Methods: Mind the Gap Webinar Series

Challenges Associated With Multiple Outcome Definitions in Clinical Research

Presented by

Evan Mayo-Wilson, D.Phil.

Johns Hopkins Bloomberg School of Public Health

Sources of support

European Social Research Council (ESRC)

*Laura and John Arnold Foundation

*National Institutes of Health (NIH)

National Institutes of Health and Care Excellence (NICE, UK)

Swedish Board of Health and Welfare (Socialstyrelsen)

*Patient Centered Outcomes Research Institute (PCORI)

*Robert Wood Johnson Foundation

*U.S. Food and Drug Administration (FDA)

Current

Multiple Data Sources (MUDS) Investigators

Steering Committee
Dickersin, Kay (KD)
Fusco, Nicole (NF)
Li, Tianjing (TL)
Mayo-Wilson, Evan (EMW)
Tolbert, Elizabeth (ET)

Protocol development, study implementation
Cowley, Terrie (TC)
Haythornthwaite, Jennifer (JH)
Hong, Hwanhee
Payne, Jennifer (JP)
Singh, Sonal (SS)
Stuart, Elizabeth (ES)
EMW, KD, TL, NF, ET, JE

Data acquisition
Bertizzolo, Lorenzo (LB)
Ehmsen, Jeffery (JE)
Gresham, Gillian (GG)
Heyward, James (JHe)
Lock, Diana (DL)
Rosman, Lori (LR)
Suarez-Cuervo, Catalina (CS)
Twose, Claire (CT)
KD, NF, EMW, TL, SV

Analysis and interpretation of data
Canner, Joseph (JC)
Guo, Nan (NG)
Hong Hwanhee (HH)
Stuart, Elizabeth (ES)
NF, EMW, KD, TL

Systematic Review Data Repository
Jap, Jens (JJ)
Lau, Joseph (JL)
Smith, Bryant (BS)

Ancillary studies
Golozar, Asieh (AG)
Hutfless, Susie (SH)
EMW, KD, TC

Overview

- Concerns about multiple outcomes & analyses (multiplicity)
- Evidence of multiplicity in clinical trials
- Strategies to address multiplicity

A replication crisis?

BuzzFeed News Here's How Cornell Scientist Brian Wansink Turned Shoddy Data Into Viral Studies About How We ...

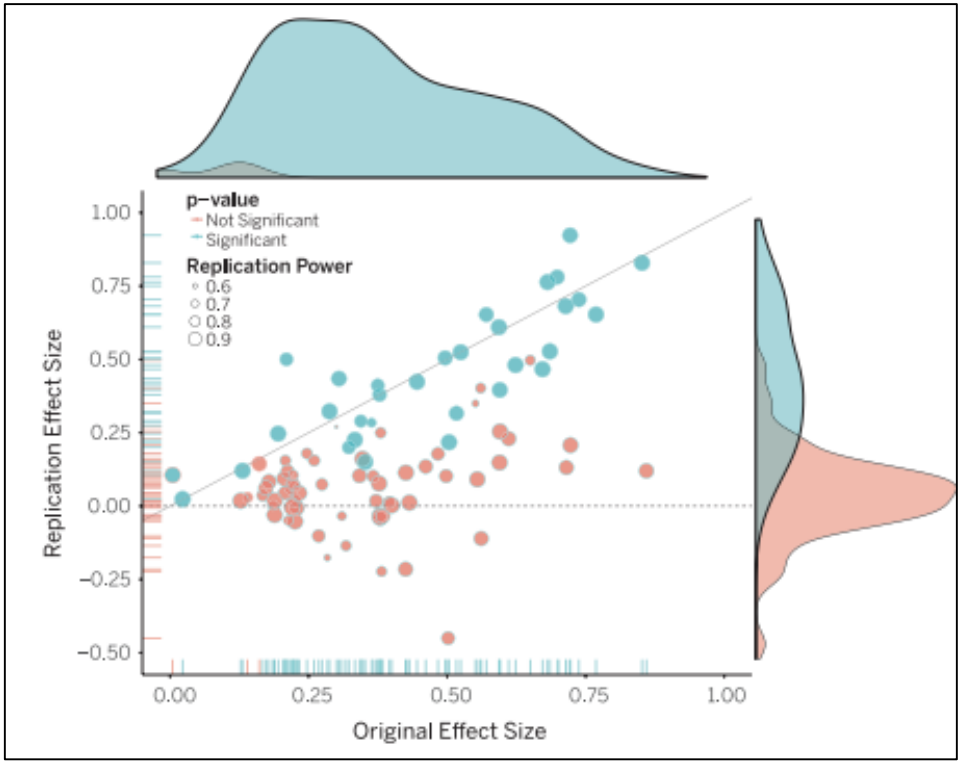
SCIENCE

Here's How Cornell Scientist Brian Wansink Turned Shoddy Data Into Viral Studies About How We Eat

