

Current use of routinely collected health data to complement randomized controlled trials: a meta-epidemiological survey

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

See also www.cmaj.ca/lookup/doi/10.1503/cmaj.150653 and www.cmaj.ca/lookup/doi/10.1503/cmaj.151470

Abstract

Background: Studies that use routinely collected health data (RCD studies) are advocated to complement evidence from randomized controlled trials (RCTs) for comparative effectiveness research and to inform health care decisions when RCTs would be unfeasible. We aimed to evaluate the current use of routinely collected health data to complement RCT evidence.

Methods: We searched PubMed for RCD studies published to 2010 that evaluated the comparative effectiveness of medical treatments on mortality using propensity scores. We identified RCTs of the same treatment comparisons and evaluated how frequently the RCD studies analyzed treatments that had not been compared previously in randomized trials. When RCTs did exist, we noted the claimed motivations for each RCD study. We also analyzed the citation impact of the RCD studies.

Results: Of 337 eligible RCD studies identified, 231 (68.5%) analyzed treatments that had already been compared in RCTs. The study investigators rarely claimed that it would be unethical (6/337) or difficult (18/337) to perform RCTs on the same question. Evidence from RCTs was mentioned or cited by authors of 213 RCD studies. The most common motivations for conducting the RCD studies were alleged limited generalizability of trial results to the “real world” (37.6%), evaluation of specific outcomes (31.9%) or specific populations (23.5%), and inconclusive or inconsistent evidence from randomized trials (25.8%). Studies evaluating “real world” effects had the lowest citation impact.

Interpretation: Most of the RCD studies we identified explored comparative treatment effects that had already been investigated in RCTs. The objective of such studies needs to shift more toward answering pivotal questions that are not supported by trial evidence or for which RCTs would be unfeasible.

Routinely collected health data (RCD) include health administrative data and data from electronic health records. They are not collected for research purposes but are often claimed to be a prime source of evidence for comparative effectiveness research.¹⁻⁶ Research using such data are currently heavily promoted with immense funding resources. Major investments have been made to build disease and patient registries, to improve clinical databases and to stimulate use of electronic health records. One example is the recent approval of \$93.5 million by the Patient-Centered Outcomes Research Institute (PCORI) to support the National Patient-Centered Clinical Research Network.⁷ Conversely, major funders have shied away from supporting randomized trials.⁸

There are different perceived uses of RCD, depending on whether randomized controlled trials (RCTs) already exist (or can be readily performed) on the same question. One may argue that RCTs are unfeasible or unrealistic to perform for each and every comparison of available medical treatments.^{4,9-11} Moreover, regulatory agencies often require only randomized

comparisons against placebo or no treatment.¹² RCD studies are advocated as being able to close this large evidence gap in a more timely fashion and with limited costs.^{6,10,11} In other cases, the contribution of routinely collected data is more incremental, and their use is intended to address questions where some evidence from randomized trials already exists. Such studies may be presented as a complement to previous RCTs, evaluating whether the trial evidence holds true in the “real world,” in different settings, with different outcomes or in populations considered to have been understudied in RCTs (e.g., women and children).^{10,13,14}

Competing interests: None declared.

This article has been peer reviewed.

Correspondence to: John Ioannidis, jioannid@stanford.edu

CMAJ Open 2016. DOI:10.9778/cmajo.20150036

Although all observational studies are limited by the lack of randomization, modern epidemiological methods such as propensity scores and marginal structural models are increasingly being used to address such biases. This may improve the reliability of RCD studies and thus their value for decision-making when evidence from clinical trials is inadequate or lacking. However, are RCD studies being performed mostly in situations where RCT evidence does not exist or clinical trials are unethical or difficult to conduct? Or do RCD studies “search under the lamppost” where trials have already taken place? Which claimed limitations of existing clinical trials motivate researchers to use routinely collected data, and what are the knowledge gaps intended to be closed? What is the scientific impact of such research, and does it differ depending on whether RCTs also exist and on what motivations are reported for the conduct of RCD studies?

We evaluated a sample of RCD studies that would be of high relevance for patients and health care decision-makers and that address patient-relevant outcomes, using a standard meta-epidemiological survey. We aimed to find out to what extent these studies examined treatment comparisons that had already been tested in RCTs. We also sought to establish reasons why investigators use RCD to answer a specific question and the impact of such studies.

Methods

We conducted a literature search to identify RCD studies that (a) evaluated the comparative effectiveness of a treatment intervention against another intervention or against no intervention, usual care or standard treatment; (b) included mortality as an assessed outcome; and (c) used propensity scores to analyze mortality. We chose to sample in this way because the large number of RCD studies published to date precludes systematic analysis of all of them, and because we wanted to focus on the most important of all outcomes (death) and to standardize the method used for analysis in the RCD studies (propensity scores). Ethical approval was not required for the study.

Literature search

We searched PubMed (last search November 2011) for eligible RCD studies published from inception to 2010. We combined various search terms for RCD (e.g., “routine*,” “registri*” and “claim*”) with terms for mortality (e.g., “mortal*”) and propensity scores (“propensity”). We considered RCD studies involving any patient population with any condition. Eligible treatment interventions were drugs, biologics, dietary supplements, devices, diagnostic procedures, surgery and radiotherapy. One reviewer (L.G.H.) screened titles and abstracts of identified studies and obtained full-text versions of potentially relevant articles to assess eligibility. Detailed inclusion criteria and the search strategy appear in Appendices 1 and 2 (available at www.cmajopen.ca/content/4/2/E132/suppl/DC1).

