Evidence reversals in primary care research: A study of randomized controlled trials

Running head

Evidence reversals in primary care research

Article category

Health Services Research

Authors

Christian Ruchon[a], Roland Grad[a], Mark H. Ebell[b,] David C. Slawson[c], Pierre Pluye[a], Kristian B. Filion[d], Mathieu Rousseau[a], Emelie Braschi[e], Soumya Sridhar[f], Anupriya Grover-Wenk[g], Jennifer Ren-Si Cheung[h], Allen F. Shaughnessy[h]

Affiliations

[a] Department of Family Medicine, McGill University, Montreal, Canada

[b] College of Public Health, University of Georgia, Athens, US

[c] Atrium Health, Charlotte, US

[d] Department of Medicine and of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Canada

[e] Department of Family Medicine, University of Ottawa, Ottawa, Canada

[f] Department of Family Medicine, University of Rochester Medical Center, Rochester, US [g] HCA Healthcare, Tufts University School of Medicine Family Medicine, Portsmouth, US [h] Tufts University School of Medicine and Cambridge Health Alliance, Boston, US

Correspondence

Mr. C. Ruchon; Department of Family Medicine, McGill University, 5858, chemin de la Côte-des-Neiges, Montreal, H3S 1Z1, Canada, christian.ruchon@mail.mcgill.ca

Key messages

- As reported elsewhere, Medical Reversals are not a rare phenomenon.
- In 408 trials relevant to primary care, we found a low rate of reversal.
- A low rate of reversal is good news for clinical decision-making.

Structured abstract

Background

Evidence-Based Medicine is built on the premise that clinicians can be more confident when their decisions are grounded in high quality evidence. Furthermore, evidence from studies involving patient-oriented outcomes is preferred when making decisions about tests or treatments. Ideally, the findings of relevant and valid trials should be stable over time, that is, unlikely to be reversed in subsequent research.

Objective

To evaluate the stability of evidence from trials relevant to primary healthcare and to identify study characteristics associated with their reversal.

Method

We studied synopses of randomized controlled trials (RCTs) published from 2002 to 2005 as "Daily POEMs" (Patient Oriented Evidence that Matters). The initial evidence (E_1) from these POEMs (2002 to 2005) was compared with the updated evidence (E_2) on that same topic in a summary resource (DynaMed 2019). Two physician-raters independently categorized each POEM-RCT as (1) reversed when $E_1 \neq E_2$, or as (2) not reversed, when $E_1 = E_2$. For all 'Evidence Reversals' ($E_1 \neq E_2$), we assessed the direction of change in the evidence.

Results

We evaluated 408 POEMs on RCTs. Of those, 35 (9%; 95% CI [6 to 12]) were identified as reversed, 359 (88%) were identified as not reversed, and 14 (3%) were indeterminate. On average, this represents about two evidence reversals per annum for POEMs about RCTs.

Conclusion

Over 12-17 years, 9% of RCTs summarized as POEMs are reversed. Information alerting services that apply strict criteria for relevance and validity of clinical information are likely to identify RCTs whose findings are stable over time.

Keywords

Contradicted findings; "Evidence-Based Medicine/trends"; evidence-based practice; evidence reversal; primary care; randomized controlled trials.

Lay summary

We studied the extent to which evidence from RCTs relevant to primary care is contradicted in subsequent research. When it was, we identified this event as an evidence reversal. In addition, we sought to identify characteristics of RCTs associated with their reversal.

From 408 RCTs published during the period 2002 to 2005, study characteristics such as sample size were identified and extracted. Subsequently, we compared the evidence reported in each of these RCTs with the evidence on that same topic in an online summary resource in 2019. This allowed us to classify each RCT in one of the following three categories; evidence confirmed, reversed or uncertain if this evidence is confirmed or reversed.

Over 12-17 years of follow up time, the findings of about 9 in ten RCTs summarized as POEMs are stable. We found no statistically significant associations between trial characteristics and their subsequent reversal. This low rate of evidence reversal is good news for the RCTs that are used to inform decision-making.