Brian Wansink won fame, funding, and influence for his science-backed advice on healthy

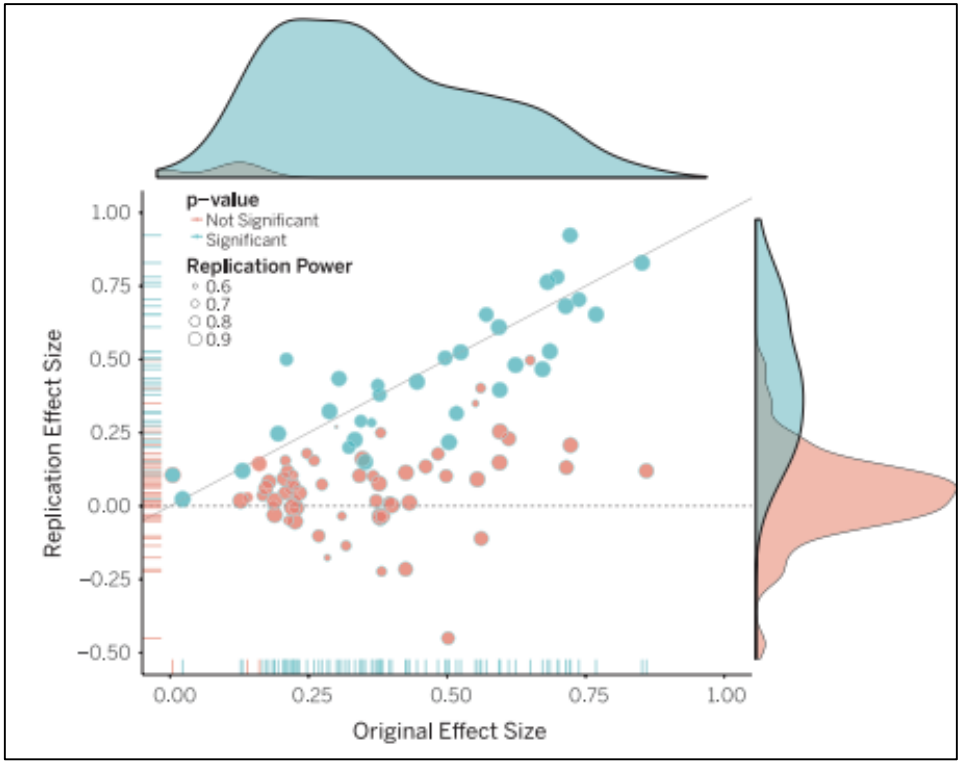
https://www.buzzfeednews.com/article/stephaniemlee/brian-wansink-cornell-p-hacking

Replication project



Open Science Collaboration, 2015. DOI: 10.1126/science.aac4716

Replication project



PERSPECTIVE

SCIENTIFIC INTEGRITY

What does research reproducibility mean?

Steven N. Goodman,* Daniele Fanelli, John P. A. Ioannidis

The language and conceptual framework of "research reproducibility" are nonstandard and unsettled across the sciences. In this Perspective, we review an array of explicit and implicit definitions of reproducibility and related terminology, and discuss how to avoid potential misunderstandings when these terms are used as a surrogate for "truth."

perimental design. Some irreproducible reports are probably the result of coincidental findings that happen to reach statistical significance, coupled with publication bias. Another pitfall is over-interpretation of creative 'hypothesis-generating' experiments, which are designed to uncover new avenues of inquiry rather than to provide definitive proof for any single question. Still, there

"Multiplicity, combined with incomplete reporting, might be the single largest contributor to the phenomenon of nonreproducibility, or falsity, of published claims."

