

AMERICAN THORACIC SOCIETY DOCUMENTS

Informing Healthcare Decisions with Observational Research Assessing Causal Effect An Official American Thoracic Society Research Statement

Andrea S. Gershon, Peter K. Lindenauer, Kevin C. Wilson, Louise Rose, Allan J. Walkey, Mohsen Sadatsafavi, Kevin J. Anstrom, David H. Au, Bruce G. Bender, M. Alan Brookhart, Raed A. Dweik, MeiLan K. Han, Min J. Joo, Valery Lavergne, Anuj B. Mehta, Marc Miravittles, Richard A. Mularski, Nicolas Roche, Eyal Oren, Kristin A. Riekert, Noah C. Schoenberg, Therese A. Stukel, Curtis H. Weiss, Hannah Wunsch, Joel J. Africk, and Jerry A. Krishnan; on behalf of the American Thoracic Society Assembly on Behavioral Sciences and Health Services Research

THIS OFFICIAL RESEARCH STATEMENT OF THE AMERICAN THORACIC SOCIETY WAS APPROVED SEPTEMBER 2020

Rationale: Decisions in medicine are made on the basis of knowledge and reasoning, often in shared conversations with patients and families in consideration of clinical practice guideline recommendations, individual preferences, and individual goals. Observational studies can provide valuable knowledge to inform guidelines, decisions, and policy.

Objectives: The American Thoracic Society (ATS) created a multidisciplinary *ad hoc* committee to develop a research statement to clarify the role of observational studies—alongside randomized controlled trials (RCTs)—in informing clinical decisions in pulmonary, critical care, and sleep medicine.

Methods: The committee examined the strengths of observational studies assessing causal effects, how they complement RCTs, factors that impact observational study quality, perceptions of observational research, and, finally, the practicalities of incorporating observational research into ATS clinical practice guidelines.

Measurements and Main Results: There are strengths and weakness of observational studies as well as RCTs. Observational studies can provide evidence in representative and diverse patient populations. Quality observational studies should be sought in the development of ATS clinical practice guidelines, and medical decision-making in general, when 1) no RCTs are identified or RCTs are appraised as being of low- or very low-quality (**replacement**); 2) RCTs are of moderate quality because of indirectness, imprecision, or inconsistency, and observational studies mitigate the reason that RCT evidence was downgraded (**complementary**); or 3) RCTs do not provide evidence for outcomes that a guideline committee considers essential for decision-making (e.g., rare or long-term outcomes; “**sequential**”).

Conclusions: Observational studies should be considered in developing clinical practice guidelines and in making clinical decisions.

Keywords: observational studies; practice guidelines; evidence synthesis

Contents	Results	Pragmatic RCTs
Overview	Strengths and Limitations of	Recommendations When
Introduction	Observational Studies	Incorporating Observational
Methods	Diversity and Health Equity	Research into Clinical Practice
Participants in the <i>Ad Hoc</i>	Observational Study Quality	Guidelines
Committee	Common Perceptions Surrounding	Recommendation Reevaluation
Evidence Review and Discussion	Observational Studies	Conclusions

ORCID IDs: 0000-0002-0246-594X (A.S.G.); 0000-0002-8414-0507 (P.K.L.); 0000-0003-1700-3972 (L.R.); 0000-0003-4685-6894 (A.J.W.); 0000-0002-0419-7862 (M.S.); 0000-0001-6452-2172 (K.J.A.); 0000-0002-4425-1288 (R.A.D.); 0000-0002-9095-4419 (M.K.H.); 0000-0003-2770-1360 (V.L.); 0000-0002-9850-9520 (M.M.); 0000-0001-6979-2542 (R.A.M.); 0000-0002-3162-5033 (N.R.); 0000-0001-7817-3516 (E.O.); 0000-0001-6892-3512 (K.A.R.); 0000-0003-3165-8529 (N.C.S.); 0000-0001-9283-8764 (T.A.S.); 0000-0002-3102-5273 (C.H.W.); 0000-0001-5525-4778 (J.A.K.).

Ⓜ You may print one copy of this document at no charge. However, if you require more than one copy, you must place a reprint order. Domestic reprint orders: amy.schrivner@sheridan.com; international reprint orders: louisa.mott@springer.com.

Supported by the American Thoracic Society.

Correspondence and requests for reprints should be addressed to Andrea S. Gershon, M.D., M.Sc., Sunnybrook Health Sciences Centre, G1 06, 2075 Bayview Avenue, Toronto, ON, M4N 3M5 Canada. E-mail: andrea.gershon@sunnybrook.ca.

Am J Respir Crit Care Med Vol 203, Iss 1, pp 14–23, Jan 1, 2021

Copyright © 2021 by the American Thoracic Society

DOI: 10.1164/rccm.202010-3943ST

Internet address: www.atsjournals.org

Overview

Decisions in medicine are made on the basis of knowledge and reasoning, often in shared conversations with patients and families in consideration of individual preferences and goals. Randomized controlled trials (RCTs) are generally considered to have the best study design for making inferences about the causal effect of an intervention on outcomes. Observational studies, however, can also offer valuable information and complement RCTs in many ways. This research statement summarizes the work of an *ad hoc* diverse multidisciplinary committee of the American Thoracic Society (ATS) to provide recommendations on the role of observational studies—alongside RCTs—in informing clinical and policy decisions in pulmonary, critical care, and sleep medicine. This statement has a specific focus on the role of observational studies for informing practice guidelines.

1. Observational studies have strengths as well as limitations. Many of their strengths complement those of RCTs. Whereas observational studies tend to have stronger external validity, RCTs tend to have stronger internal validity.
2. By studying larger and more representative samples of patients under real-world conditions, observational studies contribute to our knowledge of patients of diverse backgrounds and settings.
3. When assessing quality, the individual merits of each observational study should be considered, rather than discounting studies simply because of their observational nature.
4. In keeping with the ATS commitment to use the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework in developing clinical practice guidelines, observational studies should be included in the development of guidelines in the following circumstances:
 - When RCTs do not provide evidence for outcomes that the guideline committee considers essential (e.g., because they would be considered unethical or not feasible to conduct), observational evidence should be sought.

