Original Investigation

Reanalyses of Randomized Clinical Trial Data

Shanil Ebrahim, PhD; Zahra N. Sohani, MSc; Luis Montoya, DDS; Arnav Agarwal, BSc; Kristian Thorlund, PhD; Edward J. Mills, PhD; John P. A. Ioannidis, MD, DSc

IMPORTANCE Reanalyses of randomized clinical trial (RCT) data may help the scientific community assess the validity of reported trial results.

OBJECTIVES To identify published reanalyses of RCT data, to characterize methodological and other differences between the original trial and reanalysis, to evaluate the independence of authors performing the reanalyses, and to assess whether the reanalysis changed interpretations from the original article about the types or numbers of patients who should be treated.

DESIGN We completed an electronic search of MEDLINE from inception to March 9, 2014, to identify all published studies that completed a reanalysis of individual patient data from previously published RCTs addressing the same hypothesis as the original RCT. Four data extractors independently screened articles and extracted data.

MAIN OUTCOMES AND MEASURES Changes in direction and magnitude of treatment effect, statistical significance, and interpretation about the types or numbers of patients who should be treated.

RESULTS We identified 37 eligible reanalyses in 36 published articles, 5 of which were performed by entirely independent authors (2 based on publicly available data and 2 on data that were provided on request; data availability was unclear for 1). Reanalyses differed most commonly in statistical or analytical approaches (n = 18) and in definitions or measurements of the outcome of interest (n = 12). Four reanalyses changed the direction and 2 changed the magnitude of treatment effect, whereas 4 led to changes in statistical significance of findings. Thirteen reanalyses (35%) led to interpretations different from that of the original article, 3 (8%) showing that different patients should be treated; 1 (3%), that fewer patients should be treated; and 9 (24%), that more patients should be treated.

CONCLUSIONS AND RELEVANCE A small number of reanalyses of RCTs have been published to date. Only a few were conducted by entirely independent authors. Thirty-five percent of published reanalyses led to changes in findings that implied conclusions different from those of the original article about the types and number of patients who should be treated.

JAMA. 2014;312(10):1024-1032. doi:10.1001/jama.2014.9646

Editorial page 1002

Supplemental content at jama.com

Author Affiliations: Stanford Prevention Research Center, Department of Medicine, Stanford University, Stanford, California (Ebrahim, Thorlund, Mills, Ioannidis); Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada (Ebrahim, Sohani, Thorlund, Mills); Department of Anesthesia, McMaster University (Ebrahim); Department of Anaesthesia and Pain Medicine, Hospital for Sick Children, Toronto, Ontario, Canada (Ebrahim); Population Genomics Program, McMaster University (Sohani); University Health Network, Toronto, Ontario, Canada (Montoya); Faculty of Health Sciences, McMaster University (Agarwal); Faculty of Health Sciences, University of Ottawa, Ottawa, Ontario, Canada (Mills); Department of Health Research and Policy, Stanford University School of Medicine (Ioannidis); Department of Statistics, Stanford University School of Humanities and Sciences (Ioannidis); Meta-Research Innovation Center at Stanford (METRICS), Stanford University (Ioannidis).

Corresponding Author: John P. A. Ioannidis, MD, DSc, 1265 Welch Rd, Medical School Office Bldg, Room X306, Stanford, CA 94305 (jioannid@stanford.edu).

Whether trial investigators should be required to make patient data from randomized clinical trials (RCTs) available for reanalysis is controversial.[1-5] Since reanalyses of raw data from oseltamivir trials led to conclusions different from those in the original trials and subsequent meta-analyses, some authors have argued that a standard of data sharing and reanalysis should be more widely adopted and could have major consequences for individual and public health, and that paying consumers (the public) should have access to complete information about drugs and devices.[6]

Arguments against accessibility to raw data and reanalyses include potential risk to trial patient confidentiality[7]; inappropriate dredging of data sets, resulting in spurious findings[6]; release of commercially sensitive information[6]; the requirement for a data infrastructure for sharing data and reanalysis[8]; and "rogue" reanalysis by nonexperts or by analysts who have conflicts of interest, as in the case of the Methane Awareness Resource Group Diesel Coalition that tried to thwart a study showing association of diesel exhaust with cancer outcomes via multiple requests for raw data for reanalysis.[9]

In this study, we identified published articles that reported reanalyses of patient-level data from RCTs testing the same hypothesis as the original article. We evaluated the authorship of the reanalyses, how the findings compare with

those from the original analysis, and whether the reanalysis could modify interpretations from the original article about which patients should be treated.

Methods

Search

We searched MEDLINE from inception to March 9, 2014, for articles reporting reanalyses of previously published RCTs, in which reanalysis was defined as testing of an identical hypothesis (eg, identical population, intervention, comparator, outcome, study design).

We identified articles by using a combination of relevant MeSH terms: (replicat*[title] OR reproduc*[title] OR reapprais*[title] OR re-apprais*[title] OR re-evaluat*[title] OR reevaluat*[title] OR re-assess*[title] OR reassess*[title] OR revis*[title]) OR (re-analysis OR reanalysis OR reanalyzed OR re-analyzed) NOT reproductiv*. We limited articles to English language and clinical trials and excluded meta-analyses that used patient-level data.

