DO ROUTINELY COLLECTED HEALTH DATA COMPLEMENT RANDOMIZED EVIDENCE? A SURVEY

Lars G. Hemkens MD MPH [1,2], Despina G. Contopoulos-Ioannidis MD [3,4], John P.A. Ioannidis MD DSc [1,5,6,7]

1 Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine, Stanford, California, USA

2 Basel Institute for Clinical Epidemiology and Biostatistics, University Hospital Basel, Basel, Switzerland

3 Department of Pediatrics, Division of Infectious Diseases, Stanford University School of Medicine, Stanford, California, USA

4 Palo Alto Medical Foundation Research Institute, Palo Alto, California, USA

5 Department of Health Research and Policy, Stanford University School of Medicine, Stanford, California, USA

6 Department of Statistics, Stanford University School of Humanities and Sciences, Stanford, California, USA

7 Meta-Research Innovation Center at Stanford (METRICS)

Version 2014-10-10

CORRESPONDENCE TO:

John P A Ioannidis

Stanford Prevention Research Center

Medical School Office Building, Room X306

1265 Welch Rd

Stanford, CA 94305, USA

Tel: 650 725 5465

jioannid@stanford.edu

Word count: Main text: 2957 / Abstract: 250

ABSTRACT

Background: Routinely collected data (RCD) are proposed to complement randomized controlled trials (RCTs) for comparative effectiveness research and to inform health care decisions when RCTs would be unfeasible.

Methods: We searched Pubmed for RCD-studies published up to 2010 evaluating the comparative effectiveness of medical treatments on mortality using propensity scores. We identified RCTs on the same treatment comparison and evaluated how frequently RCD-studies analyzed treatments never compared previously in RCTs. When RCTs were already available, we noted the claimed motivations for each RCD-study. We also analyzed the citation impact of RCD-studies.

Results: Of 337 eligible RCD-studies, 231 (68.6%) analyzed only treatments that had already been compared in RCTs. Their investigators rarely claimed that it would be unethical (6/337) or difficult (18/337) to perform RCTs on the same question. RCT evidence was mentioned or cited by authors of 213 RCD-studies. Their most common motivations were alleged limited generalizability of RCT results to the "real world" (37.6%), evaluation of specific outcomes (31.9%) or specific populations (23.5%) and inconclusive or inconsistent RCT evidence (25.8%). RCD-studies on treatments never compared in RCTs before had significantly higher citation impact. RCD-studies conducted to supplement existing trials by studying different outcomes had higher impact, while studying "real word" effects had lower impact.

Interpretation: Most studies using routinely collected health data explore comparative treatment effects that have already been investigated in RCTs anyhow. The agenda of RCD-studies needs to shift more towards pivotal questions that have no randomized evidence at all and where RCTs are unfeasible to perform.

INTRODUCTION

Routinely collected data (RCD), such as administrative claims data or electronic medical records databases, are often claimed to be a prime source of evidence for comparative effectiveness research (CER)[1-5]. Research using routine data is currently heavily promoted with immense allocated funding resources. Major infrastructural investments are 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[6]. Conversely, major funders shun away from supporting randomized trials[7]. There are different perceived uses of RCD, depending on whether randomized controlled trials (RCTs) also exist (or can be readily performed) on the same question or not. One may argue[3,8-10] that RCTs are unfeasible or unrealistic to perform for each and every comparison of available medical treatments. Moreover, regulatory agencies often only require randomized comparisons against placebo or no treatment[11]. RCD-studies are touted as being able to close this large evidence gap timely with very limited cost[5,9,10]. In other cases, the contribution of RCD-data is more incremental and it pertains to addressing questions where some data from RCTs also already exist. Then, RCD-studies may be presented as a complement to previous RCT evidence, evaluating whether the results of RCTs also hold true in "real world" circumstances, different settings, with different outcomes, or in populations considered to have been understudied in RCTs (e.g. women, children, or elderly)[9,12,13].

While all observational data analyses are limited by the lack of randomization, modern epidemiological methods like propensity scores or marginal structural models are

increasingly used to address such biases. This may improve the reliability of RCD-studies and thus their value for decision-making in situations where clinical trial evidence is inadequate or lacking.

However, are RCD-studies performed mostly in situations where no RCT evidence exists and is unethical or very difficult to obtain? Or, conversely, do RCD-studies "search under the lamppost" where RCTs have already taken place? Which claimed limitations of existing clinical trials motivate researchers to use routine data and what are the knowledge gaps intended to be closed? Eventually, what is the scientific impact of such research and does it differ depending on whether RCTs also exist or not and on what motivations are reported for the conduct of RCD-studies? We aimed to answer these questions by surveying a large number of RCD-studies.