- When RCTs are appraised as being of low or very low quality, observational evidence should be sought to “**replace**” RCT evidence.
- When RCTs are appraised to be of moderate quality because of *indirectness, imprecision, or inconsistency* of the RCT evidence, observational studies should be sought to “**complement**” RCT evidence.
- When RCTs do not provide the best evidence for outcomes that the guideline committee considers essential for decision-making, such as rare or long-term outcomes, observational studies should be sought to be “**sequential**” to RCTs.

Introduction

As a patient service and advocacy organization, Respiratory Health Association is aware of the systematic underrepresentation of women and certain sub-populations (e.g., minorities, people with fewer socioeconomic resources) in randomized clinical studies. While this is a situation that needs to be remedied, increased use of observational studies can also help to address systematic underrepresentation for the benefit of the full demographic spectrum of patients.

—Joel J. Africk, President and Chief Executive Officer of Respiratory Health Association

Decisions in medicine are made on the basis of knowledge and reasoning, often in shared conversations with patients and families that consider individual preferences and goals. Clinical practice guidelines inform clinical decisions about medical care by using structured syntheses of evidence to formulate recommendations on the optimal course of action to prevent, diagnose, and manage disease—ultimately to improve health. Policies and programs also use evidence to inform best care for patients.

RCTs are generally considered to have the best study design for making inferences about the causal effect of an intervention on outcomes because there is random distribution of measured and unmeasured confounding characteristics (Table 1). Efficacy RCTs seek to enroll relatively homogeneous groups of individuals, with interventions delivered using a study

protocol to which adherence is strict. Therefore, observed treatment effects in efficacy trials may not be seen in real-world clinical conditions (1). These considerations have led to growing interest in pragmatic RCTs (also known as effectiveness or practical RCTs, which are designed to evaluate the effectiveness of an intervention in real-world practice conditions) that evaluate healthcare options in populations that more closely approximate people who receive care in routine healthcare settings. However, some clinical questions cannot be feasibly studied with RCT designs because of ethical issues or other factors (e.g., lack of time or other resources). Furthermore, RCTs can take a long time to complete and may not focus on rare or long-term outcomes. Various factors may also limit participation in clinical trials—including knowledge about clinical trials, perceived burden and personal benefit of participation, level of altruism, concerns about safety or being assigned to a less effective treatment, trust in healthcare providers, and access to transportation and health care—which may vary according to socioeconomic resources (2–9). Finally, evidence from RCTs may only be of very low, low, or moderate quality because of bias (e.g., low rates of completed follow-up visits, publication bias), lack of generalizability, imprecision, and inconsistency of evidence; in such cases, evidence from well-designed observational studies can be helpful in filling evidence gaps.

In contrast to RCTs, observational (or “nonexperimental” or “nonrandomized”) studies are those in which individuals are observed for outcomes of interest, usually in the course of routine medical care. Although researchers make no attempt to *actively* assign patients to different treatments, such studies can also estimate causal effects. Situations in which a “dose response” can be observed and/or in which plausible residual confounding would be expected to attenuate the effect estimate can contribute to compelling evidence for causal inference (10). The Hill criteria are also used to help determine whether observed associations are causal (11). Observational studies comparing the effectiveness of treatment options using existing data (e.g., claims data) can

generally be completed more quickly than RCTs. Thus, although observational studies are not without their own limitations, they can complement the shortcomings of RCTs. Observational studies also often capture more diverse patient populations and practice settings, allowing generalizability that even well-conducted RCTs rarely provide. Organizations like the Food and Drug Administration and the European Medicines Agency use population-based observational studies of real-world data to support regulatory decision-making.

The purpose of this research statement is to describe the ways in which observational studies and RCTs can be used to inform clinical decisions in pulmonary, critical care, and sleep medicine, with a specific focus on the role and inclusion of observational studies for informing practice guidelines. Although observational methods can be applied to answer a wide variety of questions (both descriptive and predictive), this research statement focuses on observational studies designed for the purpose of making causal inferences about the effects of interventions (treatments). It also focuses mainly on studies that use existing and/or secondary data—retrospectively or prospectively obtained—that were not collected to address the specific research question being studied. This includes studies comparing effectiveness of interventions (termed “comparative-effectiveness studies”). The current research statement will review strengths and weaknesses of observational studies, methods of evaluating observational study quality, and how observational studies align with pragmatic RCTs. This will provide context for the practical recommendations provided for the incorporation of observational studies into clinical practice guidelines. This document is intended for anyone who synthesizes evidence or who uses synthesized evidence and is concerned about how that synthesis is informed.

Methods

Participants in the *Ad Hoc* Committee

The *ad hoc* committee included ATS members and nonmembers of different gender and professional backgrounds with clinical expertise in pulmonary, critical care, and sleep medicine, as well as individuals with expertise in observational

research study design, RCTs, pragmatic controlled trials, clinical practice guidelines, quality improvement, population health, behavioral health, epidemiology, health services research, patient-centered care, comparative-effectiveness research, implementation science, biostatistical methods, and health economics. A caregiver also participated. Two decision-makers provided comment. Potential conflicts of interest, including intellectual and financial conflicts, were disclosed and managed in accordance with the policies and procedures of the ATS.

Evidence Review and Discussion

Committee participants were divided into working groups that focused on the following areas: 1) the strengths of observational studies and how they complement RCTs—specifically how they address diversity and health equity; 2) perceptions of observational research; and, finally, 3) the practicalities of incorporating observational research into evidence syntheses for medical decision-making, specifically for ATS clinical practice guidelines.

Participants were provided with articles from a targeted literature search to facilitate discussions within the working groups. From February to May of 2018, each working group was tasked with summarizing the literature and formulating provisional conclusions and recommendations for further discussion at an in-person meeting on May 19, 2018, during the ATS International Conference in San Diego, California. At the meeting, co-chairs led discussions by the working groups. During the meeting, participants believed that because pragmatic RCTs share some similar features with observational studies, brief discussion of their utility should be included in the final statement. After the meeting, discussions to refine ideas continued through teleconferences.

Drafts of the research statement were written, revised, and circulated to all members of the committee to seek further feedback. Additional teleconferences were held, and suggestions were incorporated until consensus was reached among the committee members. Two policy-makers were also invited to provide comments. The revised draft was submitted to the ATS. The report underwent peer review and revision, with all committee members reviewing before finalization. The final version of the

research statement was presented to the ATS Board of Directors for approval.

