; 3(2):108-12.
39. Steuten L, Vallejo-Torres L, Young T, Buxton M. Transferability of economic evaluations of medical technologies: a new technology for orthopedic surgery. Expert Rev Med Devices. 2008; 5(3):329-36.
40. Drummond M, Pang F. Transferability of economic evaluation results. In: Drummond M, McGuire A, eds, Economic Evaluation in Health Care. Merging Theory with Practice. Oxford: Oxford University Press, 2001: pp.254-76.

41. O'Brien BJ. A tale of two (or more) cities: geographic transferability of pharmacoeconomic data. Am J Manag Care. 1997 May;3 Suppl:S33-9.
42. Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Urdahl H, Davies LM, Eastwood A. Generalisability in economic evaluation studies in healthcare: a review and case studies. Health Technol Assess. 2004; 8(49):iii-iv, 1-192.
43. Barbieri M, Drummond M, Rutten F, Cook J, Glick HA, Lis J, Reed SD, Sculpher M, Severens JL; ISPOR Good Research Practices Economic Data Transferability Task Force. What do international pharmacoeconomic guidelines say about economic data transferability? Value Health. 2010; 13(8):1028-37.
44. Coyle D, Lee KM. The problem of protocol driven costs in pharmacoeconomic analysis. Pharmacoeconomics. 1998; 14(4):357-63.
45. Baltussen R, Leidl R, Ament A. Real world designs in economic evaluation. Bridging the gap between clinical research and policy-making. Pharmacoeconomics. 1999; 16(5 Pt 1):449-58.
46. Abdul Pari AA, Simon J, Wolstenholme J, Geddes JR, Goodwin GM. Economic evaluations in bipolar disorder: a systematic review and critical appraisal. Bipolar Disord. 2014; 16(6):557-82.
47. Gheorghe A, Roberts T, Pinkney TD, Morton DG, Calvert M. Rational centre selection for RCTs with a parallel economic evaluation--the next step towards increased generalisability? Health Econ. 2015; 24(4):498-504.
48. Schünemann HJ, Fretheim A, Oxman AD; WHO Advisory Committee on Health Research. Improving the use of research evidence in guideline development: 1. Guidelines for guidelines. Health Res Policy Syst; 2006 21;4:13.
49. Adam T, Koopmanschap MA, Evans DB. Cost-effectiveness analysis: can we reduce variability in costing methods? Int J Technol Assess Health Care. 2003; 19(2):407-20.

50. Berg ML, Cheung KL, Hiligsmann M, Evers S, de Kinderen RJA, Kulchaitanaroaj P, Pokhrel S. Model-based economic evaluations in smoking cessation and their transferability to new contexts: a systematic review. Addiction. 2017; 112(6):946-967.
51. Knies S, Evers SM, Candel MJ, Severens JL, Ament AJ. Utilities of the EQ-5D: transferable or not? Pharmacoeconomics. 2009; 27(9):767-79.
52. Coyle D, Drummond MF. Analyzing differences in the costs of treatment across centers within economic evaluations. Int J Technol Assess Health Care. 2001; 17(2):155-63.
53. Drummond M, Manca A, Sculpher M. Increasing the generalizability of economic evaluations: recommendations for the design, analysis, and reporting of studies. Int J Technol Assess Health Care. 2005; 21(2):165-71.
54. Drummond M, Augustovski F, Kaló Z, Yang B-M. Challenges faced in transferring economic evaluations to middle income countries. Int J Technol Assess Health Care. 2015; 31(6):442-48.
55. Zwolsman S, Kastelein A, Daams J, Roovers JP, Opmeer BC, WOMEN-UP Consortium: http://www.women-up.eu/consortium/. Heterogeneity of cost estimates in health economic evaluation research. A systematic review of stress urinary incontinence studies. Int Urogynecol J. 2019; 30(7):1045-59.
56. Gray AM, Wilkinson T. Economic evaluation of healthcare interventions: old and new directions. Oxford Review of Economic Policy. 2016; 32(1):102-21.
57. Ginsberg GM. Generalizability of cost-utility analyses across countries and settings. Best Pract Res Clin Gastroenterol. 2013; 27(6):845-52.
58. Munthe-Kaas H, Nøkleby H, Nguyen L. Systematic mapping of checklists for assessing transferability. Syst Rev. 2019; 14;8(1):22.

59. Pichon-Riviere A, Augustovski F, García Martí S, Sullivan SD, Drummond M. Transferability of health technology assessment reports in Latin America: an exploratory survey of researchers and decision makers. Int J Technol Assess Health Care. 2012; 28(2):180-6.
60. Ruggeri M, Manca A, Coretti S, Codella P, Iacopino V, Romano F, Mascia D, Orlando V, Cicchetti A. Investigating the Generalizability of Economic Evaluations Conducted in Italy: A Critical Review. Value Health. 2015; 18(5):709-20.
61. Vemer P, Rutten-van Mölken MP. Crossing borders: factors affecting differences in cost-effectiveness of smoking cessation interventions between European countries. Value Health. 2010; 13(2):230-41.
62. Vemer P, Rutten-van Mölken MP. Largely ignored: the impact of the threshold value for a QALY on the importance of a transferability factor. Eur J Health Econ. 2011; 12(5):397-404.
63. Sculpher MJ, Drummond MF. Analysis sans frontières: can we ever make economic evaluations generalisable across jurisdictions? Pharmacoeconomics. 2006; 24(11):1087-99.
64. Urdahl H, Manca A, Sculpher MJ. Assessing generalisability in model-based economic evaluation studies: a structured review in osteoporosis. Pharmacoeconomics. 2006; 24(12):1181-97.
65. Stawowczyk E, Kawalec P. A Systematic Review of the Cost-Effectiveness of Biologics for Ulcerative Colitis. Pharmacoeconomics. 2018 Apr;36(4):419-34.
66. Baltussen R, Ament A, Leidl R. Making cost assessments based on RCTs more useful to decision-makers. Health Policy. 1996 Sep;37(3):163-83.
67. Birch S, Gafni A. Economics and the evaluation of health care programmes: generalisability of methods and implications for generalisability of results. Health Policy. 2003 May;64(2):207-19.

68. Mason J. The generalisability of pharmacoeconomic studies. Pharmacoeconomics. 1997 Jun;11(6):503-14.

69. Grutters JP, Seferina SC, Tjan-Heijnen VC, van Kampen RJ, Goettsch WG, Joore MA. Bridging trial and decision: a checklist to frame health technology assessments for resource allocation decisions. Value Health. 2011 Jul-Aug;14(5):777-84.

70. Augustovski F, Iglesias C, Manca A, Drummond M, Rubinstein A, Martí SG. Barriers to generalizability of health economic evaluations in Latin America and the Caribbean region. Pharmacoeconomics. 2009;27(11):919-29.

71. Colvin CJ, Garside R, Wainwright M, Munthe-Kaas H, Glenton C, Bohren MA, Carlsen B, Tunçalp Ö, Noyes J, Booth A, Rashidian A, Flottorp S, Lewin S. Applying GRADE-CERQual to qualitative evidence synthesis findings-paper 4: how to assess coherence. Implement Sci. 2018 Jan 25;13(Suppl 1):13.

72. Noyes J, Booth A, Lewin S, Carlsen B, Glenton C, Colvin CJ, Garside R, Bohren MA, Rashidian A, Wainwright M, Tunçalp Ö, Chandler J, Flottorp S, Pantoja T, Tucker JD, Munthe-Kaas H. Applying GRADE-CERQual to qualitative evidence synthesis findings-paper 6: how to assess relevance of the data. Implement Sci. 2018 Jan 25;13(Suppl 1):4.

73. Hill SR, Olson LG, Falck-Ytter Y, Cruz AA, Atkins D, Baumann M, Jaeschke R, Woitalla T, Schünemann HJ; ATS/ERS Ad Hoc Committee on Integrating and Coordinating Efforts in COPD Guideline Development. Incorporating considerations of cost-effectiveness, affordability, and resource implications in guideline development: article 6 in Integrating and coordinating efforts in COPD guideline development. An official ATS/ERS workshop report. Proc Am Thorac Soc. 2012 Dec;9(5):251-5.

74. Klarenbach S, Sims-Jones N, Lewin G, Singh H, Thériault G, Tonelli M, Doull M, Courage S, Garcia AJ, Thombs BD; Canadian Task Force on Preventive Health Care.

Recommendations on screening for breast cancer in women aged 40-74 years who are not at increased risk for breast cancer. CMAJ. 2018 Dec 10;190(49):E1441-E1451.

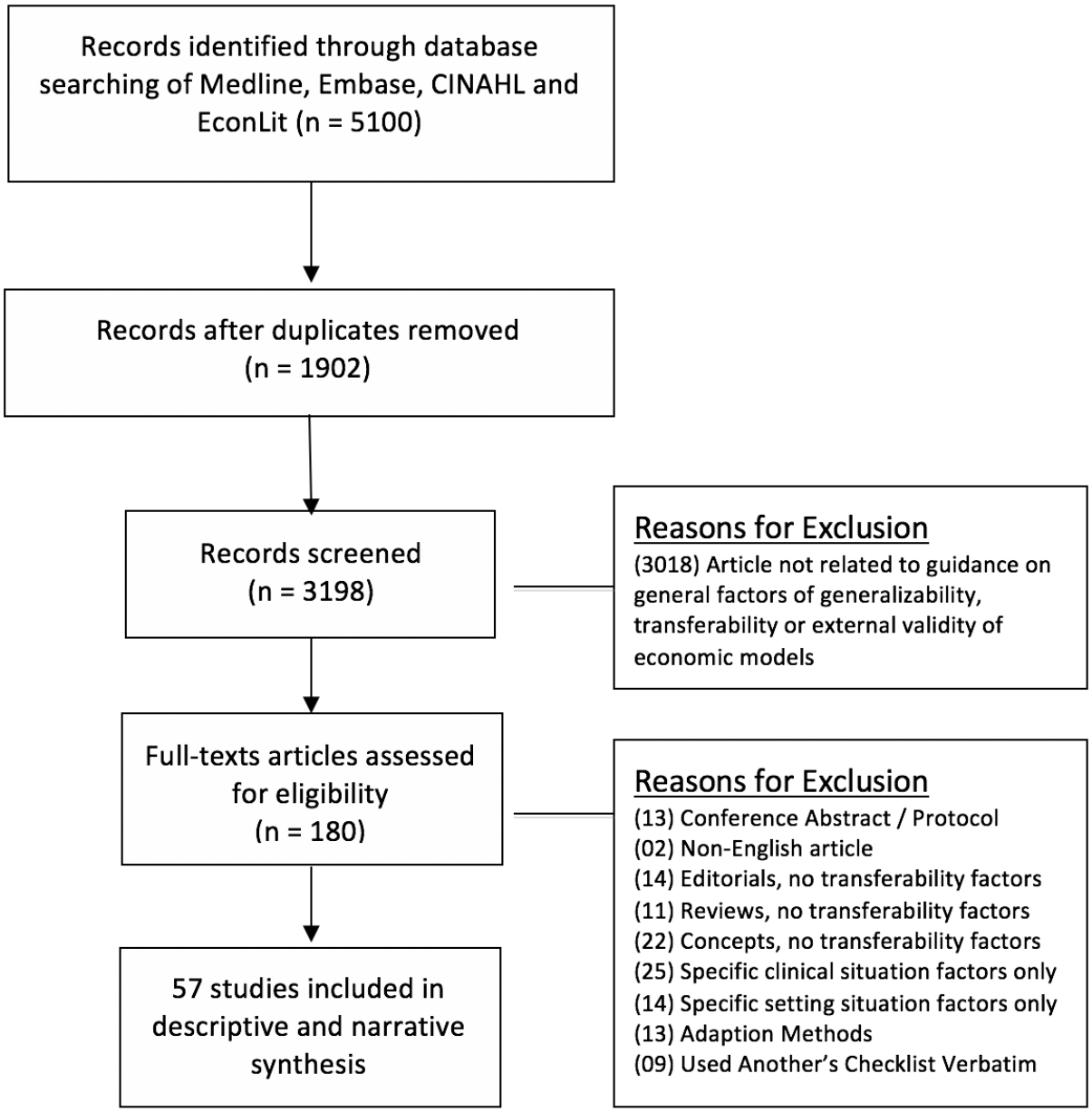
75. Witt DM, Nieuwlaat R, Clark NP, Ansell J, Holbrook A, Skov J, Shehab N, Mock J, Myers T, Dentali F, Crowther MA, Agarwal A, Bhatt M, Khatib R, Riva JJ, Zhang Y, Guyatt G. American Society of Hematology 2018 guidelines for management of venous thromboembolism: optimal management of anticoagulation therapy. Blood Adv. 2018 Nov 27;2(22):3257-91.
76. Busse JW, Craigie S, Juurlink DN, Buckley DN, Wang L, Couban RJ, Agoritsas T, Akl EA, Carrasco-Labra A, Cooper L, Cull C, da Costa BR, Frank JW, Grant G, Iorio A, Persaud N, Stern S, Tugwell P, Vandvik PO, Guyatt GH. Guideline for opioid therapy and chronic noncancer pain. CMAJ. 2017 May 8;189(18):E659-E666.
77. Gray AM, Wilkinson T. Economic evaluation of healthcare interventions: old and new directions. Oxford Review of Economic Policy. 2016; 32(1); 102-21.
78. Evers S, Goossens M, de Vet H, van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. Int J Technol Assess Health Care. 2005 Spring;21(2):240-5.
79. Chiou CF, Hay JW, Wallace JF, Bloom BS, Neumann PJ, Sullivan SD, Yu HT, Keeler EB, Henning JM, Ofman JJ. Development and validation of a grading system for the quality of cost-effectiveness studies. Med Care. 2003 Jan;41(1):32-44.
80. Nixon J, Rice S, Drummond M, Boulenger S, Ulmann P, de Pouvourville G. Guidelines for completing the EURONHEED transferability information checklists. Eur J Health Econ. 2009; 10: 157-65.
81. Wennberg JE, Barnes BA, Zubkoff M. Professional uncertainty and the problem of supplier-induced demand. Soc Sci Med. 1982;16(7):811-24.