Introduction

Concerns about the reliability of evidence, especially in terms of its trustworthiness are nothing new.[1-4] Even high-quality randomized controlled trials (RCTs) supported by robust evidence can be reversed, further proving the fluidity of evidence.[5] For example, although aspirin (ASA) is widely prescribed for the primary prevention of cardiovascular disease, [6, 7] interpretations of the ARRIVE trial[8] (and other recently published RCTs) suggest this practice is no longer justified. This shift in the evidence is associated not with ASA itself, but with changed external factors, such as a reduction of the risk for cardiovascular disease in the general population.

In seminal work, Ioannidis identified research studies that were cited more than 1000 times and compared their results to subsequent studies that were either larger or conducted with a lower risk of bias.[9] In the subsequent studies, similar results were reported 44% of the time along with results that contradicted the earlier research 16% of the time. One-quarter (24%) of the original studies had not been repeated in the subsequent one to four years.

In the context of Internal Medicine practice, Prasad and colleagues reported 11-13% of original research articles concerning any medical practice and 24-46% of original studies on already adopted medical practices were subject to a reversal or shift in evidence of effect.[10-12] Prasad and colleagues then coined the term "medical reversal"[13] - when subsequent research such as that from a newer RCT presents findings to contradict a practice that had been adopted in the absence of good quality evidence.

In contrast to “medical reversal” is the broader concept of “evidence reversal”. In the latter, an initial claim derived from research-based evidence is subsequently contradicted (or reversed) in a newer research study deemed to be of higher quality.[14]

Objectives

We sought to evaluate the frequency of Evidence Reversal in the context of Family Medicine by scrutinizing RCTs, summarized as POEMs (Patient Oriented Evidence that Matters). In addition, we sought to identify the characteristics of RCTs associated with Evidence Reversal. To our knowledge, there are no studies of the reliability (or stability) of the findings of RCTs chosen for their relevance to primary care.

Methods

Study design

This was a study of RCTs summarized as POEMs to determine whether they were reversed in subsequent research.

Sampling

POEMs are summaries of newly published research that meet criteria for low risk of bias (validity) and demonstrate an impact on patient-level clinical outcomes (relevance), which can lead to a change in practice (importance). Each POEM consists of a title and a clinical question followed by a ‘bottom line’ statement. Following this statement is further information on study design, setting, and study findings.[15] Studies that become POEMs are found in a monthly scan of 102 journals.[16] Each month, about 25 POEMs are delivered to subscribers. Once delivered, each new POEM is included in the Essential Evidence resource for retrieval.[17] In 2020, Essential Evidence contained more than 6500 POEMs. From this

resource, in September 2017, we extracted all POEMs about RCTs published from 2002 to 2007 (n=960) (see Figure 1).

Of these 960 POEM-RCTs, we selected the oldest 408 entries, published between 2002 and 2007, to maximize our opportunity to detect the occurrence of a reversal.

Variables

Our main outcome of interest, an Evidence Reversal, was deemed to occur in two situations:

1. When an initial positive RCT result (e.g., in which one intervention was shown to be better than another) was contradicted in subsequent research by findings going from positive to negative;
2. When an initial negative RCT result (e.g., one intervention was no better than the other) was contradicted in subsequent research by findings going from negative to positive.

Initial RCT results (E_1) were contained in the summary statement of each POEM we scrutinized. Thus, the variable (E_1) defined the original evidence from 2002-2007. Then, one of us (CR) extracted updated evidence from DynaMed (a summary resource), in 2019. This updated evidence was termed E_2. In all situations when $E_1 \neq E_2$, an Evidence Reversal was identified.

To find all occurrences of Evidence Reversal, two raters independently compared E_1 and E_2. We recruited nine raters for this task. Raters had between 3 and 35 years of clinical experience in Medicine and Pharmacy. Disagreements as to the occurrence of an Evidence Reversal were resolved, when possible, by a third party (RG).

To train the raters, we conducted a pilot test with four raters and 10 POEMs. This pilot test revealed the need for a codebook of definitions for the concept of Evidence Reversal and its types. Further, we learned that raters needed E_2 presented to them as a summary of the evidence. This summary included whether the intervention described in E_1 was mentioned in DynaMed and if it was, whether DynaMed evidence (E_2) was consistent with E_1 in the opinion of the first author (CR).