Study Selection and Data Extraction

We screened titles and abstracts of identified citations to flag potentially eligible studies and read the full text to identify those that met the eligibility criteria. We developed items for data extraction after review of a random sample of 10 potentially eligible studies. We completed extraction exercises for all data items by using 6 articles until 100% consensus was achieved between 4 extractors, and the remaining articles were then divided among the extractors and 1 author verified all the extracted data.

We extracted information about trial characteristics (participants' disease/condition, intervention and comparators, and definition or measurement of primary outcome), authorship (countries and overlap in authors or in research group/consortium affiliations), and analyses (differences in methods used, handling of missing data, use of intention-to-treat principle, and whether reanalysis authors identified any errors in the original data set or analysis); 1 or more type of reanalysis may have been performed for the same article. We also assessed public availability of patient-level data, and for reanalyses performed by no authors from the original article or its team or consortium, we contacted the corresponding author of the reanalysis article to clarify whether patient-level data from the RCT were publicly available or whether the authors had to undertake an approval process to obtain the individual patient data, whenever this information was not discernible from the reanalysis article.

We categorized differences between original trial and reanalysis findings as changes in direction and magnitude of treatment effect, changes in statistical significance, and changes that could lead to differences in interpretation about the types or numbers of patients who should be treated with the active intervention, or the newer or more experimental intervention when 2 active interventions were compared.

Statistical Analysis

We describe data as proportions and means or medians as appropriate. To assess changes in magnitude of treatment effect (defined as nonoverlap of CIs with reanalysis), we extracted effect estimates and associated 95% CIs, calculated within-trial differences in treatment effect expressed as standardized mean differences for articles reporting continuous outcomes and relative risks (risk ratios, odds ratios [ORs], risk differences, or hazard ratios [calculated from dividing the median survival times of relevant arms]) for articles reporting dichotomous outcomes, and evaluated whether the CIs overlapped. We used Fisher exact test to compare proportions of reanalyses that recommended a change in the number of patients who should be treated by authorship (overlapping vs independent authors).

We used a significance threshold of .05 and reported 2-tailed *P* values for this comparison.

We completed all analyses with SPSS version 22.0.0. We did not require approval from a research ethics board because we were using summary data from published trials and reanalyses and did not require individual patient data.

Results

Search Results

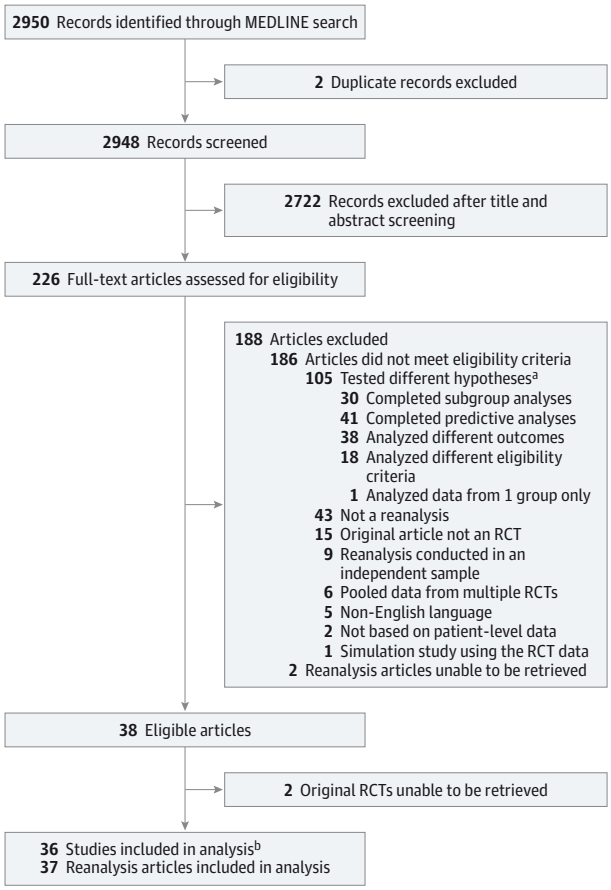
We identified 2950 citations in an initial search and screened 2948 for eligibility (**Figure**). We assessed the full text of 226 and excluded 186. We were unable to retrieve 2 articles to determine eligibility[10,11] and were unable to retrieve the original trials[10,11] for 2 reanalyses.[12,13] Our final sample comprised 36 articles (Appendix A in the Supplement), and because one article included 2 separate reanalyses of different articles,[14] our evaluation is based on 37 eligible reanalyses (Figure).

Study Characteristics

Table 1 summarizes the characteristics of the studies (Appendix B in the Supplement). Thirty-one reanalyses (84%) had an overlap of at least 1 author and 32 (86%) were published by the same research group or consortium as the original article. Of the 5 reanalyses from authors entirely independent from those in the original article, individual patient data were publicly available for 2, authors of 2 reanalyses sought and received approval from the original authors, research group, or consortium, and we could not clarify data availability for 1 reanalysis. In 3 instances, patient-level data availability was unclear and we needed to contact the corresponding author of reanalyses to clarify whether those data were publicly available or whether the authors had to undertake an approval process to obtain them.

Twelve of the original RCTs (32%) were published in general medical journals. Conversely, only 3 of the reanalyses (8%) were published in general medical journals (Table 1, Appendix C in the Supplement). Most original studies and reanalyses were completed by authors from Europe (n = 23 and n = 22, respectively) and the United States (n = 19 and n = 18, respectively). The median time between publication of the original

Figure. Systematic Search and Identification of Eligible Randomized Clinical Trials (RCTs)



[a] Articles included 1 or more different reasons for testing different hypotheses.