82. Schünemann HJ, Wiercioch W, Etxeandia I, Falavigna M, Santesso N, et al. Guidelines 2.0: systematic development of a comprehensive checklist for a successful guideline enterprise. CMAJ. 2014 Feb 18;186(3):E123-42.
83. Edejer TT. Improving the use of research evidence in guideline development: 11. Incorporating considerations of cost-effectiveness, affordability and resource implications. Health Res Policy Syst. 2006 Dec 5;4:23.
84. van Mastrigt GA, Hiligsmann M, Arts JJ, Broos PH, Kleijnen J, Evers SM, Majoie MH. How to prepare a systematic review of economic evaluations for informing evidence-based healthcare decisions: a five-step approach (part 1/3). Expert Rev Pharmacoecon Outcomes Res. 2016 Dec;16(6):689-704.
85. Garrison LP Jr. Cost-Effectiveness and Clinical Practice Guidelines: Have We Reached a Tipping Point?-An Overview. Value Health. 2016 Jul-Aug;19(5):512-5.
86. Husbands S, Jowett S, Barton P, Coast J. Understanding and Identifying Key Issues with the Involvement of Clinicians in the Development of Decision-Analytic Model Structures: A Qualitative Study. Pharmacoeconomics. 2018 Dec;36(12):1453-62.
87. Munthe-Kaas H, Bohren MA, Glenton C, Lewin S, Noyes J, Tunçalp Ö, Booth A, Garside R, Colvin CJ, Wainwright M, Rashidian A, Flottorp S, Carlsen B. Applying GRADE-CERQual to qualitative evidence synthesis findings-paper 3: how to assess methodological limitations. Implement Sci. 2018 Jan 25;13(Suppl 1):9.
88. Robinson KA, Chou R, Berkman ND, Newberry SJ, Fu R, Hartling L, et al. Twelve recommendations for integrating existing systematic reviews into new reviews: EPC guidance. Journal of clinical epidemiology. 2016;70:38-44.
89. Greiner W, Schoffski O, Graf VDS. The transferability of international economic health-economic results to national study questions. Health Econ Prevention Care. 2000;1:94-102.

90. Attinger EO, Panerai RB. Transferability of health technology assessment with particular emphasis on developing countries. Int J Technol Assess Health Care. 1988;4:545-54.
91. Wijnen B, Van Mastrigt G, Redekop WK, Majoie H, De Kinderen R, Evers S. How to prepare a systematic review of economic evaluations for informing evidence-based healthcare decisions: data extraction, risk of bias, and transferability (part 3/3). Expert Rev Pharmacoecon Outcomes Res. 2016 Dec;16(6):723-32.
92. Sanabria AJ, Kotzeva A, Selva Olid A, Pequeño S, Vernooij RWM, Martínez García L, Zhang Y, Solà I, Thornton J, Alonso-Coello P. Most guideline organizations lack explicit guidance in how to incorporate cost considerations. J Clin Epidemiol. 2019 Dec;116:72-83.
93. Zhang Y, Alonso-Coello P, Guyatt GH, Yepes-Nuñez JJ, Akl EA, Hazlewood G, Pardo-Hernandez H, Etxeandia-Ikobaltzeta I, Qaseem A, Williams JW Jr, Tugwell P, Flottorp S, Chang Y, Zhang Y, Mustafa RA, Rojas MX, Schünemann HJ. GRADE Guidelines: 19. Assessing the certainty of evidence in the importance of outcomes or values and preferences-Risk of bias and indirectness. J Clin Epidemiol. 2019 Jul;111:94-104.

Figures

Figure 1: Flow Diagram of Article Selection



Tables

Table 1: Characteristics of Included Articles

Economist Article Type (n=57)	n	%
Systematic Review	21	37
Commentary	13	23
Checklist Development	10	18
Literature Summary	7	12
Data Set Evaluation	3	5
Interview / Survey	3	5
Publication Year (n=131)*		
1981-2005	95	73
2006-2010	16	12
2011-2015	14	11
2016-2020	6	4
Country of Principle Investigator (n=57)		
International	2	3
North America	9	16
Europe	41	72
South America	2	3
Middle East	1	2
Australia	1	2
Asia	1	2
Authorship Team Size (n=57)		
Individual Author	8	14
Small (2-5 authors)	28	49
Medium (>5 authors)	17	30
Task Force	4	7
"Most Important" Transferability Factors (n=57)		
Reported	23	40
Not Reported	34	60

Abbreviations: n – sample size; (%) – percentage

* There were 86 endorsements taken from Goeree 2007 et al.[18]

Table 2: Transferability Study Characteristic Themes -
Ranked and Categorized in the GRADE Evidence-to-Decision Theoretical Framework

Study Characteristics to Consider When Deciding to Use an Economic Evaluation as Research Evidence[1]	Endorsements as a Transferability Theme, n=131 (%)[2]	Endorsements as a "Most Important" Theme, n=23 (%)	CERQual Confidence in the Evidence as "Most Important" Theme[3]
General Population and Disease		*11 (48)*	*High*
- Demographics (age, sex, race), education, socio-economic status	39 (30)	11 (48)	High
- Epidemiology (incidence/prevalence, disease progression, spread)	36 (28)	05 (22)	Moderate
- Setting of Intervention (e.g. outpatient vs in-patient)	16 (12)	04 (17)	Moderate
- Disease severity, disease duration, case mix	14 (11)	NR	
- Mortality rates, life expectancy	13 (10)		
- Risk factors, medical history, genetic factors	10 (08)		
- Disease interaction, co-morbidity, concurrent medications	08 (06)		
- Lifestyle, environmental factors	06 (05)		
General Intervention and Funding Approach		*08 (35)*	*High*
- Efficacy of Intervention	10 (08)	02 (09)	Moderate
General Comparison (Usual Care) and Funding Approach		*07 (30)*	*High*
- Clinical practice, conventions, guidelines, norms	71 (54)	03 (13)	Moderate[4]
- Available treatment options (comparators)	38 (29)	02 (09)	Moderate
- Availability of generics or substitutes	05 (04)	02 (09)	Moderate
General Outcomes		*06 (26)*	*Low*
- Clinical endpoints/outcome measures	08 (06)	02 (09)	Very Low
General Values and Preferences (Patients)		*02 (09)*	*Low*
- Population values (utilities)	16 (12)	01 (04)	Very Low
- Compliance and adherence rates, ethical standards	14 (11)	01 (04)	Very Low
- Patient attitudes toward treatment, culture, religion, hygiene, nutrition	18 (14)	NR	
General Resource Use Estimation and Methodology		*11 (48)*	*High*
- Absolute or relative prices	79 (60)	11 (48)	High
- Sensitivity considerations, analysis reported	12 (09)	03 (14)	High
- Costing methodology, estimation procedures (e.g. productivity cost)	30 (23)	04 (17)	Moderate
- Study perspective	27 (21)	06 (26)	Moderate
- Model representative or validated	NR	04 (17)	Moderate
- Study factors (trial conditions, industry-related conflict of interest)	30 (22)	02 (09)	Low
- Exchange rates, purchasing power parities	10 (08)	01 (04)	Low

- Time Horizon, Timing of the economic evaluation	11 (08)	02 (09)	Very Low
- Opportunity cost (foregone benefits)	10 (08)	01 (04)	Very Low
- Input mix (personnel, equip.), specialization of labor, joint production	08 (06)	01 (04)	Very Low
General Acceptability (Provider and Decision-Maker)		*02 (09)*	*Moderate*
- Similar incentives for institutions	14 (11)	01 (04)	Moderate
- Similar incentives for providers, liability	17 (13)	01 (04)	Very Low
- Decision-maker Discount rates	15 (12)	01 (04)	Very Low
- Decision-maker Affordability (cost-effectiveness thresholds)	08 (06)	01 (04)	Very Low
- Method of remuneration (supplier-induced demand)	06 (05)	01 (04)	Very Low
General Feasibility		NR	
- Available resources (staff, facilities, equipment), programs, services	38 (29)		
- Organization of delivery system, structure, level of competition	29 (22)		
- Experience, education, training, skills, learning curve position	23 (18)		
- Level of technological advancement, innovation and availability	13 (10)		
- Capacity utilization, economies of scale, technical efficiency	13 (10)		
- Regulatory and organizational infrastructure, licensing of products	05 (04)		

Abbreviations: n – sample size; (%) percentage; NR - not reported

[1] Only presented, if sub-theme endorsed a minimum of 5 times.

[2] The systematic review by Goeree et. al. 2007 contributes general transferability factors from 86 articles.[18]

[3] Tables 2 & 3 have itemized summaries of GRADE-CERQual assessments.

[4] Rated up confidence in the evidence for adequacy of data due 71 observations (54% of articles) as a general transferability factor.

Appendices

Appendix 1: Itemized Summary of Characteristics of Included Articles

Number	Author, Year	Country of Principal Investigator	Journal / Website Source	Type of Article	Authorship Team Established Factors
1.	Hutton 2005	Switzerland	Health Policy and Planning	Commentary	Small
2.	Drummond 2009	United Kingdom	Value in Health	Checklist development	Task Force
3.	Knies 2009a	The Netherlands	Value in Health	Checklist development	Medium
4.	Mason 2006	United Kingdom	PharmacoEconomics	Literature review	Small
5.	Goeree 2007	Canada	Current Medical Research and Opinion	Systematic review of economic evaluation geographical transferability factors	Medium
6.	Goeree 2011	Canada	ClinicoEconomics and Outcomes Research	Systematic review of economic evaluation transferability checklists	Medium
7.	Welte 2004	Germany	PharmacoEconomics	Checklist development	Small
8.	Heyland 1996	Canada	Critical Care Medicine	Checklist development	Small
9.	Spath 1999	France	Health Policy	Systematic review of cost-effectiveness methodological approaches	Small
10.	EUnetHTA 2011	European Union	NHS website: https://eunethta.eu/wp-content/uploads/2011/01/EUnet	Checklist development	Task Force

			HTA_adptation_toolkit_2011_version_5.pdf		
11.	Antonanzas 2009	Spain	Health Economics	Checklist development	Medium
12.	Boulenger 2005	France	The European Journal of Health Economics	Checklist development	Medium
13.	Dryvig 2014	Denmark	Journal of Evaluation in Clinical Practice	Systematic review of external validity checklists	Small
14.	Burchett 2013	United Kingdom	Health Promotion International	Decision-maker Interviews	Small
15.	Ademi 2018	Switzerland	Swiss Medical Weekly	Systematic review of cost-effectiveness studies	Medium
16.	Anderson 2010	United Kingdom	Health Economics	Commentary	Individual author
17.	van Haalen 2014	The Netherlands	PharmacoEconomics	Systematic review of cost-effectiveness studies	Small
18.	Gonzalez-Perez 2002	United Kingdom	European Journal of Health Economics	Checklist development	Individual author
19.	Fukuda 2011	Japan	Infection	Systematic review of cost-effectiveness studies	Small
20.	Wolfenstetter 2010	Germany	International Journal of Environmental Research and Public Health	Systematic review of economic evaluations	Small
21.	Cook 2004	United States	Statistical Methods in Medical Research	Commentary	Individual author
22.	Li 2007	Australia	Expert Opinion on	Literature review	Individual author

			Pharmacotherapy		
23.	Jaime Caro 2014	Canada	Value in Health	Checklist Development	Task Force
24.	Rutten 1996	The Netherlands	Health Policy	Review of a data set	Individual author
25.	Bryan 1998	United Kingdom	Journal of Health Services Research and Policy	Commentary	Small
26.	Steuten 2008	United Kingdom	Expert Review of Medical Devices	Literature review	Small
27.	Drummond 2001	United Kingdom	Book Chapter	Literature review	Small
28.	O'Brien 1997	Canada	American Journal of Managed Care	Commentary	Individual author
29.	Sculpher 2004	United Kingdom	Health Technology Assessment	Systematic review of Economic Evaluation Methodological Studies and Conference Abstracts Related to Generalizability	Medium
30.	Barbieri 2010	United Kingdom	Value in Health	Systematic Review of Transferability Factors across guidelines	Task Force
31.	Coyle 1998	Canada	PharmacoEconomics	Commentary	Small
32.	Baltussen 1999	The Netherlands	PharmacoEconomics	Commentary	Small
33.	Abdul Pari 2014	United Kingdom	Bipolar Disorder	Systematic review of economic evaluations	Medium
34.	Gheorghe 2015	United Kingdom	Health Economics	Commentary	Medium
35.	Schünemann 2006	Italy	Health Research Policy and Systems	Systematic review of applicability, transferability and	Small

				adaptation of guidelines	
36.	Adam 2003	World Health Organization	International Journal of Technology Assessment in Health Care	Systematic review of cost-effectiveness methodology studies	Small
37.	Berg 2017	The Netherlands	Addiction	Systematic review of cost-effectiveness studies	Medium
38.	Knies 2009b	The Netherlands	PharmacoEconomics	Review of 15 data sets	Medium
39.	Coyle 2001	Canada	International Journal of Technology Assessment in Health Care	Review of a data set	Small
40.	Drummond 2005	United Kingdom	International Journal of Technology Assessment in Health Care	Commentary	Small
41.	Drummond 2015	United Kingdom	International Journal of Technology Assessment in Health Care	Decision-maker Interviews	Medium
42.	Zwolsman 2019	The Netherlands	International Urogynecology Journal	Systematic review of cost-effectiveness studies	Medium
43.	Gray 2016	United Kingdom	Oxford Review of Economic Policy	Commentary	Small
44.	Ginsberg 2013	Israel	Best Practice & Research Clinical Gastroenterology	Literature Review	Individual author
45.	Munthe-Kaas 2019	Norway	Systematic Reviews	Systematic review of general transferability checklists, excluding ones	Small

				for economic evaluations	
46.	Pichon-Riviere 2012	Argentina	International Journal of Technology Assessment in Health Care	Decision-maker Survey	Medium
47.	Ruggeri 2005	Italy	Value in Health	Systematic review of economic evaluations	Medium
48.	Vemer 2010	The Netherlands	Value in Health	Systematic review of cost-effectiveness studies	Small
49.	Vemer 2011	The Netherlands	The European Journal of Health Economics	Literature review	Small
50.	Sculpher 2006	United Kingdom	PharmacoEconomics	Literature review	Small
51.	Urdahl 2006	United Kingdom	PharmacoEconomics	Systematic review of decision-analytical models	Small
52.	Stawowczyk 2018	Poland	PharmacoEconomics	Systematic review of cost-effectiveness studies	Small
53.	Baltussen 1996	The Netherlands	Health Policy	Commentary	Small
54.	Birch 2003	Canada	Health Policy	Commentary	Small
55.	Mason 1997	United Kingdom	PharmacoEconomics	Commentary	Individual author
56.	Grutters 2011	The Netherlands	Value in Health	Checklist Development	Medium
57.	Augustovski 2009	South America	PharmacoEconomics	Systematic Review of cost-effectiveness studies	Medium