Open Science Collaboration, 2015. DOI: 10.1126/science.aac4716

Goodman, *et al.*, 2016. DOI: 10.1126/scitranslmed.aaf5027

Publication bias & replication

PUBLICATION DECISIONS AND THEIR POSSIBLE EFFECTS ON INFERENCES DRAWN FROM TESTS OF SIGNIFICANCE —OR VICE VERSA*

THEODORE D. STERLING
University of Cincinnati

There is some evidence that in fields where statistical tests of significance are commonly used, research which yields nonsignificant results is not published. Such research being unknown to other investigators may be repeated independently until eventually by chance a significant result occurs—an "error of the first kind"—and is published. Significant results published in these fields are seldom verified by independent replication. The possibility thus arises that the literature of such a field consists in substantial part of false conclusions resulting from errors of the first kind in statistical tests of significance.

Sterling, 1959. DOI: 10.1080/01621459.1959.10501497

Publication bias & replication

PUBLICATION DECISIONS AND THEIR POSSIBLE EFFECTS ON INFERENCES DRAWN FROM TESTS OF SIGNIFICANCE —OR VICE VERSA*

THEODORE D. STERLING
University of Cincinnati

There is some evidence that in fields where statistical tests of significance are commonly used, research which yields nonsignificant results is not published. Such research being unknown to other investigators may be repeated independently until eventually by chance a significant result occurs—an "error of the first kind"—and is published. Significant results published in these fields are seldom verified by independent replication. The possibility thus arises that the literature of such a field consists in substantial part of false conclusions resulting from errors of the first kind in statistical tests of significance.

TABLE 31

OUTCOMES OF TESTS OF SIGNIFICANCE FOR FOUR PSYCHOLOGY RESEARCH JOURNALS

Journals: All Issues From January To December	Total Number of Research Reports (1)	Number of Research Reports Using Tests of Significance (2)	Number of Research Reports that Reject H_0 with $Pr(E \mid H_0) \leq .05$ (3)	Number of Research Reports that Fail to Reject H_0 (4)	Number of Research Reports That are Replication of Previously Published Experiments (5)
Experimental Psychology (1955)	124	106	105	1	0
Comparative and Physiological Psychology (1956)	118	94	91	3	0
Clinical Psychology (1955)	81	62	59	3	0
Social Psychology (1955)	39	32	31	1	0
Total	362	294	286	8	0

Sterling, 1959. DOI: 10.1080/01621459.1959.10501497

Concerns about multiplicity and reporting bias in clinical trials

RANDOMIZE THE FIRST PATIENT!

To the Editor: The issue of September 2 of the *New England Journal of Medicine* contains three articles and two editorials that beautifully illustrate the importance of randomizing the first patient when new therapies are introduced for human beings. The therapeutic trial of human leukocyte interferon[1] in four patients with chronic active

tients could have been treated in the more ethical and scientific manner of a controlled trial by now. Also, there ought to be at the very least some better method of centrally recording the sporadic individual trials now going on.

THOMAS C. CHALMERS, M.D.
Mount Sinai Medical Center
New York, NY 10029

Chalmers, 1977. DOI: 10.1056/NEJM197701132960214

Concerns about multiplicity and reporting bias in clinical trials

RANDOMIZE THE FIRST PATIENT!

To the Editor: The issue of September 2 of the *New England Journal of Medicine* contains three articles and two editorials that beautifully illustrate the importance of randomizing the first patient when new therapies are introduced for human beings. The therapeutic trial of human leukocyte interferon[1] in four patients with chronic active

tients could have been treated in the more ethical and scientific manner of a controlled trial by now. Also, there ought to be at the very least some better method of centrally recording the sporadic individual trials now going on.

THOMAS C. CHALMERS, M.D.
Mount Sinai Medical Center
New York, NY 10029

The "File Drawer Problem" and Tolerance for Null Results

Robert Rosenthal
Harvard University

For any given research area, one cannot tell how many studies have been conducted but never reported. The extreme view of the "file drawer problem" is that journals are filled with the 5% of the studies that show Type I errors, while the file drawers are filled with the 95% of the studies that show nonsignificant results. Quantitative procedures for computing the tolerance for filed and future null results are reported and illustrated, and the implications are discussed.

Chalmers, 1977. DOI: 10.1056/NEJM197701132960214
Rosenthal, 1979. DOI: 10.1037/0033-2909.86.3.638

Concerns about multiplicity and reporting bias in clinical trials

RANDOMIZE THE FIRST PATIENT!