Data extraction

For each article, we identified all intervention comparisons with any result reported in the abstract, which indicated that

they were of primary interest of the authors. One reviewer (L.G.H.) extracted all data using an electronic extraction sheet. We formulated the primary research questions of the RCD studies following the PICO scheme except for outcome (e.g., “In patients with hypertension [P], what is the effect of diuretics [I] compared to beta-blockers [C]”). For each research question, we perused the complete publication for reported comparative effects of these treatments on mortality derived from propensity score analyses. Only research questions with such results were considered for further analyses. If there were several research questions, we considered them separately. We excluded articles without such treatment comparisons. Clinically relevant treatment variations (e.g., substantial changes of timing or dosage) or patient conditions (e.g., comorbidities) were considered separately. We also considered specific sub-questions separately (e.g., the main research question compared antihypertensive drugs with no antihypertensive treatment, and subanalyses compared separately diuretics and β -blockers). Evaluations of specific age groups within adult patient populations and demographic subpopulations (e.g., sex, race/ethnicity) were not considered separately.

We categorized eligible studies by type of analyzed disease or condition, interventions and type of RCD. The study type was categorized as “registry data” for studies in which the authors described the data as “registry” or “registered data” (solely or linked with other data); “administrative data” for studies using solely administrative data; “electronic medical or health records” for studies clearly reporting the sole use of electronic medical or health records; or “other” for studies using other types of RCD, RCD that could not be clearly allocated to the other categories or combinations of nonregistry data sources.

Identification of clinical trial evidence

We perused the main text of each article and the cited literature to identify existing RCTs on each extracted primary research question. If no such trial was mentioned or cited in an article, we searched PubMed and the Cochrane Library (last search December 2013) for RCTs or for systematic reviews or meta-analyses of RCTs (we updated searches of existing evidence syntheses or directly searched RCTs without time restrictions; details are in Appendix 3, available at www.cmajopen.ca/content/4/2/E132/suppl/DC1) and recorded whether there were any RCTs published before the year in which the RCD was published. One reviewer (L.G.H.) conducted these processes. He marked all RCTs identified in our searches where he felt there was some uncertainty about their eligibility. This was discussed with a second reviewer (D.C.I.), who also confirmed eligibility of all identified pertinent RCTs and who spot-checked the excluded full-length articles for verification. Discrepancies were resolved by discussion.

Evaluation of research motivation

For the RCD studies, we recorded how often the authors claimed that performing a clinical trial on their research questions would be unfeasible for ethical reasons or would be difficult (for any reason) and how often they claimed that performing a trial would be necessary.

For RCD studies whose authors knew that existing RCTs had already compared the treatments examined in their own study (as indicated by direct mention in the text or by citing an RCT or meta-analysis or systematic review of RCTs), we evaluated the motivation that the authors claimed for performing their study and which gaps in clinical trial evidence they aimed to close. We grouped the authors' motivations into 4 prespecified categories: to assess effects on outcomes different from those reported in existing trials or on outcomes they felt were not adequately studied in existing trials (e.g., because of low power); to assess effects in specific demographic populations (e.g., women and children) or populations with specific conditions (e.g., comorbidities) that they felt were not adequately studied in clinical trials; to assess effects outside of controlled trials because they felt that RCTs did not, or did not adequately, reflect the "real world"; or because findings from previous RCTs were inconclusive or inconsistent compared with other randomized or nonrandomized evidence. Types of motivation that fell outside these categories were also systematically extracted.

One reviewer (L.G.H.) marked all articles that he felt clearly reported a research motivation. A second reviewer (D.C.I.) evaluated all other articles where the first reviewer could not identify a research motivation or felt that there was some uncertainty about its categorization. Discrepancies were resolved by discussion.

Evaluation of citation impact of RCD studies

One reviewer (L.G.H.) extracted the bibliographic information for each eligible article and recorded the impact factor of the publishing journal (ISI Web of Knowledge 2012), the 5-year impact factor and the number of times the article was cited until June 2014 (ISI Web of Science). We compared the citation impact metrics of RCD studies that mentioned or cited at least 1 previous RCT on the same research question with the citation impact metrics of studies where no RCT existed. We also compared the citation impact metrics of RCD studies according to whether specific motivations for their conduct were mentioned. Finally, we compared RCD studies for which RCTs on the same research question existed (regardless of the awareness of the authors) and those for which RCTs did not exist.

Statistical analysis

We used Stata 13.1 (Stata Corp) for all statistical analyses. We report results as medians with interquartile ranges if not otherwise stated. We tested differences between continuous variables using the Mann–Whitney *U* test; for categorical data, we used the Fisher exact test. The *p* values are two-tailed.

Results

Our literature search yielded 929 references. After the screening of titles and abstracts, 420 records were selected for full-text evaluation, and 337 were included in our analysis (Figure 1). The median publication year was 2008. The diseases or conditions most frequently evaluated were cardiovascular disease (63.2%), cancer (11.6%) and transplantation (4.5%). As

for types of treatment, most of the studies evaluated drugs (48.1%) or coronary revascularization procedures (27.9%). About half (51.9%) used an active comparator. Most of the studies relied on registry data (64.4%), and 13.6% used solely administrative data (Table 1).

Existence of RCT comparisons

In total, 231 (68.5%) of the studies assessed the comparative effectiveness of interventions that had already been compared in RCTs. In most (213, 63.2%), there was some mention or reference to at least 1 RCT or to a meta-analysis or systematic review including such a trial. We identified at least 1 previous RCT in another 18 cases where the study had not mentioned or referenced any clinical trial evidence.