Results

Strengths and Limitations of Observational Studies

Beyond generally taking less time and money to conduct, observational studies have a number of strengths, as outlined in Table 1 (12). They have the potential to provide high-quality evidence. They can complement evidence from efficacy and pragmatic RCTs when concerns exist about the representativeness of patient and provider experience or conditions under which the trial was conducted (e.g., intervention rigidly implemented or study cohort not representative of disease populations). Observational studies have the ability to incorporate experiences of vast numbers of patients treated across a wide range of real-world practice settings, including academic and community settings. Observational studies can also be used to assess interventions that would be unethical to study with an RCT because of lack of equipoise or prevailing restrictions and attitudes toward the risk of harms and/or potential benefits. Furthermore, they can address questions that are not under the control of the investigator, such as questions about genetic markers. Finally, the large sample sizes and extended follow-up periods typical of many observational studies provide the time and statistical power needed to identify rarer exposures or outcomes—like adverse effects or later-emerging benefits/harms that could affect clinical practice guideline recommendations.

Observational studies also have limitations that are important to recognize (Table 1). First, without randomization, comparison groups—including those who have and have not received an intervention—are likely to differ in ways that are associated with the outcomes of interest. Although methods exist that can adjust for observed differences between treated and untreated individuals, it is challenging for them (although not impossible through methods such as instrumental variable analysis and Mendelian randomization) to account for unobserved differences or unmeasured confounders (12, 13). Observational studies may also select patients for study entry or

exposure classification in ways that cause spurious effect estimates, even in the absence of confounding. Some common examples of biases that are not due to confounding include collider-stratification bias induced by study entry based on an effect of the exposure of interest, immortal time bias arising from improper classification of time before exposure, and unintentional conditioning on effect mediators (14, 15). Challenges to identifying, measuring, and adjusting for all potential confounders and accounting for all selection biases make observational studies more susceptible to bias than RCT designs. Second, many large observational studies rely on secondary data that were not collected for research purposes. Such data may lack desirable granular clinical detail—such as pulmonary-function and other test-result data—that would help identify patient characteristics or patient-reported outcomes, such as symptoms and quality of life. This could make these types of studies vulnerable to misclassification (if the investigators tried to categorize a variable that was not reliably measured because of lack of detail), unmeasured confounding, and conclusions based on outcomes that are not the most relevant to patients.

Observational studies have strengths and weaknesses that tend to complement those of RCTs. For example—as discussed above—whereas high-quality observational studies tend to have potential for stronger external validity, RCTs tend to have stronger internal validity. Other complementary features are shown in Table 1. Thus, the best medical evidence on which to base clinical decisions is that which is determined to be of the highest quality in both designs.

Diversity and Health Equity

Because active participation is not required, and a waiver of informed consent can be obtained from research ethics boards, observational studies are more likely to include representative and diverse patient populations. With larger and more representative samples, observational studies provide an opportunity to explore treatment effect heterogeneity at multiple levels, including by sex, race/ethnicity, comorbidity, adherence to therapy, and access to care. Such information is fundamental to understand and overcome factors that contribute to health inequity.

In contrast to observational studies, RCTs often include participants who are not representative of the populations affected by the diseases they examine. Studies have shown that approximately 6% of patients with asthma (16) and 27% of patients with chronic obstructive pulmonary disease (17, 18) meet eligibility criteria of contemporary RCTs in these areas. A recent systematic review found that African Americans—a group disproportionately at risk of asthma morbidity and mortality—were underrepresented in RCTs assessing adherence to asthma medications (19). Barriers that prevent people from being recruited and participating in RCTs, such as strict eligibility criteria (20–22), difficulties contacting potential participants or their surrogate decision-makers, large participant burden in terms of time and effort, beliefs about research, and need for informed consent (23), tend to disproportionately impact minorities, patients of lower socioeconomic status, and patients with mental health conditions or other comorbidities (24). Having limited access to healthcare providers who participate in clinical trials and structural racism and research abuses that have led people to, understandably, distrust the medical establishment also reduce the involvement of underrepresented groups in research (25). Although we must strive to overcome such obstacles in the conduct of RCTs, observational studies may overcome such barriers to participation by underrepresented groups.

Observational Study Quality

Weaknesses of observational studies may be mitigated with rigorous study design and the use of causal diagrams. Observational study methods that reduce confounding and strengthen causal inference have developed greatly in the past 15 years and can be conducted by knowledgeable researchers (12, 26, 27). One of these is targeted trial emulation, which is the application of design principles from randomized trials to the design and analysis of observational studies. This has been shown to help researchers identify and avoid unnecessary biases and provide a clear means for articulating the trade-offs that need to be made in observational studies (28, 29). In addition, reports indicate that outcomes of high-quality observational studies and RCTs often do not differ (30–32).

When assessing study quality, the individual merits of each observational study should be considered, rather than discounting studies simply because of their observational nature. Tools for assessing the quality of observational studies—including those that investigate causal inference (26)—include the Newcastle-Ottawa Scale and ROBINS-I (Risk of Bias in Non-Randomized Studies—of Interventions) tool. The Newcastle-Ottawa Scale was developed to assess the quality of observational studies (http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp), and the ROBINS-I tool was developed to address the risk of bias in observational studies. Other guides also exist.

Common Perceptions Surrounding Observational Studies

Although observational studies are becoming better understood and more accepted by the scientific community, there is still some mistrust of their validity by clinicians that can lead to exclusion or discounting of their results during evidence syntheses and clinical decision-making (33). Such distrust has been perpetuated by generalizations that, although they might apply to some observational studies, are taken by some to be absolute. Table 2 outlines some of the perceptions surrounding observational studies.

Pragmatic RCTs

Pragmatic RCTs occupy a space between traditional efficacy RCTs and observational studies. This trial design prioritizes design decisions such that study results may more closely mimic those observed in routine clinical conditions (34). Such design decisions include eligibility criteria that rely on data collected during routine care, research embedded in clinical practice, and some flexibility in intervention delivery. Although pragmatic RCTs attempt to reflect real-world circumstances, studies employing pragmatic trial designs often require informed consent and other forms of active patient participation (e.g., completing study questionnaires) that could limit the applicability of study results to real-world clinical populations. Observational studies can also evaluate interventions in real-world conditions. When done well, because they can include more people, they have the potential to address external validity in a way that pragmatic RCTs cannot.

