

Routinely collected data and comparative effectiveness evidence: promises and limitations

Lars G. Hemkens MD MPH, Despina G. Contopoulos-Ioannidis MD, John P.A. Ioannidis MD DSc

See also page www.cmaj.ca/lookup/doi/10.1503/cmaj.160410, www.cmaj.ca/lookup/doi/10.1503/cmaj.151470 and *CMAJ Open* article www.cmajopen.ca/content/4/2/E132

Competing interests: None declared.

This article has been peer reviewed.

Correspondence to:
John Ioannidis,
jioannid@stanford.edu

CMAJ 2016. DOI:10.1503/cmaj.150653

Routinely collected data (RCD) are increasingly used for biomedical research. Extensive resources have been invested in this field: they include the set-up of disease registries and clinical databases at regional, national or international levels; the promotion of the use of electronic health records; and making use of wearable devices for the collection of health data. Analysis of this data can inform on descriptive features (prevalence or incidence of disease, treatments and risk factors), associations with putative risk factors and/or treatment effects of interventions (e.g., drugs, surgery, psychotherapy or medical devices).

Although descriptive estimates and associations offer interesting information, treatment effects are most important for clinical decision-making. They are the core of comparative effectiveness research. In this article, we focus primarily on RCD for determining treatment effects, because they are increasingly considered mainstream options for building evidence on treatment choices. The promises and hype of personalized medicine (or precision medicine, predictive medicine, participatory medicine, 4P or stratified medicine) are also similarly fueled by the widespread use of RCD. We do not use these terms here, because these promises have the same major challenges that are faced by traditional comparative effectiveness research —

even to a higher degree — because they try to identify best options for single patients or small subgroups rather than larger populations. In this overview, we contrast the expectations many have of the use of RCD versus their limitations, discuss which expectations can be met and suggest potential changes in the research agenda for RCD.

Main strengths and weaknesses of routinely collected data

Big data studies with enormous sample sizes or real-world analyses of near-perfect representations of routine care fuel tremendous expectations for RCD in clinical decision-making. Although the traditional limitations of observational research remain, such extremes amplify strengths and weaknesses. The latter may increase exponentially by challenges specifically related to the very nature of data not collected for the purpose of research (e.g., additional biases or errors occurring when gigantic datasets have to be assembled, cleaned, processed, linked and retrospectively analyzed).

In theory, RCD have several advantages. Data collection under real-world circumstances maximizes representativeness and generalizability, minimizes costs and effort, and allows the capture of information in large populations and many clinical events in large datasets that are continuously updated and cover long periods.

However, these theoretical advantages should be viewed cautiously. First, many RCD are collected in situations where populations, diseases, settings and/or interventions are not representative (e.g., when data are collected in tertiary referral hospitals or in health care systems where the population or use of specific interventions are selected by ability to pay or other filters). Evaluation of newly approved drugs may be difficult because there are few existing routine data, and barriers to access innovative drugs may create strong confounding by indication. Second, costs are not necessarily low in all cases (e.g., many hospitals and

KEY POINTS

- Routinely collected data (RCD) are increasingly used for biomedical research; however, their utility for understanding treatment effects is probably overestimated.
- Many of the perceived advantages of RCD should be viewed cautiously, because of the inevitable biases of observational research and specific biases due to the nature of these data.
- Improvements may increase the utility of RCD but require resources for implementation; they include improvements in research priority setting, transparency of data and protocols, and collaborative research networks.
- Although many evidence gaps may be better addressed by an improved randomized controlled trial (RCT) agenda, RCD studies may be required in situations where RCTs are difficult or impossible to perform; interpretation of these studies should be cautious.

health care systems make large investments in infrastructure and maintenance because of the increasing popularity of electronic health records). Fragmentation of efforts escalates cost compared with centralized systems that include all health care facilities in a country (e.g., the health care system in Taiwan¹). Third, large sample sizes without thorough analytical safeguards can result in statistically significant false-positive and false-negative results.