METHODS

Identification of routine data studies

The number of RCD-studies published to-date is too large (probably many thousands) to allow systematic analysis of all of them. Conversely, we aimed to evaluate a reproducible 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 epidemiological method. For consistency, we thus focused here on studies that (1) evaluated the comparative effectiveness of a treatment intervention against another intervention or no intervention/usual care/standard treatment, (2) included mortality as an assessed outcome, and (3) used propensity scores to analyze mortality.

We searched PubMed (last search November 2011) for eligible RCD-studies published up to 2010. We considered RCD-studies in any patient population with any condition. Eligible treatment interventions were drugs, biologics, dietary supplements, devices, diagnostic procedures, surgery, or radiotherapy. Titles and abstracts were screened and potentially relevant articles were obtained as full-text to assess eligibility. Detailed inclusion criteria and the search strategy appear in Webappendix 2.

Data extraction

For each article, we identified all intervention comparisons with any result reported in the abstract, indicating that they were of primary interest of the authors. We formulated the primary research questions of the RCD-studies following the PICO scheme (but ignoring the outcome), for example: "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, these were considered separately. Articles without any such treatment comparison were excluded. 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 sub-analyses compared separately diuretics and beta-blockers). Evaluations of specific age groups within adult populations and demographic subpopulations (sex, race/ethnicity) were not separately considered.

We categorized eligible studies by the type of analyzed disease/condition, interventions, and type of RCD. The RCD type was categorized as follows: "Registry": studies using RCD described by the authors as "registry" or "registered data" (solely or linked with other data); "Administrative data": studies using solely administrative data; "Electronic medical or health records": studies clearly reporting the solely use of electronic medical or health records; "Other": studies using other types of RCD, RCD that could not be clearly allocated to the other categories, or combinations of non-registry data sources.

Identification of RCT evidence

We perused the main text of each article and the cited literature to identify existing RCTs on each extracted primary research question. When we identified no mentioned/cited RCT, we searched PubMed and the Cochrane Library (last search December 2013) for RCTs or systematic reviews or meta-analyses of RCTs (details in Webappendix 3) and recorded whether there were any RCTs published up until the year before the year of publication of the RCD.

Evaluation of research motivation

For all RCD-studies, we recorded how often their authors claimed that performing RCTs on their research questions would be unfeasible due to ethical reasons or difficult (due to any reasons) and how often they claimed that performance of RCTs would be necessary.

For RCD-studies where the authors knew that existing RCTs have already compared the treatments examined in their own study (as indicated by direct mentioning in the text or citing an RCT, a meta-analysis or a review of such RCTs), we evaluated the motivation that authors claimed for performing this research and which gaps in clinical trial evidence they aimed to close. We evaluated whether the authors aimed to assess effects on different

outcomes than that reported from RCTs or outcomes that have in their opinion not been adequately studied in RCTs (e.g. because of low power); effects in specific demographic populations (e.g. children, elderly) or populations with specific conditions (e.g. comorbidities) which have in their opinion not been adequately studied in RCTs; or effects outside of controlled trials because they felt that RCTs did not or not adequately reflect the "real world"; and whether the authors deemed previous RCTs inconclusive or inconsistent compared to other randomized or non-randomized evidence. Any other types of motivation that did not fall in these four pre-specified categories were also systematically extracted. One reviewer performed all extractions and literature searches (LGH). This reviewer marked all articles having in his opinion clearly reported any research motivations. A second reviewer (DCI) 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

We extracted bibliographic information of 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 citations received by the article until June 2014 (ISI Web of Science). We compared the citation impact metrics of RCD-studies where at least one previous RCT was mentioned or cited on the same research question versus those where no such RCT existed and according to the presence or not of specific motivations for their conduct.

Statistical analysis

We used Stata 13.1 (Stata Corp, College Station, TX, USA). Results are reported as medians with interquartile ranges (IQR) if not otherwise stated. We tested differences between

continuous variables using the Mann–Whitney U test and for categorical data we used Fisher's exact test. P-values are 2-tailed.

RESULTS

Sample of routine data studies

Our literature search resulted in 929 references. After screening titles and abstracts, 420 references were selected for full-text evaluation and 337 studies were eligible. The median publication year was 2008. These studies evaluated most frequently patients with cardiovascular conditions (63.2%), followed by cancer and transplantation settings. Most studies evaluated drugs (48.1%) or coronary revascularization procedures (27.9%). About half used an active comparator (51.9%). Most studies relied on routine data described as "registry" (64.4%) and 13.7% used solely administrative data (Table 1).