Table 1. Potential Strengths and Limitations of RCTs and Observational Studies That Examine Causal Associations

Strengths of RCTs	Strengths of Observational Studies
Random distribution of measured and unmeasured confounding factors reduces bias Blinding of RCTs minimizes performance bias and assessment bias Generally simpler to understand Accepted by the medical community	Produce results with higher levels of generalizability/external validity with regard to research participants and practice settings Able to capture diverse patient populations Able to study clinical questions and reduce potential harms associated with interventional research when equipoise is unclear Better suited to the study of rare outcomes and those that require long periods of follow-up Data sources frequently used to conduct observational studies of intended treatment effects typically include very large numbers of patients, providing more power than is achieved in most RCTs and allowing evaluation of treatment-effect heterogeneity Often require less time and/or are less costly to conduct Can be conducted in situations in which randomization is not feasible In retrospective studies, less alteration of behavior due to awareness of being studied
Limitations of RCTs	Limitations of Observational Studies
Limited generalizability because of the recruitment of select populations cared for under optimal conditions and/or often do not reflect real-world circumstances Unethical to study clinical questions that do not have equipoise Take a long time to complete Often expensive Difficult for rare diseases, rare outcomes, and long follow-up Need for informed consent, and stringent exclusion criteria might limit external validity Alteration of behavior due to awareness of being studied (Hawthorne effect)	Difficult to control for unmeasured confounding or other bias Often use secondary data sources that are vulnerable to missing data and misclassification error, often do not provide patient-reported outcomes, and can have poor-quality data or lack data needed to establish causal effect Not always accepted by the medical community

Definition of abbreviation: RCT = randomized controlled trial.

Recommendations When Incorporating Observational Research into Clinical Practice Guidelines

According to the National Academy of Medicine, clinical practice guidelines make recommendations informed by the best available evidence identified by systematic reviews (35). It is imperative that clinical guidelines that influence the care of millions of people reflect evidence of best-known care. Appropriate inclusion of observational studies in evidence syntheses can contribute to this.

Currently, many guideline groups use a stepwise approach to identify the best evidence. RCTs are initially sought and, as a group, assessed for quality; if RCTs are identified and are determined to be of “sufficient” or “adequate” quality, they are used to inform a recommendation. What constitutes sufficient quality is left to the discretion of the guideline panel. If no RCTs are identified or if the available

RCTs are determined to be of insufficient quality, then observational studies are sought. Observational studies of sufficient quality are first used to inform the recommendation. If no observational studies are identified or if the available observational studies are determined to be of insufficient quality, then indirect evidence is sought. Finally, if sufficient indirect evidence does not exist, then either no recommendation is made or uncontrolled studies or expert opinions (i.e., clinical knowledge and experience) are used to inform recommendations.

This stepwise approach, which saves time and effort, means that observational studies are considered only on an as-needed basis, rather than being considered routinely.

Although this “as-needed” approach to the inclusion of observational studies in practice guidelines is practical and efficient, it means that the totality of evidence is often not used to inform patient care.

Varying approaches used by other guideline developers may lead to the selection of different studies, resulting in inconsistent estimates of effects and different recommendations across guidelines. A more unified, widely accepted approach to the use of observational studies is desirable.

In 2005, the ATS adopted the GRADE (36) approach, a dynamic paradigm for appraising and summarizing evidence, as well as for formulating, writing, and grading recommendations (37). GRADE is also endorsed by the World Health Organization, Cochrane Collaboration, American College of Physicians, and other guideline-developing organizations (<http://gradeworkinggroup.org>). GRADE recognizes that study type is not the only indicator of study quality, as all study designs can be of variable quality. GRADE uses study design to make an initial assumption about the quality of evidence and then provides criteria that warrant upgrading the quality of a body of

Table 2. Perceptions and Generalizations Surrounding Observational Research by Some in the Medical Community

Perception/Generalization	Reality	Additional Comments
Study quality can be determined by the “hierarchy of evidence” in which observational studies are always of inferior quality compared with RCTs.	Study design is only one factor that determines study quality.	The traditional hierarchy of evidence has been updated by more accurate frameworks that consider study design (e.g., GRADE) and other factors. Different study designs are suited to studying different aspects of medicine.
Observational studies cannot determine causal association.	Minimal risk-of-bias associations shown by observational studies support causal association.	Methods are available to determine how well a study establishes causal effect, regardless of study type. For example, GRADE recognizes that an observational study supports causal association if there is a large effect size, a “dose–response” gradient, and/or if all plausible residual confounding results in an underestimate of an apparent association (10).
Because randomization does not occur, unmeasured confounding limits the interpretability of observational studies.	Confounding can be minimized through careful study design and appropriate analyses and can be further addressed through sensitivity analyses.	Assessing the quality of study designs means scrutinizing them for different types of bias. Sensitivity analyses offer ways to address the likelihood of bias if it exists (12).
Conflicting results from observational studies and RCTs that address similar research questions prove that observational studies are of poor quality.	Differences in observational studies and RCTs addressing similar research questions are commonly explained by factors other than study design, such as differences in types of patients being studied, definitions of study variables, and/or study settings (ideal vs. real-world conditions) (28).	Disagreement rates between RCTs and observational studies are no greater than disagreement rates between different RCTs addressing the same research question (12–14).
Observational studies, unlike RCTs, can be manipulated to produce results of interest.	Observational studies and RCTs can be manipulated. Researchers are encouraged to submit study protocols before analyses begin (e.g., to clinicaltrials.gov or to the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance).	The development of tools to ensure reliability and prespecification of study procedures in observational studies lags behind that in RCTs, but these tools do exist in observational research. For example, the STROBE statement and the RECORD statement are tools to assess completeness of reporting of observational studies (41).
Because of randomization, RCTs are free from bias.	RCTs can have many biases.	Some possible biases of RCTs include selection bias, performance bias, detection bias, attrition bias, and reporting bias (42).