To the Editor: The issue of September 2 of the *New England Journal of Medicine* contains three articles and two editorials that beautifully illustrate the importance of randomizing the first patient when new therapies are introduced for human beings. The therapeutic trial of human leukocyte interferon[1] in four patients with chronic active

tients could have been treated in the more ethical and scientific manner of a controlled trial by now. Also, there ought to be at the very least some better method of centrally recording the sporadic individual trials now going on.

THOMAS C. CHALMERS, M.D.
Mount Sinai Medical Center
New York, NY 10029

The "File Drawer Problem" and Tolerance for Null Results

Robert Rosenthal
Harvard University

For any given research area, one cannot tell how many studies have been conducted but never reported. The extreme view of the "file drawer problem" is that journals are filled with the 5% of the studies that show Type I errors, while the file drawers are filled with the 95% of the studies that show nonsignificant results. Quantitative procedures for computing the tolerance for filed and future null results are reported and illustrated, and the implications are discussed.

A problem in evaluating different therapies from a review of clinical trials is that the published clinical trial literature may be biased in favor of positive or promising results. In this report, a model is proposed for reviewing clinical trial results which is free from publication bias based on the selection of trials registered in advance in a registry. The value of a registry is illustrated by comparing a review of *published* clinical trials located by a literature search with a review of *registered* trials contained in a cancer trials registry. Two therapeutic questions are examined: (1) the survival impact of initial alkylating agent (AA) v

ence in survival is demonstrated based on a pooled analysis of registered trials (median survival ratio, 1.05; $P = .25$). For multiple myeloma, a pooled analysis of published trials also demonstrates a significant survival advantage for CC (median survival ratio, 1.26; $P = .04$), especially for poor risk patients (ratio, 1.66; $P = .002$). A pooled analysis of registered trials also shows a survival benefit for patients receiving combination chemotherapy (all patients, $P = .06$; poor risk, $P = .03$), but the estimated magnitude of the benefit is reduced (all patients: ratio, 1.11; poor risk: ratio, 1.22). These examples illustrate an ap-

Editorial

Toward Prospective Registration of Clinical Trials

Curtis L. Meinert, PhD

Chalmers, 1977. DOI: 10.1056/NEJM197701132960214
Rosenthal, 1979. DOI: 10.1037/0033-2909.86.3.638
Simes, 1986. DOI: 10.1200/JCO.1986.4.10.1529
Meinert, 1988. DOI: 10.1016/0197-2456(88)90002-5

Concerns about multiplicity and reporting bias in clinical trials

RANDOMIZE THE FIRST PATIENT!

To the Editor: The issue of September 2 of the *New England Journal of Medicine* contains three articles and two editorials that beautifully illustrate the importance of randomizing the first patient when new therapies are introduced for human beings. The therapeutic trial of human leukocyte interferon[1] in four patients with chronic active

tients could have been treated in the more ethical and scientific manner of a controlled trial by now. Also, there ought to be at the very least some better method of centrally recording the sporadic individual trials now going on.

THOMAS C. CHALMERS, M.D.
Mount Sinai Medical Center
New York, NY 10029

DATA TORTURING

"IF you torture your data long enough, they will tell you whatever you want to hear" has become a popular observation in our office. In plain English, this means that study data, if manipulated in enough different ways, can be made to prove whatever the investigator wants to prove. Unfortunately, this is generally true. Because every investigator wants to present results in the most exciting way, we all look for the most dramatic, positive findings in our data. When this process goes beyond reasonable interpretation of the facts, it becomes data torturing. The unfortunate result of torturing data is the dis-

The "File Drawer Problem" and Tolerance for Null Results

Robert Rosenthal
Harvard University

For any given research area, one cannot tell how many studies have been conducted but never reported. The extreme view of the "file drawer problem" is that journals are filled with the 5% of the studies that show Type I errors, while the file drawers are filled with the 95% of the studies that show nonsignificant results. Quantitative procedures for computing the tolerance for filed and future null results are reported and illustrated, and the implications are discussed.