Existence of randomized treatment comparisons

In total, 231 (68.6%) RCD-studies assessed the comparative effectiveness of interventions that had already been compared in RCTs. In most studies (213, 63.2%), there was some mention or reference to at least one RCT or a meta-analysis or review including such RCT. Our electronic searches identified at least one previous RCT in another 18 cases where the RCD-study had not mentioned or referenced any RCT evidence.

Across the 337 eligible RCD-studies, in 6 their authors claimed that RCTs were unethical to perform to address the question of interest (2 of the 124 RCD-studies without mention/reference to an RCT, and 4 of the other 213 RCD-studies). Authors of 18 RCD-studies claimed that performing RCTs would be difficult, but RCTs already existed in 11/18.

Authors of 101/337 RCD-studies deemed RCTs necessary to conduct in the future. This included 56 RCD-studies where the authors were aware of existing RCTs and called for additional RCT evidence and 45 RCD-studies where the authors were not aware of any RCTs (RCTs actually already existed in 7 cases) and called for novel RCT evidence.

Motivation of research efforts when clinical trial evidence was present

For the 213 studies where the authors of RCD-studies were aware of RCTs, Table 2 summarizes how the authors described why their research was necessary and which limitations of existing clinical trials they aimed to address. Examples of typical statements for each of the most frequent justifications/motivations are given in Table 3.

In most studies, we identified a single motivation (125 of 213), and some studies had two (n=60) or more (n=9) (Figure 1). Most frequently (37.6%), authors felt that available RCTs provided insufficient knowledge on the value of the compared treatments outside of controlled trials 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 of inconsistent or inconclusive findings in previous 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).

Authors of 9.9% of RCD-studies reported other claimed gaps in RCT evidence that encouraged them to analyze routine data. This included that authors deemed the circumstances outdated under which the existing RCTs were conducted (e.g. no modern background treatments were used), methodological limitations of the RCTs (e.g. early

discontinuation for benefit; high treatment cross-over rates), indications of potentially subgroup effects or effect-modifications identified in previous RCTs that merit closer investigation, or factors making new trials unfeasible or unethical to be performed (e.g. due to established benefits or harm of one comparator). In 8.9%, we could not identify any specified motivation or related rationale for the research efforts that would be juxtaposed to the RCT evidence. Other motivations of the RCD analyses that were not related to claimed problems with RCT evidence included utilization issues (13.1%; 28 of 213) or the evaluation of risk factors, predictors, or effect modifiers (10.8%; 23 of 213). In 2.8% (6 of 213) no rationale was listed for the research efforts, either related to RCT evidence or not.

Citation impact of RCD-studies

RCD-studies without mention or reference to prior existing RCT evidence and those of authors who were aware of such RCT evidence were published in journals with similar impact factors (median 4.5 versus 4.5 p=0.21) and similar 5-year impact factors (median 4.8 versus 4.6, p=0.14). However, RCD-studies without prior RCTs had significantly more subsequent citations than studies supplementing existing RCT evidence. When RCTs were present, 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 4).

DISCUSSION

Our analysis of 337 CER studies using routinely collected health data shows that about 70% of this research supplements in an incremental fashion available existing randomized trials

on the very same question, but does not provide fundamentally novel answers on the comparative effectiveness of treatments that have never been compared in clinical trials. RCD-studies are only rarely conducted using as rationale the fact that RCTs are unfeasible or unrealistic to perform. The most frequently reported research motivation for RCD-studies, the alleged limited generalizability of randomized trial results 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. Some other research motivations were also 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 pre-specified. Occasionally, they may have been post-hoc justifications trying to buttress that it was important to perform this work. Regardless, on the whole the published motivations reflect how investigators perceive eventually their RCD-studies and their importance.

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 the pre-existing RCTs were necessarily mentioned and cited. We did not evaluate whether the cited RCTs or their reviews were an incomplete sample of the existing RCTs since 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[14]. Thus, the citation record of previous RCTs by RCD-studies may be better, but still not perfect. We suggest that a systematic review of existing evidence should precede any new

RCD-study. This systematic review may actually need to assess not only previous RCTs, but also previous RCD-studies on the same topic.