Definition of abbreviations: GRADE = Grading of Recommendations Assessment, Development and Evaluation; RECORD = Reporting of Studies Conducted Using Observational Routinely-Collected Health Data; RCT = randomized controlled trial; STROBE = Strengthening the Reporting of Observational Studies in Epidemiology.

evidence (e.g., large magnitude of effect, dose–response gradient, plausible confounders contributing to opposite effect) or downgrading the quality of a body of evidence (e.g., risk of bias, indirectness, inconsistency, imprecision, and publication bias).

The GRADE working group is developing guidance for the use of observational studies in the development of clinical practice guidelines (10, 38, 39). An algorithm for including observational studies in the development of guidelines is provided in Figure 1.

An observational study can provide higher-quality evidence than an RCT. When

no RCTs are identified or RCTs are appraised as being of low or very low quality, observational studies are sought; in this case, observational evidence is considered “**replacement**” evidence because it may substitute for RCT evidence or its quality may surpass RCT evidence (38). When moderate-quality RCT evidence exists, the justification for the moderate-quality rating should be reviewed. Observational studies may be sought when concerns such as indirectness, imprecision, or inconsistency have led to downgrading the RCT evidence from being of high quality to being of moderate quality; in this instance, observational evidence is considered

“**complementary**” because it provides additional information.

Indirectness refers to situations in which the studies included patient populations, interventions, comparators, or outcomes that were different from those in the question posed by the guideline panel (36). Indirectness is often used to refer to a lack of generalizability. As an example, if a guideline panel asks about vaccinations in the elderly, but all relevant studies enrolled younger volunteers, indirectness of the population exists. *Imprecision* indicates that the confidence interval (CI) of the estimated effect is too wide to definitively answer the question asked by the guideline panel (i.e., the ends

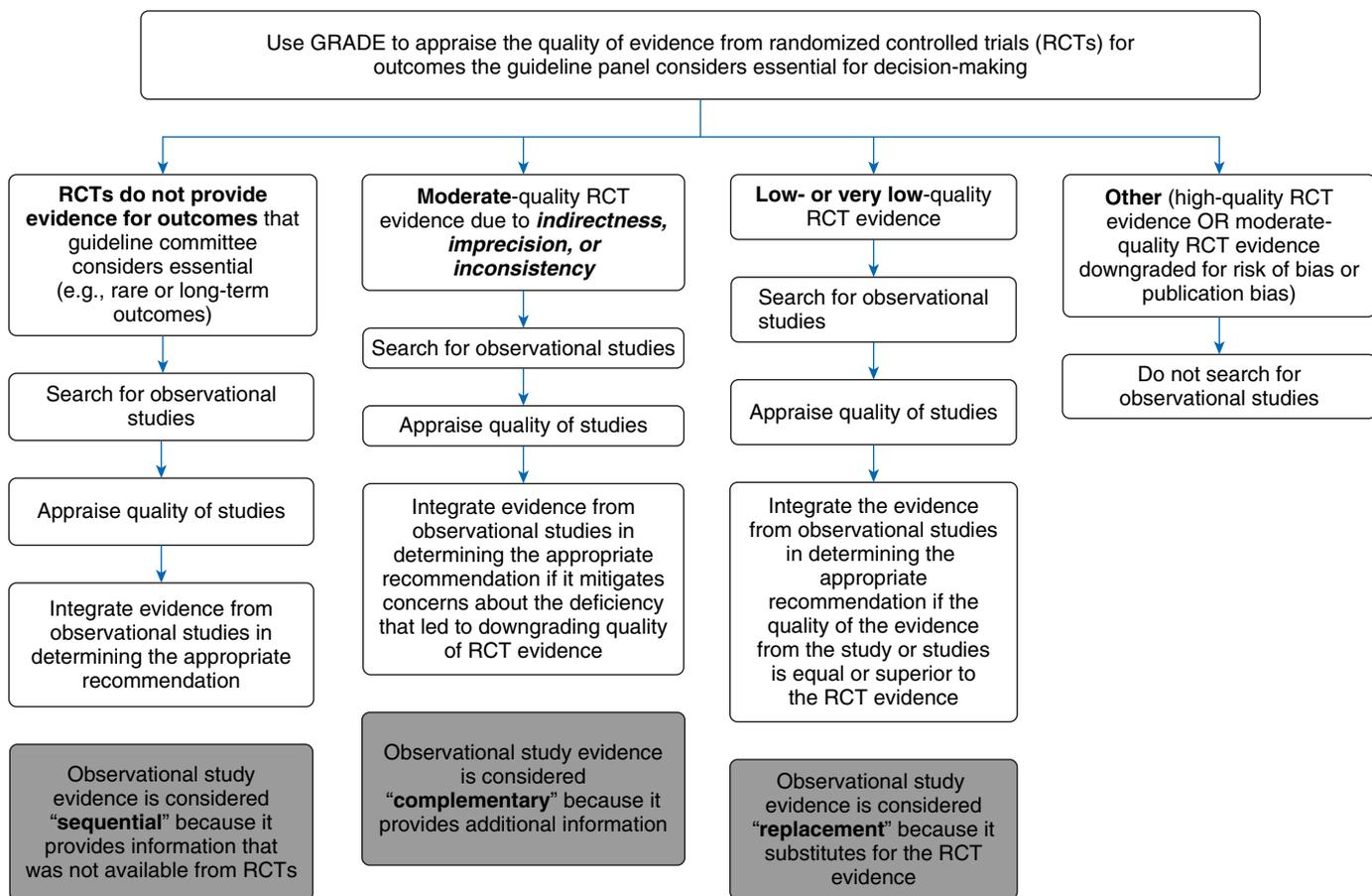


Figure 1. Algorithm for including observational research in medical decision-making. GRADE = Grading of Recommendations Assessment, Development and Evaluation.

of the CI would lead to different clinical decisions). As an example, if a guideline panel decided *a priori* that a 5% mortality reduction is necessary to use a drug and the studies estimate that the drug reduces mortality by 7% with a 95% CI of 3–11%, then imprecision exists because at one end of the CI you would use the drug and at the other end you would not. *Inconsistency* exists if there is variability in the direction or magnitude of effect across studies; this determination may be subjective or may use the I^2 statistic or a *P* value for heterogeneity. Indirectness, imprecision, and inconsistency are the causes of downgrading evidence from having high quality to having moderate quality that warrant seeking complementary evidence because these are the limitations that may be overcome by observational evidence. As examples, consider the following: if RCTs are limited by indirectness, then observational studies that directly address the guideline question may be found; if RCTs are limited by imprecision,

then large observational studies with narrow CIs may be found; and if RCTs are limited by inconsistency, then multiple consistent observational studies may be found. In contrast—although this is controversial—it is less certain that observational studies can overcome RCTs with risk of bias because observational studies, according to GRADE, also have a risk of bias.