The observational nature of RCD is an inherent limitation for the study of treatment effects. Which treatment is chosen depends on various known (e.g., severity of disease) or unknown factors that may be associated with the outcome. Such confounding by indication can invalidate real-world observations. Multiple statistical methods are used to reduce these biases (e.g., propensity scores and instrumental variables analyses),^{2,3} but only properly designed randomized controlled trials (RCTs) can pre-emptively overcome such biases.

Multiple errors and biases may interfere with routine data collection and processing (e.g., data linkage problems, misclassification bias and underreporting).^{3,4} This further reduces the validity of RCD. Additional steps, such as manual reviews of patient records, are sometimes incorporated to improve the quality of the RCD data. However, this adds to the cost and does not solve misclassification problems that occur when risk exposures and/or outcomes are ascertained in a nonstandardized way and when differences in coding practices also exist. Differences in management practice within and across institutions can reflect differences in several other confounding factors (e.g., disease severity).

Studies of RCD or better RCTs?

To understand how to best use RCD for health care decision-making, we should revisit the limitations of RCTs (the gold standard for studying treatment effects) and whether overcoming these limitations needs a better RCT agenda or use of RCD.

Generalizability and real-world relevance of clinical studies, in particular those that are used for drug approval, are often limited by narrow inclusion and exclusion criteria,⁵ and trial participants may have different characteristics than non-participants. Trials are frequently conducted under artificial conditions that differ from routine care (e.g., use of run-in periods, structured follow-up visits or standardized cotreatments). Certain populations are frequently underrepresented in RCTs, including children, women, older adults or patients with comorbidities and polypharmacy.⁶⁻¹⁰ Drug-drug interactions or adverse effects occurring in routine care may be overlooked. Cost con-

siderations prohibit large studies that would be informative for subgroup-specific effects.

Some of these deficiencies may be best solved by improving the RCT agenda rather than turning to RCD. For example, the cost of RCTs can be reduced substantially, allowing very large sample sizes and better representativeness of the enrolled populations, if simple, pragmatic megatrials are adopted and RCD are used for collecting outcome information.^{11,12} Nevertheless, such megatrials are uncommon, and thus observational RCD studies are used to fill the evidence gap. For uncommon conditions, even megatrials would have few patients to inform on outcomes in these subgroups. Studies using RCD can reach sample sizes that are 100- to 1000-fold bigger than the sample sizes of large trials. However, the planning and reporting for claims of subgroup differences in clinical research have been dismal, and most claims are not validated.¹³ For example, it remains unknown whether the treatment effect suggested by RCD studies involving patients over 80 years of age with modest renal impairment, hypertension and taking three other drugs would be more reliable than the average treatment effect suggested by an RCT that involved patients with none or few of these characteristics.

Given the limited funds for RCTs, many important health care questions are not studied. Such evidence gaps could be addressed by a better RCT research agenda that prioritizes the use of pragmatic, patient-important outcomes¹⁴ and relevant head-to-head comparisons.^{5,15,16} Some comparative effectiveness evidence may also be accommodated by network meta-analyses of RCTs.^{5,15,17} However, even then, an exhaustive evaluation of treatment effects on mortality and other patient-important outcomes (including major harms) with RCTs alone is unrealistic. Here, RCD could fill many evidence gaps. One may then decide that the RCD evidence is strong enough to lead to policy or guideline changes, or the RCD evidence may be used to guide the design of future RCTs. There are also situations where conducting RCTs would be unrealistic or perceived as unethical.¹⁸

Randomized controlled trials currently differ from RCD studies in many features besides randomization. Many of the features that improve the validity of RCTs, either directly or indirectly, may also contribute to the perceived practical disadvantages of this type of research. For example, the regulatory requirements that need to be fulfilled before a trial may start are often cumbersome.¹⁹ These requirements are a direct result of the experimental nature and ethical implications of randomization.²⁰ They include thorough reflec-

tions about the intended purpose of the research to justify randomization, study protocols clearly stating assumptions, hypotheses and calculations of sample size, and submission of protocols to regulatory authorities. Working in larger collaborative groups of researchers with various backgrounds and exchanges with involved stakeholders, ethics committees or data–safety monitoring boards generates feedback loops that may improve initial RCT research plans.