[b] One article included 2 separate reanalyses of a different article.

trial and its partner reanalysis was 48 months (interquartile range, 23-98 months).

Differences in Methods

There were 46 differences in methods identified in the 37 reanalyses (differences in statistical or analytical methods [n = 18], definition or measurement of outcomes [n = 12], approaches to handling missing data [n = 8], use of intention to treat [n = 2], and other [n = 6]) (**Table 2**); numbers were comparable when counting what authors identified as the most important difference per reanalysis (statistical or analytical methods [n = 17], definition or measurement of outcomes [n = 11], approaches to handling missing data [n = 2], use of intention to treat [n = 1], and other [n = 6]).

Four reanalyses addressed errors in the data set or analysis of the original article (1 article reporting 2 reanalyses excluded patients who should have been ineligible in the original article,[14] 1 reanalysis identified misclassified cases in the original article caused by errors at collection of clinical data and by lack of blood sample validation,[15] and 1 reanalysis reported a misinterpretation of findings based on assumptions of the original analysis[16]). All errors were identified by authors from the same group.

Differences in Findings

Fifteen reanalyses (41%) reported only *P* values without treatment effect sizes, only treatment effects without *P* values or measures of precision, or effects in units not comparable to those in the original analyses (Appendix Table D in the Supplement). Of 42 comparisons reported in the remaining 22 reanalyses, the direction of treatment effect in the original and reanalysis was the same in 38 cases and different in 4 (in 2 reanalyses a previously non-null treatment effect became null,[17-19] in 1 reanalysis a previously null effect became non-null,[20] and in 1 the direction of treatment effect was on the opposite side of the null compared with that in the original[21]).

We were able to calculate standardized effect differences for 32 comparisons in 22 study pairs (Appendix D in the Supplement) to assess changes with reanalysis in magnitude of treatment effect and found 2 reanalyses with nonoverlapping CIs compared with those of the original trials.

One reanalysis of a trial comparing Holter monitoring with electrophysiologic testing used 2 revised criteria for predicting drug efficacy; there was a decrease in drug efficacy prediction (OR, 0.59; 95% CI, 0.41-0.85) with one of the criteria in the reanalysis compared with that in the original trial (OR, 0.24; 95% CI, 0.16-0.36).[22]

A second reanalysis motivated by changes in erythropoiesis-stimulating agent recommendations for the hemoglobin level at which to initiate therapy with the agent showed a decrease in benefit of fixed-dose darbepoetin alfa every 3 weeks vs weight-based weekly dosing, using a threshold hemoglobin level of less than 10 g/dL (OR, 0.88; 95% CI, 0.85-0.88) compared with the original trial hemoglobin threshold of less than 11 g/dL (OR, 0.77; 95% CI, 0.76-0.80).[23]

Two reanalyses showed a loss in statistical significance[14,24] and 2 showed a gain.[25,26]

Thirteen reanalyses (35%) reported a change in findings that implied a difference in interpretation about who should be treated (**Table 3**). Eight of the 13 changes in interpretation were accompanied by changes in direction of effect or in gain or loss of nominal statistical significance and 5 by changes in size of the treatment effect.

Three studies (8%) implied that different patients should be treated because there was a change in the understanding of the reasons of benefit or the types of patients benefiting more from the treatment. For example, a treatment trial for esophageal varices showed a reduction in mortality but not rebleeding with sclerotherapy, with proportional hazard modeling, whereas its reanalysis suggested a reduction in rebleeding but not mortality, based on a multistage competing risk model.[28] One reanalysis (3%) concluding that fewer patients should be treated reversed the conclusion that homeopathic treatment was effective for fibrositis by disaggregating a composite end point comprising pain and sleep.[24] Nine reanalyses (24%) were interpreted as showing that more patients should be treated. For example, a trial comparing mycophenolate mofetil and azathioprine in heart transplant patients showed no difference be-

tween treatments at preventing growth of intravascular ultrasonographically measured intimal medial thickness, whereas its reanalysis suggested superiority of mycophenolate mofetil when data were matched by site.[20]

Reanalyses by Different Authors

Only 5 reanalyses (13.5%) were performed by completely independent authors.[24,26,31-33] Three of the 5 used different analytical methods,[26,32,33] 1 considered the original analysis of a crossover RCT invalid and reanalyzed the first treatment period only,[24] and 1 used a different definition for the primary outcome.[31] Two of these 5 independent reanalyses did not change the original trial interpretation,[32,33] 2 suggested that more patients should be treated,[26,31] and 1 suggested that fewer patients should be treated compared with the original article.[24] We found no statistically significant difference in proportion of reanalyses leading to different conclusions about who should be treated when reanalyses were performed by overlapping vs independent authors (10/32 vs 3/5; OR, 0.30; 95% CI, 0.04-2.11; P = .32).