Statistical analyses

POEMs were classified as reversed or not, then analyzed to identify characteristics associated with their reversal, using four statistical modeling approaches. These four approaches were a multiple logistic regression analysis, a least absolute shrinkage and selection operator, a classification tree, and a random forest analysis. For this analysis, we excluded POEMs whose Evidence Reversal status was classified as 'uncertain' (meaning raters could not decide if it was reversed) or 'cannot be resolved and not reversed' (meaning raters could not determine if the intervention was reversed, e.g. when the drug was removed from the market after the publication of E_1).

Several variables were transformed to facilitate the interpretation of model outputs. Total sample size and the sample size of the intervention group were combined into a single variable, the sample size ratio. The rationale for the sample size ratio was to facilitate the interpretation of the output of statistical models. The higher the ratio, the closer the size of the intervention group to the total sample size. Sample size was divided into 4 categories, informed by the quartiles of the distribution of this variable: 0 to 99, 100 to 249, 250 to 499, and 500 to 39,999 participants. The number of trial arms was summarized in 3 groups: two-

arm trial; three-arm trial; and trials with more than three arms. Finally, a 'Level of evidence" assigned to each POEM-RCT in line with the Oxford Centre for Evidence-Based Medicine rating scale was transformed into a binary variable: (1) 1b and 1b-; or (2) 2b, 2b- and 2c.

Results

Of the 408 double-blind POEM-RCTs that we assessed, published from 2002 to 2005, we found 35 occurrences (9%; 95% CI 6 - 12%) of an Evidence Reversal (Figure 1). The characteristics of these 408 RCTs are summarized in Table 1. Most RCTs studied an adult population (76%) in an outpatient setting (74%). These RCTs used a parallel design with two-arms (74%); three-arms (11%), or four-arms (11%). In our statistical modeling, we found no relationship between groups based on reversal status and the index study in terms of level of evidence, sample size, or use of concealed allocation (see supplementary material).

Of the 35 reversed POEM-RCTs, 31 (89%) studied a drug treatment while 4 (11%) studied devices. Observing 35 evidence reversals over 17-years of follow up time represents a rate of about 2 reversals per year for these primary care relevant RCTs. Eighteen of 35 reversals failed to confirm the superiority of the intervention as demonstrated in the index study (i.e., direction of effect from positive to negative). Another 17 reversals were in the opposite direction, where one treatment was later found to be superior to the other in a subsequent RCT (negative to positive) (Figure 3).

In total, 14 POEM-RCTs (3%) were rated as 'not reversed and cannot be resolved or uncertain' (Figure 2). For example, one of these involved a drug which was subsequently withdrawn from the market and therefore could not be re-evaluated for any reversal of effect.[18]

As an example of one Evidence Reversal, we offer the following. In 2003, a double-blind placebo-controlled trial of dexamethasone 0.6 mg/kg in children (n=184) aged 5 to 16 years in the emergency department with acute pharyngitis found no clinically important effect for the outcome of time to onset of pain relief.[19] In 2009, evidence from a systematic review and meta-analysis suggested a reversal with respect to the effect of dexamethasone. This updated evidence included eight trials, and 369 children.[20] For the outcome of time to onset of pain relief, this occurred on average 6.3 hours earlier with corticosteroids than without. In the supplementary material, we list all 35 Evidence Reversals.

Discussion

In a consecutive sample of 408 RCTs summarized as POEMs, 9% were reversed in subsequent research when scrutinized from 12 to 17 years later. In other words, RCTs with good internal validity, focusing on relevant and important outcomes for primary care produce findings that are relatively stable over time.

Of the Evidence Reversals we identified, one-half suggested a practice should be stopped, as the change in direction of effect went from positive (in favour of a practice) to negative (against that practice). We found 18 reversals of this type, for an estimated rate of one POEM-RCT per year among the 250 or so POEMs published annually. This finding supports physicians who wish to implement a new intervention in their practice, even when this intervention is supported by one RCT summarized as a POEM. These findings also have implications for editors of knowledge resources. As an updating task, editors should consider flagging studies in their summary resource that have been identified as reversed. In addition, physicians should be aware of the phenomenon of Evidence Reversal, as they attempt to

make sense of new evidence that contradicts the findings of earlier research. In the same vein, teachers of evidence-based medicine may want to update their curricula to raise awareness of this phenomenon.