A problem in evaluating different therapies from a review of clinical trials is that the published clinical trial literature may be biased in favor of positive or promising results. In this report, a model is proposed for reviewing clinical trial results which is free from publication bias based on the selection of trials registered in advance in a registry. The value of a registry is illustrated by comparing a review of *published* clinical trials located by a literature search with a review of *registered* trials contained in a cancer trials registry. Two therapeutic questions are examined: (1) the survival impact of initial alkylating agent (AA) v

ence in survival is demonstrated based on a pooled analysis of registered trials (median survival ratio, 1.05; $P = .25$). For multiple myeloma, a pooled analysis of published trials also demonstrates a significant survival advantage for CC (median survival ratio, 1.26; $P = .04$), especially for poor risk patients (ratio, 1.66; $P = .002$). A pooled analysis of registered trials also shows a survival benefit for patients receiving combination chemotherapy (all patients, $P = .06$; poor risk, $P = .03$), but the estimated magnitude of the benefit is reduced (all patients: ratio, 1.11; poor risk: ratio, 1.22). These examples illustrate an ap-

Editorial

Toward Prospective Registration of Clinical Trials

Curtis L. Meinert, PhD

Chalmers, 1977. DOI: 10.1056/NEJM197701132960214
Rosenthal, 1979. DOI: 10.1037/0033-2909.86.3.638
Simes, 1986. DOI: 10.1200/JCO.1986.4.10.1529
Meinert, 1988. DOI: 10.1016/0197-2456(88)90002-5
Mills, 1993. DOI: 10.1056/NEJM199310143291613

Concerns about multiplicity and reporting bias in clinical trials

RANDOMIZE THE FIRST PATIENT!

To the Editor: The issue of September 2 of the *New England Journal of Medicine* contains three articles and two editorials that beautifully illustrate the importance of randomizing the first patient when new therapies are introduced for human beings. The therapeutic trial of human leukocyte interferon[1] in four patients with chronic active

tients could have been treated in the more ethical and scientific manner of a controlled trial by now. Also, there ought to be at the very least some better method of centrally recording the sporadic individual trials now going on.

THOMAS C. CHALMERS, M.D.
Mount Sinai Medical Center
New York, NY 10029

DATA TORTURING

"IF you torture your data long enough, they will tell you whatever you want to hear" has become a popular observation in our office. In plain English, this means that study data, if manipulated in enough different ways, can be made to prove whatever the investigator wants to prove. Unfortunately, this is generally true. Because every investigator wants to present results in the most exciting way, we all look for the most dramatic, positive findings in our data. When this process goes beyond reasonable interpretation of the facts, it becomes data torturing. The unfortunate result of torturing data is the dis-

The "File Drawer Problem" and Tolerance for Null Results

Robert Rosenthal
Harvard University

For any given research area, one cannot tell how many studies have been conducted but never reported. The extreme view of the "file drawer problem" is that journals are filled with the 5% of the studies that show Type I errors, while the file drawers are filled with the 95% of the studies that show nonsignificant results. Quantitative procedures for computing the tolerance for filed and future null results are reported and illustrated, and the implications are discussed.

Summary of Costs of HARKing

Several potential costs of HARKing have been identified, including

1. Translating Type I errors into hard-to-eradicate theory.
2. Propounding theories that cannot (pending replication) pass Popper's disconfirmability test.
3. Disguising post hoc explanations as a priori explanations (when the former tend also be more ad hoc, and consequently, less useful).
4. Not communicating valuable information about what did not work.
5. Taking unjustified statistical licence.
6. Presenting an inaccurate model of science to students.

A problem in evaluating different therapies from a review of clinical trials is that the published clinical trial literature may be biased in favor of positive or promising results. In this report, a model is proposed for reviewing clinical trial results which is free from publication bias based on the selection of trials registered in advance in a registry. The value of a registry is illustrated by comparing a review of *published* clinical trials located by a literature search with a review of *registered* trials contained in a cancer trials registry. Two therapeutic questions are examined: (1) the survival impact of initial alkylating agent (AA) v

ence in survival is demonstrated based on a pooled analysis of registered trials (median survival ratio, 1.05; $P = .25$). For multiple myeloma, a pooled analysis of published trials also demonstrates a significant survival advantage for CC (median survival ratio, 1.26; $P = .04$), especially for poor risk patients (ratio, 1.66; $P = .002$). A pooled analysis of registered trials also shows a survival benefit for patients receiving combination chemotherapy (all patients, $P = .06$; poor risk, $P = .03$), but the estimated magnitude of the benefit is reduced (all patients: ratio, 1.11; poor risk: ratio, 1.22). These examples illustrate an ap-

Editorial

Toward Prospective Registration of Clinical Trials

Curtis L. Meinert, PhD

Chalmers, 1977. DOI: 10.1056/NEJM197701132960214
Rosenthal, 1979. DOI: 10.1037/0033-2909.86.3.638
Simes, 1986. DOI: 10.1200/JCO.1986.4.10.1529
Meinert, 1988. DOI: 10.1016/0197-2456(88)90002-5
Mills, 1993. DOI: 10.1056/NEJM199310143291613
Kerr, 1998. DOI: 10.1207/s15327957pspr0203_4

Concerns about multiplicity and reporting bias in clinical trials

RANDOMIZE THE FIRST PATIENT!