Discussion

In this review, we identified 37 reanalyses of patient-level data from previously published RCTs (reported in 36 articles). Most of the reanalyses were completed by authors involved in the original trial, and most assessed the effect of different analytical methods or a change of outcome definition on the trial's estimate of effect. Original RCT data sets were publicly available in 2 of 5 instances when trial data were reanalyzed by independent authors. Five of 42 comparisons in reanalyses resulted in a change in treatment effect, 2 reanalyses resulted in a loss of statistical significance, and 2 resulted in a decrease in estimate of treatment effect. Approximately a third (35%) of the published reanalyses led to changes in findings that implied conclusions different from those of the original article about the types and number of patients who should be treated. We performed a search of MEDLINE from inception to June 19, 2014, using a combination of search terms and review, and found no studies evaluating reanalyses of RCTs. Thus, we believe that our study represents the first empirical evaluation of reanalyses of RCT data and of changes emerging from those reanalyses.

The uncovering of distortion and bias in the reporting of trials of rofecoxib and oseltamivir explains why efforts to improve access to RCT data have attracted substantial interest from researchers, regulators, funders, and pharmaceutical companies.[4] Reproducibility is an important step to ensure that the findings of original trials are not distorted, biased, or incomplete.[4] Evidence is limited in the current biomedical literature about whether the results of RCTs can be reproduced by independent analysts, perhaps partly because lack of publicly available data sets prevents reanalyses by independent authors.

As a result, there have been increasing calls within the medical community for access to raw data from published trials so that reanalyses can take place. Although some large companies have committed to making trial data available, there is no consensus about the optimal data format, what data ought to be shared, and who can access them and when.[4] There have been emerging data-sharing initiatives, including the Yale University Open Data Access project,[34] the National Institutes of Health Data Sharing Requirements, and the International Stroke Trial.[35] Some journals have also created policies that support public accessibility of data and protocols as prerequisites of publication for some types of research.[36] Furthermore, the Office of Science and Technology Policy recently released a memo to develop a plan to support public access to data.[8,37] However, a standard of data nonsharing by investigators still remains common.[38-40]

Involving authors of the original article in reanalyses may be a condition for providing access to data and may ensure that direct knowledge of study nuances is accounted for in a reanalysis. Involving such authors might also limit the independence of any coauthors to refute initial results if

Table 1. Study Characteristics

Variables	No. (%)	
	Original Articles (n = 37)	Reanalyses (n = 37)
Same authors[a]		
Yes		32 (86)
No		5 (14)
Data availability		
Not publicly available		
Reanalysis authors were in the same group/consortium/research organization		32 (86)
Reanalysis authors received approval from the original authors		2 (5)
Publicly available		2 (5)
Not reported		1 (3)
Publication date		
Pre-1991	10 (27)	6 (16)
1991-2000	13 (35)	10 (27)
2001-2005	8 (22)	5 (14)
2006-2010	5 (14)	17 (30)
2011-2013	1 (3)	5 (14)
General medical journal		
Yes[b]	12 (32)	4 (11)
No[c]	25 (68)	33 (89)
Country(ies)[d]		
Europe	23 (62)	22 (59)
United States	19 (51)	18 (49)
Canada	8 (22)	6 (16)
United Kingdom	6 (16)	9 (24)
Oceania	5 (14)	4 (11)
Asia	3 (8)	4 (11)
Africa	1 (3)	1 (3)
South America	1 (3)	0

[a] Original trial and reanalysis authors from the same group/consortium/research organization.

[b] *BMJ*, *JAMA*, *Lancet*, *New England Journal of Medicine*.

[c] Journals provided in Appendix Table C.

[d] Numbers exceed n = 37 because studies may have >1 country of affiliation.

the original authors have commitments to their findings. Reanalyses by independent authors might obviate those conflicts but be equally problematic if they have competing interests. For example, a trial comparing acupuncture and amitriptyline in human immunodeficiency virus-infected patients reported no effect for either intervention on

Table 2. Differences in Methods Used in the Reanalysis

Differences Cited in the Reanalysis	No. (%)				
		Did the Reanalysis Modify Inferences of the Original Trial?			
	Reanalyses (n = 37)[a]	No (n = 29)	Treat Different Patients (n = 3)	Treat More Patients (n = 13)[b]	Treat Fewer Patients (n = 1)
Differences in statistical or other analytical methods	18 (48.6)	11 (61)	3 (17)	3 (17)	1 (5.5)
Nonparametric statistical technique	1	1			
Separation of composite end points for analysis	1				1
Measure of clinical significance to confirm original findings	2	2			
Informative censoring approach	3	3			
Competing risks model	1		1		
Nonlinear model	2	1	1		
Triangular and restricted sequential design	1	1			
Multivariate techniques	1	1			
Matched site-to-site image analysis between trial centers	1			1	
Linear transformation of scores	1	1			
Adjustment for confounders[c]	1	1			
Bayesian methods	1			1	
Additional Poisson models	1		1		
Wilcoxon and Mann-Whitney *U* tests to compare treatment groups	1			1	
Differences in the definition or measurement of same outcome	12 (32.4)	6 (50)		6 (50)	
Computer-assisted method for measurement of outcome	1			1	
New criteria for the assessment of outcome[d]	7	4		3	
Use of rate of change of the outcome as end point	1			1	
Different measurement to assess the same construct	3	2		1	
Differences in the handling of missing data	8 (21.6)	5 (63)		3 (37)	
Single imputation (baseline or last observation carried forward)[e]	3	3			
Multiple imputation[d]	2	1		1	
Use of associations between predictor and outcome for imputations[c]	1	1			
Excluded patients in reanalysis[c]	2			2	
Differences in the intention-to-treat or on-treatment principle	2 (5.4)	2 (100)			
Original without ITT; reanalysis with ITT[c]	1	1			
Original with modified ITT; reanalysis with standard ITT	1	1			
Differences in any other aspect of the analysis or methods	6 (16.2)	5 (83.3)	0	1 (16.7)	
Correction of errors—exclusion of patients	2	2			
Testing sensitivity of excluding 1 or more sites	1			1	
Testing differences in study design	1	1			
Central site reanalysis	1	1			
Exclusion of 1 site because of protocol inconsistencies	1	1			

Abbreviation: ITT, intention to treat.

