Journal of Clinical Epidemiology 86 (2017) 39–50

Journal of Clinical Epidemiology

# Multiple outcomes and analyses in clinical trials create challenges for interpretation and research synthesis

Evan Mayo-Wilson[a,*], Nicole Fusco[a], Tianjing Li[a], Hwanhee Hong[b], Joseph K. Canner[c], Kay Dickersin[a], for the MUDS investigators

[a]*Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205, USA*
[b]*Department of Mental Health, Johns Hopkins University Bloomberg School of Public Health, 624 N Broadway, Hampton House, Baltimore, MD 21205, USA*
[c]*Department of Surgery, Johns Hopkins University School of Medicine, 600 North Wolfe Street, Blalock Building, Baltimore, MD 21287, USA*

Accepted 9 May 2017; Published online 18 May 2017

## Abstract

**Objective:** To identify variations in outcomes and results across reports of randomized clinical trials (RCTs).

**Study Design and Setting:** Eligible RCTs examined gabapentin for neuropathic pain and quetiapine for bipolar depression, reported in public (e.g., journal articles) and nonpublic (e.g., clinical study reports) sources by 2015. We prespecified outcome domains. From each source, we collected ''outcomes'' (i.e., domain, measure, metric, method of aggregation, and time point); ''treatment effect'' (i.e., outcome plus the methods of analysis [e.g., how missing data were handled]); and results (i.e., numerical contrasts of treatment and comparison groups). We assessed whether results included sufficient information for meta-analysis.

**Results:** We found 21 gabapentin (68 public, 6 nonpublic reports) and seven quetiapine RCTs (46 public, 4 nonpublic reports). For four (gabapentin) and seven (quetiapine) prespecified outcome domains, RCTs reported 214 and 81 outcomes by varying four elements. RCTs assessed 605 and 188 treatment effects by varying the analysis of those outcomes. RCTs reported 1,230 and 661 meta-analyzable results, 305 (25%) and 109 (16%) in public reports.

**Conclusion:** RCTs included hundreds of outcomes and results; a small proportion were in public reports. Trialists and meta-analysts may cherry-pick what they report from multiple sources of RCT information. 

*Keywords:* Clinical trials; Systematic reviews; Meta-analysis; Outcomes; Selective outcome reporting

## 1. Introduction

Although randomized clinical trials (RCTs) are considered the reference standard for examining effectiveness and safety of treatments, it is rare that a single RCT provides sufficient evidence to merit adoption of a treatment for any given condition. Furthermore, clinicians and others can no longer stay abreast of rapidly growing knowledge, including the findings of all RCTs pertinent to their treatment decisions. Accordingly, they look to summaries of knowledge, such as clinical practice guidelines, that depend in part on evidence syntheses (e.g., systematic reviews, meta-analyses); evidence syntheses combine information from similar studies, often focusing on RCTs for treatment decisions.

In many systematic reviews, ''outcomes'' are not well defined. Although ''outcomes'' are often described by a ''name'' such as ''pain intensity,'' this name is actually the ''outcome domain,'' one of five elements comprising an outcome [1]. The five elements are as follows: (1) outcome domain; (2) measure (e.g., McGill Pain Questionnaire, Montgomery Åsberg Depression Rating Scale); (3) metric (e.g., value at a time point, change from baseline); (4) method of aggregation (e.g., mean value for continuous data, percent with an outcome for categorical data); and (5) time point at which the assessment was made (e.g., 8 weeks after starting treatment). Thus, for a single outcome domain, one RCT may include many defined outcomes

The MUDS investigators include Lorenzo Bertizzolo, Terrie Cowley, Peter Doshi, Jeffrey Ehmsen, Gillian Gresham, Nan Guo, Jennifer Haythornthwaite, James Heyward, Diana Lock, Jennifer Payne, Lori Rosman, Elizabeth Stuart, Catalina Suarez-Cuervo, Elizabeth Tolbert, Claire Twose, and Swaroop Vedula.

Ethics approval: The study received an exemption from the Johns Hopkins Bloomberg School of Public Health Institutional Review Board (IRB no: 00006324).

Funding: Supported by contract ME 1303 5785 from the Patient-Centered Outcomes Research Institute (PCORI) and a fund established at Johns Hopkins for scholarly research on reporting biases by Greene LLP. The funders were not involved in the design or conduct of the study, article preparation, or the decision to submit the article for publication.

* Corresponding author. Tel.: 443-287-5042.
*E-mail address*: evan.mayo-wilson@jhu.edu (E. Mayo-Wilson).

http://dx.doi.org/10.1016/j.jclinepi.2017.05.007

**What is new?**

**Key findings**

- Trials of the same intervention and condition included hundreds of different outcomes and results, and much of this information was available only using nonpublic sources.