Finally, observational studies may be sought when a guideline panel surmises that RCTs do not provide the best evidence for outcomes that the guideline committee considers essential for decision-making, such as when rare or long-term outcomes are judged as critical. In this situation, observational evidence is considered “**sequential**” because necessary information is not available from RCTs, so it must be obtained from observational studies. Sequential evidence and replacement evidence are frequently confused. Generally speaking, replacement

evidence is from observational studies that are sought because there is no RCT evidence or very poor RCT evidence, whereas sequential evidence is sought because, although adequate quality RCT evidence exists, the RCT evidence may be incomplete or too narrow to be informative. As an example, if a guideline committee was addressing a bronchoscopic intervention for asthma, there might be high- or moderate-quality RCTs reporting short-term outcomes, but the guideline committee might also be interested in long-term outcomes not reported by the RCTs and, therefore, might seek observational studies as sequential evidence.

Observational evidence is unnecessary when RCT evidence that examines outcomes that the guideline committee considers essential for decision-making (i.e., critical outcomes) is appraised as being of high quality (i.e., observational evidence

is not needed to replace, complement, or provide sequential evidence).

Recommendation Reevaluation

Our recommendations on when to integrate observational study evidence in clinical guidelines and evidence synthesis should be evaluated, and updates should be made as we learn about its benefits (e.g., how often observational evidence changes guideline recommendations). In these evaluations, consideration should be given to the evolution of the GRADE approach as well as to the added resources needed to search for and identify observational studies,

review and assess their quality, and make decisions about their suitability for inclusion.

Conclusions

Observational research is important to guide medical decisions about patient care, programs, and policy. Its importance will likely grow as we seek knowledge to guide personalized medicine; as real-world data and real-world evidence are increasingly requested by decision-makers (40); as rich data sources from electronic medical records,

for example, become more plentiful; as RCTs get more expensive; and as we increase our commitment to diversity and health equity. Quality clinical practice guidelines are instrumental in synthesizing evidence and making recommendations that improve health outcomes for millions of people around the world. The strongest medical evidence is often supported by both observational studies and RCTs; thus, both observational and randomized studies are key to informing decisions and maximizing the health of our patients. ■

This official research statement was prepared by an *ad hoc* subcommittee of the ATS Assembly on Behavioral Science and Health Services Research.

Members of the subcommittee are as follows:

ANDREA S. GERSHON, M.D., M.Sc.¹ (*Co-Chair*)
 JERRY A. KRISHNAN, M.D., Ph.D.² (*Co-Chair*)
 PETER K. LINDENAUER, M.D., M.Sc.³
 (*Co-Chair*)

JOEL J. AFRICK, A.B., J.D.⁴

KEVIN J. ANSTROM, Ph.D.⁵

DAVID H. AU, M.D., M.Sc.⁶

BRUCE G. BENDER, M.D.⁷

M. ALAN BROOKHART, Ph.D.⁸

RAED A. DWEIK, M.D.⁹

MEILAN K. HAN, M.D., M.S.¹⁰

MIN J. JOO, M.D., M.P.H.²

VALERY LAVERGNE, M.D., M.Sc.¹¹

ANUJ B. MEHTA, M.D.¹²

MARC MIRAVITLLES, M.D.¹³

RICHARD A. MULARSKI, M.D., M.S.H.S, M.C.R.¹⁴

EYAL OREN, M.S., Ph.D.¹⁵

KRISTIN A. RIEKERT, Ph.D.¹⁶

NICOLAS ROCHE, M.D.¹⁷

LOUISE ROSE, R.N., B.N., Ph.D.¹⁸

MOHSEN SADATSAFAVI, M.D., Ph.D.¹⁹

NOAH C. SCHOENBERG, M.D.²⁰

THERESE A. STUKEL, Ph.D.^{21,22}

ALLAN J. WALKEY, M.D., M.Sc.²³

CURTIS H. WEISS, M.D., M.S.²⁴

KEVIN C. WILSON, M.D.²³

HANNAH WUNSCH, M.D., M.Sc.^{25,26}

¹Department of Medicines, Sunnybrook Health Sciences Centre and University of Toronto, Toronto, Ontario, Canada; ²Division of Pulmonary, Critical Care, Sleep and Allergy, Department of Medicine, University of Illinois at Chicago, Chicago, Illinois; ³Institute for Healthcare Delivery and Population Science, University of Massachusetts Medical School-Baystate, University of Massachusetts, Springfield, Massachusetts;

⁴Respiratory Health Association, Chicago, Illinois; ⁵Department of Biostatistics and Bioinformatics and ⁸Department of Population Health Sciences, Duke University, Durham, North Carolina; ⁶Division of Pulmonary, Critical Care and Sleep Medicine, Department of Medicine, University of Washington, Seattle, Washington; ⁷Division of Pediatric Behavioral Health, Center for Health Promotion, and ¹²Division of Pulmonary, Critical Care, and Sleep Medicine, Department of Medicine, National Jewish Health, Denver, Colorado; ⁹Respiratory Institute, Cleveland Clinic, Cleveland, Ohio; ¹⁰Division of Pulmonary & Critical Care, University of Michigan, Ann Arbor, Michigan; ¹¹Department of Microbiology, Infectious Medicine, and Immunology, University of Montreal, Montreal, Quebec, Canada; ¹³Department of Pneumology, Vall d'Hebron University Hospital, Barcelona, Spain; ¹⁴Department of Pulmonary and Critical Care Medicine, Center for Health Research, Kaiser Permanente Northwest, Portland, Oregon; ¹⁵Division of Epidemiology & Biostatistics, School of Public Health, San Diego State University, San Diego, California; ¹⁶Division of Pulmonary and Critical Care Medicine, Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, Maryland; ¹⁷Department of Respiratory Medicine, Cochin Hospital and Institute, APHP Center University, Paris, France; ¹⁸Florence Nightingale Faculty of Nursing, Midwifery and Palliative Care, King's College London, London, United Kingdom; ¹⁹Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, British Columbia, Canada; ²⁰Department of Internal Medicine, Pulmonary and Critical Care Division, Beth Israel Deaconess Medical Center, Boston, Massachusetts; ²¹ICES, Toronto, Ontario, Canada; ²²Institute of Health Policy, Management and Evaluation, ²⁵Department of Anesthesia, and ²⁶Interdepartmental Division of Critical Care, University of Toronto, Toronto, Ontario, Canada; ²³Division of Pulmonary, Allergy, Sleep, and Critical Care Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts; and ²⁴NorthShore University Health System, Evanston, Illinois