Of the total 337 RCD studies, the authors of 6 studies (2 of the 124 without mention or citation of an RCT, and 4 of the other 213) claimed that RCTs were unethical to perform to address the question of interest. Authors of 18 studies claimed that performing RCTs would be difficult, but we found that RCTs had already compared the treatments examined in 11 of the 18 studies. Authors of 101 studies deemed RCTs necessary to conduct in the future. The authors of 56 of these studies were aware of existing RCTs and called for additional clinical trial evidence; the authors of the other 45 studies were not aware of RCTs (trials already existed in 7 cases) and called for novel clinical trial evidence.

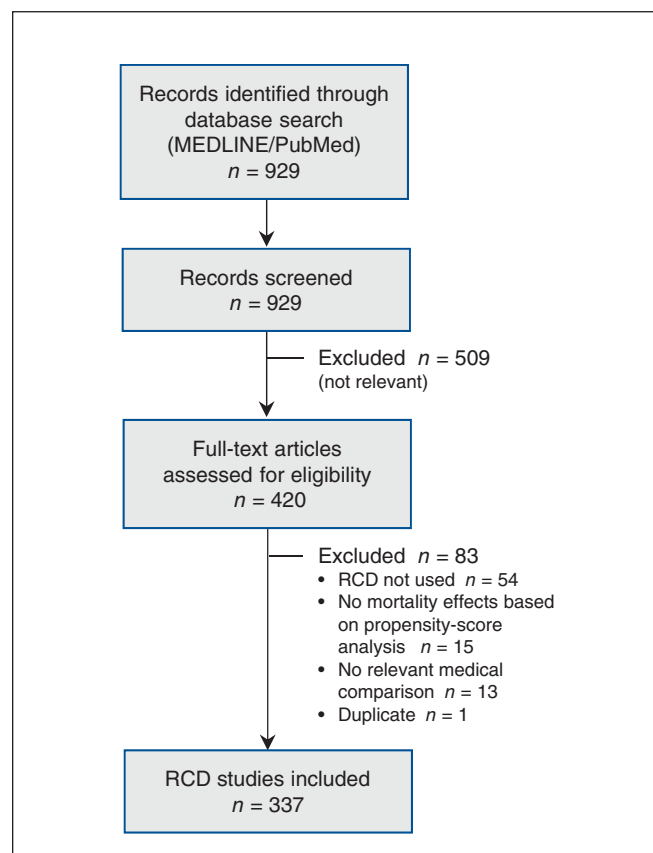


Figure 1: Selection of RCD studies for the analysis. RCD = routinely collected health data.

Table 1: Characteristics of studies using routinely collected health data (RCD studies)

Characteristic	No. (%) of RCD studies*			p value
	All studies n = 337	Studies with existing RCTs† n = 231	Studies without existing RCTs n = 106	
Publication year, median (IQR)	2008 (2006–2009)	2008 (2007–2009)	2008 (2005–2009)	0.4
Type of condition/disease/specialty				0.06
Chronic kidney disease	10 (3.0)	4 (1.7)	6 (5.7)	
Cardiovascular disease	213 (63.2)	155 (67.1)	58 (54.7)	
Cancer	39 (11.6)	26 (11.3)	13 (12.3)	
Diabetes mellitus	8 (2.4)	6 (2.6)	2 (1.9)	
Pediatrics	3 (0.9)	3 (1.3)	0	
Pregnancy	1 (0.3)	1 (0.4)	0	
Psychiatry	11 (3.3)	8 (3.5)	3 (2.8)	
Pulmonology	10 (3.0)	6 (2.6)	4 (3.8)	
Surgery	10 (3.0)	4 (1.7)	6 (5.7)	
Transplantation	15 (4.5)	6 (2.6)	9 (8.5)	
Other	17 (5.0)	12 (5.2)	5 (4.7)	
Type of treatment				< 0.001
Coronary revascularization	94 (27.9)	74 (32.0)	20 (18.9)	
Devices	9 (2.7)	3 (1.3)	6 (5.7)	
Drugs	162 (48.1)	124 (53.7)	38 (35.8)	
Radiation	10 (3.0)	5 (2.2)	5 (4.7)	
Surgery‡	35 (10.4)	15 (6.5)	20 (18.9)	
Different types§	8 (2.4)	3 (1.3)	5 (4.7)	
Other	19 (5.6)	7 (3.0)	12 (11.3)	
Type of comparator				0.2
Active intervention	175 (51.9)	125 (54.1)	50 (47.2)	
Usual care	162 (48.1)	106 (45.9)	56 (52.8)	
Type of routine data				0.6
Registry data	217 (64.4)	147 (63.6)	70 (66.0)	
Administrative data	46 (13.6)	34 (14.7)	12 (11.3)	
EMR/EHR	7 (2.1)	6 (2.6)	1 (0.9)	
Other	67 (19.9)	44 (19.0)	23 (21.7)	

Note: EHR = electronic health record, EMR = electronic medical record, IQR = interquartile range, RCT = randomized controlled trial.
 *Unless stated otherwise.
 †Studies in which RCTs (or meta-analyses or systematic reviews of RCTs) on the same research question were mentioned or cited, or cases where at least 1 previous RCT was identified in the literature searches.
 ‡Excludes coronary artery bypass graft surgery, which is included under coronary revascularization.
 §Comparison of different types of interventions (e.g., drug therapy v. radiation).

Motivation for research efforts

For the 213 studies where the authors mentioned or cited existing RCT evidence, Table 2 summarizes the motivations that the authors claimed for performing their research and which limitations of existing clinical trials they aimed to address. Examples of typical statements for each of the most frequent justifications or motivations are given in Box 1.