Several limitations need to be considered. First, we included only studies reporting on mortality. Mortality is typically the most patient-relevant outcome. While this ensures that all included studies are highly relevant for patients and decision makers, this probably also leads to an overestimation of the proportion of studies aiming to assess specific outcomes. Mortality tends to be more uncommon than other main outcomes and RCTs may not be able to definitively address mortality effects. Second, we included only studies using propensity scores what ensures that our sample represents studies using a widely used, standard CER method[15,16]. Propensity methods are probably the most popular type of methodology involved in CER, but many other methods are increasingly used[15-17]. 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 conducted when RCTs on the same question are available might be even higher. Fourth, we only included RCD-studies published until the end of 2010. This is because our literature search protocol was aimed to serve concurrently also another project that we are conducting and where we are assessing whether RCTs were subsequently performed, when no RCTs were available by the time the RCD-study was published, and to determine the results of these RCTs. This required a minimum window of a few years of follow-up after the publication of the RCD-study. Preliminarily, we found very few subsequent RCTs (only for n=19 topics covered by RCD-studies). This suggests that RCD-studies do have a unique opportunity to cover evidence

gaps that are unlikely to be covered by RCTs in the current circumstances. It is unlikely that RCD-studies published in the last 3 years have markedly changed the profile of their motivations. Finally, citation impact of single papers is not a perfect measure of quality. However, it gives a measure of how much the study results have been used by the subsequent scientific literature.

Our results suggest that studies using routinely collected health data are relatively rarely used to address health care problems when randomized trials would be unfeasible. Closing serious clinical evidence gaps with this data source when no RCT data exist or are easy to obtain seems to be rather the exception than the rule. This is unfortunate because currently there is a wealth of CER questions where RCTs would be unfeasible or impractical to perform. For example, in many diseases, many interventions have been approved based only on comparisons against placebo or no treatment[11]. Head-to-head comparisons are relatively rare in the RCT literature[18,19] and demonstrating differences in mortality or other major outcomes for all these approved treatments is impractical in the RCT setting. RCD-studies would be welcome for such CER applications. Conversely, RCD-studies seem to be searching currently mostly under the lamppost of RCT evidence rather than answering new important questions. While a justification can almost always be invoked for such incremental studies, it is unclear whether this makes these studies worthy their effort. RCD studies may need to be emancipated and undertake research into more bold territories where no RCT evidence exists and where RCTs may not be possible to perform.

REFERENCES

1. *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide*. Rockville MD: Agency for Healthcare Research and Quality.; 2013.
2. Cox E, Martin BC, Van Staa T, Garbe E, Siebert U, Johnson ML. 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.
3. Howie L, Hirsch B, Locklear T, Abernethy AP. Assessing the value of patient-generated data to comparative effectiveness research. *Health Aff (Millwood).* 2014;33:1220-8.
4. Rassen JA, Schneeweiss S. Newly marketed medications present unique challenges for nonrandomized comparative effectiveness analyses. *J Comp Eff Res.* 2012;1:109-11.
5. Berger ML, Mamdani M, Atkins D, Johnson ML. 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.
6. Patient-Centered Outcomes Research Institute. PCORI Awards $93.5 Million to Develop National Network to Support More Efficient Patient-Centered Research. 2014; http://www.pcori.org/2013/pcori-awards-93-5-million-to-develop-national-network-to-support-more-efficient-patient-centered-research/. Accessed 17 July, 2014.
7. Ioannidis JP, Nicholson JM. NIH funding: The critics respond. *Nature.* 2013;493:26.
8. 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.
9. Dreyer NA, Tunis SR, Berger M, Ollendorf D, Mattox P, Gliklich R. Why observational studies should be among the tools used in comparative effectiveness research. *Health Aff (Millwood).* 2010;29:1818-25.
10. Lewsey JD, Leyland AH, Murray GD, Boddy FA. Using routine data to complement and enhance the results of randomised controlled trials. *Health Technol Assess.* 2000;4:1-55.
11. Downing NS, Aminawung JA, Shah ND, Krumholz HM, Ross JS. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012. *JAMA.* 2014;311:368-77.
12. 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.
13. U.S. Department of Health & Human Services. Federal Coordinating Council for Comparative Effectiveness Research: Report to the President and the Congress, June 30, 2009. 2009;