Appendix 2: GRADE-CERQual Evidence Profile and Coding of "Most Important" Transferability Factors

Author, Year	GRADE CERQual Evidence Profile					Quotes from Articles Related to Most Important Factors to Transferability	Coding of Factors
	Methodological Limitations [1]	Coherence [2]	Adequacy of Data [3]	'Most Important' Definition Relevance [4]	Preliminary Judgement of Single Article Confidence in the Evidence		
Drummond 2009	No or very minor concerns	No or very minor concerns	Minor concerns due to lack of clarity around technology relevance and methodological quality	No or very minor concerns with: situations where results are not transferable	High	(1) if either the experimental technology or the comparator(s) are not relevant in the jurisdiction of interest; (2) if the methodological quality of the studies doesn't meet local standards, which is similar to Welte's general 'knock-out' criteria; (3) if the study population is different between jurisdictions.	M, M1 P I C
Welte 2004	No or very minor concerns	No or very minor concerns	No or very minor concerns	No or very minor concerns with: knock-out criteria	High	(1) The evaluated technology is not comparable to the one that shall be used in the decision country. (2) The comparator is not comparable to the one that is relevant to the decision country. An example is a comparator drug that is not licensed in the decision country. (3) The study does not possess an acceptable quality, i.e. it does not live up to the standards required the decision context, e.g. there is double counting of costs.	M I C RU, RU1
Heyland 1996	No or very minor concerns	No or very minor concerns	No or very minor concerns	No or very minor concerns with: minimum	High	(1) comprehensive description of competing alternatives,	M2 T P I1

				methodological standard		(2) sufficient evidence of clinical effectiveness or efficacy, (3) appropriate identification, (4) measurement and (5) valuation of all important costs, and appropriate sensitivity analysis that takes into account all estimates of uncertainty.	C1 RU, RU1
Spath 1999	No or very minor concerns	Minor concerns as fit between underpinning data and summary unclear	Minor concerns with how criteria were described	No or very minor concerns with: critical criteria for methodological internal validity	Moderate	(1) perspective of the study from national level, (2) comparison of two or more options, (3) description of the evaluated therapies, and (4) the assessed therapies and/or its comparators are used in the health system of interest) that need to be satisfied before consideration of transferability indicators.	M3 T I C, C1
EUnet HTA 2011	No or very minor concerns	Moderate concerns as fit between underpinning data and summary unclear	No or very minor concerns	Minor concerns with indirect relevance: speedy sifting criteria	Moderate	(1) Are the policy and research questions being addressed relevant to your questions? (2) What is the language of this HTA report? Is it possible to translate this report into your language? (3) Is there a description of the health technology being assessed? (4) Is the scope of the assessment specified? (5) Has the report been externally reviewed? (6) Is there any conflict of interest? (7) When was the work that underpins this report done? Does this make it out of date for your purposes? (8) Have the methods of the assessment been described in the HTA report?	M, M4, M5, M6 T, T1 P I C O A1

Antonanzas 2009	Moderate concerns as only a literature summary provided	No or very minor concerns	Minor concerns due to lack of clarity around low methodological quality factors	Moderate concerns with partial relevance: relevant parameters needed to calculate cost-effectiveness ratio	Low	1. The relevant parameters needed to calculate the ratio cost/effectiveness are given in the study. 2. The quality of the study is acceptable: a. The study objectives are presented in a clear, specific, and measurable manner. b. The variable estimates used in the analysis come from the best available source. c. The measurement of cost is appropriate and the methodology for the estimation of quantity and unit costs is clearly described. d. The health outcome measures are based on valid and reliable scales, when available. Otherwise, the scales used in the study must be fully justified. e. The economic model (including its structure), study methods, and components of the costs and effectiveness are presented in a clear manner. f. The conclusions and recommendations for the study are justified and based on the study results.	M T, T1 O RU, RU1
Boulenger 2005	Minor concerns as methods to synthesize participant feedback and develop questionnaire unclear	No or very minor concerns	No or very minor concerns	No or very minor concerns with: concise sub-set of Items considered to be most important for assessing transferability	High	HT1a. Is the intervention described in sufficient detail? HT2a. Is (are) the comparator(s) described in sufficient details? SE2a. Is (are) the country (ies) in which the economic study took place clearly specified? P1a. Did the authors correctly state which perspective they adopted for the economic analysis?	M, M2, M3, M7 T, T2, T3 P I, I1 C RU, RU2

						SP1a. Is the target population of the health technology clearly stated by the authors or when it is not done can it be inferred by reading the article? SP3a. Does the article provide sufficient detail about the study sample(s)? E5a. Have the principal estimates of effectiveness measures been reported? E7a. Does the article provide the results of a statistical analysis of the effectiveness results? B5a. Is the level of reporting of benefit data adequate (incremental analysis, statistical analyses)? C1a. Are the cost components/items used in the economic analysis presented? C9a. Is the currency unit reported? S1a. Are quantitative and/or descriptive analysis conducted to explore variability from place to place? O1a. Did the authors discuss caveats regarding the generalisability of their results?	
Dryvig 2014	No or very minor concerns	No or very minor concerns	No or very minor concerns	Serious concerns with unclear relevance: most used 6 Items from review of 21 external validity checklists	Low	(1) Baseline characteristics and risk (2) Eligible setting (3) Description of usual care (4) Population, enrolled, declined, excluded (5) Relevant outcomes (6) Valid outcome measures and presentation of them	T M P, P1, P2 C2 O, O1 VP2

Burchett 2013	No or very minor concerns	No or very minor concerns	Minor concerns due to lack of clarity provided around population and setting factors	No or very minor concerns with: top 3 rated factors by 69 decision makers	High	(1) Relevant Population (2) Relevant Setting (3) Acceptability (Intervention, Politically)	P, P2 A, A1 VP
Ademi 2018	No or very minor concerns	Moderate concerns as fit between underpinning data and summary unclear	Moderate concerns due to lack of clarity provided around important base case, intervention and outcome elements	Moderate concerns with partial relevance: general factors from CHEERS items 4,7,10 are met prior to further assessment	Very Low	(1) Describe characteristics of the base case population and subgroups analysed, including why they were chosen. (2) Describe the interventions or strategies being compared and state why they were chosen. (3) Describe what outcomes were used as the measure(s) of benefit in the evaluation and their relevance for the type of analysis performed.	T P I C O, O1
Anderson 2010	Serious concerns as only literature examples provided	Minor concerns as fit between underpinning data and summary somewhat unclear	Serious concerns due to a lack of authorship team	Serious concerns with unclear relevance: evaluation focuses that are 'rather uninformative material'	Very Low	If economic evaluations remain mostly intervention-focussed (with little detailed description of context and patient characteristics), and exclusively descriptive in aim (that is, to measure and report the total costs and total effects for a particular comparison), then they will probably be rather uninformative material for such a review.	M3 T P RU
van Haalen 2014	No or very minor concerns	Minor concerns as fit between underpinning data and summary	Minor concerns due to lack of clarity around minimum methodological and structural	Minor concerns with: minimum methodological and structural requirements	Moderate	As the first step in determining which models are potentially suitable to be transferred, we assessed the conceptual validity of the identified models, i.e. whether the individual models adequately represented the concept of the	M, M3, M7 P, P1

		somewhat unclear	requirements			disease and its clinical context in their modelling framework. For this purpose, it was necessary to specify a list of disease- specific minimal methodological and structural requirements that were considered important for securing the conceptual validity of health economic evaluation models.	
Gonzalez-Perez 2002	No or very minor concerns	No or very minor concerns	Serious concerns due to a lack of authorship team	Minor concerns with indirect relevance: worst case scenario critical factors results in zero transferability score	Low	The worst scenario for the transferability of results to the NHS context is where the disease and the interventions are not relevant, the study population is not comparable, and the view point considered is restricted to the provider of care, for example, the hospital or the general practitioner. A score of zero is awarded in such cases.	M3 P, P1 P2, I
Fukuda 2011	No or very minor concerns	Minor concerns as fit between underpinning data and summary somewhat unclear	No or very minor concerns	Minor concerns with indirect relevance: scale of transparency of costing reporting, which without, readers would unable to assess applicability	Moderate	– Level A: all components of costs were described and data for both quantity and unit price of resources were reported for each component. – Level B: all components of costs were described and data for costs in each component were reported. This included studies that used graphical presentations of the aforementioned data. – Level C: all components of costs were described but data for costs in each component were not reported. – Level D: only the scope of costing was described but the components of costs were not described. For example, studies that only reported terms such as ‘‘hospital stay’’	T RU

						or ‘‘direct costs’’ without further exposition were evaluated at Level D.	
Wolfenstetter 2010	No or very minor concerns	No or very minor concerns	Minor concerns as lack of clarity around transparency and comprehensible descriptions	No or very minor concerns with: prerequisite factors for examining transferability	High	A prerequisite for examining the transferability of international studies to Germany requires that methods, data resources and study results are transparently and comprehensible described [22,26,35].	M T RU O
Cook 2004	Moderate concerns as only a literature summary provided	Minor concerns as fit between underpinning data and summary somewhat unclear	Minor concerns with small authorship team	Serious concerns with unclear relevance: assessment of extent effectiveness deviates from regular practice	Very Low	It was mentioned earlier that one of the advantages of clinical trials as a source of economic data is their high internal validity. However, another important characteristic of economic data is the need for external validity, or relevance to other settings. One threat to generalizability is the potential difference between the clinical care practice in trials and that in regular clinical care. As mentioned above, it is therefore important to make an assessment of the extent to which the trial deviates from regular practice prior to embarking on extensive economic data collection.	P2 I C2
Li 2007	Moderate concerns as only a literature summary provided	No or very minor concerns	Moderate concerns due to lack of authorship team and lack of clarity around the methodolo	Minor concerns with indirect relevance: two major barriers that hamper	Very Low	Related to: (1) the standardisation of methodology in the process of collecting pharmacoeconomic data and performing the analysis; and (2) the acceptability of these data on the part of the decision makers.	M, M7 RU, RU1 A

			gy standardization	generalisability			
Jaime Caro 2014	Moderate concerns as methods to synthesize participant feedback and develop questionnaire unclear	No or very minor concerns	Moderate concerns due to lack of clarity around a sufficiently validated model	No or very minor concerns with: fatal flaw factors	Low	Just because a modeling analysis has been published does not mean that it is credible— much depends on the quality of the peer review process. If a model has not been sufficiently validated, or the decision maker cannot tell this information, then the results of the model should not be trusted (i.e., this is a fatal flaw). External validation is essential in establishing the credibility of the model, and a "No" answer to this question should be considered a fatal flaw.	M5, M7 T
Rutten 1996	Serious concerns as only literature examples provided	Moderate concerns as fit between underpinning data and summary unclear	Moderate concerns due lack of authorship team and lack of clarity around patterns of resource use and the way healthcare is funded	Moderate concerns with partial relevance: differences making model often not suitable	Very Low	Major differences in unit costs, the patterns of resource use and the way in which health care is funded.	RU, RU3 A2
Bryan 1998	Moderate concerns as only a literature summary provided	Moderate concerns as fit between underpinning data and summary unclear	Minor concerns due to small authorship team	No or very minor concerns with: key parameter, one whose change in value may change policy implicatio	Very Low	A 'key parameter' is one whose change in value may change the policy implications of the results. The following list indicates possible sources of variation in the value of key parameters that should be assessed when attempting to extrapolate from an economic evaluation:	P1 C RU

				ns of results		• unit costs or prices of the resources used which are likely to differ between settings; • prevalence, incidence and natural history of many diseases which are known to vary from region to region both within and between countries; and • comparators used in the published study which may not be directly relevant locally.	
Steuten 2008	Moderate concerns as only a literature summary provided	No or very minor concerns	No or very minor concerns	No or very minor concerns with: has to be comparable to directly transfer the conclusions	Moderate	Nonetheless, even when incremental cost per quality-adjusted-life-years gained is the preferred outcome measure both in study and target country, then still the 'willingness to pay' for one quality- adjusted life year (which lies between £20,000 and £30,000 in the UK) [11] has to be comparable in order to directly transfer the conclusions. Otherwise, it may happen that a technology falls well below the threshold of the study country and is thus deemed cost-effective, while it actually exceeds the threshold in the target country and should therefore not be reimbursed.	A3
Drummond 2001	Moderate concerns as only a literature summary provided	No or very minor concerns	Minor concerns due to small authorship team	Minor concerns with indirect relevance: inclusion criteria for considering a review on economic	Low	Inclusion criteria: (paraphrased from Jefferson 1996) -clear aim, viewpoint and timespan -design consistent with study aim -coherent methods, results and conclusions -itemized costs	M, M2, M3, M8 T O RU

				evaluations		-marginal and sensitive analysis -all of the above clearly and unequivocally stated in text	
O'Brien 1997	Serious concerns as only literature examples provided	No or very minor concerns	Serious concerns due to a lack of authorship team	Minor concerns with indirect relevance: threats to transferability	Very Low	Threats to transferring data involve differences among countries with regard to demography and epidemiology of disease, clinical practice and conventions, incentives to and regulation of health care providers, relative prices levels, consumer preferences, and opportunity cost of resources. The article describes specific details of each of the suggested threats.	P C2 VP, VP1 RU4 A2, A3, A4

[1] Limitations through underpinning review (example articles, literature summary, systematic review) or clarity of interview methods.
[2] Fit between article finding and underpinning approach through clear rationale from data to reporting.
[3] Quantity of data through number of participants or authors informing transferability factors; Richness of data through clarify of transferability study characteristic descriptions.
[4] Threats to relevance through indirect, partial relevance or unclear relevance of transferability study characteristic definition of "most important".