**What this adds to what was known?**

- Multiple outcome definitions and multiple methods of analysis lead to challenges for interpreting clinical trials, particularly because they create opportunities for cherry-picking by both clinical trialists and systematic reviewers.
- Variation in outcomes, and incomplete results reporting, makes it difficult to compare clinical trials and to translate knowledge into practice.

**What is the implication and what should change now?**

- Clinical trials and systematic reviews should define their outcomes and methods of analysis completely and report their results transparently.
- Guidance is needed for using multiple outcomes and results in systematic reviews.

because different measures, metrics, time points, and methods of aggregation were used (Fig. 1).

Investigators performing evidence syntheses usually prespecify eligibility criteria for including RCTs and outcome domains that will be examined. It is not unusual for investigators to find, however, that even when many trials are eligible for a systematic review, only a few trials can be combined using meta-analysis [2,3]. Consequently, many trials that are eligible for systematic reviews are not included in the meta-analyses they contain; those trials thus contribute little information to the overall conclusions of systematic reviews. This may occur because the included RCTs did not assess the same outcome domains because different outcomes within the same domains could not be combined in meta-analysis (i.e., one or more elements differed), because the included RCTs assessed but did not report the same outcomes, or because RCTs reported the same outcomes but did not report sufficient statistical information to allow combination of the numerical results [4,5]. Furthermore, if systematic reviewers assume that outcomes within RCTs can be used interchangeably, even if those outcomes are not defined using all elements, reviewers may be making assumptions that lead to errors when synthesizing overall results [6,7].

The fact that RCTs may assess multiple outcomes for the same domain leads to challenges for systematic reviewers, regardless of whether they conduct meta-analyses [8,9]. First, if an RCT reports multiple outcomes, which outcome should be used to determine whether the intervention "works"? Second, a single RCT might report different results for the same outcome by using multiple methods of analyses (e.g., methods for handling missing data) [10–14]. If there are multiple results for an outcome, which estimate should the meta-analyst use? Third, even when it is possible to combine multiple RCTs, synthesized results (e.g., the combined standardized mean difference [SMD]) may be difficult to interpret if studies used different outcome definitions or different methods of analysis [12]. All of these situations pose challenges to the proper interpretation of RCTs and evidence syntheses, and they may lead to innocent errors.

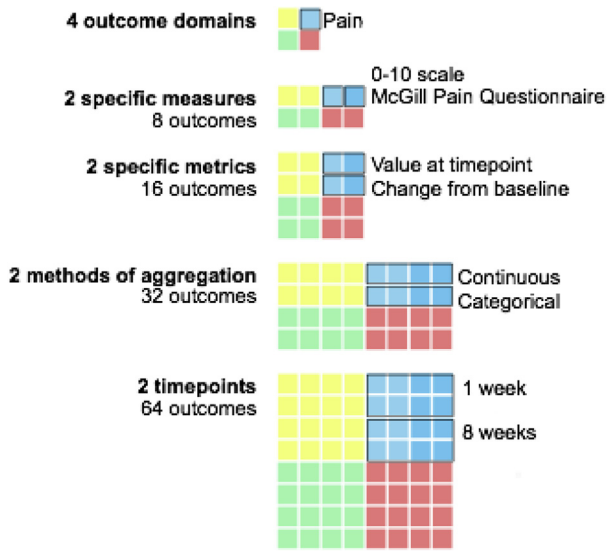


**Fig. 1.** The number of outcomes in a trial is a function of the number of definitions of each of the five elements. In this hypothetical example, the number of treatment effects for a single trial is the product of definitions for each element. In a trial with four outcome domains, introducing two definitions for each of the other elements will result in 64 unique defined outcomes.

Defining multiple outcomes under the same domain may also be associated with deliberate efforts (e.g., by trialists or systematic reviewers) to conceal findings and to mislead readers. For example, in RCTs that include many outcomes, trialists might report statistically significant results selectively [14–16]. In systematic reviews, investigators might cherry-pick results to include in meta-analyses [17,18]. Furthermore, when only some outcomes are reported publicly, it is impossible for the systematic reviewer or other interpreter of the trial findings to know for sure whether there has been selective reporting.

Few studies have explored the number of results that investigators could select to include in meta-analyses [7,13,19]. We know of no studies that have used both public and nonpublic data sources for RCTs to quantify the number of outcomes and results reported across RCTs, the number of reported outcomes that are defined, or the number of results that are meta-analyzable.