[a] Numbers may exceed 37 because of multiple differences per reanalysis.

[b] Thirteen differences among 9 reanalyses implying treatment of more patients.

[c] Not cited as the primary reason for reanalysis.

[d] Not cited as the primary reason for reanalysis in 1 study.

[e] Not cited as the primary reason for reanalysis in 2 studies.

Table 3. Reanalyses Producing a Change in Whom to Treat

Source	Patient Population	Intervention Comparators	Primary Outcome	Original Trial Interpretation	Differences in Methods Used in the Reanalysis	Change in Finding	Change in Interpretation
Johnston et al,[27] 1985	Coronary artery disease with regional left ventricular dysfunction	Pindolol vs propranolol	Left ventricular ejection fraction at rest and exercise	No difference between pindolol and propranolol	Regional wall motion abnormalities reanalyzed with a computer-assisted rather than visual method	Pindolol superior to propranolol	Treat with pindolol rather than propranolol (more patients to be treated with the newer treatment)
Brooks et al,[17] 1998	Alzheimer disease	Acetyl L-carnitine vs placebo	Performance on the cognitive subscale of the Alzheimer Disease Assessment Scale	No difference between acetyl L-carnitine and placebo	Cognitive subscale reanalyzed as rate of change; analysis included test for interaction between drug effect and age	Test of drug × age interaction statistically significant, with younger patients benefiting more from treatment than older patients	Treat younger patients (different patients)
Thomsen et al,[28] 1998	Cirrhosis and esophageal varices	Medical treatment vs medical treatment plus sclerotherapy	Cumulative overall mortality	Borderline statistically significant reduction of overall mortality with sclerotherapy, no effect on rebleeding	Reanalysis used a multistage competing-risks model accounting for varying risk of mortality at stages of bleeding and nonbleeding	Reduction in rebleeding with sclerotherapy late in disease, especially after first episode; no reduction in mortality	Treat patients with rebleeding rather than patients at higher risk of mortality (different patients)[a]
Kobashigawa et al,[20] 2006	Heart transplant	Mycophenolate mofetil vs azathioprine	Change in IVUS-measured maximal intimal thickness first year after transplant (possible surrogate marker for adverse posttransplant outcomes)	No difference between mycophenolate mofetil and azathioprine	Reanalysis completed a matched site-to-site image analysis between trial centers and excluded patients with poor-quality data	Mycophenolate mofetil superior to azathioprine	Treat patients with mycophenolate mofetil rather than azathioprine (more patients to be treated with the newer treatment)
Lachin et al,[16] 2008	Type I diabetes	Intensive therapy aimed at maintaining glycemic levels as close as possible to the nondiabetic range vs conventional therapy with the goal of maintaining clinical well-being with no specific glucose targets	Retinopathy progression (sustained 3-step progression)	Intensive superior to conventional treatment	Reanalysis used additional Poisson models	Hemoglobin A_{1C} explains virtually all the difference between intensive and conventional treatment	Treat with intensive or conventional therapy according to patients' A_{1C} levels (different patients)[a]
Johnson et al,[29] 2009	Scleroderma	Methotrexate vs placebo	Modified Rodnan skin score,[b] UCLA skin score, physician global assessment of disease activity	No difference between methotrexate and placebo	Reanalyzed with bayesian methods and used multiple imputation to address missing data	High probability that methotrexate results in better mean outcomes than placebo	Treat more patients with methotrexate
Paradis et al,[18] 2010	Adults with out-of-hospital cardiac arrest	Mechanical (AutoPulse) CPR vs traditional manual CPR	Survival to hospital discharge	No difference between mechanical and manual CPR in 4-h survival; worse survival to hospital discharge with mechanical CPR	Reanalysis excluding 1 site that switched intervention protocols, resulting in intervention delays	Mechanical superior to manual CPR at improving 4-h survival	Treat with mechanical rather than manual CPR (more patients to be treated with the newer treatment)
Chmielewski et al,[21] 2011	Pelvic organ prolapse	3 Surgical techniques: standard anterior vs absorbable-mesh augmented vs ultralateral colporrhaphy	Resolved (stage 0) prolapse	Success 30%-42%, with no difference between surgical approaches	Reanalyzed with clinical instead of anatomic criteria for success	Success 88%, with no difference between surgical approaches	More patients should be treated with any of the approaches
Nagakane et al,[25] 2011	Acute ischemic stroke	Alteplase vs placebo	MRI infarct growth attenuation	No difference between alteplase and placebo	Reanalyzed with a different measurement to assess the same construct	Alteplase superior to placebo	Treat more patients with alteplase
McCann et al,[30] 2012	Metham-phetamine dependence	Bupropion vs placebo	Change in percentage of participants with methamphetamine-free urine each week	No difference between bupropion and alteplase	Reanalyzed with new criteria for the assessment of outcome	Bupropion superior to placebo	Treat more patients with bupropion