That just 9% of 408 POEM-RCTs were reversed in our study should be considered in light of the findings of others. For example, Prasad found that 24-46% of original studies on already adopted medical practices were reversed over time. There are two differences between our work and that of Prasad. First, POEM-RCTs are selected after an assessment of their validity and relevance using established criteria.[15] For example, POEM synopses on hypertension must include studies in which outcomes were patient-oriented, such as effects on mortality or morbidity. Second, Prasad studied the reversal of medical practices ('Medical Reversal') which had been implemented in the absence of high-quality evidence.

A recent editorial in this journal defined meta-research as a new discipline that aims to understand what makes research trustworthy and what can be done to strengthen both research methods and the evidence they generate.[21] More specifically, the authors alluded to the importance of subjecting RCTs to empirical evaluation and improvement. As pillars of evidence, RCTs are considered the best test of the effect of a new intervention.[22, 23] For this reason, we conducted this empirical evaluation of RCTs summarized as POEMs for primary care.

Limitations

For reasons of feasibility, we analyzed the first 408 POEMs in our data set. It is unclear whether random sampling of all POEMs would have resulted in a different rate of reversal. However, the ability to identify an Evidence Reversal likely increases with time, and we

evaluated the earliest POEM-RCTs in our sample. According to Donald T. Campbell's evolutionary perspective on science, evidence is evolving over time, and reproducing this study at a later time with the same subset of POEM-RCTs may result in slightly different estimates of reversal.[24] In the same vein, we did not distinguish POEM-RCTs that were not reversed (when E_2 confirmed E_1) from POEM-RCTs on topics where new evidence has not yet emerged (no E_2).

Finally, we cannot say whether any single reversal was due to the particular characteristics of interventions tested in that RCT, given the limited number of reversals we identified. Innovation in science and technology create external factors that affect the outcomes of clinical research, unrelated to trial design. For example, for decades ASA was recommended for the primary prevention of cardiovascular disease. Subsequently, we observed a decline in the population risk for cardiovascular disease due to external factors such as a reduced prevalence of smoking. Concurrent to this, we see an Evidence Reversal with respect to the use of ASA in primary prevention, as the gastrointestinal harms are now perceived to outweigh the potential to prevent cardiovascular events.[25] In future research, it would be of interest to develop and test a model to predict the probability of Evidence Reversal. Such a tool could help to improve healthcare delivery and medical education. Indeed, if a clinician knew the probability of reversal associated with any single RCT, then s/he could consider this issue as a metric of uncertainty in a shared decision-making context.

Conclusion

Findings of RCTs fitting criteria for relevance and validity of clinical information have a high likelihood of being stable over time. Information alerting services that apply strict criteria for

relevance and validity of clinical information are likely to identify RCTs whose findings are stable over time.

Declaration

Ethical approval: none

Funding: Joule Inc. (a subsidiary of the Canadian Medical Association) provided a grant to Roland Grad at McGill University in support of this work. These funds provided a graduate student scholarship to Christian Ruchon. Dr. Filion is supported by a senior salary support award from the *Fonds de recherche du Québec – santé* and a William Dawson Scholar award from McGill University.

Conflict of interest: Drs. Ebell, Slawson, and Shaughnessy are paid as editorial consultants by Wiley-Blackwell to write POEMs.

Data availability

The data underlying this article is available by request to the corresponding author.