Most of these steps are often not undertaken for RCD research. Some of the perceived practical advantages of RCD studies may actually be limitations. Available datasets may be rapidly analyzed by small teams or a single researcher. Studies of RCD are largely overpowered to obtain nominally significant effects, however small they may be.²¹ Post hoc explanations are easily invoked, increasing confidence in spurious findings.^{22,23} Results can remain unpublished, or results may be published depending on the plausibility of explanations, preconceived hypotheses, commercial interests or the researcher's personal need for scientific reward.

In Table 1, we summarize some of the limitations of current RCTs, beginning with those that may be the most amenable to improvement of the current RCT agenda. We list ways to bypass these limitations with RCD and highlight residual caveats of RCD studies.

The status quo of routinely collected data

We recently conducted an empirical analysis on how RCD studies try to complement RCTs to understand treatment effects.²⁴ We assessed 337 RCD studies that investigated the comparative effectiveness of medical treatments on mortality. Seventy percent of these studies were incremental research that supplemented existing RCTs but did not fill fundamental knowledge gaps (i.e., questions never evaluated in RCTs). In only six (1.8%) of these RCD studies did the authors state that conducting RCTs on their research topic would be unethical, and in only 18 (5.3%) did they state that it would be difficult. Typically, investigators conducting the RCDs reasoned that RCT results had limited generalizability (37.6%), did not adequately address specific outcomes (31.9%) or certain populations (23.5%), or were inconclusive or inconsistent (25.8%).

Most RCD studies focus on questions that have been addressed by RCTs or could be definitively addressed by RCTs.²⁴ Agreement between the results of such RCD studies and the results of the RCTs offers some incremental reassurance,

but the benefit for clinical decision-making is limited or nonexistent. When RCTs and observational studies disagree,²⁵ the situation becomes complicated. Much of the interpretation of inconsistent results between such sources of evidence is currently a case-by-case discussion. Eventually, residual bias owing to nonrandomization or the artificial RCT setting may be used as arguments for almost any disagreement. Consensus becomes difficult to reach.

In areas without evidence from RCTs, studies of RCTs may provide the only guidance on a critical health care question, albeit with recognizable limitations. Policy or guideline changes based on RCD should acknowledge the limitations of RCD, and strategic plans should be in place to monitor the clinical impact of these changes. Unfortunately, current RCD studies do not focus on the large numbers of critical health care questions that do not have evidence from RCTs.²⁴ For example, comparisons of drug and nondrug treatments, and evaluations of inexpensive drugs are lacking. Evidence from RCD studies would be useful in providing answers to these vital questions.

Changes in the RCD research agenda and practices

Overall, expectations about the utility of RCD studies for understanding treatment effects are probably overestimated. We discuss what improvements can be made in RCD studies and what resources would be required (Table 2).

Selecting priorities

In selecting research questions, prior evidence must be systematically reviewed. Another study or analysis may not be necessary. Routinely collected data studies should focus more on questions that have not been addressed or are difficult or impossible to address with other study designs.

Protocols and prespecification

Research using RCD may or may not use explicit protocols and prespecified analyses. It is important to know what was not prespecified. Exploratory analyses should be described as such; they need further prospective validation with protocol-based, prespecified studies. Wherever prespecification is not feasible, transparent and complete documentation of the conduct of the study is still useful. The validity of RCD and their proper interpretation can be improved by using falsification end points (negative controls of known null associations),²⁷ validation datasets²⁸ and prespeci-

Table 1 (part 1 of 2): Limitations of RCTs and whether they can be amended by RCD studies