Author Disclosures: J.A.K. served on a data safety and monitoring board for Sanofi; and received research support from Inogen, PCORI, National Institutes of Health, and ResMed. K.J.A. served on an advisory committee for Promedior; served as a consultant for AstraZeneca, Janssen, and Promedior; served on a data safety and monitoring board for Boehringer Ingelheim and Promedior; and received research support from Boehringer Ingelheim and Bristol Myers Squibb. D.H.A. served as a consultant for Gilead; served on a data safety and monitoring board for Novartis; and received research support from the American Lung Association. M.A.B. served on an advisory committee for AbbVie, Amgen, Atara Biotherapeutics, Brigham and Women's Hospital, Merck, and Vertex; and has equity interest in NovSci. M.K.H. served as a consultant for AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Merck, and Mylan; served on a data and safety monitoring board and as a speaker for Novartis; and received research support from Novartis and Sunovion. V.L. received research support from Fonds de Recherche du Québec-Santé and is employed as the Senior Methodologist and Editor for the department of Clinical Affairs and Practice Guideline. M.M. served as a consultant for AstraZeneca, Bial, Boehringer Ingelheim, Chiesi, CSL Behring, Ferrer, Gebro, GlaxoSmithKline, Grifols, Kamada, Laboratorios Esteve, Mereo, Novartis, pH Pharma, Sanofi, Spin Therapeutics, Teva, and Verona; served as a speaker for AstraZeneca, Bial, Boehringer Ingelheim, Chiesi, Cipla, CSL Behring, Grifols, Menarini, Novartis, Rovi, Sandoz, and Zambon; and received research support from GlaxoSmithKline and Grifols. R.A.M. received research support from GlaxoSmithKline. K.A.R. served on an advisory committee for Gilead; and received royalties from Springer Publishing. N.R. served on an advisory committee for AstraZeneca, Boehringer Ingelheim, Chiesi, Novartis,

Sanofi, and Teva; served as a speaker for AstraZeneca, Chiesi, Novartis, Teva, and Zambon; received research support from Boehringer Ingelheim, Novartis, Pfizer, and Zambon; and received personal fees from GlaxoSmithKline and Pfizer. M.S. served on an advisory committee for AstraZeneca and GlaxoSmithKline; served as a speaker for Boehringer Ingelheim and Teva; and received research support from AstraZeneca and Boehringer Ingelheim. A.J.W. received royalties from UpToDate. C.H.W. served on an

advisory committee and as a speaker for Mylan/Theravance; and received research support from the National Institutes of Health. K.C.W. is employed by the American Thoracic Society as Chief of Documents and Documents Editor with interest in the success of ATS guidelines. A.S.G., P.K.L., J.J.A., B.G.B., R.A.D., M.J.J., A.B.M., E.O., L.R., N.C.S., T.A.S., and H.W. reported no commercial or relevant noncommercial interests.

Acknowledgment: The authors thank Ms. Anne Hayes, Director, Research, Analysis and Evaluation Branch, Ontario Ministry of Health, and Ms. Sarah Burke Dimitrova, Lead, Quality Standards, Ontario Health, for their thoughtful comments and suggestions. They thank Dr. Carlos A. Cuello-Garcia, Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada, for his guidance.