In most of the studies (125/213), we identified a single motivation; some had 2 ($n = 60$) or more ($n = 9$) motivations (Figure 2). Most frequently (37.6%), authors felt that available RCTs provided insufficient knowledge on the value of the

compared treatments in the “real world.” In 31.9%, the authors aimed to assess effects on outcomes that were not, or in their opinion not adequately, studied in RCTs (this included mortality or long-term clinical outcomes in 94.1%, $n = 64$). In 25.8%, the authors deemed their research necessary because findings from existing RCTs were inconsistent or inconclusive compared with other randomized or nonrandomized evidence. In 23.5%, the authors aimed to assess effects in specific populations (specific demographic populations or ethnic groups in 13.6%, $n = 29$; populations characterized by specific diseases or conditions in 9.9%, $n = 21$).

Table 2: Motivation for research efforts reported by authors aware of existing RCT evidence

RCT-related motivation	No. (%) studies	
	Studies with only 1 RCT-related motivation <i>n</i> = 125	All studies* <i>n</i> = 213
Assess effects in the “real world”	41 (32.8)	80 (37.6)
Assess different outcomes	34 (27.2)	68 (31.9)
RCT evidence inconclusive or inconsistent	23 (18.4)	55 (25.8)
Compared with randomized evidence	6 (4.8)	15 (7.0)
Compared with randomized and nonrandomized evidence	17 (13.6)	40 (18.8)
Study specific patient population	20 (16.0)	50 (23.5)
Other	7 (5.6)	21 (9.9)
None	–	19 (8.9)

Note: RCT = randomized control trial.
*For study details, see Appendix 4 (available at www.cmajopen.ca/content/4/2/E132/suppl/DC1).

Authors of 9.9% of the RCD studies claimed other gaps in RCT evidence that encouraged them to analyze routine data, including (a) outdated circumstances under which the existing RCTs were conducted (e.g., no modern background treatments); (b) methodological limitations of the RCTs (e.g., early discontinuation for benefit; high treatment crossover rates); (c) signals in previous RCTs for potential subgroup effects or effect modifications that would merit further investigation; and (d) factors making new RCTs unfeasible or unethical (e.g., due to clearly established benefits or harms of one comparator).

In 8.9% of the studies, we could not identify a specified RCT-related motivation or rationale for the research efforts. Other motivations that were not related to claimed problems with RCT evidence included utilization issues (13.1%, *n* = 28) and the evaluation of risk factors, predictors or effect modifiers (10.8%, *n* = 23). In 2.8% (6/213), no rationale was listed for the research efforts, either related to RCT evidence or not related.

Citation impact of RCD studies

The RCD studies without mention or reference to existing RCT evidence and those by authors who were aware of such RCT evidence were published in journals with similar impact factors (median 4.5 v. 4.5, *p* = 0.2) and similar 5-year impact factors (median 4.8 v. 4.6, *p* = 0.1). Studies without mention or reference to prior RCTs had significantly more subsequent citations than studies conducted to supplement existing clinical trial evidence. The subsequent citation impact depended on the type of evidence gap the authors aimed to close. Studies conducted to supplement RCT knowledge on certain outcomes had more subsequent impact, and studies conducted with the justification to explore “real world” effects had significantly lower citation impact than other studies (Table 3).

When we compared RCD studies without RCTs before publication (i.e., neither mentioned in the articles nor identified in our searches) with the other studies, the differences in citation impact were smaller (Table 3).

Interpretation

In our analysis of 337 RCD studies, about 70% of the research supplemented existing clinical trial evidence and did not provide fundamentally novel answers on the comparative effectiveness of treatments never compared before in clinical trials. Only rarely were RCD studies published with a claim by the authors that RCTs were unfeasible or unrealistic to perform. The most frequently reported research motivation for RCD studies — the alleged limited generalizability of clinical trial evidence to the “real world” — was associated with the lowest citation impact. On the other hand, studies venturing into areas where no RCT existed had significantly more citation impact.

We focused primarily on the claimed motivation of research related to previously existing trial evidence. Other motivations were reported occasionally, but the vast majority of RCD studies had at least one reported motivation related to the respective RCT evidence. Moreover, the claimed motivations may not necessarily have been prespecified. Occasionally, they may have been post-hoc justifications to try to explain the importance of the work. It would be impossible to determine definitely the prespecification of motivations without preregistration of protocols for RCD studies. However, the credibility of authors’ statements regarding the prespecification of research questions in RCTs was shown to be low.²⁷ Nevertheless, on the whole, the motivations reported in their publications reflect how investigators perceive eventually the value and relevance of their RCD studies.

In most situations where RCTs already existed, the authors of RCD studies did mention or cite at least one of them. This does not mean, however, that all of the pre-existing RCTs were necessarily mentioned and cited. We did not evaluate whether the cited RCT evidence was an incomplete sample of the existing RCTs, because this would have required performing systematic reviews on hundreds of topics. There is evidence that RCTs only sparingly cite previous trials, with more

than 75% of existing RCTs not being cited.²⁸ The record for previous RCTs being cited in RCD studies may be better, but still not perfect. We suggest that a systematic review of existing evidence be done before any new RCD study is conducted. The systematic review may need to assess not only previous RCTs, but also previous RCD studies on the same topic.