Potential limitations of RCTs		What RCD can offer	Challenges and remaining caveats in RCD studies
Substantial improvement by an amended RCT agenda			
Understudied health care questions	No direct comparison of relevant treatments or use of pragmatic patient-important outcomes (i.e., mortality); possible and feasible to conduct but not prioritized	Selection of almost any research topic and treatment comparison and of many patient-important clinical events and mortality	Some outcomes typically require deviation from routine care (e.g., evaluation of patient-reported outcomes, such as pain or quality of life, by surveying patients) and are often unavailable.
Data access/publication bias	Data often generated and collected by industrial sponsors without sharing raw data; reproduction often impossible without infrastructure	Unknown	Access issues and publication bias likely do not improve with studies using RCD when compared with RCTs.
Considerable improvement by an amended RCT agenda			
Generalizability and real-world relevance	Study populations differ from real-world target population because inclusion and exclusion criteria are too strict; treatment circumstances differ from routine care because of trial setting.	Liberal inclusion criteria; real-world data with high external validity; no interference with routine care	Some outcomes typically require deviation from routine care. External validity not necessarily high when collection of data depends on other factors (e.g., collected in tertiary centres, for patients with certain insurance plans).
Specific conditions/subgroup effects	Patients from specific demographic populations or patients with complex conditions are often underrepresented.	Large populations; liberal inclusion criteria	Importance of subgroup claims and consequences are frequently unclear and might be overrated; high risk of false-positive findings
Conflicts of interest/sponsorship bias	Evidence generated, analyzed and published by researchers or trial sponsors who have an economic conflict of interest; almost always for novel drugs	Unknown, often fewer conflicts and nonconflicted sponsors	Financial and scientific conflicts due to strong beliefs or preconceived hypotheses may be prominent even for analyses using RCD.
Modest improvement by an amended RCT agenda			
Costs	Logistic costs, and efforts for data generation and collection	Much lower costs for data generation and collection	High investments in data infrastructures and maintenance, although some are not directly research-related (e.g., for electronic health records). Nonstandardized efforts that are often fragmented across teams waste resources, increase costs and create false leads that further waste resources.
Speed	Time needed for planning, protocol development, regulatory issues, and patient recruitment; time of follow-up until outcomes are observed	No need to wait for outcomes in analyses of existing data; time for prespecification not required for exploratory analyses; analyses can be run by small teams or one investigator	Lack of prespecification and protocols may reduce validity because of increased risk of findings that are false-positive or false-negative and bias (e.g., selective reporting bias and modelling biases). Thorough reflections about research and involvement of larger teams may improve initial research plans and provide a wider perspective and increase research usefulness and value.
Regulations	Randomization requires ethical and regulatory approval and the process can be cumbersome.	No or fewer requirements for ethical and/or regulatory approval	Less oversight; more opportunities for unnoticed errors and biases
Late outcomes	Length of follow-up too short for detecting long-term effects	Long observation periods	Missing data; no consistent outcome ascertainment across patients; crossover; poor adherence common with long-term follow-up

Table 1 (part 2 of 2): Limitations of RCTs and whether they can be amended by RCD studies

Potential limitations of RCTs		What RCD can offer	Challenges and remaining caveats in RCD studies
Modest improvement by an amended RCT agenda			
Uncommon conditions	Trial populations small; recruitment difficult	Recruitment usually not difficult; very diverse treatment settings because of liberal inclusion criteria and large populations	Confounding by indication; referral biases
Minimal improvement by an amended RCT agenda			
Uncommon outcomes	Insufficient statistical power for detecting effects on uncommon outcomes because trial populations are too small or follow-up is too short.	Many events because of large populations and long observation periods	Spurious findings and significant findings that are false-positive because of overpowered studies and lack of analytical safeguards. High risk of confounding could also lead to spurious nullification of true treatment effects (false-negative results).
Superseded/outdated/unusual treatment setting or unfeasible conduct	Outdated or unusual circumstances under which existing RCTs were conducted (e.g., no modern background treatments); new trials can be done, but they would be expensive and take a long time. Perceived disadvantage of one treatment making recruitment of patients difficult.	Gathering relevant data on very diverse treatment settings is feasible because of liberal inclusion criteria and large populations.	High risk for confounding by indication (i.e., strong indications required for treatments that are superseded or perceived as inferior); generalizability limited to settings with similar circumstances
No improvement by an amended RCT agenda			
Unethical conduct	Proven disadvantage or anticipated harm with one treatment (lack of equipoise), making randomization unethical ¹⁸	Size of the disadvantage or harm can be documented	If the treatment is clearly inferior, maybe it may not have been used even in RCD settings, and it would be of no or little clinical relevance.
Note: RCD = routinely collected data, RCT = randomized controlled trial.			

fied rules when the study hypotheses should be considered confirmed or rejected.