To the Editor: The issue of September 2 of the *New England Journal of Medicine* contains three articles and two editorials that beautifully illustrate the importance of randomizing the first patient when new therapies are introduced for human beings. The therapeutic trial of human leukocyte interferon[1] in four patients with chronic active

tients could have been treated in the more ethical and scientific manner of a controlled trial by now. Also, there ought to be at the very least some better method of centrally recording the sporadic individual trials now going on.

THOMAS C. CHALMERS, M.D.
Mount Sinai Medical Center
New York, NY 10029

DATA TORTURING

"IF you torture your data long enough, they will tell you whatever you want to hear" has become a popular observation in our office. In plain English, this means that study data, if manipulated in enough different ways, can be made to prove whatever the investigator wants to prove. Unfortunately, this is generally true. Because every investigator wants to present results in the most exciting way, we all look for the most dramatic, positive findings in our data. When this process goes beyond reasonable interpretation of the facts, it becomes data torturing. The unfortunate result of torturing data is the dis-

The "File Drawer Problem" and Tolerance for Null Results

Robert Rosenthal
Harvard University

For any given research area, one cannot tell how many studies have been conducted but never reported. The extreme view of the "file drawer problem" is that journals are filled with the 5% of the studies that show Type I errors, while the file drawers are filled with the 95% of the studies that show nonsignificant results. Quantitative procedures for computing the tolerance for filed and future null results are reported and illustrated, and the implications are discussed.

Summary of Costs of HARKing

Several potential costs of HARKing have been identified, including

1. Translating Type I errors into hard-to-eradicate theory.
2. Propounding theories that cannot (pending replication) pass Popper's disconfirmability test.
3. Disguising post hoc explanations as a priori explanations (when the former tend also be more ad hoc, and consequently, less useful).
4. Not communicating valuable information about what did not work.
5. Taking unjustified statistical licence.
6. Presenting an inaccurate model of science to students.

A problem in evaluating different therapies from a review of clinical trials is that the published clinical trial literature may be biased in favor of positive or promising results. In this report, a model is proposed for reviewing clinical trial results which is free from publication bias based on the selection of trials registered in advance in a registry. The value of a registry is illustrated by comparing a review of *published* clinical trials located by a literature search with a review of *registered* trials contained in a cancer trials registry. Two therapeutic questions are examined: (1) the survival impact of initial alkylating agent (AA) v

ence in survival is demonstrated based on a pooled analysis of registered trials (median survival ratio, 1.05; $P = .25$). For multiple myeloma, a pooled analysis of published trials also demonstrates a significant survival advantage for CC (median survival ratio, 1.26; $P = .04$), especially for poor risk patients (ratio, 1.66; $P = .002$). A pooled analysis of registered trials also shows a survival benefit for patients receiving combination chemotherapy (all patients, $P = .06$; poor risk, $P = .03$), but the estimated magnitude of the benefit is reduced (all patients: ratio, 1.11; poor risk: ratio, 1.22). These examples illustrate an ap-

Editorial

Toward Prospective Registration of Clinical Trials

Curtis L. Meinert, PhD

Empirical Evidence for Selective Reporting of Outcomes in Randomized Trials

Comparison of Protocols to Published Articles

An-Wen Chan, MD, DPhil
Asbjørn Hróbjartsson, MD, PhD
Mette T. Haahr, BSc
Peter C. Gøtzsche, MD, DrMedSci