References

- Roche N, Reddel HK, Agusti A, Bateman ED, Krishnan JA, Martin RJ, *et al.*; Respiratory Effectiveness Group. Integrating real-life studies in the global therapeutic research framework. *Lancet Respir Med* 2013;1:e29–e30.
- Chu SH, Kim EJ, Jeong SH, Park GL. Factors associated with willingness to participate in clinical trials: a nationwide survey study. *BMC Public Health* 2015;15:10.
- Comis RL, Miller JD, Aldigé CR, Krebs L, Stoval E. Public attitudes toward participation in cancer clinical trials. *J Clin Oncol* 2003;21:830–835.
- Flink M, Brandberg C, Ekstedt M. Why patients decline participation in an intervention to reduce re-hospitalization through patient activation: whom are we missing? *Trials* 2019;20:82.
- Hussain-Gambles M, Atkin K, Leese B. South Asian participation in clinical trials: the views of lay people and health professionals. *Health Policy* 2006;77:149–165.
- Kravitz RL, Paterniti DA, Hay MC, Subramanian S, Dean DE, Weisner T, *et al.* Marketing therapeutic precision: potential facilitators and barriers to adoption of n-of-1 trials. *Contemp Clin Trials* 2009;30:436–445.
- Shavers VL, Lynch CF, Burmeister LF. Knowledge of the Tuskegee study and its impact on the willingness to participate in medical research studies. *J Natl Med Assoc* 2000;92:563–572.
- Shaya FT, Gbarayor CM, Yang HK, Agyeman-Duah M, Saunders E. A perspective on African American participation in clinical trials. *Contemp Clin Trials* 2007;28:213–217.
- Wallington SF, Luta G, Noone A-M, Caicedo L, Lopez-Class M, Sheppard V, *et al.* Assessing the awareness of and willingness to participate in cancer clinical trials among immigrant Latinos. *J Community Health* 2012;37:335–343.
- Guyatt G, Akl EA, Oxman A, Wilson K, Puhan MA, Wilt T, *et al.*; ATS/ERS Ad Hoc Committee on Integrating and Coordinating Efforts in COPD Guideline Development. Synthesis, grading, and presentation of evidence in guidelines: article 7 in integrating and coordinating efforts in COPD guideline development: an official ATS/ERS workshop report. *Proc Am Thorac Soc* 2012;9:256–261.
- Fedak KM, Bernal A, Capshaw ZA, Gross S. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. *Emerg Themes Epidemiol* 2015;12:14.
- Gershon AS, Jafarzadeh SR, Wilson KC, Walkey AJ. Clinical knowledge from observational studies: everything you wanted to know but were afraid to ask. *Am J Respir Crit Care Med* 2018;198:859–867.
- Emdin CA, Khera AV, Kathiresan S. Mendelian randomization. *JAMA* 2017;318:1925–1926.
- Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;15:615–625.
- Suissa S. Inhaled steroids and mortality in COPD: bias from unaccounted immortal time. *Eur Respir J* 2004;23:391–395.
- Travers J, Marsh S, Williams M, Weatherall M, Caldwell B, Shirlcliffe P, *et al.* External validity of randomised controlled trials in asthma: to whom do the results of the trials apply? *Thorax* 2007;62:219–223.
- Halpin DM, Kerkhof M, Soriano JB, Mikkelsen H, Price DB. Eligibility of real-life patients with COPD for inclusion in trials of inhaled long-acting bronchodilator therapy. *Respir Res* 2016;17:120.
- Pahus L, Burgel PR, Roche N, Paillasseur JL, Chanez P; Initiatives BPCO scientific committee. Randomized controlled trials of pharmacological treatments to prevent COPD exacerbations: applicability to real-life patients. *BMC Pulm Med* 2019;19:127.
- Riley IL, Murphy B, Razouki Z, Krishnan JA, Apter A, Okelo S, *et al.* A systematic review of patient- and family-level inhaled corticosteroid adherence interventions in black/African Americans. *J Allergy Clin Immunol Pract* 2019;7:1184–1193, e3.
- Paskett ED, Reeves KW, McLaughlin JM, Katz ML, McAlearney AS, Ruffin MT, *et al.* Recruitment of minority and underserved populations in the United States: the Centers for Population Health and Health Disparities experience. *Contemp Clin Trials* 2008;29:847–861.
- Rivers D, August EM, Sehovic I, Lee Green B, Quinn GP. A systematic review of the factors influencing African Americans' participation in cancer clinical trials. *Contemp Clin Trials* 2013;35:13–32.
- Hamel LM, Penner LA, Albrecht TL, Heath E, Gwede CK, Eggly S. Barriers to clinical trial enrollment in racial and ethnic minority patients with cancer. *Cancer Contr* 2016;23:327–337.
- Tu JV, Willison DJ, Silver FL, Fang J, Richards JA, Laupacis A, *et al.*; Investigators in the Registry of the Canadian Stroke Network. Impracticability of informed consent in the Registry of the Canadian Stroke Network. *N Engl J Med* 2004;350:1414–1421.
- Prieto-Centurion V, Rolle AJ, Au DH, Carson SS, Henderson AG, Lee TA, *et al.*; CONCERT Consortium. Multicenter study comparing case definitions used to identify patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2014;190:989–995.
- Durant RW, Wenzel JA, Scarinci IC, Paterniti DA, Fouad MN, Hurd TC, *et al.* Perspectives on barriers and facilitators to minority recruitment for clinical trials among cancer center leaders, investigators, research staff, and referring clinicians: enhancing minority participation in clinical trials (EMPaCT). *Cancer* 2014;120:1097–1105.
- Lederer DJ, Bell SC, Branson RD, Chalmers JD, Marshall R, Maslove DM, *et al.* Control of confounding and reporting of results in causal inference studies: guidance for authors from editors of respiratory, sleep, and critical care journals. *Ann Am Thorac Soc* 2019;16:22–28.
- Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003;158:915–920.
- García-Albéniz X, Hsu J, Hernán MA. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. *Eur J Epidemiol* 2017;32:495–500.
- Labrecque JA, Swanson SA. Target trial emulation: teaching epidemiology and beyond. *Eur J Epidemiol* 2017;32:473–475.
- Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000;342:1878–1886.
- Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev* 2014;(4):MR000034.
- Lodi S, Phillips A, Lundgren J, Logan R, Sharma S, Cole SR, *et al.*; INSIGHT START Study Group and the HIV-CAUSAL Collaboration. Effect estimates in randomized trials and observational studies: comparing apples with apples. *Am J Epidemiol* 2019;188:1569–1577.
- Wang MT, Bolland MJ, Grey A. Reporting of limitations of observational research. *JAMA Intern Med* 2015;175:1571–1572.
- Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. *BMJ* 2015;350:h2147.
- Kahn JM, Gould MK, Krishnan JA, Wilson KC, Au DH, Cooke CR, *et al.*; ATS Ad Hoc Committee on the Development of Performance Measures from ATS Guidelines. An official American Thoracic Society

- workshop report: developing performance measures from clinical practice guidelines. *Ann Am Thorac Soc* 2014;11: S186–S195.
36. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, *et al.*; GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–926.
37. Schünemann HJ, Jaeschke R, Cook DJ, Bria WF, El-Solh AA, Ernst A, *et al.*; ATS Documents Development and Implementation Committee. An official ATS statement: grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. *Am J Respir Crit Care Med* 2006;174: 605–614.
38. 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. *Res Synth Methods* 2013;4:49–62.
39. Cuello-Garcia CA. The role of randomized and non-randomized studies in knowledge synthesis of health interventions [doctoral dissertation]. Hamilton, ON, Canada: McMaster University; 2017.
40. U.S. Food & Drug Administration. Silver Spring, MD: U.S. Food and Drug Administration [accessed 2019 Nov 15]. Available from: www.fda.gov.
41. Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, *et al.*; RECORD Working Committee. The reporting of studies conducted using observational routinely-collected health data (RECORD) statement. *PLoS Med* 2015;12: e1001885.
42. Mansournia MA, Higgins JPT, Sterne JAC, Hernán MA. Biases in randomized trials: a conversation between trialists and epidemiologists. *Epidemiology* 2017;28:54–59.