Registration

Registration of RCD studies that have prospective design and/or analysis elements and explicit protocols would help shape a more efficient research agenda and reduce selective reporting of methods and findings. For explorative research, it may be best to register datasets; this would facilitate planning a concerted research agenda, data-sharing activities and using datasets for validation.^{29,30}

Reporting

Incomplete or unusable reporting wastes research resources.³¹ Studies using RCD have a low rate of reporting.³² Recently, the RECORD (REporting of studies Conducted using Observational Routinely-collected health Data) statement was published,³³ which aims to improve the reporting quality specifically of observational RCD studies by providing an extension to the STROBE

(STrengthening the Reporting of OBServational studies in Epidemiology) statement.³⁴ In addition to transparent reporting, the results need to be embedded into a systematic review of the available evidence. Journals, peer reviewers, funders and authorities can help to improve the reporting quality of RCD studies.

Access to raw data

Lack of access to raw data makes it impossible to independently assess analytic errors and biases, and limits opportunities for joint analyses. Facilitated availability of different datasets would support external validation and improve standardization and efforts to enhance quality. Patients should be asked for explicit consent up front for prospective data sharing of RCD, as is required for RCTs. The misleading view that health information is not really protected data when it is routinely collected creates serious problems.³⁵ Consent issues would be best decided during database building. Data deidentification should also be carefully planned.

Research networks

Large research networks can foster the joint use of RCT and RCD datasets. Research networks may be in the best position to face the challenges involved in establishing harmonized/standardized research. This includes outcome definitions (e.g., by developing and validating universally accepted lists of diagnostic codes for

specific outcomes), time points of outcome assessments, risk exposures to be analyzed, subgroup analyses to be explored, and predetermined effect sizes and other criteria for clinically significant outcome differences. Standardized guidance can be developed for organizing and implementing data sharing. Collaborators with various levels of expertise and backgrounds

Table 2: Options to improve the value of routinely collected health data

Process	Options	Resources needed
Selecting priorities	<ul style="list-style-type: none"> • Systematic review of all available evidence and description of potential research consequences • Consideration of novelty, incremental value and usefulness of research in context of systematic literature review • Focus on health care questions that have not been addressed and are difficult or impossible to address with other designs 	Some funding resources are needed to conduct systematic reviews, but the pay back should be much greater because efficiency is improved, important questions are addressed and uninformative and redundant research is avoided.
Protocols and prespecification	<ul style="list-style-type: none"> • Clear statement on which analyses are exploratory (post hoc) and which are the main study analyses planned a priori For planned nonexploratory research: <ul style="list-style-type: none"> • Prespecified hypotheses, research questions, definitions, detailed statistical analysis plans, model assumptions • Predetermination of effect sizes that are of clinical significance • Falsification end points • Validation (split or multiple datasets) • Decision rules for the consequences of RCD–research findings 	Best practices may be promoted by funders (as requirements for funding) and endorsed by journals and research communities
Registration	<ul style="list-style-type: none"> • Registration of datasets to improve research agenda and to support data sharing and validation activities For planned nonexploratory research: <ul style="list-style-type: none"> • Registration of protocols and planned analyses to reduce selective reporting bias 	Some resources are needed to establish and maintain suitable registries; existing registries for clinical trials also include observational studies but may need to be modified for maximal relevance; registration should be informative and nonbureaucratic
Reporting	<ul style="list-style-type: none"> • Transparent and complete reporting • Results reported and interpreted in context of all evidence derived from systematic literature review 	Journals, peer reviewers, funders and authorities (e.g., ethics committees) may consider requiring reporting guidelines for preparation of reports and manuscripts
Raw data availability	<ul style="list-style-type: none"> • Consideration of ethical and privacy issues • Pre-emptive planning on consent issues • Address deidentification issues • Promotion of sharing data and providing access • Promotion of research networks and joint analyses to allow evaluation of internal and external validity 	Preparation, cleaning, deposition, curation and meaningful sharing of datasets needs committed resources and standardization efforts
Research networks	<ul style="list-style-type: none"> • Establishment of large research networks involving various stakeholders to consider various perspectives • Harmonization/standardization of research conduct and data-sharing efforts (e.g., protocols for exchanging RCD, codes or datasets) 	Resources are needed to build and maintain networks, such as OHSDI, but may lead to a multiplier effect on efficiency and major quality improvements
Research on research	<ul style="list-style-type: none"> • Research on methods to synthesize evidence from various data sources • Research on reliability of RCD studies • Validation of methods used in RCD studies to minimize confounding by indication biases • Validation of findings in validation datasets across datasets and/or compared with designs of other studies • Better understanding of and tools to measure and improve risk of bias, data validity and generalizability 	Resources needed to perform metaresearch projects