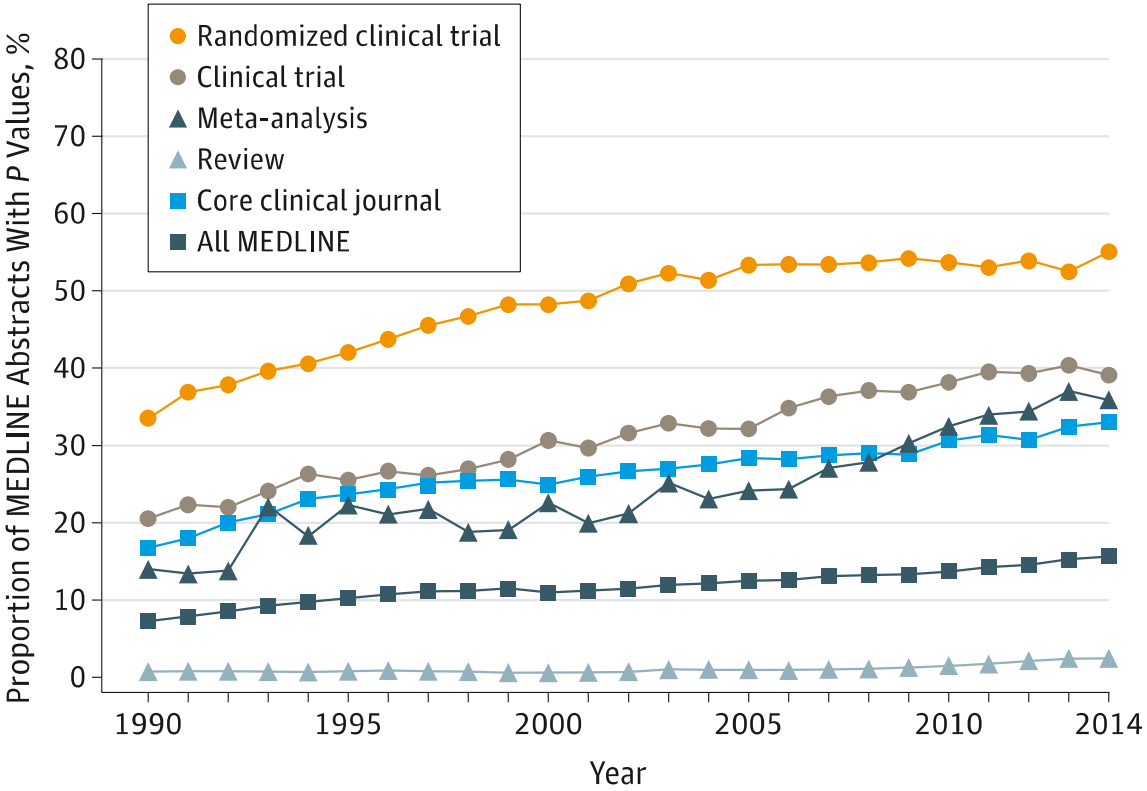
Context Selective reporting of outcomes within published studies based o
ture or direction of their results has been widely suspected, but direct evidenc
bias is currently limited to case reports.

Objective To study empirically the extent and nature of outcome reportir
a cohort of randomized trials.

Chalmers, 1977. DOI: 10.1056/NEJM197701132960214
Rosenthal, 1979. DOI: 10.1037/0033-2909.86.3.638
Simes, 1986. DOI: 10.1200/JCO.1986.4.10.1529
Meinert, 1988. DOI: 10.1016/0197-2456(88)90002-5
Mills, 1993. DOI: 10.1056/NEJM199310143291613
Kerr, 1998. DOI: 10.1207/s15327957pspr0203_4
Chan, 2004. DOI: 10.1001/jama.291.20.2457

Emphasis on "significance" over time

Figure 1. Proportion of MEDLINE Abstracts Reporting at Least 1 *P* Value in the Period 1990-2015