Coding List for Appendix 2:

M - General Methodological Quality M2 – sensitivity analysis M3 – study perspective M7 – model representative M5 – peer reviewed M1 – local standards M4 – language translatable M6 – not out of date M8 – time horizon T - General Transparent Reporting T1 – conflicts of interest reported T2 – country specified T3 – generalisability discussed P - General Population P1 – baseline risk P2 – setting I - General Intervention I1 – intervention effectiveness C - General Comparison C1 – competing alternatives C2 – usual care O - General Outcomes O1 – valid outcome measure	VP - General Patient Values & Preferences VP1 – health states (utilities) VP2 – patient compliance RU - General Resource Use Inputs RU1 – measurement RU2 – currency unit provided RU3 – patterns of use RU4 – opportunity costs A - General Acceptability A1 – policy acceptable A2 – funding models A3 – provider incentives A4 – discount rates A3 – willingness to pay thresholds F - General Feasibility E - General Equity

Appendix 3: GRADE-CERQual Summary of Findings for "Most Important" Factors when Assessing Transferability

Themes and Sub-Themes	Frequency of Endorsement	Overall GRADE CERQual Rating	Articles Informing Factors	Rationale for Rating Down
General Methodological Quality	10	High	Drummond 2009, Welte 2004, EUnetHTA 2011, Antonanzas 2009, Boulenger 2005, Dryvig 2014, van Haalen 2014, Wolfenstetter 2010, Li 2007, Drummond 2001	Only literature summary in 3 articles; unclear fit between summary and data in one article; partial relevance in 2 articles
- sensitivity analysis	3	High	Heyland 1996, Boulenger 2005, Drummond 2001	Only literature summary in one article
- study perspective	6	Moderate	Spath 1999, Boulenger 2005, Anderson 2010, van Haalen 2014, Gonzalez-Perez 2002, Drummond 2001	Only literature summary in 3 articles; lack of authorship team in 2 articles; unclear relevance in one article
- model representative	4	Moderate	Boulenger 2005, van Haalen 2014, Li 2007, Jaime Caro 2014	Only literature summary in one article; small authorship team in one article; lack of clarity around factor in 2 articles
- peer reviewed	2	Low	EUnetHTA 2011, Jaime Caro 2014	Unclear fit between summary and data across studies, reported only in 2 articles; lack of clarity around factor in one article; indirect relevance in one article
- meets local standards	1	Low	Drummond 2009	Only literature summary; small authorship team; reported only in one article; indirect relevance
- not out of date	1	Very Low	EUnetHTA 2011	Unclear fit between summary and data;

				reported only in one article; indirect relevance
- time horizon	1	Very Low	Drummond 2001	Only literature summary; small authorship team; reported only in one article; indirect relevance
General Transparent Reporting	12	High	Heyland 1996, Spath 1999, EUnetHTA 2011, Antonanzas 2009, Boulenger 2005, Dryvig 2014, Ademi 2018, Anderson 2010, Fukuda 2011, Wolfenstetter 2010, Jaime Caro 2014, Drummond 2001	Only literature summary in 4 articles; unclear fit between summary and data in 2 articles; lack of authorship team in one article; partial relevance in 3 articles
- country specified	1	Low	Boulenger 2005	Only reported in one article
- generalisability discussed	1	Low	Boulenger 2005	Only reported in one article
- language translatable	1	Very Low	EUnetHTA 2011	Unclear fit between summary and data; reported only in one article; indirect relevance
General Population	11	High	Drummond 2009, Heyland 1996, EUnetHTA 2011, Boulenger 2005, Dryvig 2014, Burchett 2013, Ademi 2018, Anderson 2010, van Haalen 2014, Gonzalez-Perez 2002, O'Brien 1997	Only reported literature examples in 2 articles; unclear fit between summary and data in one article; lack of authorship team in 2 articles; partial relevance in 4 articles
- epidemiology (baseline risk, incidence/prevalence)	5	Moderate	Dryvig 2014, van Haalen 2014, Gonzalez-Perez 2002, Bryan 1998, O'Brien 1997	Only literature summary in 2 articles; unclear fit between summary and data in one article; small authorship team in 2 articles; indirect relevance in 2 articles
- setting	4	Moderate	Dryvig 2014, Burchett 2013,	Only literature summary in one article; small authorship team in one

			Gonzalez-Perez 2002, Cook 2004	article; unclear relevance in two articles
General Intervention	8	High	Drummond 2009, Welte 2004, Spath 1999, EUnetHTA 2011, Boulenger 2005, Ademi 2018, Gonzalez-Perez 2002, Cook 2004	Only literature summary in 2 articles; unclear fit between summary and data in 2 articles; lack of authorship team in one article; lack of clarity around factor in one article; partial relevance in 2 articles
- intervention effectiveness	2	Moderate	Heyland 1996, Boulenger 2005	Only reported in two articles
General Comparison	7	High	Drummond 2009, Welte 2004, Spath 1999, EUnetHTA 2011, Boulenger 2005, Ademi 2018, Bryan 1998	Only literature summary in 2 articles; unclear fit between summary and data in 3 articles; partial relevance in one article
- competing alternatives	2	Moderate	Heyland 1996, Spath 1999	Only literature summary in one article; only reported in 2 articles
- usual care similarity	3	Low	Dryvig 2014, Cook 2004, O'Brien 1997	Only literature summary in 2 articles; partial relevance across 3 articles
General Outcomes	6	Low	EUnetHTA 2011, Antonanzas 2009, Dryvig 2014, Ademi 2018, Wolfenstetter 2010, Drummond 2001	Only literature summary in 2 articles; unclear fit between summary and data in two articles; lack of clarity around factor in one article, partial relevance in 3 articles
- valid outcome measure	2	Very Low	Dryvig 2014, Ademi 2018	Unclear fit between summary and data in one article; lack of clarity around factor in one article; only reported in 2 articles; partial relevance across articles
General Patient Values & Preferences	2	Low	Burchett 2013, O'Brien 1997	Only reported in 2 articles; only literature examples in one article; lack of authorship team in one article

- health states (utilities)	1	Very Low	O'Brien 1997	Only literature examples; lack of authorship team; indirect relevance
- patient compliance	1	Very Low	Dryvig 2014	Only reported in one article; unclear relevance
General Resource Use Inputs	11	High	Welte 2004, Heyland 1996, Antonanzas 2009, Boulenger 2005, Anderson 2010, Fukuda 2011, Wolfenstetter 2010, Li 2007, Rutten 1996, Bryan 1998, Drummond 2001	Only literature summary in 6 articles; unclear fit between summary and data in one article; lack of authorship team in 2 articles; partial relevance in 3 articles
- input measurement	4	Moderate	Welte 2004, Heyland 1996, Antonanzas 2009, Li 2007	Only literature summary in 2 articles; partial relevance in 2 articles
- biases: industry conflicts of interest, trial conditions	2	Low	EUnetHTA 2011, Antonanzas 2009	Only literature summary in one article; unclear fit between summary and data in one article; reported only in 2 articles; indirect relevance across articles
- currency unit provided	1	Low	Boulenger 2005	Only reported in one article
- patterns of resource use	1	Very Low	Rutten 1996	Only reported literature examples; unclear fit between summary and data; no authorship team; lack of clarity around factor; only reported in one article; partial relevance
- opportunity costs	1	Very Low	O'Brien 1997	Only literature examples; lack of authorship team; indirect relevance
General Acceptability	2	Moderate	Burchett 2013, Li 2007	Only literature summary in one article; only reported in two articles
- policy acceptable	2	Moderate	EUnetHTA 2011, Burchett 2013	Unclear fit between summary and data in one article; only reported in two articles

- funding models	1	Very Low	Rutten 1996	Only reported examples; unclear fit between summary and data; no authorship team; lack of clarity around factor; only reported in one article; partial relevance
- willingness to pay thresholds	1	Very Low	Steuten 2008	Only literature summary; only reported in one article
- provider incentives	1	Very Low	O'Brien 1997	Only literature examples; lack of authorship team; indirect relevance
- discount rates	1	Very Low	O'Brien 1997	Only literature examples; lack of authorship team; indirect relevance
General Feasibility	-	-	-	-
General Equity	-	-	-	-

Appendix Table 4: Coding of General Transferability Comparability Factors

Author Year	Country of PI	Quotes from Articles Related to General Transferability Factors	Coding of Factors
Goeree 2007	Canada	Patient characteristics Demographics (age, gender, race), education, socio-economic status Risk factors, medical history, genetic factors Lifestyle, environmental factors Mortality rates, life expectancy Attitudes toward treatment, culture, religion, hygiene, nutrition Compliance and adherence rates, ethical standards Population values (utilities) Population density, immigration, emigration, traveling patterns Income, employment rates, productivity, work loss time, friction time Type of insurance coverage, user fees, co-payments, deductibles Incentives for patients Disease characteristics Epidemiology (incidence/prevalence, disease progression, spread) Disease severity, case mix Disease interaction, co-morbidity, concurrent medications Mortality due to disease Provider characteristics Clinical practice, conventions, guidelines, norms Experience, education, training, skills, learning curve position Quality of care provided Method of remuneration (supplier-induced demand) Patient identification Cultural attitudes Incentives for providers, liability Health care system characteristics Absolute or relative prices Available resources (staff, facilities, equipment), programs, services Organization of delivery system, structure, level of competition Level of technological advancement, innovation and availability Available treatment options (comparators) Capacity utilization, economies of scale, technical efficiency	Used to develop start codes and frequencies mapped to GRADE domains. Contains factors from 86 articles.

		Input mix (personnel, equip.), specialization of labor, joint production Access to programs and services, gatekeepers, historical differences Waiting lists, referral patterns Regulatory and organizational infrastructure, licensing of products Availability of generics or substitutes Market form of suppliers, payment of suppliers, supplier incentives Incentives for institutions <u>Methodological characteristics</u> Costing methodology, estimation procedures (e.g. productivity cost) Study perspective Study factors (artificial trial conditions, industry-related bias) Timing of the economic evaluation Clinical endpoints/outcome measures Discount rates Exchange rates, purchasing power parities Opportunity cost (foregone benefits) Affordability (CE thresholds)	
Goeree 2011	Canada	No new themes presented. For example, the critical transferability factors that have been proposed seem to focus on issues of study quality, transparency of methods, the level of reporting of methods and results, and the applicability of the treatment comparators to the target country. The proposed list of noncritical factors has been much more extensive and perhaps future research might focus on narrowing or refining this list. More recently, indices have been promised to measure transferability potential. However, due to the complexities of identifying appropriate weights for each of the noncritical factors, it is still uncertain whether the assessment and calculation of an overall transferability score or index will be practical or useful for transferability considerations in the future	A B2 B10 E2
Welte 2004	Germany	- Methodological characteristics – Perspective, Discount rate, Medical cost approach, Productivity cost approach - Healthcare system characteristics (supply of technology) - Absolute and relative prices in healthcare, Direct medical costs, Costs and effects Costs, Practice variation Technology availability - Population characteristics (demand for technology) – Disease incidence/prevalence, case-mix, life expectancy, health-status preferences, acceptance, compliance, incentives to patients, productivity and work-loss time, disease spread	Did not code to avoid double-counting.
Sculpher 2004	United Kingdom	<u>Transparency of Reporting</u> Q2.1 Study setting specified? Q2.2 Patient population specified?	A B11 D14