Note: OHSDI = Observational Health Data Sciences and Informatics program,²⁶ RCD = routinely collected data.

would provide diverse perspectives to maximize research applicability.

Research on research

More research on the reliability of RCD results is necessary (e.g., on the performance of approaches to deal with confounding by indication, such as propensity scores, instrumental variables or the use of falsification end points). Compared with RCTs, there is little empirical guidance on the interpretation of RCD evidence. We need to develop a better understanding of and tools for assessment of risk of bias, generalizability and data validity.

Conclusion

Research using RCD is becoming increasingly popular, but its limitations cannot be overstated. Several suggested improvements may increase the utility of this research but would require additional resources. Studies using RCD should be prioritized for situations where RCTs cannot be conducted. Nevertheless, interpretation of RCD must be done with caution.

References

- Hsing AW, Ioannidis JP. Nationwide population science: lessons from the Taiwan National Health Insurance Research Database. *JAMA Intern Med*. 2015;175:1527-9.
- Hernán MA, Robins JM. Instruments for causal inference: An epidemiologist's dream? *Epidemiology* 2006;17:360-72.
- Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005;58:323-37.
- Bohensky MA, Jolley D, Sundararajan V, et al. Data linkage: a powerful research tool with potential problems. *BMC Health Serv Res* 2010;10:346.
- Naci H, Ioannidis JP. Comparative effectiveness of exercise and drug interventions on mortality outcomes: metaepidemiological study. *BMJ* 2013;347:f5577.
- Geller SE, Adams MG, Carnes M. Adherence to federal guidelines for reporting of sex and race/ethnicity in clinical trials. *J Womens Health (Larchmt)* 2006;15:1123-31.
- Dodd KS, Saczynski JS, Zhao Y, et al. Exclusion of older adults and women from recent trials of acute coronary syndromes. *J Am Geriatr Soc* 2011;59:506-11.
- Heiat A, Gross CP, Krumholz HM. Representation of the elderly, women, and minorities in heart failure clinical trials. *Arch Intern Med* 2002;162:1682-8.
- Konrat C, Boutron I, Trinquart L, et al. Underrepresentation of elderly people in randomised controlled trials. The example of trials of 4 widely prescribed drugs. *PLoS ONE* 2012;7:e33559.
- Van Spall HG, Toren A, Kiss A, et al. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *JAMA* 2007;297:1233-40.
- Eisenstein EL, Lemons PW II, Tardiff BE, et al. Reducing the costs of phase III cardiovascular clinical trials. *Am Heart J* 2005;149:482-8.
- Vickers AJ. Clinical trials in crisis: four simple methodologic fixes. *Clin Trials* 2014;11:615-21.
- Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet* 2005;365:176-86.
- Chalmers I, Bracken MB, Djulbegovic B, et al. How to increase value and reduce waste when research priorities are set. *Lancet* 2014;383:156-65.
- Lathyrus DN, Patsopoulos NA, Salanti G, et al. Industry sponsorship and selection of comparators in randomized clinical trials. *Eur J Clin Invest* 2010;40:172-82.
- Rizos EC, Salanti G, Kontoyiannis DP, et al. Homophily and co-occurrence patterns shape randomized trials agendas: illustration in antifungal agents. *J Clin Epidemiol* 2011;64:830-42.
- Mills EJ, Thorlund K, Ioannidis JP. Demystifying trial networks and network meta-analysis. *BMJ* 2013;346:f2914.