http://www.med.upenn.edu/sleepctr/documents/FederalCoordinatingCoucilforCER_2009.pdf. Accessed July 17, 2014.
14. 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.
15. Sox HC, Goodman SN. The methods of comparative effectiveness research. *Annu Rev Public Health.* 2012;33:425-45.
16. Johnson ML, Crown W, Martin BC, Dormuth CR, Siebert U. 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.
17. Hlatky MA, Winkelmayer WC, Setoguchi S. Epidemiologic and statistical methods for comparative effectiveness research. *Heart Fail Clin.* 2013;9:29-36.
18. Lathyris DN, Patsopoulos NA, Salanti G, Ioannidis JP. Industry sponsorship and selection of comparators in randomized clinical trials. *Eur J Clin Invest.* 2010;40:172-82.
19. Salanti G, Kavvoura FK, Ioannidis JP. Exploring the geometry of treatment networks. *Ann Intern Med.* 2008;148:544-53.
20. Groeneveld PW, Yang L, Greenhut A, Yang F. Comparative effectiveness of carotid arterial stenting versus endarterectomy. *J Vasc Surg.* 2009;50:1040-8.
21. Brener SJ, Ellis SG, Schneider J, Apperson-Hansen C, Topol EJ. Abciximab-facilitated percutaneous coronary intervention and long-term survival--a prospective single-center registry. *Eur Heart J.* 2003;24:630-8.
22. 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.
23. 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.
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 623.
25. 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.
26. 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.
27. Lindenauer PK, Pekow P, Wang K, Mamidi DK, Gutierrez B, Benjamin EM. Perioperative beta-blocker therapy and mortality after major noncardiac surgery. *N Engl J Med.* 2005;353:349-61.

28. 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.
29. 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:II1-6.
30. Elkin EB, Hurria A, Mitra N, Schrag D, Panageas KS. 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.
31. 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. *International journal of cardiology.* 2011;153:291-5.

ACKNOWLEDGMENTS

Declaration of competing interests

All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf and declare: DCI and JPAI had no financial support for this project.; LGH had support from The Commonwealth Fund and from Santésuisse for the submitted work. All authors declare no financial relationships with any organization that might have an interest in the submitted work in the previous 3 years and no other relationships or activities that could appear to have influenced the submitted work.

Contributors

LGH, DCI, JPAI conceived the study, analyzed the data, and interpreted the results. LGH wrote the first draft and all authors made revisions on the manuscript. LGH and DCI extracted the data. All authors read and approved the final version of the paper. JPAI is the guarantor.

Funding

Supported by Santésuisse, the umbrella association of Swiss social health insurers. Supported 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, its directors, officers, or staff. The Meta-Research Innovation Center at Stanford is funded by a grant by the Laura and John Arnold Foundation.

Role of the funding source

The funders had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript or its submission for publication.

Ethical approval

Not required for this study.