References

1. Ioannidis JP. Evidence-based medicine has been hijacked: a report to David Sackett. *J Clin Epidemiol*. May 2016;73:82-6. doi:10.1016/j.jclinepi.2016.02.012
2. Cosgrove L, Vannoy S, Mintzes B, Shaughnessy AF. Under the Influence: The Interplay among Industry, Publishing, and Drug Regulation. *Account Res*. 2016;23(5):257-79. doi:10.1080/08989621.2016.1153971
3. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ*. 2003;326(7400):1167. doi:10.1136/bmj.326.7400.1167
4. Every-Palmer S, Howick J. How evidence-based medicine is failing due to biased trials and selective publication. *J Eval Clin Pract*. Dec 2014;20(6):908-14. doi:10.1111/jep.12147
5. Greene P, Prasad V, Cifu A. Should Evidence Come with an Expiration Date? journal article. *Journal of General Internal Medicine*. May 06 2019;doi:10.1007/s11606-019-05032-4
6. Final Report on the Aspirin Component of the Ongoing Physicians' Health Study. *New England Journal of Medicine*. 1989;321(3):129-135. doi:10.1056/nejm198907203210301
7. Sanmuganathan PS, Ghahramani P, Jackson PR, Wallis EJ, Ramsay LE. Aspirin for primary prevention of coronary heart disease: safety and absolute benefit related to coronary risk derived from meta-analysis of randomised trials. *Heart*. Mar 2001;85(3):265-71. doi:10.1136/heart.85.3.265
8. Grobman WA, Rice MM, Reddy UM, et al. Labor Induction versus Expectant Management in Low-Risk Nulliparous Women. *The New England journal of medicine*. Aug 9 2018;379(6):513-523. doi:10.1056/NEJMoa1800566
9. Ioannidis JP. Contradicted and initially stronger effects in highly cited clinical research. *JAMA*. 2005;294doi:10.1001/jama.294.2.218
10. Prasad V, Gall V, Cifu A. The frequency of medical reversal. *Archives of internal medicine*. Oct 10 2011;171(18):1675-6. doi:10.1001/archinternmed.2011.295
11. Prasad V, Vandross A, Toomey C. A decade of reversal: an analysis of 146 contradicted medical practices. *Mayo Clinic proceedings*. 2013;88doi:10.1016/j.mayocp.2013.05.012
12. Herrera-Perez D, Haslam A, Crain T, et al. A comprehensive review of randomized clinical trials in three medical journals reveals 396 medical reversals. *eLife*. 2019/06/11 2019;8:e45183. doi:10.7554/eLife.45183
13. Prasad V, Cifu A. Medical reversal: why we must raise the bar before adopting new technologies. *The Yale journal of biology and medicine*. Dec 2011;84(4):471-8.
14. Sutton D, Qureshi R, Martin J. Evidence reversal - when new evidence contradicts current claims: a systematic overview review of definitions and terms. *J Clin Epidemiol*. 2018;94:76-84. doi:10.1016/j.jclinepi.2017.10.004
15. Grad R, Pluye P, Tang D, Shulha M, Slawson DC, Shaughnessy AF. Patient-oriented evidence that matters (POEMs) suggest potential clinical topics for the Choosing Wisely campaign. *J Am Board Fam Med*. Mar-Apr 2015;28(2):184-9. doi:10.3122/jabfm.2015.02.140226
16. Essential evidence plus. Wiley InterScience. https://www.essentialevidenceplus.com/content/poems
17. Smith R. A POEM a week for the BMJ. *BMJ (Clinical research ed)*. 2002;325(7371):983-983. doi:10.1136/bmj.325.7371.983

18. Lebwohl M, Tyring SK, Hamilton TK, et al. A novel targeted T-cell modulator, efalizumab, for plaque psoriasis. *The New England journal of medicine*. Nov 20 2003;349(21):2004-13. doi:10.1056/NEJMoa030002
19. Bulloch B, Kabani A, Tenenbein M. Oral dexamethasone for the treatment of pain in children with acute pharyngitis: a randomized, double-blind, placebo-controlled trial. *Ann Emerg Med*. 2003;41(5):601-608.
20. Hayward G, Thompson M, Heneghan C, Perera R, Del Mar C, Glasziou P. Corticosteroids for pain relief in sore throat: systematic review and meta-analysis. *BMJ*. 2009;339
21. Tatsioni A, Ioannidis J. Meta-research: bird's eye views of primary care research. *Fam Pract*. 2020;
22. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343doi:10.1136/bmj.d5928
23. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg*. 2011;128(1):305-310. doi:10.1097/PRS.0b013e318219c171
24. Cziko GA, Campbell DT. Comprehensive evolutionary epistemology bibliography. *J Soc Biol Struct*. 1990/01/01/ 1990;13(1):41-82. doi:https://doi.org/10.1016/0140-1750(90)90033-3
25. Moriarty F, Ebell MH. A comparison of contemporary versus older studies of aspirin for primary prevention. *Fam Pract*. Jul 23 2020;37(3):290-296. doi:10.1093/fampra/cmz080