(continued)

Table 3. Reanalyses Producing a Change in Whom to Treat (continued)

Source	Patient Population	Intervention Comparators	Primary Outcome	Original Trial Interpretation	Differences in Methods Used in the Reanalysis	Change in Finding	Change in Interpretation
Colquhoun,[24] 1990	Fibrositis	Homeopathic treatment vs placebo	The number of tender spots, number of patients with improved pain or sleep	Homeopathic treatment superior to placebo	Reanalysis separated composite pain and sleep end points	No difference between homeopathic treatment and placebo	Treat fewer patients with homeopathic therapy
Rothwell et al,[31] 2003	Carotid stenosis	Immediate carotid endarterectomy plus best medical treatment vs best medical treatment alone	Any first stroke or surgical death (≥70% degree of stenosis without near occlusion)	Carotid endarterectomy beneficial only in patients with ≥80% stenosis	Reanalyzed with criteria for severity of stenosis and assessment of outcome used in a similar trial	Carotid endarterectomy beneficial in patients with >70% stenosis	Treat more patients with carotid endarterectomy
Welling and Nagaraja,[26] 2000	Meniere disease	Endolymphatic mastoid shunt vs mastoidectomy (placebo)	Combined score for nausea and vomiting, dizziness, tinnitus, subjective and objective hearing impairment, pressure sensation in the ears	No difference between endolymphatic mastoid shunt and mastoidectomy (placebo)	Reanalyzed with Wilcoxon and Mann-Whitney *U* tests to compare treatment groups	Endolymphatic mastoid shunt superior to mastoidectomy (placebo) for nausea and vomiting, dizziness, tinnitus	Treat more patients with endolymphatic mastoid shunt

Abbreviations: CPR, cardiopulmonary resuscitation; IVUS, intravascular ultrasound; MRI, magnetic resonance imaging.

[a] Implied by findings but not directly stated by authors.

[b] Modified Rodnan skin score evaluates improvements in skin.

neuropathic pain.[41] A reanalysis by authors who were proponents of complementary and alternative medicine analyzed mortality and attrition (not the primary or secondary outcomes in the original article) and concluded that acupuncture had a lower attrition rate and lower (zero) mortality rate, with $P = .05$ for both comparisons.[42] Thus, reanalyses with discrepant results may under some circumstances raise as many questions as the original trial.

Ideally, authors completing the reanalysis should not have conflicting financial, ideological, or political interests.[43] At the least, when a reanalysis is completed, authors of the original article should be provided with the opportunity to review and comment on it before publication.[43] In our review, we found no statistically significant difference in the proportion of reanalyses resulting in a change in recommendation about the number of patients who should be treated when the reanalysis was conducted by original trial vs independent authors, but there were only 13 reanalyses that could result in a change in recommendation, so the comparison was underpowered to detect a difference and the CIs were wide enough to be inconclusive.

In our evaluation, 65% of reanalyses were successfully reproduced without changing the interpretation of the results, which may be encouraging in that the majority of the trials' findings and conclusions were reproducible. However, 35% of the published reanalyses could alter the conclusions of the original trial on which or how many patients should be treated. It is difficult to assess whether these changes in trial conclusions led eventually to major changes in clinical practice and, if so, how large these changes were. Clinical practice choices depend only partly on trial evidence, and sometimes multiple additional trials exist that inform the same question. Nevertheless, when contradicting messages exist, it is unclear which of the 2 discrepant articles will have more influence: the original is usually published in more influential journals, but the subsequent reanalysis may be viewed as a more correct appraisal of the data.

Our study has limitations. Authors of reanalyses that led to changes in findings did not always specify how they thought the differences should be interpreted in regard to alterations in who should be treated, so we used subjective judgment to translate the change in findings into categories of changes in interpretation (treat more, fewer, or different patients). Also, we excluded meta-analyses. Authors of meta-analyses using patient-level data may routinely reanalyze data from studies they include, but whether the results of single trials have been verified or contradicted remains unclear because the authors do not typically publish each as a reanalysis, the publication emphasis is on summary results, and many data sets differ from those used in the original articles, eg, they have longer follow-up. Moreover, typically in such meta-analyses, the authors of the original articles also coauthor the meta-analysis, so accounting for trials included in patient-level meta-analyses might not increase the small number of independent reanalyses by different authors. We focused on reanalyses of single trials, but there is increasing interest in reanalyses and meta-analysis of multiple trials on the same topic, as in the case of human bone morphogenetic protein 2.[44]

We excluded non-English-language trials (n = 3); treatment effect estimates or associated CIs were missing in several published articles, which did not allow fully standardized comparison of effect sizes; the study was underpowered to detect a difference in the proportion of reanalyses resulting in a change in recommendation about the number of patients who should be treated when conducted by original trial vs independent authors; and our search may have missed some articles that were in fact reanalyses but were not named as reanalyses (or replications, reevaluations, reappraisals, reproductions, or related terms) by their authors.

Finally, we cannot exclude the possibility that many other reanalyses might have been performed that were never published, especially those with results and conclusions identical to those of the original article. Authors of confirmatory reanalyses may choose not to publish the results or, alternatively, they may have difficulty publishing their article because many journals may not consider it interesting. Thus, our observed estimate of different conclusions (35%) is probably an overestimate.