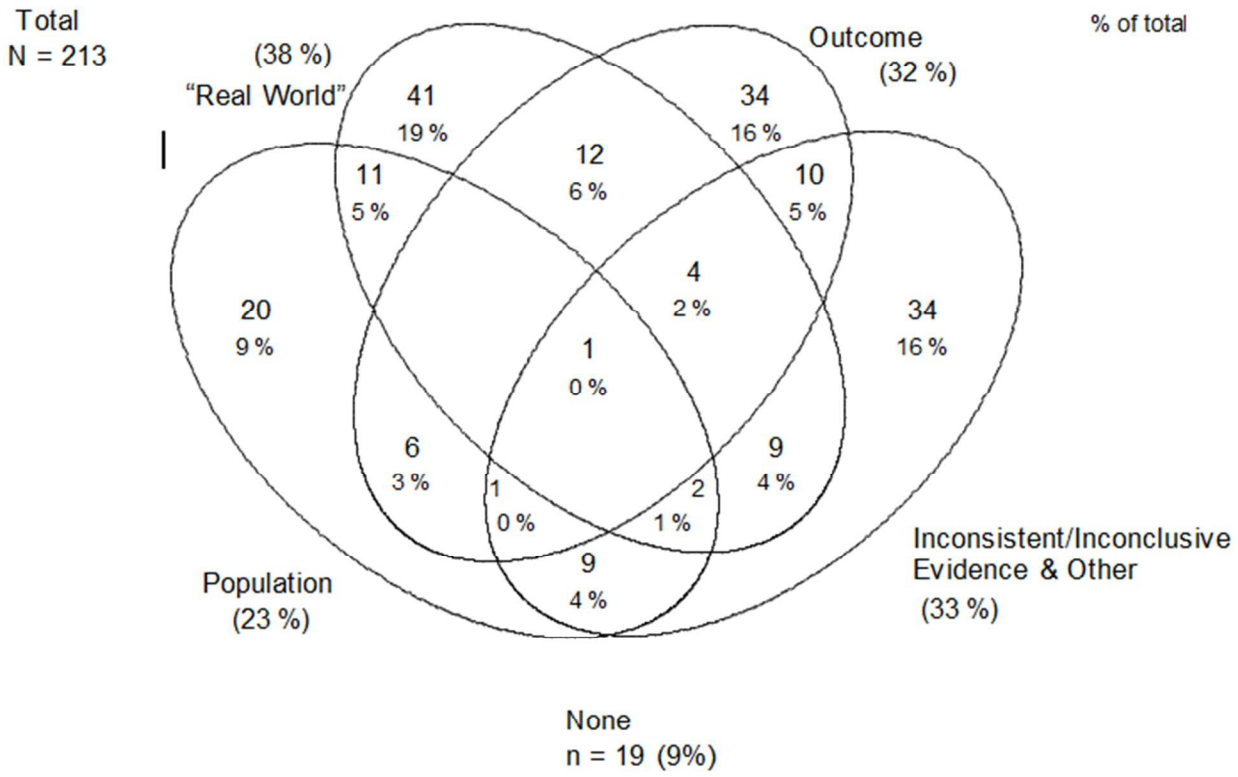

Figure 1: Most frequent motivations of research efforts in RCD-studies where authors were aware of existing RCTs (Venn diagram)

Table 1: Characteristics of Routinely Collected Data Studies

	All Studies n (%)	RCD-studies with RCTs before publication n (%)		
		Yes	No	p-value
Number of studies	337 (100)	231 (100)	106 (100)	
Publication year (median, IQR)	2008 (2006;2009)	2008 (2007;2009)	2008 (2005;2009)	0.39
Type of condition or disease				0.06
CKD	10 (3.0)	4 (1.7)	6 (5.7)	
CVD	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 (0)	
Pregnancy	1 (0.3)	1 (0.4)	0 (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 treatments				<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 (35.9)	
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				0.24
Active intervention	175 (51.9)	125 (54.1)	50 (47.2)	
No treatment beyond usual care	162 (48.1)	106 (45.9)	56 (52.8)	
Type of routine data				0.64
Registry data	217 (64.4)	147 (63.6)	70 (66.0)	
Administrative data	46 (13.7)	34 (14.7)	12 (11.3)	
EMR/EHR	7 (2.1)	6 (2.6)	1 (0.9)	
Other	67 (19.9)	44 (19.1)	23 (21.7)	

CKD: Chronic Kidney Disease. CVD: Cardiovascular Disease. EMR: Electronic Medical Record. EHR: Electronic Health Record. Surgery: excludes CABG which is categorized under Coronary revascularization. Different types: RCD-studies compared different types of interventions (e.g. drug therapy vs. radiation).

Table 2: Motivation of research efforts of routinely collected data studies

	Studies with only one RCT-related motivation n (%)	All Studies n (%)
Total	125 (100)	213 (100)
RCT-related research motivation		
Effects in the "real world"	41 (32.8)	80 (37.6)
Other outcomes	34 (27.2)	68 (31.9)
Specific Population	20 (16.0)	50 (23.5)
Inconclusive or inconsistent evidence	23 (18.4)	55 (25.8)
RCTs only	6 (4.8)	15 (7.0)
RCTs and non-randomized evidence	17 (13.6)	40 (18.8)
Other	7 (5.6)	21 (9.9)
No RCT-related rationale	-	19 (8.9)

Table 3: Examples for motivation of research efforts

Limited generalizability of clinical trials: not adequate reflection of the "real world"
▪ "[…] it remains uncertain how CAS performs in comparison to CEA outside the context of clinical trials." [20] *[CAS: Carotid arterial stent; CEA: Carotid endarterectomy]*
▪ "[…] it 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 […]". [21]
▪ "It is well known that the results of randomized clinical trials do not necessarily apply to the results observed in everyday's 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." [22]
Outcomes not adequately studied in clinical trials
▪ "[…] limited data exist regarding the long-term outcomes of coronary stenting, as compared with standard CABG […] the 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." [23] *[CABG: Coronary artery bypass graft. DES: Drug eluting stent]*
▪ "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." [24]
▪ "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." [25]
Previous clinical trials inconclusive or inconsistent compared to other randomized or non-randomized 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." [26]
▪ "[…] 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. […] the 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, two 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." [27]
▪ "[…] the 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." [28] *[CAS: Carotid arterial stent]*
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. " [29] *[CABG: Coronary artery bypass graft. OPCABG: on-pump CABG]*
▪ "[...] women age 70 and older have been under represented in most breast cancer treatment trials. [...] this study did not address the benefit of any chemotherapy versus no chemotherapy in older women. [...] the 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." [30]
▪ "[...] A recent meta-analysis of seven randomized trials [...] demonstrated an improved clinical outcome among patients with unstable angina or NSTEMI [...] data 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." [31]