Tables

Table 1. Characteristics of 394 Patient Oriented Evidence that Matters - Randomized Controlled Trials (POEM-RCTs) from 2002 to 2005*

	Reversed	Not Reversed
POEM-RCT	35	359
	N (%)	N (%)
Publication Year		
2002	2 (6%)	112 (31%)
2003	12 (34%)	101 (28%)
2004	11 (31%)	77 (21%)
2005	10 (29%)	69 (19%)
RCT characteristics		
Total Sample Size		
Mean	1831	2417
Standard Deviation	6714	5809
Median	275	326
Range	[39 ; 39,876]	[12 ; 39,876]
Intervention Group Size		
Mean	870	1084
Standard Deviation	3354	2642
Median	122	139
Range	[13 ; 19,934]	[6 ; 19,937]
Setting		
Outpatient	26 (74%)	260 (72%)
Inpatient	3 (9%)	60 (17%)
Emergency department	3 (9%)	14 (4%)
Population-based	3 (9%)	18 (5%)
Other	0	7 (2%)
Age group		
Adults	27 (77%)	272 (76%)
Children	5 (14%)	39 (11%)
Both adults and children	3 (9%)	48 (13%)
Allocation Concealment		
Concealed	25 (71%)	229 (64%)
Uncertain	10 (29%)	130 (36%)

*Excluding 14 POEMs: 11 where Evidence Reversal was uncertain and 3 that were not reversed and cannot be resolved