		Q2.3 Alternative interventions stated and justified? Q2.4 Model structure clearly stated? Q2.5 Main assumptions clearly stated and justified? <u>Information Sources:</u> Q3.1a Sources of clinical data provided? Q3.2b Sources of resource use provided? Q3.1c Sources of unit costs provided? Q3.1d Sources of preferences/ utilities provided? <u>Relevance of the model input to the stated decision-maker</u> Q3.2a Clinical data sources relevant to decision-maker? Q3.2b Resource use data relevant to decision-maker? Q3.2c Unit costs relevant to decision-maker? Q3.2d Preferences/ utilities relevant to decision-maker? <u>Sensitivity analysis of estimates of clinical effectiveness and health state valuation</u> Q4.1a Was the robustness of the effect estimate explored? Q4.1b Did the model accommodate for potential difference in compliance rates? Q4.1c Did the model accommodate for differences in utilities? Q4.1d External consistency: were the results compared with other relevant studies? <u>Sensitivity analysis of resource use and valuation</u> Q4.2a Did the model reflect variation in costs nationally? Q4.2b Did the model reflect variation in resource use patterns nationally? Q4.2c Did the model reflect variation in costs internationally? Q4.2d Did the model reflect variation in resource use patterns internationally? <u>Authors' comments on the generalisability of the results</u> Did the author explicitly address the issue of transferability of the results to other jurisdictions?	Did not code to avoid double-counting; but, found 3 additional emerging codes A, B11 and D14.
Barbieri 2010	United Kingdom	The results on the degree of transferability of the five key data inputs (baseline risk, treatment effect, health state utilities, resource use, and unit costs) are presented. Two data inputs were considered on the grounds that there is still some debate about whether they are transferable or not: 1. baseline risk 2. health utilities NICE's methods guidelines recommend that baseline risk has "to be relevant to UK practice and patients, and to compare all relevant treatment options for the relevant patient groups." Nevertheless, it	A C2 C5 B11 D13

		is added that "evidence on effectiveness might come from outside the UK health care system . . . . Despite such weaknesses in the evidence base, decisions still have to be made about the use of technologies. Therefore, analyses should use the best evidence available, be explicit about data limitations and any attempts to overcome these, and quantify as fully as possible how the limitations of the data are reflected in the uncertainty in the results of the analysis." With respect to utility values, the NICE guidelines state that: "The valuation of changes in HRQL reported by patients should be based on public preferences, elicited using a choice-based method in a representative sample of the UK population" and "The EQ-5D is the preferred measure of HRQL in adults. When EQ-5D data are not available or are inappropriate for the condition or effects of treatment, the valuation methods should be fully described and comparable to those used for the EQ-5D." For example, "methods can be used to estimate EQ-5D utility data by mapping (also known as "cross-walking") EQ-5D utility data from other HRQL measures included in the relevant clinical trial(s)." Different methods and issues exist for these two situations. In the case of studies based on a clinical trial, issues might arise for adapting data to the country of interest from a clinical trial performed in another country or a multinational study with or without the inclusion of a substantial number of patients from the reference country. Different methods have been suggested (e.g., statistical tests of homogeneity, fixed-effect models, and multilevel models) with different advantages and disadvantages. The methods recommended to address the issue of external validity of data obtained from clinical trials have also been reviewed [2]. In the case of modeling, issues might arise about the structural adaptation of the model among countries because differences in comparators or patient populations might deter- mine differences in practice patterns or natural history of the disease. Nevertheless, the main issues for economic evaluations based on decision models relate to methods used to extract and synthesize data obtained from several sources, and the methods to deal with the potential variability in the sources of data.	
Author Year	Country of PI		
Heyland 1996	Canada	Clinical Generalizability: Are the patients described in the analysis similar to those patients you see in your setting? There is little difficulty in passing this judgment if patients you care for meet the inclusion and exclusion criteria of the study. Healthcare Systems Generalizability: 1) Is the viewpoint of the analysis relevant to your clinical setting/situation? Thus, the results of an analysis of a hospital program from a third-party payer viewpoint may be different than the results of	C1 B1 B3 B4 B6 D1 D14

		an analysis of the same program from a government viewpoint, partly because the inclusion and valuation of costs and benefits between the respective viewpoints may differ. 2) Is the intervention/program under study generalizable to your setting?. Even if there is good evidence to support the use of a program or intervention, the reader has to assess whether there exists enough resources, infrastructure, trained personnel, demand, etc. to support such technology. Often, local budget constraints determine the availability of such interventions. 3) Are the costing methods generalizable to the healthcare system in which you work? Different unit prices for physician fees, laboratory tests, drugs, etc. will affect total costs, as well as result in a different mix of resources consumed to perform a given task. Differing patient volumes will result in a different average and/or marginal cost. Finally, converting costs using exchange rates across countries represents a formidable challenge since exchange rates do not, in most cases, reflect the relative difference in costs of resources consumed; rather, exchange rates reflect government monetary policy. 4) Are the outcomes measured appropriate to your setting? Here, we want to consider whether the most appropriate instrument relevant to the local setting was used to measure the primary outcome, whether we are dealing with natural units (e.g., cases of pneumonia), utilities (quality-adjusted life years), or benefits (dollars). Furthermore, when appropriate preference- based measures were used, one needs to ask whether there is evidence that the preferences in your society, or of your patients, are the same as those preferences expressed in the analysis. 5) Is the discount rate applicable to your setting? Because as individuals and as society, we typically prefer to have dollars or resources and benefits and good consequences now as opposed to later, future costs and benefits are discounted or reduced to reflect the fact that, for example, dollars saved or spent in the future are not valued as highly as dollars spent or saved today [10]. There is general agreement that costs and consequences that occur in the future should be discounted to present values. However, there is no agreement on what should be the discount rate. Since the discount rate reflects time preference (i.e., the relative value of cost and consequences which occur at different points in time), one needs to be assured that the time preferences are similar across health systems or societies [19].	F2
Spath 1999	France	2.3.1. Potential users We examined the perspective in order to define the potential users of the economic evaluation. This was done on the basis of: (1) the perspective mentioned by the authors; (2) the range of cost data included in the study; and (3) the source(s) of cost data. 2.3.2. Characteristics of the treated patient population We assessed whether patients in the FHCS with the same characteristics (age and medical history) as the study population would receive the therapies being investigated.	B3 B4 C1 D13

		2.3.3. Health outcome data Health outcome data comprises: (1) the efficacy of a therapy: the therapy's proven ability to do more good than harm established under strictly controlled conditions (randomized controlled trials); (2) the effectiveness of a therapy: the therapy's performance in actual clinical use (determined by non-randomized trials including either contemporaneous or historical controls); and (3) preferences (in cost-utility analyses). 2.3.4. Health-care resources We listed the resources (laboratory tests, consultations, etc.) reported in the articles. We then determined whether the resources included in the economic evaluation study were reported in detail, i.e. whether all resources were identified and quantified. 2.3.5. Unit prices of health-care resources and discount rates Furthermore, cost data concerning health care resources depends on the source of the data; it is well known that 'costs' and 'charges' are different [23]. Cost data also depends on the discount rates used. We first listed the sources of cost data used by the authors of the economic evaluations. We then determined whether the papers reported: (1) the unit prices of all resources included in the economic evaluation study; and (2) the discount rates used.	
EUnetHTA 2011	European Union	6. Is there any consideration of when and how technical characteristics affect outcomes? 7. Are there any differences in the use of this technology within the target setting (compared to the uses described in the HTA report for adaptation)? 15. Does the population described for eligibility match the population to which it is targeted in the target setting? 16. Are there any reasons to expect differences in complication rates (e.g. epidemiology, genetic issues, healthcare system (quality of care, surveillance))? 17. Are the requirements for its use (special measures needed for use/implementation, maintenance etc.) available in the target setting? 18. Is the necessary expertise (knowledge and skills) available in the target setting? 19. a) Is safety particularly dependent on training? b) Are there types of teams to which the procedure should be limited for safety reasons? c) Is there a need for special training or certification to deliver the intervention properly. d) Would it be possible (affordable) to organise such training, if any? 14. Would you expect the baseline risk of patients within your own setting to be the same as the baseline risk of those patients considered within the HTA report for adaptation? (assuming that patients receive the same treatment and same comparator). We would expect the relative risk to be the same	B1 B3 B4 B11 C1 C2 D1 D3 D5 D6 D8 D13 D14 E1 E3 F1 F2

		and baseline risk different. The user needs to consider the impact of local epidemiological and demographic data on the baseline risk. 27. How generalisable and relevant are the results, and validity of the data and model to the relevant jurisdictions and populations? 28. a) Are there any differences in the following parameters? Perspective, Preferences, Relative costs, Indirect costs, Discount rate, Technological context, Personnel characteristics, Epidemiological context (including genetic variants), Factors which influence incidence and prevalence Demographic context, Life expectancy, Reproduction, Pre- and post-intervention care, Integration of technology in health care system Incentives b) If differences exist, how likely is it that each factor would impact the results? In which direction? Of what magnitude? c) Taken together, how would they impact the results and of what magnitude? d) Given these potential differences, how would the conclusions likely change in the target setting? Are you able to quantify this in any manner? 29. Does the economic evaluation violate your national/regional guidelines for health economic evaluation?	
Boulenger 2005	France	Q1. Is the study question clearly stated? Q2. Are the alternative technologies justified by the author(s)? SE1. Did the authors correctly specify the setting in which the study took place (e.g. primary care, community)? SP2. Are the population characteristics described? (e.g. age, sex, health status, socio-economic status, inclusion/exclusion criteria) SP4. Does the paper provide sufficient information to assess the representativeness of the study sample with respect to the target population? M1. If a model is used is it described in detail? M2. Are the origins of the parameters used in the model given? E1. If a single study is used is the study design described (sample selection, study design, allocation, follow-up)? E2. If a single study is used are the methods of data analysis described (ITT/per protocol or observational data)? E3. If based on a review/synthesis of previous published studies, are review methods described (search strategy, inclusion criteria, sources, judgement criteria, combination, investigation of differences)? E4. If based on opinion, are the methods used to derive estimates described?	A B4 B7 B9 B10 C1 D13 D14 E2 F2

		E6. Are the side effects or adverse effects addressed in the analysis? B1. Do the authors specify any summary benefit measure(s) used in the economic analysis? B2. Do the authors report the basic method of valuation of health states or interventions? B4. Do the authors specify the valuation tool used? C2. Are the methods used to measure costs components/items provided? C3. Are the sources of resource consumption data provided? C4. Are the sources of unit price data provided? C8. Is the time horizon given for each element of the cost analysis? C10. Is a currency conversion rate given? C11. Does the article provide the results of a statistical analysis of cost results? D1. Was the summary benefit measure(s) discounted? D2. Were the cost data discounted? D3. Do the authors specify the rate(s) used in discounting costs and benefits? D4. Were discounted and not discounted results reported?	
Burchett 2013	United Kingdom	-Congruence - With previous experience, With beliefs and values, With other evidence - Knowledge of similar project/programme, Inherent value of intervention's approach or content Findings from other studies - Ease of implementation of the intervention - Intervention characteristics, Capacity to implement - Content or approach, cost, implementation challenges, Acceptability, affordability, human resources, political will Ability to maintain implementation over time - Setting of Intervention - Intervention need, Country-level influences, Population-level -influences - Focused on pertinent health problem, addressed determinants of health problem Geographical location and proximity, development level, within-country differences, 'Culture', urban-rural settings, women's status, religion, social structures, literacy -Effectiveness of Intervention - Original study findings, Potential effectiveness - Outcomes presented, relevance of outcomes to Ghanaian context, interpretations of statistics; Based on: perception of Ghanaian situation, intervention approach, perceived ease of implementation or experience with similar interventions -Research-specific factors - Methods/study design, Results General quality - Sampling methods, scale or coverage of intervention, methods of analysis; Additional information about findings Internal validity, 'soundness' of the study	B4 B5 C1 C2 C6 D1 D2 D6 D13 D14 E2 E5 F1 F4
Ademi 2018	Australia	CHEERS checklist items: 5,6,8,13,14,19:	B3 B4

		5. Setting and location - State relevant aspects of the system(s) in which the decision(s) need(s) to be made. 6. Study perspective - Describe the perspective of the study and relate this to the costs being evaluated. 8. State the time horizon(s) over which costs and consequences are being evaluated and say why appropriate. 13. Describe approaches used to estimate resource use associated with the alternative interventions. Describe primary or secondary research methods for valuing each resource item in terms of its unit cost. Describe any adjustments made to approximate to opportunity costs. 14. Currency, price date, conversion - Report the dates of the estimated resource quantities and unit costs. Describe methods for adjusting estimated unit costs to the year of reported costs if necessary. Describe methods for converting costs into a common currency base and the exchange rate. 19. incremental costs and outcomes - For each intervention, report mean values for the main categories of estimated costs and outcomes of interest, as well as mean differences between the comparator groups. If applicable, report incremental cost-effectiveness ratios	B7 B9 B10 D14 E2 E5 F2
Anderson 2010	United Kingdom	Firstly, compared with effectiveness studies, there is a much wider range of factors that limit the generalisability of cost–effectiveness results, over time and between health systems and service settings, including the context-dependency of resource use and opportunity costs, and different decision contexts and budget constraints. Secondly, because economic evaluations are more explicitly intended to be decision-informing, the requirements for generalisability take primacy, and considerations of internal validity become more secondary. Thirdly, since one of the two main forms of economic evaluation – decision analytic modelling – is itself a well- developed method of evidence synthesis, in most cases the need for a comprehensive systematic review of previous economic evaluations of a particular health technology or policy choice is unwarranted. To inform the development of an economic decision model - What are the key theoretical trade-offs (between levels and types of resources, and levels and types of outcome) implicit in a given treatment/policy choice? What do previously published empirical eco- nomic studies (with patient-level cost and outcome data) reveal or refute about such trade-offs? What are the strengths and weaknesses of previously used decision model structures and modelling approaches for evaluating similar decision problems? Are any of the previously developed models fit-for-purpose for analysing the current decision problem? To identify the one or two most relevant existing studies to inform a particular decision - Are there any currently published economic evaluations of the decision problem, which might be transferred, or	A B1 B4 D14

		adapted and updated, to reliably inform our present policy choice? In what ways are the cost–effectiveness results from this/these studies likely to differ for this jurisdiction at this point in time? To identify the key economic (causal) trade-offs implicit in a given treatment/policy choice or patient group - What are the key theoretical trade-offs (between levels and types of resources, and levels and types of outcome) implicit in a given treatment/policy choice? What do previously published empirical economic studies (with patient-level cost and outcome data) reveal or refute about such trade-offs?	
van Haalen 2014	The Netherlands	2.2.2 Step 2: Which Models Fit? Next, we specified a list of criteria to assess the model fit, thereby evaluating whether the model structure is appropriate in the context of a particular health care setting. The Dutch health care setting was used as an illustrative case. These criteria were based on the 'specific knock-out criteria' proposed by Welte et al. [4]. We distinguished between specific knock-out criteria that concerned parameter or structural uncertainty within a model. Parameter uncertainty relates to the uncertainty surrounding the input data and probabilities that govern the outcomes because of, for example, multiple (conflicting) studies, lack of internal or external validity of the study data, or lack of data [12]. Structural uncertainty is present when there is uncertainty about the functional form of the model, i.e. whether the model adequately reflects reality surrounding the decision problem. Generally, in the presence of structural uncertainty, one cannot be certain that the produced results are valid, even if the true values of all input parameters are known [12].	A B1 B4 D14
Gonzalez-Perez 2002	United Kingdom	Relevance to the UK but readily adaptable to another setting - The additional item concerns the transferability of the results, specifically: "Is the paper relevant for the NHS and are the results sufficiently transparent for them to be replicated in an NHS setting?" Three questions are relevant when considering the application of the results of a study to the NHS. First, are the technologies or diseases relevant for the UK population? Second, how similar is the study population to that covered by the NHS? Third, are all the costs and consequences relevant to the health sector or to society included?	A B4 D6 D14 C1
Fukuda 2011	Japan	Therefore, the second axis of evaluation was used to identify costing methodologies. The optimal choice of costing methods is the use of micro-costing or quasi-micro-costing, i.e., activity-based costing, in which the measurement of actual resource consumption is attempted. The second costing method involves the use of relative value units (RVUs). As there is a strong political dimension in the determination of charge data, the external validity of such estimates to different contexts is drastically reduced. The final category of the evaluation of the method of cost estimation includes studies that offer no information to readers about the methodology used.	A B1 B4 F4