Limitations

Several limitations need to be considered. First, we included only studies reporting on mortality. Because mortality is typically the most clinically important outcome, it is expected that all included studies are highly relevant for health care decision-making. However, this probably leads

Box 1: Examples of motivation for research efforts

Limited generalizability of clinical trials: not adequate reflection of the “real world”

- “... [I]t remains uncertain how CAS performs in comparison to CEA outside the context of clinical trials.”¹⁵
- “... [I]t remained unclear whether the data accumulated in randomized clinical trials apply to patients with different baseline and procedural characteristics treated in routine practice. Thus, we compared the longterm survival of patients treated with and without abciximab ...”¹⁶
- “It is well known that the results of randomized clinical trials do not necessarily apply to the results observed in everyday clinical practice. Therefore, the aim of our analysis was to determine the effectiveness and safety of enoxaparin in unselected patients with STEMI in clinical practice in the German Acute Coronary Syndromes (ACOS) registry.”¹⁷

Outcomes not adequately studied in clinical trials

- “... [L]imited data exist regarding the long-term outcomes of coronary stenting, as compared with standard CABG. ... [T]he long-term safety of DES has been questioned by recent reports suggesting increased risk of late stent thrombosis, mortality, or myocardial infarction (MI). ... Therefore, very-long-term follow-up after DES implantation in a large patient cohort ... is important.”¹⁸
- “Although nonantipsychotic psychiatric medications ... are also used for management of neuropsychiatric symptoms of dementia, there is little research support for their efficacy for this indication. ... Because psychotropic agents for neuropsychiatric symptoms are frequently used for long periods, it is also important to compare mortality risks during both acute and maintenance treatment. The purpose of this study was to compare 12-month mortality risks among patients who had recently had prescriptions filled for conventional antipsychotics, atypical antipsychotics, or nonantipsychotic psychiatric medications in outpatient settings following a dementia diagnosis.”¹⁹
- “This gap in our knowledge is due to the paucity of controlled clinical trials evaluating potential therapies. Moreover, the few randomized clinical trials that have been completed focused on regulatory end points and have lacked the power to assess the effect of current intravenous therapies on hospital mortality rates. ... The analyses presented here were undertaken to evaluate the safety (mortality and worsening renal function) of the use of vasodilators and inotropes (INO) during hospitalization for decompensated heart failure.”²⁰

Previous clinical trials inconclusive or inconsistent compared with other randomized or nonrandomized evidence

- “Results of randomised trials on the survival benefits of early revascularisation after acute coronary syndromes are inconsistent. ... Our aim was, therefore, to investigate the effect on 1-year mortality of revascularisation within 14 days after an acute myocardial infarction in a large cohort of unselected patients.”²¹
- “... In the past decade, two influential randomized trials found that treatment with beta-blockers can decrease the incidence of myocardial infarction and death after noncardiac surgery. ... [T]he Agency for Healthcare Research and Quality identified the perioperative use of beta-blockers among intermediate- and high-risk patients as one of the nation’s ‘clear opportunities for safety improvement.’ ... Yet, 2 recent randomized trials ... reported no benefit from perioperative beta-blocker therapy and raised questions about the generalizability of earlier studies. While awaiting the results of large randomized trials ... we evaluated the use and effectiveness of perioperative beta-blocker therapy in routine clinical practice.”²²
- “... [T]he safety and efficacy of CAS are controversial. ... The 2005 Cochrane review concluded that CAS conferred a significant reduction in cranial nerve injury and was no different from CEA for the end points of 30-day death/any stroke, death/disabling stroke, death, stroke, or myocardial infarction (MI) ... The 2007 Cochrane review concluded that CAS conferred significant reductions in cranial nerve injury but that it was associated with a significant increase in the 30-day risk of death/stroke and any stroke. There was no difference in 30-day death and death/disabling stroke or the risk of late stroke. ... The 2009 Cochrane review found that CAS conferred significant reductions in not only cranial nerve injury but also MI and that it was associated with a significant increase in 30-day death/stroke, which was no longer significant in a random-effects model.”²³

Specific patient populations not adequately studied in clinical trials

- “Because no randomized trial of OPCABG versus on-pump CABG in women exists, we retrospectively reviewed CABG outcomes in women in a large hospital system database.”²⁴
- “... [W]omen age 70 and older have been under represented in most breast cancer treatment trials. ... [T]his study did not address the benefit of any chemotherapy versus no chemotherapy in older women. ... [T]he paucity of such data for adjuvant chemotherapy in older breast cancer patients suggests that we examine other possible data sources. ... Our goal was to assess the relationship between adjuvant chemotherapy use and survival in a large population-based cohort of older women with hormone receptor (HR)–negative breast cancer.”²⁵
- “... A recent meta-analysis of seven randomized trials ... demonstrated an improved clinical outcome among patients with unstable angina or NSTEMI. ... [D]ata regarding gender differences in hospital and long-term outcomes after acute NSTEMI are scarce. ... The purpose of the present subanalysis of the Acute Coronary Syndrome (ACOS) registry is to examine differences in patient characteristics, acute therapy, hospital course and one-year outcome of women presenting with NSTEMI treated with an invasive vs conservative strategy.”²⁶

Note: CABG = coronary artery bypass graft surgery, CAS = carotid arterial stent, CEA = carotid endarterectomy, DES = drug-eluting stent(s), NSTEMI = non-ST-segment elevation myocardial infarction, OPCABG = off-pump CABG, STEMI = ST-segment elevation myocardial infarction.

to an overestimation of the proportion of studies conducted to assess specific outcomes (i.e., mortality).