Figures

Figure 1. Flow Chart – Selection of POEM-RCTs for analysis of evidence reversal

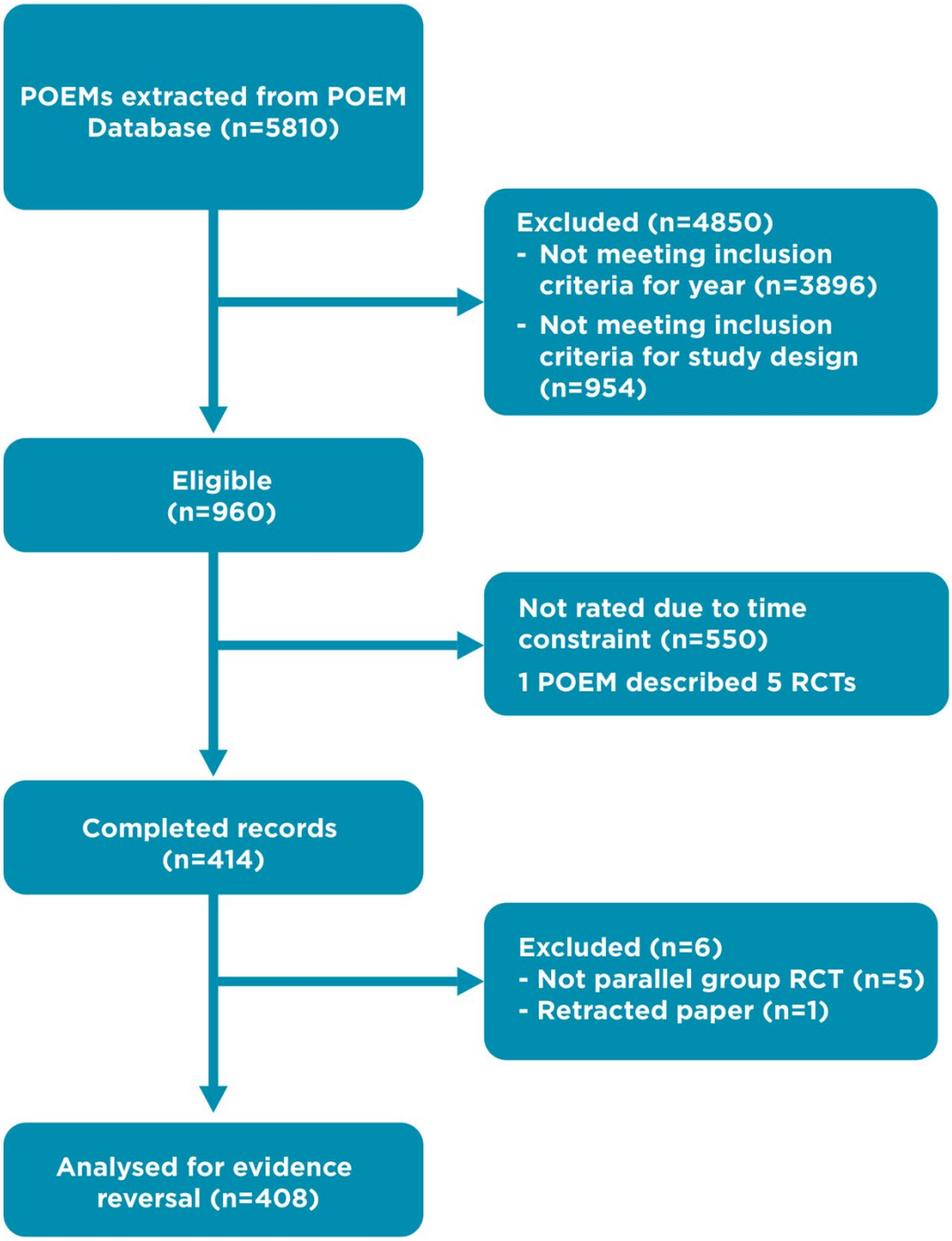

Figure 2. Distribution of disagreements per group

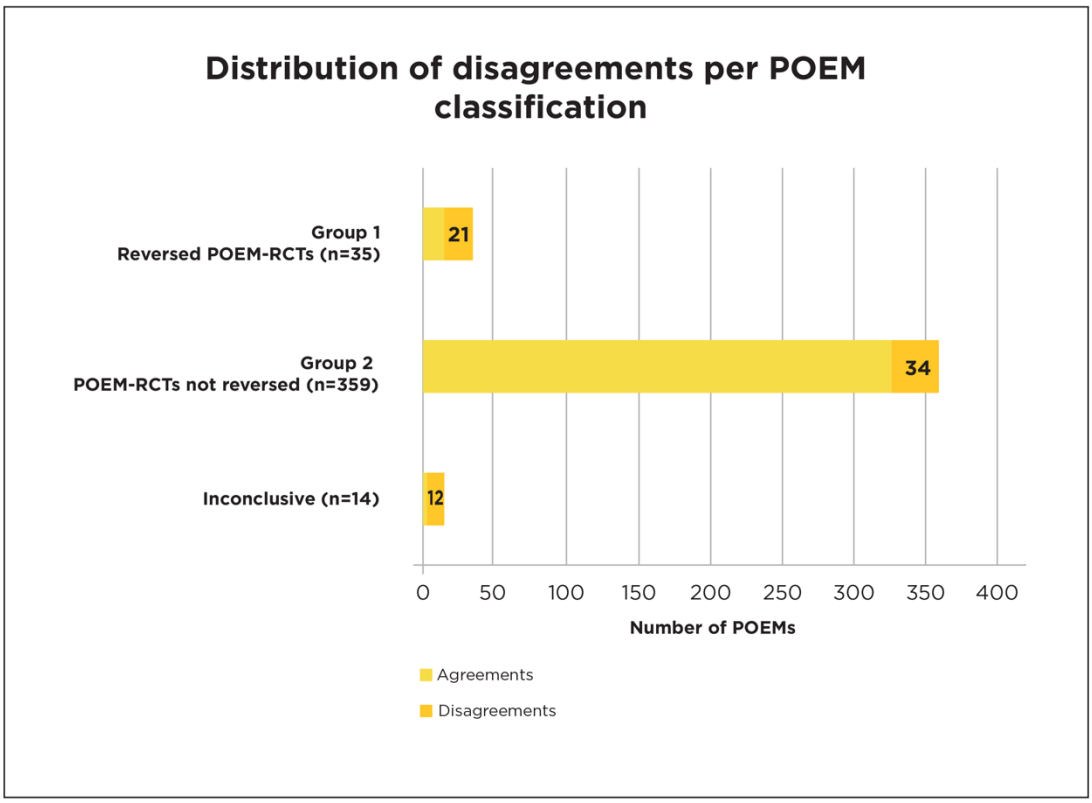

Figure 3. Direction of shift in the evidence