Wolfenstetter 2010	Germany	Welte et al. systematically identified the factors that may influence the transferability of health economic study results between countries. These transferability factors can be differentiated into three categories: methodological characteristics (perspective, discount rate, medical cost approach, productivity cost approach), healthcare system characteristics (absolute and relative prices in health care, practice variation, technology availability) and population characteristics (disease incidence / prevalence, case-mix, life expectancy, health status preferences, acceptance, compliance, incentives to patients, productivity and work-loss time, and disease spread). All potential transferability factors have to possess four characteristics: influence on outcomes of economic evaluations, international variation, measurability and being distinguishable from other factors [35].	B1 B3 B8 C2 C6 C9 D1 D5 D6 E3 F1 F2
Li 2007	Australia	For the technical and methodology related barrier, two major issues emerge: one is the generalisability of efficacy data and the other is the generalisability of cost data. It would appear that the first issue of generalisability in treatment efficacy is comparatively easier to handle, particularly in view of the globalisation and digitalisation of communication, leading to the standardisation of healthcare and clinical practice around the world. The idea that healthcare decisions should be based on best evidence is generally well accepted and not controversial. The general consensus is that the evidence of efficacy should be obtained from scientifically robust evidence, with the hierarchy of evidence generally established through the discipline of evidence-based medicine.	B1 B2 B4 D13
Jaime Caro 2014	Canada	Model validation factors: Points to think about in general: - Did the model builders have a formal process for validating their model? [10] - Has a report of the validation been made available? - Is the validation process well described? - Have the types of validations performed been detailed? - Were the approaches to finding data sources for the validation reasonably comprehensive? - Were the data sources used for validation appropriate for the proposed uses of the model? - Were the methods for setting up the simulation of each source adequately described? Do they seem reasonable? For example, how well was the simulated population matched to the validation one? - Were those performing the validations blind to the results of the model? - Were results of the validations provided in sufficient detail? - Were the implications of the validations discussed adequately?	A B2 B4

		- Were there quantitative measures of how well the model's results match the outcomes observed in the data source?	
Coyle 1998	Canada	The standard design of trials involves direct interference with the clinical management of patients which can result in atypical and ungeneralisable estimates of resource use.[10] The additional costs included as a result of clinical trial design are referred to as protocol driven costs. The need to consider the issue of protocol driven costs has been long established in the guidelines for conducting economic evaluations alongside clinical trials.[11-15]	B2
Baltussen 1999	The Netherlands	This involves the consideration of context-specific factors such as the various procedures included in the economic evaluation, and specific physician, hospital and healthcare system characteristics. Also, indications of possible levels of noncompliance and their impact on costs are important in this regard. Furthermore, future developments that may have an impact on the cost effectiveness of the intervention under scrutiny should be considered. Possible approaches to these analyses are: Checklist - This approach allows policy-makers to make an assessment of the relevance of results in their own context. Sensitivity Analysis - By applying sensitivity analysis to those factors that seem to most affect cost effectiveness in the real world, and by screening the range of parameters relevant to the specific decision context, policy-makers can assess the specified impact.	B11 C3 D1 D2 D6
Abdul Pari 2014	United Kingdom	6. We recommend presenting results of economic evaluation in both a disaggregated and aggregated manner. This would ensure transparency of results while providing decision makers with opportunities to contrast clinically relevant out- comes together with economic endpoints.	A B1
Gheorghe 2015	United Kingdom	Finally, we propose two concepts that may advance generalisability research. First, we distinguish between the 'research space' and the 'policy space' and argue that policy makers are interested in the latter, while current methods describe the former.	E1 F1 F4
Schünemann 2006	Italy	Factors influencing the applicability or transferability of guidelines across different settings 1. Is there important variation in need (prevalence, baseline risk or health status) that might lead to different decisions? 2. Is there important variation in the availability of resources that might lead to different decisions? 3. Is there important variation in costs (e.g. of drugs or human resources) that might lead to different decisions? 4. Is there important variation in the presence of factors that could modify the expected effects (e.g. resistance patterns of microbiological pathogens), which might lead to different decisions? 5. Is there important variation in the relative values of the main benefits and downsides that might lead to different decisions?	B1 C2 C6 D1 D13

Adam 2003	World Health Organization	Methodological reporting areas: (Table 1) -perspective -choice of comparator (new vs doing nothing; new vs current practice) -types of costs (overhead, shared costs and provider time, indirect costs such as volunteers or lost time in seeking care, unrelated illnesses due to increased life-years) -data collection (reliable and valid sources, bottom-up vs top-down, price adjustments e.g. market price is a reflection of opportunity cost, exchange rates, time costs e.g. lost productivity, capital costs e.g. rental vs annualized costs, prices valued on costs vs valued on charges) -data analysis – discounting costs with range rates provided (2.5%-14% was range), capacity utilization, sensitivity analysis - reporting results – ingredient approach and transparency	A B1 B3 B4 B7 B11 C9 D1 E1 E2 F2
Berg 2017	The Netherlands	Some authors argue that the factors affecting the perception of applicability (the process question) and transferability (the outcome question) together might be broader than the factors associated with external validity [13]. Notwithstanding this difference, the EURONHEED method relies heavily upon the quality of reporting to ascertain transferability [32]. Therefore, such scores can be limited in use by the end-users for two reasons. First, a poorly constructed model could have been reported well scoring high on the transferability scale and vice versa. Secondly, without a threshold score, it is hard to judge a study or to rank and compare across the studies. Nixon et al. [32] argue that the EURONHEED score should, rather, be used as a general guide in making decisions, but also note that the explicit assessment of transferability using this method will introduce an educational element, helping researchers to improve the design, conduct and reporting of future studies.	B4 F4
Knies 2009a	The Netherlands	The third general knockout criterion states that the study should possess an acceptable quality, but it is not clear how the quality should be assessed by Welte [6]. Some common quality criteria for economic evaluations, like perspective and discount rate, are now used as transferability factors. Nevertheless, it is known that between countries, the guidelines for economic evaluations and the quality criteria differ. The last problem deals with the lack of attention for the transferability of effects. Almost all specific knockout criteria discuss the transferability of cost parameters. Only the criterion “health-status preference” focuses solely on the transferability of effects. Nevertheless, some criteria have an influence on both costs and effects. Furthermore, more attention should be given to how and with which instrument the effects are measured in a study, because this could influence the effect parameter. It is known from the literature that the valuation of health states and the instruments used vary from	A B2 B3 C5 D8 F2

		country to country [51]. There- fore, this should be taken into account when transferring effect data to other countries. This could be a new factor in an improved version of Welte's model. At this moment Welte's model is mainly focused on the idea to assess the transferability of whole studies. In the case that a study as a whole is not transferable, it could be that a study section is still usable. Therefore, a second option to improve Welte's model is to give more attention to the possibility to assess the transferability of a section or sections of a study.	
Knies 2009b	The Netherlands	Conclusion: All results indicate that the differences between the EQ-5D value sets are considerable and should not be ignored. Therefore, further research should focus on investigating the transferability of utilities across countries or agreeing on a standard to perform valuation studies. For the time being, transferring utilities from one country to another without any adjustment is not advisable.	C5
Coyle 2001	Canada	Economic evaluations based on data from more than one treatment center are important because they can allow for the identification of different hospital practices and differences in the provision of services, which may influence both the absolute and incremental costs of treatment. This can then assist in interpolating the study results to other settings and inform issues surrounding the adoption of the new technology in a given country.	D2 D14
Drummond 2001	United Kingdom	- Basic Demography – particularly population based interventions (screening/immunization), which are affected by incidence levels per age group (e.g. age, lifestyle, medical history) -Availability of resources and variations in clinical practice – wait times, availability of alternatives for treatment and/or providers who can offer the treatment -Incentives to providers and institutions – fee for service (may order more tests) vs capitation (may defer demand), hospital fixed amount per case (will want to free up bed sooner) or global budget -Relative prices or costs – do the relative prices differ, if other healthcare resources differ – things will appear more CE if the relative prices of hospitals, surgeries, providers, investigations is more expensive or based on different clinical care pathway -Population values – health states may vary country to country	B1 C1 C2 C5 C7 D1 D2 D4 D5 D9 E1 E2 E6
Drummond 2005	United Kingdom	Transparent Report Factors: Describe the characteristics of the centers participating in the trial. If these are from different countries, also report the relevant features of the various health-care systems. Report the types of patients excluded from the trials and the percentage of the normal caseload that these represent. Comparison with the relevant patient population outside the trial centers.	A B1 B2 B11 C3

		Describe the alternatives in detail, so that study users can assess the relevance to their own setting. Report costs and benefits by each relevant perspective. Report quantities separately from prices/unit costs. Report the source of the values and any instrument used. Provide details of quantitative analysis of variability by location. Ideally, this will be based on statistical analysis (such as multilevel modeling), but should at least incorporate standard sensitivity analysis. Provide details on the extent of incomplete observations (i.e., missing and censored data). Detail the characteristics of patients with incomplete data. Describe the methods used to address the problem.	C6 D14 E2 E6
Drummond 2015	United Kingdom	Use of transferability checklists - All of the organizations studied reported that the economic evaluations submitted to them contained data generated in other jurisdictions. Several checklists have been developed to assist those wrestling with the challenges of adapting studies or data from other jurisdictions. Respondents were asked indicate whether they had consulted any of those checklists published in the literature. In general, the checklists were not used, with the EUnetHTA Adaptation Toolkit (www.hta.ac.uk/links/finaladaptationtoolkitnetscc.pdf) being the most frequently mentioned (i.e. by 3 of the organizations studied).	A
Zwolsman 2019	The Netherlands	Transfer between countries -In this review, differences among countries are accounted for using transparent methods to adjust cost estimates. Variation in economic estimates that are attributable to differences between countries seems of low significance [58], but study outcomes are not generalizable when economic circumstances and differences in health systems across countries are not taken into account [53]. Oppong et al. has given more information about differences among countries and how these affect generalizability. Oppong et al. propose that overcoming systematic differences due to economic circumstances and health systems and improving generalizability can be achieved by: 1. Carefully selecting countries for inclusion in studies 2. Using a checklist to overcome heterogeneity 3. Use protocols on treatment patterns 4. Reporting costs from different perspectives	A B2 B3 B11 D2
Gray 2016	United Kingdom	But perhaps a greater impediment to transferring the result of a cost-effectiveness analysis is the assumption that the expected health benefits from the intervention are greater than the health gains that could have been achieved had the money required to pay for it been spent in some other way in the health system. This unavoidable consideration of local opportunity cost, which lies at the heart of	B8 D14

		whether an intervention can be judged to be cost-effective, also means that the determination of what is cost- effective can only ever be made at a local level.	
Hutton 2005	Switzerland	Input prices are available at the disaggregate level - When studies have used the ingredient approach, thus reporting quantities and prices of all factor inputs separately, analysts in other settings could adjust for identified differences in these quantities and prices. For example, if separate information is available on wages, then the approach is to identify the wage differentials between the source and destination countries, and adjust by the factor identified.	A B1 B4
Ginsberg 2013	Israel	Costs: (a) Costs of non-tradables, that have to be converted using exchange rates based on PPP (purchasing power parity), which take into account cost differences in the source and recipient country. These include labour, disposable equipment (like cotton wool), hospital laundry services etc. In health promotion interventions these can exceed 90% of costs, while in surgical interventions they usually account for between 70% and 80% of costs. Care should be taken to ensure that labour costs in the source article include social overheads such as the employers contribution to pension fund, educational funds and national insurance. If one is converting costs from a country where the employer does not contribute to social overheads then this would give a downwardly biased estimate of interventions in a recipient country where the employer does contribute to social overheads. (b) Cost of goods that are potentially tradable, these include pharmaceutical costs and fixed equipment costs. The use of published exchange rates can be used to convert foreign currency costs into local currency. Efficacy: The success rate (or efficacy or complication rate) of say bariatric surgery is dependent on the expertise of the operating surgeon as well as other factors. Generalizability of results is problematic between countries with different medical care levels. Coverage of Services: In most industrialised countries all (100%) of the population enjoy potential access to medical services. However in many developed countries, coverage is not universal, especially in out-lying regions, so potential DALY gains from such programs may be reduced. Compliance: However, in general imputing efficacy from clinical drug or vaccination trials to another country is not such a big threat to generalizability. But problems of generalizability occur with health education/ promotion interventions where overall efficacy will be subject to cultural and religious biases that affect efficacy and compliance. -Treatment savings: Here generalizability is hindered by possible differences in medical intervention styles, which can also vary within developed countries. Some advanced countries provide more	B1 B3 B9 C2 C3 C6 C8 D1 D3 D6 D10 D13 E3 F1 F2 F4

		services (eg; chemo- therapy, syphilis treatment) on a cheaper out-patient or day hospitalisation basis (eg; haemorrhoids or hernia operations). - Discount Rate. This is used to bring the stream of costs, treatment savings and also DALYS which occur in the future back to present values. Care should be taken that the discount rate used in the source article (say 3% per annum) is applicable now to your country. In some countries the treasury sets a recommended discount rate that should be used on publicly funded projects (including health). It is likely that discount rates will vary between countries because central bank rates, which are a rough indicator of discount rates vary considerably, from less than 0.25% (Japan, USA, Switzerland, Denmark and Chech Republic) to 7.4% (India), 9.0% (Argentina), 25% (Malawi) and 28% (Belarus) [12]. - Time Horizon. Ideally interventions should be evaluated for as long as their effects persist. Many projects are evaluated over a 100 year time horizon. For example, the saving of an infant's life today (whether from an operation or vaccination) will have implications far into the future. So care should be taken to make generalizations from studies which do not have evaluations that are long enough to capture all the effects of the intervention. - Different Perspectives. A societal perspective may be ideal (capturing costs such as lost productivity from work and transport costs) especially if the decision-makers and budget holders are responsible for the broader social impacts. However, for many publically funded systems, the health service perspective is taken, especially if funding is received for that specific system. Sometimes the health service perspective is reported since data may be unobtainable on costs falling outside the health system (eg; on lost work productivity). - Different WTP thresholds: If QALYs are bought at a cost higher than thrice the GDP per head, then the project is not cost-effective. So use of country specific thresholds is recommended. However it should be noted that there are interventions where treatment savings are larger than the intervention costs in which case the intervention is called cost-saving in effect a win–win situation where DALYs are reduced at no additional cost. -Disease Incidence: -Duration of Illness: This variable does not need to be adjusted as the duration of illness should not vary much across countries with similar levels of treatment provision. However care should be made in making generalizations in diseases such as stroke, since the availability of rehabilitation facilities will greatly influence not only the duration of the side-effects of the illness but also its severity - Life expectancy in country	