Second, we included only studies that used propensity scores, to ensure that our sample represents studies that applied a widely used, standardized method of comparative effectiveness research.^{29,30} The use of propensity scores is probably the most popular type of methodology involved in comparative effectiveness research, but many other methods are increasingly being used.²⁹⁻³¹ It remains speculative whether researchers applying other methods might be more or less likely to venture on assessing research topics that are entirely novel.

Third, we used a relatively specific search strategy to identify existing RCTs comparing the same treatments as in the RCD studies. Thus, the proportion of studies for which RCTs on the same research question already exist may be even higher.

Fourth, we included only RCD studies published until the end of 2010. This was because our literature search protocol was designed to serve a concurrent project of ours designed to assess whether RCTs have been performed subsequent to RCD studies that had no pre-existing RCT evidence and to determine the results of those RCTs. This study design required a minimum window of a few years of follow-up after the publication of the RCD studies. Thus far, we found very few RCTs published subsequent to RCD studies (only for 18

topics covered by RCD studies). This suggests that RCD studies have a unique opportunity to close evidence gaps that are unlikely to be closed by RCTs in the current circumstances of research prioritization. It is unlikely that RCD studies published after 2010 have markedly changed the profile of their research motivations.

Fifth, we evaluated only published RCD studies and cannot rule out that motivations of unpublished studies were different from those we identified.

Sixth, it would have been interesting to explore whether contradictory results between intention-to-treat and per-protocol analyses (e.g., owing to time-varying effects such as treatment switches) in pre-existing RCTs might have motivated the performance of subsequent RCD studies. Such motivation was not acknowledged in our database, even in RCD studies performed to study “real world” effects. Analyses taking into account time-varying confounders would require specific models (e.g., marginal structural models with inverse probability weighting); however, this was beyond the scope of our study because we analyzed studies using propensity scores.

Finally, citation impact of individual studies is not a perfect measure of quality. However, it gives a measure of how much the study results have affected the subsequent scientific literature.

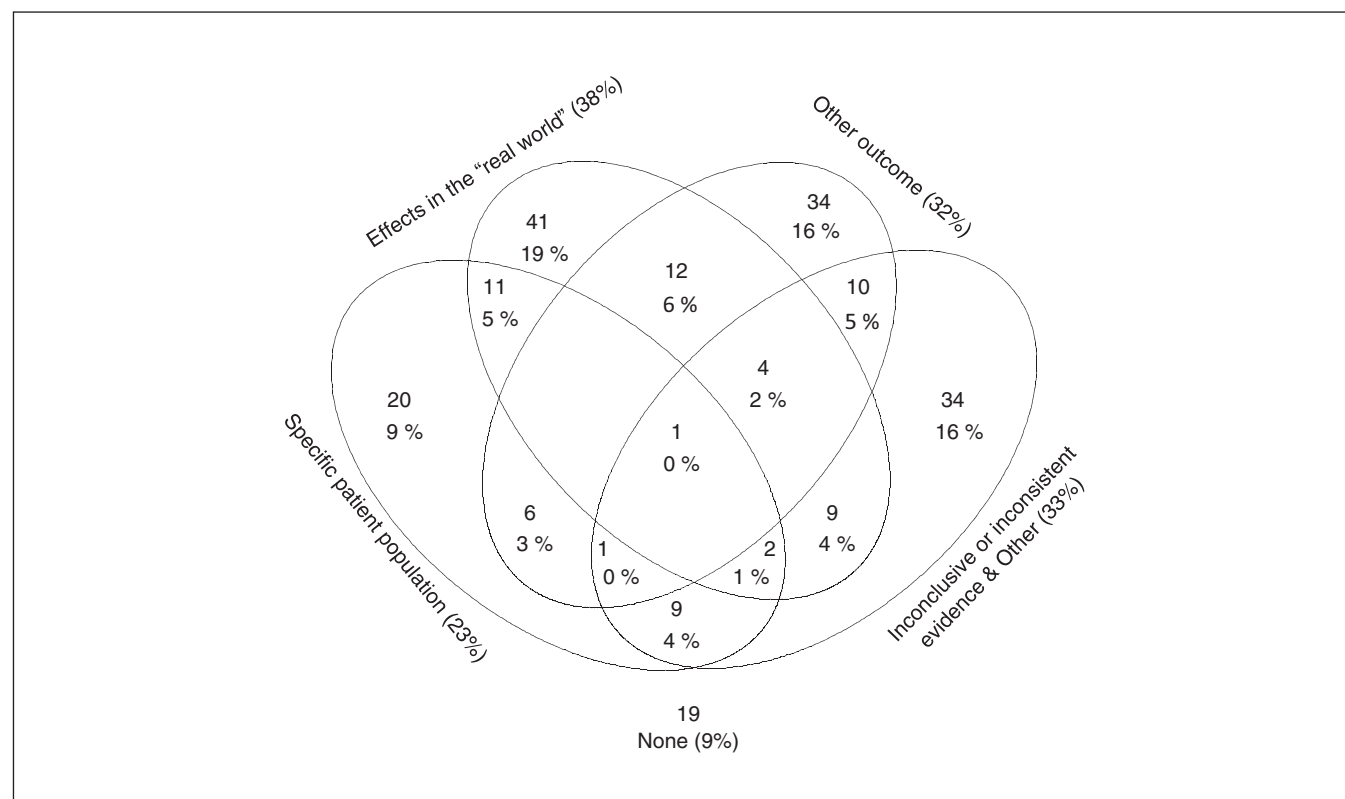


Figure 2: The most frequent motivations for performing observational studies using routinely collected health data (RCD studies) claimed by authors who were aware of existing RCT evidence. Each circular area corresponds to one research motivation category; numbers of studies with multiple motivations are depicted in the corresponding overlapping areas (e.g., 11 RCD studies [5% of 213] reported both the assessment of a specific patient population and the study of “real world” effects as motivations). The diagram is schematic; percentages do not correspond to the size of circular areas.