Conclusions

A small number of reanalyses of RCTs have been published to date; of these, only a few were conducted by entirely independent authors. Thirty-five percent of published reanalyses led to changes in findings that implied conclusions different from those of the original article about which patients should be treated.

ARTICLE INFORMATION

Author Contributions: Drs Ebrahim and Ioannidis had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Ebrahim, Thorlund, Mills, Ioannidis.
Acquisition, analysis, or interpretation of data: Ebrahim, Sohani, Montoya, Agarwal, Ioannidis.
Drafting of the manuscript: Ebrahim, Ioannidis.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Ebrahim, Ioannidis.
Administrative, technical, or material support: Ebrahim, Sohani, Montoya, Agarwal.
Study supervision: Ioannidis.

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

Funding/Support: Dr Ebrahim reports receiving support from MITACS Elevate and SickKids Restracomp Postdoctoral Awards. Dr Mills reports receiving support via a Canadian Institutes of Health Research Canada Research Chair.

Role of the Funders/Sponsors: Sponsors providing individual financial support to authors did not have a role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. Sharing public health data: necessary and now. *Lancet*. 2010;375(9730):1940.

2. Pisani E, AbouZahr C. Sharing health data: good intentions are not enough. *Bull World Health Organ*. 2010;88(6):462-466.

3. Samet JM. Data: to share or not to share? *Epidemiology*. 2009;20(2):172-174.

4. Doshi P, Goodman SN, Ioannidis JPA. Raw data from clinical trials: within reach? *Trends Pharmacol Sci*. 2013;34(12):645-647.

5. Krumholz HM, Ross JS. A model for dissemination and independent analysis of industry data. *JAMA*. 2011;306(14):1593-1594.

6. Doshi P, Jefferson T, Del Mar C. The imperative to share clinical study reports: recommendations from the Tamiflu experience. *PLoS Med*. 2012;9(4):e1001201.

7. Paroxetine and pediatric and adolescent patients. 2012. http://www.gsk.com/media/paroxetine.htm.

8. Berman F, Cerf V. Science priorities: who will pay for public access to research data? *Science*. 2013;341(6146):616-617.

9. Monforton C. Weight of the evidence or wait for the evidence? protecting underground miners from diesel particulate matter. *Am J Public Health*. 2006;96(2):271-276.

10. Koch HJ, Raschka C. Anthropometric data and acetylsalicylic acid pharmacokinetics. *Int J Clin Pharmacol Ther*. 2002;40(1):30-34.

11. Teschan PE, Beck GJ, Dwyer JT, et al. Effect of a ketoacid-aminoacid-supplemented very low protein diet on the progression of advanced renal disease: a reanalysis of the MDRD feasibility study. *Clin Nephrol*. 1998;50(5):273-283.

12. Gallo C, De Maio E, Di Maio M, et al; CLIP (Cancer of the Liver Italian Programme) Investigators. Tamoxifen is not effective in good prognosis patients with hepatocellular carcinoma. *BMC Cancer*. 2006;6:196.

13. Huff JC, Drucker JL, Clemmer A, et al. Effect of oral acyclovir on pain resolution in herpes zoster: a reanalysis. *J Med Virol*. 1993;(suppl 1):93-96.

14. Holly JM, Goodwin FJ, Evans SJ, VandenBurg MJ, Ledingham JM. Re-analysis of data in two *Lancet* papers on the effect of dietary sodium and potassium on blood pressure. *Lancet*. 1981;2(8260-61):1384-1387.

15. Amer-Wåhlin I, Kjellmer I, Maršál K, Olofsson P, Rosén KG. Swedish randomized controlled trial of cardiotocography only versus cardiotocography plus ST analysis of fetal electrocardiogram revisited: analysis of data according to standard versus modified intention-to-treat principle. *Acta Obstet Gynecol Scand*. 2011;90(9):990-996.

16. Lachin JM, Genuth S, Nathan DM, Zinman B, Rutledge BN; DCCT/EDIC Research Group. Effect of glycemic exposure on the risk of microvascular complications in the diabetes control and complications trial—revisited. *Diabetes*. 2008;57(4):995-1001.

17. Brooks JO III, Yesavage JA, Carta A, Bravi D. Acetyl L-carnitine slows decline in younger patients with Alzheimer's disease: a reanalysis of a double-blind, placebo-controlled study using the trilinear approach. *Int Psychogeriatr*. 1998;10(2):193-203.

18. Paradis NA, Young G, Lemeshow S, Brewer JE, Halperin HR. Inhomogeneity and temporal effects in AutoPulse Assisted Prehospital International Resuscitation—an exception from consent trial terminated early. *Am J Emerg Med*. 2010;28(4):391-398.

19. van der Linden YM, Lok JJ, Steenland E, et al; Dutch Bone Metastasis Study Group. Single fraction radiotherapy is efficacious: a further analysis of the Dutch Bone Metastasis Study controlling for the influence of retreatment. *Int J Radiat Oncol Biol Phys*. 2004;59(2):528-537.

20. Kobashigawa JA, Tobis JM, Mentzer RM, et al. Mycophenolate mofetil reduces intimal thickness by intravascular ultrasound after heart transplant: reanalysis of the multicenter trial. *Am J Transplant*. 2006;6(5 pt 1):993-997.