- Djulbegovic B, Hozo I. At what degree of belief in a research hypothesis is a trial in humans justified? *J Eval Clin Pract* 2002;8:269-76.
- Chan AW, Song F, Vickers A, et al. Increasing value and reducing waste: addressing inaccessible research. *Lancet* 2014;383:257-66.
- Rosner F. The ethics of randomized clinical trials. *Am J Med* 1987;82:283-90.
- Siontis GC, Ioannidis JP. Risk factors and interventions with statistically significant tiny effects. *Int J Epidemiol* 2011;40:1292-307.
- Mynatt CR, Doherty ME, Tweney RD. Confirmation bias in a simulated research environment: an experimental study of scientific inference. *Q J Exp Psychol* 1977;29:85-95.
- Nickerson RS. Confirmation bias: a ubiquitous phenomenon in many guises. *Rev Gen Psychol* 1998;2:175.
- Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis J. Do routinely collected health data complement randomized evidence? A survey. *CMAJ Open*. In press.
- Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001;286:821-30.
- Richardson WS, Wilson MC, Nishikawa J, et al. The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 1995;123:A12-3.
- Prasad V, Jena AB. Prespecified falsification end points: Can they validate true observational associations? *JAMA* 2013;309:241-2.
- Young SS, Karr A. Deming, data and observational studies. A process out of control and needing fixing. *Significance* 2011;8:116-20.
- Lash TL, Vandenbroucke JP. Should preregistration of epidemiologic study protocols become compulsory? Reflections and a counterproposal. *Epidemiology* 2012;23:184-8.
- Ioannidis JP. The importance of potential studies that have not existed and registration of observational data sets. *JAMA* 2012;308:575-6.
- Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet* 2014;383:267-76.
- Hemkens LG, Benchimol EI, Langan SM, et al. Reporting of studies using routinely collected health data: systematic literature analysis [oral abstract presentation]. 2015 REWARD / EQUATOR Conference: Increasing value and reducing waste in biomedical research; 2015 Sept. 28-30; Edinburgh (UK).
- Benchimol EI, Smeeth L, Guttman A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Med* 2015;12:e1001885.
- von Elm E, Altman DG, Egger M, et al. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med* 2007;4:e296.
- Ioannidis JP. Informed consent, big data, and the oxymoron of research that is not research. *Am J Bioeth* 2013;13:40-2.

Affiliations: Stanford Prevention Research Center (Hemkens, Ioannidis), Department of Medicine, Stanford University School of Medicine, Stanford, Calif.; Basel Institute for Clinical Epidemiology and Biostatistics (Hemkens), University Hospital Basel, Basel, Switzerland; Division of Infectious Diseases, Department of Pediatrics (Contopoulos-Ioannidis), Stanford University School of Medicine; Meta-Research Innovation Center at Stanford (METRICS) (Contopoulos-Ioannidis, Ioannidis); Department of Health Research and Policy (Ioannidis), Stanford University School of Medicine; Department of Statistics (Ioannidis), Stanford University School of Humanities and Sciences, Stanford, Calif.

Contributors: Lars Hemkens wrote the first draft of the article. All of the authors contributed to the writing and editing of the manuscript, revised it critically for intellectual content, approved the final version to be published and agreed to act as guarantors of the work.