Table 4: Scientific Impact of Routinely Collected Data Studies

		Journal Impact factor 2012 (median, IQR)	Citations per year (median, IQR)	Total citations, (median, IQR)
All studies (n=337) *		4.5 (3.2;11)	4 (1.5;8.0)	22 (9;55)
not mentioning or citing RCT evidence (n=124) *		4.5 (3.5;11)	4.6 (2.5;9.3)	29 (13;61)
mentioning or citing RCT evidence (n=213) *		4.5 (3.2;11)	3 (1.3;7.5)	19 (8;49)
	p-value	0.21	0.013	0.013
RCT-related research motivation (n=213)*				
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	0.0005	0.002
Other outcomes	Yes	5 (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	0.13	0.01	0.06
Specific Population	Yes	5 (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	0.70	0.42	0.40
Inconclusive or inconsistent trial evidence	Yes	4.8 (2.9;9.1)	3 (1.3;6.6)	14 (6;38)
	No	4.1 (3.2;14)	3.1 (1.3;8)	19 (8;54)
	p-value	0.92	0.66	0.26
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	0.23	0.37	0.39
No RCT-related rationale	Yes	3.5 (3.2;6.2)	5 (2.2;11)	29 (18;48)
	No	4.5 (3.2;14)	2.9 (1.2;7.3)	18 (7;52)
	p-value	0.61	0.09	0.06

*) Data on number of citations was missing for 3 studies (n=334; with RCT evidence n=210, without n=124); there was no impact factor available for 6 studies (n=331; with RCT evidence n=208, without n=123)

Webappendix 1: Detailed inclusion criteria

Criterion	Description
IC 1: Patients	Any patient population was eligible.
IC 2: Intervention	The following medical interventions were eligible: drugs, biologics, dietary supplements, devices, diagnostic procedures, surgery, radiotherapy. Studies investigating structural or organizational interventions, or care management questions were not included (e. g. effects of rehabilitation, of specific service providers, of specialist consultation, of disease management programs et cetera).
IC 2: Comparator	We included the following active comparisons: drugs, biologics, dietary supplements, devices, diagnostic procedures, surgery, radiotherapy Comparisons with usual care / standard treatment were also included. We accepted comparisons of treatment variations such as different dosing schemes, ways of application, or timing of application when these variations were concretely defined (e.g. i.m. vs. i.v., continuous infusion vs. bolus infusion, initiation of drug treatment within a defined period of time after an event vs. later initiation, open vs. endoscopic surgery, on-pump vs. off-pump cardiac surgery). Comparisons of not clearly defined treatment variations, regimens, or concepts were not eligible (e.g. implementation of guidelines).
IC 3: Outcome	Studies reporting treatment effects on mortality were included. Any definition of mortality was accepted. Composite outcomes including mortality were not accepted. Studies with zero fatal events were excluded.
IC 4: Propensity scores	We included all studies reporting at least one treatment effect on mortality analyzed using approaches based on propensity scores.
IC 5: Data source	We included studies based only on data which is routinely collected for purposes of health care and not for the purpose of a specific study that is specified before data collection. We also included studies based on data from registries. The data used for the analyses must have been available in any kind of database or other electronically accessible form. Typical eligible data sources fulfilling this criterion are: databases of administrative and/or health care utilization data (such as claims or prescription data), data derived from registries, electronic health records or electronic patient records. Studies with treatments allocated by the investigator (experimental studies like RCTs) and post-hoc analyses of such studies were excluded. During title/abstract screening we accepted any description of the data source indicating (1) use of health care utilization data (i.e. by mentioning terms like “claims”, “prescription data”, “administrative data”, “reimbursement”, “insurance”, “routinely collected”); (2) use of registry data; (3) use of “records” (when any indication was that they were “electronic” or in any form digitalized or that information technology was used). In addition, any article describing the use of any kind of “database” as well as any otherwise eligible observational study which included more than 1000 participants has been scrutinized in full text.
IC 6: Language	Articles in English