Mason 2006	United Kingdom	The technical elements of best practice are comparatively uncontroversial: choosing relevant alternatives; transparent reporting of methods and findings; accessing and applying the best-quality evidence; using best methods to synthesise data; and using deterministic sensitivity analysis to explore potential systematic bias whilst employing probabilistic sensitivity analysis to explore the influence of random error at the whole model level. The applicability of economic findings within their original policy context (e.g. national analyses based on generalisable within-country data) can be determined, provided that best practice guidelines for economic modelling are adhered to. The transferability of economic findings (from one policy setting to another, e.g. country, region, clinical setting or patient population) requires careful exploration of changes in resource implications, unit prices and outcomes, a process facilitated again by transparent reporting of methods, adjustment for baseline risk and potentially by recent statistical developments intended to deal with hierarchically structured data.	A B1 B3 B11 C2 D1 D14 E2
Munthe-Kaas 2019	Norway	One surprising result of the content analysis is that none of the identified checklists included factors related to religion, family structure, social equality, or welfare services. Within social care and public health, such factors could be considered important to the transferability of some review findings.	F1
Pichon-Riviere 2012	Argentina	Table 3: -Low methodological quality -Lack of transparency in the HTAs published -Differences in healthcare costs -Different epidemiological contexts -Different healthcare systems characteristics -Different scope of the reports	A B1 B3 B4 C2 D1
Ruggeri 2005	Italy	To evaluate the generalizability of the studies, authors should describe in detail the setting of care in which the study was carried out. The general objective is to help the final users to decide whether a given study is relevant to their own setting [9]. In particular, what should be reported are the structural and organizational requirements, the available equipment, the professionals' skills and expertise, and the reimbursement system. In conclusion, our results showed that a trade-off between relevance and generalizability of HEEs exists: generalizable studies are often less relevant to the target policy debate, whereas studies relevant to a specific context are usually scarcely generalizable. As a matter of fact, in Italy, HEEs are often carried out at a local level (hospital, organizational level) and are run to inform local decisions. Thus, the choice of data and methods is some- how affected by this specific goal, regardless of any transferability and generalizability issue. Therefore, this "local relevance-generalizability" trade-off	A B2 B4 D2 D3 D9 D14

		should be adequately faced, addressing these issues at the design, analysis, and reporting stage of HEEs.	
Steuten 2008	United Kingdom	It is acknowledged that boundaries between categories (in particular the patient, provider and healthcare system categories) are not exclusively set, as factors from these different categories may interact. - Demographics, such as factors, education, socioeconomic status, lifestyle, life- expectancy, attitudes towards new technologies, acceptance and compliance rates, types of insurance coverage and copayments. - Acceptability, such as culture and religion are typical parameters that might influence acceptance and can depend on the patient's compliance. The also is a presence or absence of direct monetary incentives for consumers, by means of copayments or coverage schemes and moral hazard if the costs of a technology are (partly) covered by someone else, for example, a health insurer or the public. The magnitude of moral hazard may depend on parameters such as the level of insurance coverage, copayments and deductibles on the use of healthcare [7]. - Disease, such as incidence and prevalence of disease, disease severity, case-mix and disease- specific mortality. In case of a technology that has, for example, high fixed costs but rather low variable costs, the amount of utilization will have a significant impact on its (average) unit costs. For variable capacity, utilization frequency is also linked with returns to scale [6]. Therefore, the cost of using a technology will to some extent be dependent on the disease incidence and prevalence in a particular country. The case-mix of the target population, in terms of comorbidity, severity of disease and risk factors, may have a significant impact on the effectiveness of technologies. - Provider, such as practice variation and treatment guidelines, leading to differences in rates of hospitalizations, rates of diagnostic testing, length of hospital stay and so on. Other provider characteristics relate to experience, education, skills, efficiency and where professionals stand on the learning curve of using new technologies [7]. Furthermore, incentives for providers, such as liability and method of remuneration, are likely to vary between jurisdictions and can have an important effect on how patients are managed [7]. - Healthcare system, such as differences in absolute and relative unit prices of healthcare that can have an important impact on cost–effectiveness [6,7]. Second, countries differ with respect to the types and magnitude of resources and services that are available. Conflict of Interest potential: Since economic evaluations consider incremental costs and effects of a new technology over another, any technology will prove cost effective when compared with a cost-ineffective, but perhaps irrelevant, comparator. Given the relatively large variation in terms of user	B1 B2 B3 B5 B7 B8 B9 B10 C1 C2 C6 C11 D1 D3 D4 D5 D6 D7 D9 E1 E2 E3 F1

		access to medical technologies [1] and the differential diffusion of technologies [2], even within developed countries, careful identification of the most appropriate comparator is necessary. Finally, as long as health- care systems have no uniform regulatory, reimbursement or procurement infrastructures for medical technologies, the potential impact of such healthcare system-related factors on the cost–effectiveness of technologies needs to be considered when one wants to transfer such data across jurisdictions. - Methodology, such as costing methodology that has been applied, the adopted study perspective, currency exchange rates, and opportunity costs. Of particular importance in the case of transferring economic evaluations of technologies, however, is the timing of the evaluation. As stated before, lifecycles of medical technologies are relatively short, which increases the desirability of transferring economic evaluation results instead of duplicating them in every jurisdiction. Notwithstanding these short lifecycles, there will still be a learning effect in using the technology, which should be taken into account. Another methodological factor that needs specific attention in transferring economic evaluation results of medical technologies is the clinical end points or outcome measures that are chosen for the evaluation. In some countries, as in the UK, the recommended outcome measure of an economic evaluation is 'incremental cost per quality-adjusted-life-year gained' [10,11] and national regulatory bodies largely base their recommendations regarding, for example, product reimbursement on this outcome measure. The target country, however, may well base such recommendations on other outcomes measures, for example pure cost-savings, and not even consider patient outcomes in case of technologies that are primarily aimed to benefit care providers, as might be the case with surgical technologies.	
Vemer 2010	The Netherlands	Identified 8 factors: Demographics, Prevalence, All-cause mortality in population, Epidemiology of disease, costs of disease, resource use (intervention costs, medications, providers), utility weights (general population and disease-specific utilities), Discount Rates (4% for reference case and 1.5% for outcomes)	C1 C2 C5 D1 E3 F2
Vemer 2011	The Netherlands	The role of the threshold value for a QALY has been given little consideration in these checklists, even though the importance of a factor as a cause of between country differences in CE depends on this threshold. We concluded that, when judging the transferability of a CE study we should consider the between country differences in WTP threshold values.	F4

Sculpher 2006	United Kingdom	The first step could be for each jurisdiction to develop a 'reference case' along the lines of that developed by NICE in the UK. Then discussions could take place to distinguish between 'justifiable' differences where opinions genuinely differ on, say, the choice of health state descriptive system or valuation method, and differences that are not easy to justify because the approach suggested in one or more jurisdiction is inconsistent with the objectives and constraints of the system. Table 1: -Consistent perspective on costs Failure to require a generic measure of health -Full set of comparators based on, for example, not the 'most clinically effective' options -Specification of the (sub)populations of interest -Inclusion of all relevant evidence through systematic identification -Full specification of parameter and structural uncertainty -Presentation of decision uncertainty	A B1 B3 B4 B11 C6 E2
Urdahl 2006	United Kingdom	(1) Definition of target decision-maker or jurisdiction—Being aware of the target decision-making audience for a model is important to a judgement about the appropriateness of the model and its inputs, and the review attempted to elicit the target decision-making audience or jurisdiction. (2) Transparent reporting of model specification—Transparent reporting of a model is a prerequisite to understanding the relevance of the model to the target decision- maker, as well as to assessing its generalisability to other decision-makers and jurisdictions. The specification of study setting (e.g. country, and primary or secondary care) and patient population was therefore extracted. In addition, the description and justification of alternative interventions was considered (3) Relevance of data inputs to target decision-maker or jurisdiction—The ease with which model inputs can be traced, and the relevance of those inputs to the stated decision-maker, will influence the degree to which a model is considered applicable in the target setting. The reporting of sources and the relevance of key data inputs to the model was therefore assessed, ranging from clinical data and their valuation to resource use and unit costs. Models that reported and referenced both baseline risks and risk reductions were considered as having provided sources of clinical data. (4) Assessment of robustness of model to variation in data inputs within and between jurisdictions—The use of sensitivity analysis to explore the robustness of model results to variation in data inputs that may exist within and between jurisdictions was assessed.	A B1 B3 B4 B11 C2 D13 D14 E2
Stawowczyk 2018	Poland	Clinical outcomes can be transferred to other countries and generalized; however, cost inputs are largely country-specific, which in turn limits the transferability and generalizability of the results and conclusions to other countries. Another issue is the differences in healthcare systems and	B4 B6 D9 E1

		reimbursement policies between countries, and also the methods of inpatient or outpatient care, which may have a significant influence on the results and final conclusions of economic evaluations.	
Baltussen 1996	The Netherlands	Is epidemiological information reported? * Are the patient streams (as determined by incidence and prevalence figures) indicated? * Are indications about interactions with diseases (risk reduction or risk enhancements) indicated? * Is additional information with respect to the presence of externalities provided? Are context-specific factors explicitly reported? * Are the various procedures undertaken in the RCT explicitly reported? * Are volume and price components reported separately? * Are fixed and variable costs related to the intervention specified? * Is the extent to which the case-mix of the patients participating in the RCT is representative of the patient case-mix in the 'real world' indicated? * Are the specific physician and hospital characteristics (such as experience) that seem most likely to most influence the cost estimation reported? * Are the characteristics of the health care system in which the RCT is carried out and that seem most likely to most influence the cost estimation reported? * Are any learning effects indicated? 3. Is there an indication of the impact of non-compliance? 4. Has 'real world' sensitivity analysis been applied? * Has sensitivity analysis been carried out regarding those factors which are most likely to cause deviations between RCT results and the 'real world' 5. Are future developments indicated? * Are future patient flows (as determined by the incidence and prevalence of the disease) indicated? * Is there an assessment of what kinds of patient groups will receive treatment in the future, as the intervention is diffused? * Is there an indication of possible future price changes for the equipment? The table can be considered as a checklist which decision-makers can use to assess the relevance of economic evaluations for their own decision context and which economic evaluators may review and use in their documentation. Many of these matters are related to the collection and use of additional data which will not always be available.	A B11 Did not code to avoid double-counting; but, found 2 additional emerging codes A and B11.
Birch 2003	Canada	This lack of generalisability is not confined to the results of economic evaluations, but includes the validity of the methods used to analyse the subjective component of the evaluation exercise. Particular	Did not code to avoid

		attention must be given to establishing the validity of a particular method of valuation for the setting in which it is to be used as well as the way in which the value scores are then used.	double counting
Mason 1997	United Kingdom	There are a number of reasons why cost-effectiveness data may not readily transfer from one setting to another. Factors include: the availability of alternative treatments; appropriate choice of comparison; local clinical practice and supporting care patterns; relative prices of alternative treatments and components of care; and incentives placed upon clinicians, hospitals and patients. If the relative prices of 2 or more technologies being evaluated differ between countries, their relative cost effectiveness will differ. Differences may be caused by the cost of the technologies them- selves or the relative use of associated procedures of diagnosis or care	Did not code to avoid double counting
O'Brien 1997	Canada	Threats to transferring data involve differences among countries with regard to demography and epidemiology of disease, clinical practice and conventions, incentives to and regulation of health care providers, relative prices levels, consumer preferences, and opportunity cost of resources.	Did not code to avoid double counting
Grutters 2011	The Netherlands	1. Objective: How will the HTA be used? (e.g., contribute to evidence, inform adoption decision) 2. Audience: What is the audience (principal users) for the HTA? (e.g., government, pharmaceutical companies, insurance companies, patient groups, jurisdiction) 3. Perspective: Which viewpoint or perspective is relevant for the HTA? (e.g., societal, health care, insurer, payer) 4. Population: What is the patient population relevant for the decision problem? (e.g., age, health status, sex, other characteristics) 5. Comparators: What are relevant comparators for the decision problem? (e.g., care as usual, alternative technologies) 6. Clinical practice: How are the technologies embedded in clinical practice? (e.g., diagnostics, clinical instead of research protocol) 7. Time horizon: Which time horizon is relevant for the decision problem? (e.g., lifetime, one year) 8. Consequences: Which consequences are relevant for the decision problem? (e.g., final versus intermediate outcomes, indirect and/or rare consequences) 9. Patient use: What is the patient use that is relevant for the decision problem? (e.g., uptake, compliance, adherence) 10. Professional use: What is the use of the technology by health care professionals that is relevant for the decision problem? (e.g., skills, experience, beliefs)	B1 B3 B5 B9 B10 C1 C2 C5 D3 D6 E1 E2 F1 F3