Table 3: Scientific impact of studies using routinely collected health data (RCD studies)

Variable	Citation impact metric; median (IQR)		
	Journal impact factor in 2012	No. of citations per RCD study per year	Total no. of citations per RCD study
All studies (n = 337)*	4.5 (3.2–11)	4.0 (1.5–8.0)	22 (9–55)
RCT evidence mentioned or cited			
No (n = 124)*	4.5 (3.5–11)	4.6 (2.5–9.3)	29 (13–61)
Yes (n = 213)*	4.5 (3.2–11)	3.0 (1.3–7.5)	19 (8–49)
p value	0.2	0.01	0.01
RCTs existed before publication			
No (n = 106)*	4.5 (3.5–11)	4.2 (2.3–8)	26 (12–58)
Yes (n = 231)*	4.5 (3.2–11)	3.5 (1.3–7.9)	20 (8–53)
p value	0.2	0.1	0.09
Studies with RCT-related research motivation (n = 213)†			
Assess effects in the “real world”			
Yes	3.9 (2.9–6.5)	1.9 (0.95–5.2)	11 (5–36)
No	4.7 (3.2–14)	4.4 (1.9–9.6)	24 (10–60)
p value	0.08	< 0.001	0.002
Assess different outcomes			
Yes	5.0 (3.5–14)	4.7 (2–13)	24 (10–76)
No	4.1 (3.1–9.1)	2.3 (1.3–6)	17 (7–36)
p value	0.1	0.01	0.06
RCT evidence inconclusive or inconsistent			
Yes	4.8 (2.9–9.1)	3.0 (1.3–6.6)	14 (6–38)
No	4.1 (3.2–14)	3.1 (1.3–8)	19 (8–54)
p value	0.9	0.7	0.3
Study specific patient population			
Yes	5.0 (3.2–14)	2.3 (1.3–5.4)	18 (7–38)
No	4.5 (3.2–11)	3.1 (1.3–8)	19 (8–52)
p value	0.7	0.4	0.4
Other			
Yes	4.5 (3.6–13)	3.7 (1.6–9.2)	22 (9–55)
No	4.5 (3.2–11)	2.8 (1.3–7.3)	18 (8–47)
p value	0.2	0.4	0.4
No RCT-related motivation			
Yes	3.5 (3.2–6.2)	5.0 (2.2–11)	29 (18–48)
No	4.5 (3.2–14)	2.9 (1.2–7.3)	18 (7–52)
p value	0.6	0.09	0.06

Note: IQR = interquartile range, RCT = randomized controlled trial.
 *Data on number of citations were missing for 3 studies; there was no impact factor available for 6 studies.
 †Studies mentioning or citing RCT evidence; for study details, see Appendix 4 (available at www.cmajopen.ca/content/4/2/E132/suppl/DC1).

Conclusion

Most of the studies we identified explored comparative treatment effects that had already been investigated in RCTs. Only rarely were RCD studies published with a claim by the researchers that RCTs were unfeasible or unrealistic to perform. Closing serious gaps in clinical evidence with the use of routinely collected health

data when RCT evidence does not exist or is not easy to obtain seems to be the exception rather than the rule. To be clear, we do not deprecate incrementing RCD studies, but there is an urgent need for prioritization of research in this field in favour of new knowledge rather than incremental knowledge generation. Currently, there are numerous clinical questions with no

RCT guidance at all. There is a wealth of comparative effectiveness research questions for which RCTs would be unfeasible or impractical to perform. There may be a need for RCD studies to explore more daring territories where no RCT evidence exists and where RCTs may not be possible to perform.