21. Chmielewski L, Walters MD, Weber AM, Barber MD. Reanalysis of a randomized trial of 3 techniques of anterior colporrhaphy using clinically relevant definitions of success. *Am J Obstet Gynecol*. 2011;205(1):e1-e8.

22. Reiffel JA, Reiter MJ, Freedman RA, et al; ESVEM Investigators. Influence of Holter monitor and electrophysiologic study methods and efficacy criteria on the outcome of patients with ventricular tachycardia and ventricular fibrillation in the ESVEM trial. *Prog Cardiovasc Dis*. 1996;38(5):359-370.

23. Vansteenkiste J, Hedenus M, Gascon P, et al. Darbepoetin alfa for treating chemotherapy-induced anemia in patients with a baseline hemoglobin level < 10 g/dL versus > or = 10 g/dL: an exploratory analysis from a randomized, double-blind, active-controlled trial. *BMC Cancer*. 2009;9:311.

24. Colquhoun D. Re-analysis of clinical trial of homoeopathic treatment in fibrositis. *Lancet*. 1990;336(8712):441-442.

25. Nagakane Y, Christensen S, Brekenfeld C, et al; EPITHET Investigators. EPITHET: positive result after reanalysis using baseline diffusion-weighted imaging/perfusion-weighted imaging co-registration. *Stroke*. 2011;42(1):59-64.

26. Welling DB, Nagaraja HN. Endolymphatic mastoid shunt: a reevaluation of efficacy. *Otolaryngol Head Neck Surg*. 2000;122(3):340-345.

27. Johnston DL, Manyari DE, Kostuk WJ. The influence of intrinsic sympathomimetic activity on regional left ventricular function:; use of regional ejection fraction to demonstrate a beneficial action by pindolol over propranolol on hypokinetic segments. *Can J Cardiol*. 1985;1(3):174-179.

28. Thomsen BL, Sørensen TI. Analysis of the treatment effect on recurrent bleeding and death in patients with cirrhosis and esophageal varices: multistage competing-risks model compared to conventional methods:. Tthe Copenhagen Esophageal Varices Sclerotherapy Project. *J Hepatol*. 1998;28(1):107-114.

29. Johnson SR, Feldman BM, Pope JE, Tomlinson GA. Shifting our thinking about uncommon disease trials: the case of methotrexate in scleroderma. *J Rheumatol*. 2009;36(2):323-329.

30. McCann DJ, Li S-H. A novel, nonbinary evaluation of success and failure reveals bupropion

efficacy versus methamphetamine dependence: reanalysis of a multisite trial. *CNS Neurosci Ther*. 2012;18(5):414-418.

31. Rothwell PM, Gutnikov SA, Warlow CP; European Carotid Surgery Trialist's Collaboration. Reanalysis of the final results of the European Carotid Surgery Trial. *Stroke*. 2003;34(2):514-523.

32. Diener HC, Rahlfs VW, Danesch U. The first placebo-controlled trial of a special butterbur root extract for the prevention of migraine: reanalysis of efficacy criteria. *Eur Neurol*. 2004;51(2):89-97.

33. Witkiewitz K, van der Maas HLJ, Hufford MR, Marlatt GA. Nonnormality and divergence in posttreatment alcohol use: reexamining the Project MATCH data "another way." *J Abnorm Psychol*. 2007;116(2):378-394.

34. Doshi P, Vedula SS, Li T. YODA and truth seeking in medicine. *BMJ*. 2013;347:f4251.

35. Ross JS, Lehman R, Gross CP. The importance of clinical trial data sharing: toward more open science. *Circ Cardiovasc Qual Outcomes*. 2012;5(2):238-240.

36. Ioannidis JP, Khoury MJ. Improving validation practices in "omics" research. *Science*. 2011;334(6060):1230-1232.

37. Holdren JP. Public access memorandum from the Office of Science and Technology Policy 2013. http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

38. Blumenthal D, Campbell EG, Anderson MS, Causino N, Louis KS. Withholding research results in academic life science: evidence from a national survey of faculty. *JAMA*. 1997;277(15):1224-1228.

39. Campbell EG, Clarridge BR, Gokhale M, et al. Data withholding in academic genetics: evidence from a national survey. *JAMA*. 2002;287(4):473-480.

40. Blumenthal D, Campbell EG, Gokhale M, et al. Data withholding in genetics and the other life sciences: prevalences and predictors. *Acad Med*. 2006;81(2):137-145.

41. Shlay JC, Chaloner K, Max MB, et al; Terry Beirn Community Programs for Clinical Research on AIDS. Acupuncture and amitriptyline for pain due to HIV-related peripheral neuropathy: a randomized controlled trial. *JAMA*. 1998;280(18):1590-1595.

42. Shiflett SC, Schwartz GE. Effects of acupuncture in reducing attrition and mortality in HIV-infected men with peripheral neuropathy. *Explore (NY)*. 2011;7(3):148-154.

43. Christakis DA, Zimmerman FJ. Rethinking reanalysis. *JAMA*. 2013;310(23):2499-2500.

44. Simmonds MC, Brown JV, Heirs MK, et al. Safety and effectiveness of recombinant human bone morphogenetic protein-2 for spinal fusion: a meta-analysis of individual-participant data. *Ann Intern Med*. 2013;158(12):877-889.