		11. Price and resource use: What price level and resource use are relevant for the decision problem? (e.g., personnel providing the intervention)	
Augusto viski 2009	South America	Quality of Reporting (Table II) -Reported economic evaluation perspective -Alternatives being compared were clearly described -Measure of health was specified -Year of costing was reported -Unit cost sources were reported -Unit costs and resources were reported separately -Compared with other health economic evaluations -Discussed results in context of other interventions -Caveats reported -Funding sources not stated -Public or international agencies -Industry	A B2 B3 B9 E2 D13

Appendix 5: Summary of General Codes Merged with Goeree et al. 2007[18] Start Codes and Frequencies

Themes and Sub-themes of Comparability When Deciding to use an Economic Evaluation	Total Endorsements from 131 articles	Endorsements From Newly Identified 45 Articles	Endorsements From 86 Articles from Goeree 2007[18] Systematic Review
Transparent Reporting of Factors Below	23	23	-
General Population and Disease Characteristics)			
- Demographics (age, gender, race), education, socio-economic status	39	10	29
- Epidemiology (incidence/prevalence, disease progression, spread)	36	13	23
- Setting of Intervention (e.g. outpatient vs in-patient)	16	16	-
- Disease severity, disease duration, case mix	14	7	7
- Mortality rates, life expectancy	13	5	8
- Risk factors, medical history, genetic factors	10	-	19
- Disease interaction, co-morbidity, concurrent medications	8	1	7
- Lifestyle, environmental factors	6	-	6
General Intervention Characteristics			
- Efficacy of Intervention	10	10	-
General Comparison (Usual Care) Characteristics			
- Available treatment options (comparators)	38	13	25
- Availability of generics or substitutes	5	2	3
- Clinical practice, conventions, guidelines, norms	71	7	64
General Outcome Characteristics			
- Clinical endpoints/outcome measures	8	5	3
General Patient Values and Preferences Characteristics			
- Patient Attitudes toward treatment, culture, religion, hygiene, nutrition	18	8	10
- Population values (utilities)	16	6	10
- Compliance and adherence rates, ethical standards	14	3	11
General Resource Use - Cost (Estimation and Methodology) Characteristics			
- Absolute or relative prices	79	20	59
- Sensitivity considerations, analysis	12	11	-
- Costing methodology, estimation procedures (e.g. productivity cost)	30	20	10
- Study perspective	27	16	11
- Study factors (artificial trial conditions, industry-related bias)	30	10	20
- Exchange rates, purchasing power parities	10	4	6
- Time Horizon, Timing of the economic evaluation	11	6	5
- Input mix (personnel, equip.), specialization of labor, joint production	8	2	6
- Opportunity cost (foregone benefits)	10	4	6
General Acceptability (Providers and Decision-maker) Characteristics			
- Incentives for institutions	14	4	10
- Decision-maker Affordability (cost-effectiveness thresholds)	8	6	2
- Method of remuneration (supplier-induced demand)	6	4	2

- Incentives for providers, liability	17	2	15
- Decision-maker Discount rates	15	9	6
General Feasibility Characteristics			
- Available resources (staff, facilities, equipment), programs, services	38	13	25
- Organization of delivery system, structure, level of competition	29	6	23
- Experience, education, training, skills, learning curve position	23	5	18
- Level of technological advancement, innovation and availability	13	8	5
- Capacity utilization, economies of scale, technical efficiency	13	3	10
- Regulatory and organizational infrastructure, licensing of products	5	1	4
General Equity Characteristics	-		

Coding List for All Papers – Ranking of Codes for Observations

Total n=131 n=45 n=86

A – Transparency of Reporting

23 total 23 new 0 from Goeree 2007 Review[18]

B – Certainty of Evidence (Methodological Quality & Cost Inputs)

B1 79 total 20 new 59 Absolute or relative prices

B4 30 total 20 new 10 Costing methodology, estimation procedures (e.g. productivity cost)

B2 30 total 10 new 20 Study factors (artificial trial conditions, industry-related bias)

B3 27 total 16 new 11 Study perspective

B5 13 total 03 new 10 Capacity utilization, economies of scale, technical efficiency

B11 12 total 12 new Sensitivity analysis

B7 10 total 04 new 06 Exchange rates, purchasing power parities

B8 10 total 04 new 06 Opportunity cost (foregone benefits)

B9 11 total 06 new 05 Time Horizon, Timing of the economic evaluation

B6 08 total 02 new 06 Input mix (personnel, equip.), specialization of labor, joint production

B10 08 total 05 new 03 Clinical endpoints/outcome measures

C – Population (Patient and Disease Characteristics)

C1 39 total 10 new 29 Demographics (age, gender, race), education, socio-economic status

C2 36 total 13 new 23 Epidemiology (incidence/prevalence, disease progression, spread)

C5 16 total 06 new 10 Population values (utilities)

C3 14 total 03 new 11 Compliance and adherence rates, ethical standards

C6 14 total 07 new 07 Disease severity, disease duration, case mix

C4 10 total none 10 Risk factors, medical history, genetic factors

C7	08 total	01 new	07 Disease interaction, co-morbidity, concurrent medications
C9	04 total	02 new	02 Income, employment rates, productivity, work loss time, friction time
C8	03 total	01 new	02 Population density, immigration, emigration, traveling patterns
C11	03 total	01 new	02 Mortality due to disease
C10	01 total	none	01 Type of insurance coverage, user fees, co-payments, deductibles Incentives for patients
D – Intervention (Provider and Health Care System Characteristics)			
D1	38 total	13 new	25 Available resources (staff, facilities, equipment), programs, services
D2	29 total	06 new	23 Organization of delivery system, structure, level of competition
D3	23 total	05 new	18 Experience, education, training, skills, learning curve position
D4	17 total	02 new	15 Incentives for providers, liability
D14	16 total	16 new	Setting of Intervention (e.g. outpatient vs in-patient)
D5	14 total	04 new	10 Incentives for institutions
D6	14 total	08 new	05 Level of technological advancement, innovation and availability
D13	10 total	10 new	Efficacy of Intervention
D9	06 total	04 new	02 Method of remuneration (supplier-induced demand)
D7	05 total	01 new	04 Regulatory and organizational infrastructure, licensing of products
D8	04 total	02 new	02 Quality of care provided
D10	03 total	01 new	02 Access to programs and services, gatekeepers, historical differences
D11	01 total	none	01 Market form of suppliers, payment of suppliers, supplier incentives
D12	01 total	none	01 Patient identification
E – Comparison (Usual Care Characteristics)			
E1	71 total	07 new	64 Clinical practice, conventions, guidelines, norms
E2	38 total	13 new	25 Available treatment options (comparators)
E3	13 total	05 new	08 Mortality rates, life expectancy
E4	06 total	none	06 Lifestyle, environmental factors
E5	05 total	02 new	03 Availability of generics or substitutes
E6	03 total	02 new	01 Waiting lists, referral patterns
F – Acceptability (Patient, Provider and Decision-maker Factors)			
F1	18 total	08 new	10 Patient Attitudes toward treatment, culture, religion, hygiene, nutrition
F2	15 total	09 new	06 Decision-maker Discount rates
F4	08 total	06 new	02 Decision-maker Affordability (CE thresholds)
F3	03 total	01 new	02 Provider Cultural Attitudes

CHAPTER 5. DISCUSSION

Summary of findings

This dissertation includes three studies highlighting important considerations when selecting indirect evidence for use by guideline development groups that apply GRADE. In Chapter 2, we integrated direct and indirect GRADE evidence profile evidence into one common absolute measure of association (absolute risk increase) from a previous established approach[1,2] – we used the pooled prevalence across ten low risk studies as our baseline risk, which we felt together simplified panel clinical decision-making. As this specific population was understudied in the literature, we applied prespecified decision rules for including indirect evidence; that being the inclusion of those indirect studies with opioid naive status[3] and an unknown proportion of participants with acute musculoskeletal injury - the nature of the registries that study authors used that did not always have a specific diagnosis, aside from an acute pain status not requiring hospitalization. In that work, we tested for potential differences between direct and indirect populations through sub-group analysis using a test of interaction to confirm whether results were importantly different between the two groups.[4,5] Including indirect evidence raised the number of included studies from 6 to 13 included studies and from approximately 60,000 participants to over 13,000,000. This highlights that the use of clear decision-rules may foster the merging of direct and indirect evidence together, which improves the simplicity of reporting for guideline panels and also the quantity of available data.

In Chapter 3, we described the current status of economic information reporting across published GRADE guideline evidence-to-decision frameworks in the absence of well formulated reporting guidance. We found that the overall rate of reporting of at least some indirect economic

information was high (91%); but, there was variability across pre-defined EtD Likert-type justification completion (70%), economic summary information reported or noted as not identified across EtD framework domains (57%), and remarks used to justify recommendations (38%). However, no decision rules for selection of this information was described or reported across our convenience sample of 1625 EtDs from 142 guidelines. With a very high reporting of information and only some guidelines using this indirect economic information to justify their clinical practice recommendations, this may suggest situations where evidence was too indirect to be useful to contribute to decision-making.

In Chapter 4, we built on our findings from Chapter 3 to explore the scientific literature about what study characteristics were most important to economists when selecting indirect evidence (referred to in economics as 'transferability') for use in a another situation, such as a clinical practice guideline. We found that moderate-to-high confidence in the evidence suggested that the GRADE EtD domains of Population, Intervention and Comparison elements of a research question, Resource Use estimation and methodology, and provider and decision-maker Acceptability are the most important indirectness (transferability) study characteristics that economists consider when choosing to use apply the results of an economic evaluation in another context. The application of these decision-criteria, act as a basis to potentially reduce the variability and magnitude of less helpful indirect economic information reported across GRADE guidelines.

Overall, we have found that use of indirect evidence is necessary in guideline development in many situations, such as communicable diseases, pain management, cardiovascular disease, primary care (e.g. pediatrics, pregnancy), cancer, allergic disease or venousthromboembolism. Selection of any available indirect evidence may be better than no evidence, as this acts as a basis

for decision-making beyond clinical equipoise. However, including all potentially relevant indirect evidence may represent an overuse of evidence, acting as a burden to panels and developers, with the potential for decision-making confusion, extra time demands and higher guideline funder costs.

Strengths and limitations

The strength of this work is in the rigorous, comprehensive and structured methods used that involve duplication of steps for screening, abstraction or coding steps across projects. Specifically, in Chapter 2 we presented pooled measures (direct and indirect studies) of association as relative and absolute risk increases, which strengthens inferences about importance. Additionally, in Chapter 4 we assessed the confidence in the evidence for the summarized comment information of most important indirectness (transferability) factors from articles.

Limitations were more project-specific. In Chapter 2, while we did include indirect evidence, there were still 34 of 47 (72%) predictors with sparse information from only one study. Also, the registries that authors used in the included studies could not always conclude definitively the reason for the initial or continued opioid prescription. For Chapter 3, while the convenience sample we obtained was large, many GRADE EtDs are not typically published and they were mostly a reflection of GRADE working group members that offered suggestions on sources. Additionally, Google Translate was used to convert languages, which may have introduced some language translation errors. For Chapter 4, a limitation was that we excluded non-English articles. Also, there was a dearth of empirically based articles to inform indirectness study characteristics considered most important by economists to assess. Lastly, we relied to the

coding of the previously published systematic review for some of the transferability study characteristics reported in older literature.[6]

Implications for future research

With the addition of the evidence-to-decision framework domains,[7] the demands for evidence of various types, beyond traditional benefits and harms, will continue to increase. Future research could examine how and when to best include various evidence competencies in the GRADE guideline development process. In our example of economic information indirectness, the decision-criteria tended to be rather specific to the methodology. A next step would be to formalize these economic evaluation study characteristics into an indirectness instrument that balances rigour and simplicity of use. Such an instrument may rely on aspects already explored in indirectness instruments that deal with values and preferences and models. The concept of establishing generic economic models may also be possible. However, considerations around indirectness of evidence informing the GRADE EtD acceptability, feasibility, and equity criteria is lacking and may have a bearing on these economic considerations. Exploring indirectness frameworks for these criteria may better help to streamline the process for future guideline developers and panel members. This could be partially informed through a review of published decision-criteria, if any, that previous GRADE guideline authors have used when selecting indirect evidence.

Conclusion

As interest in the GRADE evidence-to-decision frameworks increases, guideline developers, with the help of their panels, should work with the appropriate evidence competency (e.g. economistic

for economic literature) to establish and report clear decision-rules and the rationale for indirect evidence that they select for their clinical practice guidelines. Applying specific and transparently reported decisions for how indirect information is selected has the potential to optimize the presentation of this evidence for panels and developers, as well as reduce decision-making confusion, time demands and guideline funder costs.

References

1. Wang L, Guyatt GH, Kennedy SA, Romerosa B, Kwon HY, et al. Predictors of persistent pain after breast cancer surgery: a systematic review and meta-analysis of observational studies. CMAJ. 2016 Oct 4;188(14):E352-E361.
2. Wang Z. Converting odds ratio to relative risk in cohort studies with partial data information. journal of statistical software. 2013; 55(5):1-11.
3. Karmali RN, Bush C, Raman SR, Campbell CI, Skinner AC, Roberts AW. Long-term opioid therapy definitions and predictors: A systematic review. Pharmacoepidemiol Drug Saf. 2020 Mar;29(3):252-269.
4. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, et al; GRADE Working Group. GRADE guidelines: 8. Rating the quality of evidence--indirectness. J Clin Epidemiol. 2011 Dec;64(12):1303-10.
5. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, et al; GRADE Working Group. GRADE guidelines: 7. Rating the quality of evidence--inconsistency. J Clin Epidemiol. 2011 Dec;64(12):1294-302.

6. Goeree R, Burke N, O'Reilly D, Manca A, Blackhouse G, Tarride JE. Transferability of economic evaluations: approaches and factors to consider when using results from one geographic area for another. Curr Med Res Opin. 2007 Apr;23(4):671-82.
7. Alonso-Coello P, Schünemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, et al.; GRADE Working Group. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: Introduction. BMJ. 2016 Jun 28;353:i2016.