References

- Spasoff RA. *Epidemiologic methods for health policy*. New York: Oxford University Press; 1999.
- Developing a protocol for observational comparative effectiveness research: a user's guide*. Rockville (MD): Agency for Healthcare Research and Quality; 2013.
- Cox E, Martin BC, Van Staa T, et al. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report—Part II. *Value Health* 2009;12:1053-61.
- Howie L, Hirsch B, Locklear T, et al. Assessing the value of patient-generated data to comparative effectiveness research. *Health Aff (Millwood)* 2014;33:1220-8.
- Rassen JA, Schneeweiss S. Newly marketed medications present unique challenges for nonrandomized comparative effectiveness analyses. *J Comp Eff Res* 2012;1:109-11.
- Berger ML, Mamdani M, Atkins D, et al. Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report — Part I. *Value Health* 2009;12:1044-52.
- PCORI awards \$93.5 million to develop national network to support more efficient patient-centered research. Washington (DC): Patient-Centered Outcomes Research Institute; 2013. Available: www.pcori.org/2013/pcori-awards-93-5-million-to-develop-national-network-to-support-more-efficient-patient-centered-research/ (accessed 2016 Mar. 24).
- Ehrhardt S, Appel LJ, Meinert CL. Trends in National Institutes of Health funding for clinical trials registered in ClinicalTrials.gov. *JAMA* 2015;314:2566-7.
- Dreyer NA, Schneeweiss S, McNeil BJ, et al. GRACE principles: recognizing high-quality observational studies of comparative effectiveness. *Am J Manag Care* 2010;16:467-71.
- Dreyer NA, Tunis SR, Berger M, et al. Why observational studies should be among the tools used in comparative effectiveness research. *Health Aff (Millwood)* 2010;29:1818-25.
- Lewsey JD, Leyland AH, Murray GD, et al. Using routine data to complement and enhance the results of randomised controlled trials. *Health Technol Assess* 2000;4:1-55.
- Downing NS, Aminawung JA, Shah ND, et al. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005–2012. *JAMA* 2014;311:368-77.
- Bartlett C, Doyal L, Ebrahim S, et al. The causes and effects of socio-demographic exclusions from clinical trials. *Health Technol Assess* 2005;9:iii-iv, ix-x, 1-152.
- Federal Coordinating Council for Comparative Effectiveness Research: report to the President and the Congress. Washington (DC): US Department of Health and Human Services; 2009. Available: www.med.upenn.edu/sleepctr/documents/FederalCoordinatingCouncilforCER_2009.pdf (accessed 2016 Mar. 24).
- Groeneveld PW, Yang L, Greenhut A, et al. Comparative effectiveness of carotid arterial stenting versus endarterectomy. *J Vasc Surg* 2009;50:1040-8.
- Brener SJ, Ellis SG, Schneider J, et al. Abciximab-facilitated percutaneous coronary intervention and long-term survival — a prospective single-center registry. *Eur Heart J* 2003;24:630-8.
- Zeymer U, Gitt A, Junger C, et al. Efficacy and safety of enoxaparin in unselected patients with ST-segment elevation myocardial infarction. *Thromb Haemost* 2008;99:150-4.
- Park DW, Seung KB, Kim YH, et al. Long-term safety and efficacy of stenting versus coronary artery bypass grafting for unprotected left main coronary artery disease: 5-year results from the MAIN-COMPARE (Revascularization for Unprotected Left Main Coronary Artery Stenosis: Comparison of Percutaneous Coronary Angioplasty Versus Surgical Revascularization) registry. *J Am Coll Cardiol* 2010;56:117-24.
- Kales HC, Valenstein M, Kim HM, et al. Mortality risk in patients with dementia treated with antipsychotics versus other psychiatric medications. *Am J Psychiatry* 2007;164:1568-76, quiz 1623.
- Costanzo MR, Johannes RS, Pine M, et al. The safety of intravenous diuretics alone versus diuretics plus parenteral vasoactive therapies in hospitalized patients with acutely decompensated heart failure: a propensity score and instrumental variable analysis using the Acutely Decompensated Heart Failure National Registry (ADHERE) database. *Am Heart J* 2007;154:267-77.
- Stenestrand U, Wallentin L. Early revascularisation and 1-year survival in 14-day survivors of acute myocardial infarction: a prospective cohort study. *Lancet* 2002;359:1805-11.
- Lindenauer PK, Pekow P, Wang K, et al. Perioperative beta-blocker therapy and mortality after major noncardiac surgery. *N Engl J Med* 2005;353:349-61.
- Bangalore S, Bhatt DL, Rother J, et al. Late outcomes after carotid artery stenting versus carotid endarterectomy: insights from a propensity-matched analysis of the Reduction of Atherothrombosis for Continued Health (REACH) Registry. *Circulation* 2010;122:1091-100.
- Mack MJ, Brown P, Houser F, et al. On-pump versus off-pump coronary artery bypass surgery in a matched sample of women: a comparison of outcomes. *Circulation* 2004;110(Suppl 1):III-6.
- Elkin EB, Hurria A, Mitra N, et al. Adjuvant chemotherapy and survival in older women with hormone receptor-negative breast cancer: assessing outcome in a population-based, observational cohort. *J Clin Oncol* 2006;24:2757-64.
- Kleopatra K, Muth K, Zahn R, et al. Effect of an invasive strategy on in-hospital outcome and one-year mortality in women with non-ST-elevation myocardial infarction. *Int J Cardiol* 2011;153:291-5.
- Kasenda B, Schandelmaier S, Sun X, et al. Subgroup analyses in randomised controlled trials: cohort study on trial protocols and journal publications [published erratum in *BMJ* 2014;349:4921]. *BMJ* 2014;349:g4539.
- Robinson KA, Goodman SN. A systematic examination of the citation of prior research in reports of randomized, controlled trials. *Ann Intern Med* 2011;154:50-5.
- Sox HC, Goodman SN. The methods of comparative effectiveness research. *Annu Rev Public Health* 2012;33:425-45.
- Johnson ML, Crown W, Martin BC, et al. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part III. *Value Health* 2009;12:1062-73.
- Hlatky MA, Winkelmayer WC, Setoguchi S. Epidemiologic and statistical methods for comparative effectiveness research. *Heart Fail Clin* 2013;9:29-36.

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

Contributors: All of the authors conceived the study, analyzed the data and interpreted the results. Lars Hemkens and Despina Contopoulos-Ioannidis extracted the data. Lars Hemkens drafted the manuscript. All of the authors revised the manuscript critically for important intellectual content, approved the final version to be published and agreed to act as guarantors of the work.

Funding: The study was supported by Santésuisse, the umbrella association of Swiss social health insurers, and by The Commonwealth Fund, a private independent foundation based in New York City. The views presented here are those of the authors and not necessarily those of The Commonwealth Fund or its directors, officers or staff. The Meta-Research Innovation Center at Stanford is funded by a grant from the Laura and John Arnold Foundation. The funders had no role in the design or conduct of the study; the collection, management, analysis or interpretation of the data; or the preparation, review or approval of the manuscript or its submission for publication.

Supplemental information: For reviewer comments and the original submission of this manuscript, please see www.cmajopen.ca/content/4/2/E132/suppl/DC1