Webappendix 2: Search Strategy for RCD-Studies

Search	Most Recent Queries
#50	Search #43 AND #47 AND #48

#49	Search #43 AND #47
#48	Search death*[tiab] OR dead*[tiab] OR die[tiab] OR died[tiab] OR dying[tiab] OR mortal*[tiab] OR fatal*[tiab] OR surviv*[tiab]
#47	Search #45 OR #46
#46	Search propensity[All fields]
#45	Search "Propensity Score"[mh]
#44	Search #38 AND #43
#43	Search #39 OR #40 OR #41 OR #42
#42	Search database*[tiab] OR "health care databases"[All fields] OR "healthcare databases"[All fields] OR "health care database"[All fields] OR "healthcare database"[All fields] OR "healthcare data"[All fields] OR "health care data"[All fields] OR "national database"[All fields]
#41	Search "Registries"[mh] OR register*[All fields] OR registr*[All fields]
#40	Search record*[All fields] OR electronic*[All fields] OR "Personal Health"[All fields] OR EMR[All fields] OR EHR[All fields] OR PHR[All fields] OR "Electronic Health Records"[mh] OR "Health Records, Personal"[mh] OR "Medical Record Linkage"[mh] OR "Medical Records Systems, Computerized"[mh]
#39	Search Administrative[All fields] OR claim*[All fields] OR "Insurance, Health, Reimbursement"[mh] OR reimbursement[All fields] OR utilization[All fields] OR utilisation[All fields] OR "Drug Utilization Review"[mh] OR Prescription*[All fields] OR Insurance[mh] OR Insur*[All fields] OR medicare[All fields] OR medicare[mh] OR medicaid[All fields] OR medicaid[mh] OR routine*[All fields] OR certificat*[All fields]
#38	Search #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37
#37	Search "national database"[All fields]
#36	Search "health care data"[All fields]
#35	Search "healthcare data"[All fields]
#34	Search "healthcare database"[All fields]
#33	Search "health care database"[All fields]
#32	Search "healthcare databases"[All fields]
#31	Search "health care databases"[All fields]
#30	Search database*[tiab]
#29	Search registr*[All fields]
#28	Search register*[All fields]
#27	Search "Registries"[mh]
#26	Search "Medical Records Systems, Computerized"[mh]
#25	Search "Medical Record Linkage"[mh]
#24	Search "Health Records, Personal"[mh]
#23	Search "Electronic Health Records"[mh]
#22	Search PHR[All fields]
#21	Search EHR[All fields]
#20	Search EMR[All fields]
#19	Search "Personal Health"[All fields]
#18	Search electronic*[All fields]
#17	Search record*[All fields]
#16	Search certificat*[All fields]
#15	Search routine*[All fields]

#14	Search medicaid[mh]
#13	Search medicaid[All fields]
#12	Search medicare[mh]
#11	Search medicare[All fields]
#10	Search Insur*[All fields]
#9	Search Insurance[mh]
#8	Search Prescription*[All fields]
#7	Search "Drug Utilization Review"[mh]
#6	Search utilisation[All fields]
#5	Search utilization[All fields]
#4	Search reimbursement[All fields]
#3	Search "Insurance, Health, Reimbursement"[mh]
#2	Search claim*[All fields]
#1	Search Administrative[All fields]

Search: Date: 23 Nov 2011; MEDLINE; Interface: PubMed

Webappendix 3: Search approach for RCTs

Databases and stepwise approach:
We searched systematic reviews or meta-analyses of pertinent RCTs in MEDLINE (via PubMed), the Cochrane Database of Systematic Reviews (CDSR) and Database of Abstracts of Reviews of Effects (DARE). We searched MEDLINE and the Cochrane Central Register of Controlled Trials (CENTRAL) for matching RCTs published after the time period covered by the newest pertinent evidence synthesis (without time restrictions when no synthesis was identified). For RCD-studies published >= 2008 we directly searched RCTs in MEDLINE (we found no additional matching RCT in any CENTRAL search). When we identified more than 3 potentially pertinent RCTs during abstract screening, we switched to the approach as for older RCD-studies to allow consideration of existing evidence syntheses.
Search terms:
Search terms for the intervention, comparator, and condition were combined. For identification of RCTs and for systematic reviews on topics with diagnostic interventions, we added terms for mortality (which were the same as used in the search strategy of the RCD-studies) to focus on studies reporting clinical outcomes and to increase specificity. We ensured that the terms of each individual search strategy would identify the respective RCD-study (in PubMed/MEDLINE) before we applied additional filters.
Filters:
We used PubMed's systematic review filter (the clinical queries subset "systematic[sb]") or when there were more than 50 hits, we used PubMed's more specific standard filter for meta-analyses. For RCTs, PubMed's standard RCT filter was used. No such filters were necessary for CENTRAL or CDSR/DARE. The English language filter was applied in PubMed.
Other:
We used the same search terms for CDSR and DARE or CENTRAL, respectively. Searches for related topics were conducted together. When we searched explicitly for effects in specific subpopulations, we perused the full texts of RCTs only when title or abstract indicated analyses for such subgroup.