Chavalarias, 2016. DOI: 10.1001/jama.2016.1952

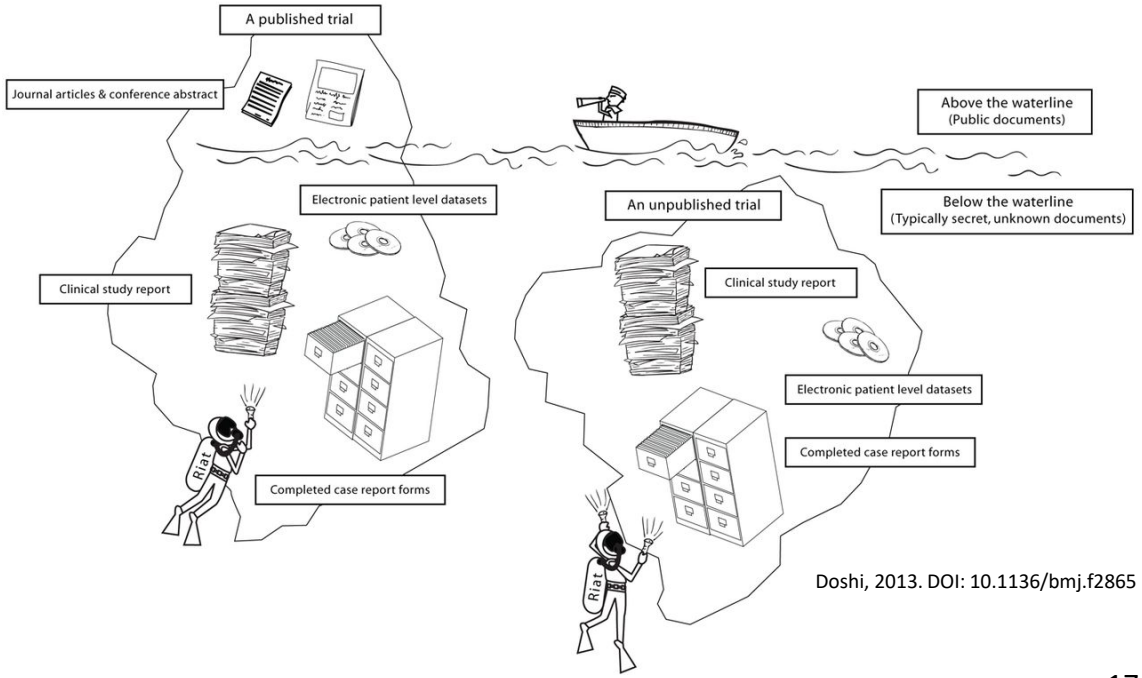
Multiple data sources

Public data sources

- Short report (e.g., letter, conference abstract)
- Journal article
- Trial registration
- Results on trial registry
- Information from regulators

Non-public data sources

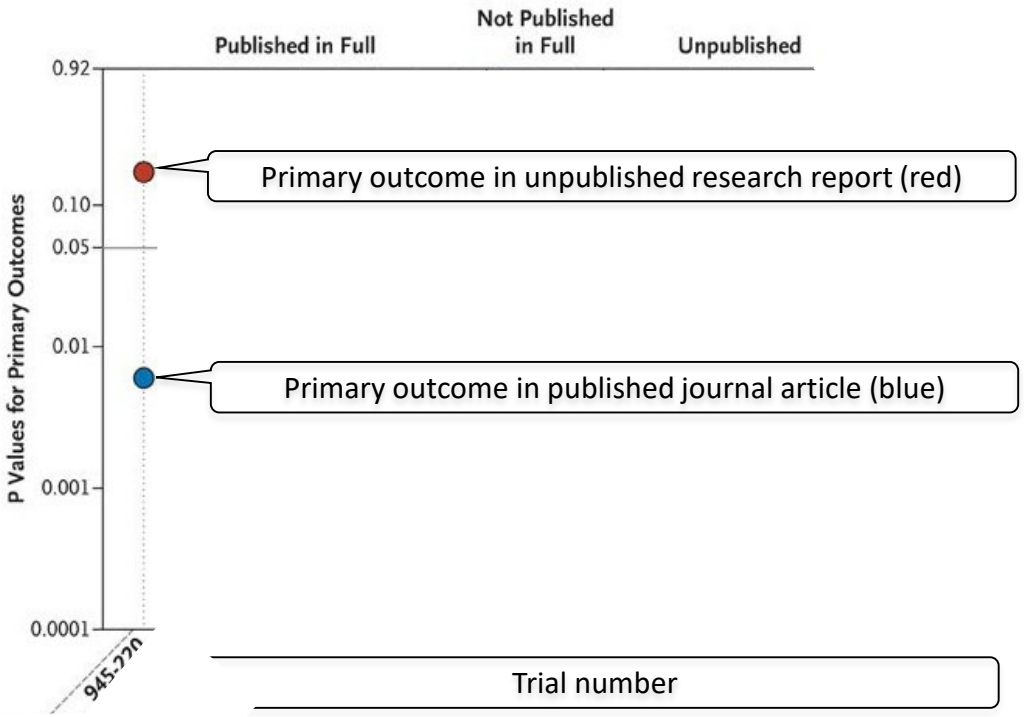
- Unpublished manuscript
- Individual participant data (IPD)
- Grant proposal
- Study protocol
- Case report form
- Memos and emails



Doshi, 2013. DOI: 10.1136/bmj.f2865

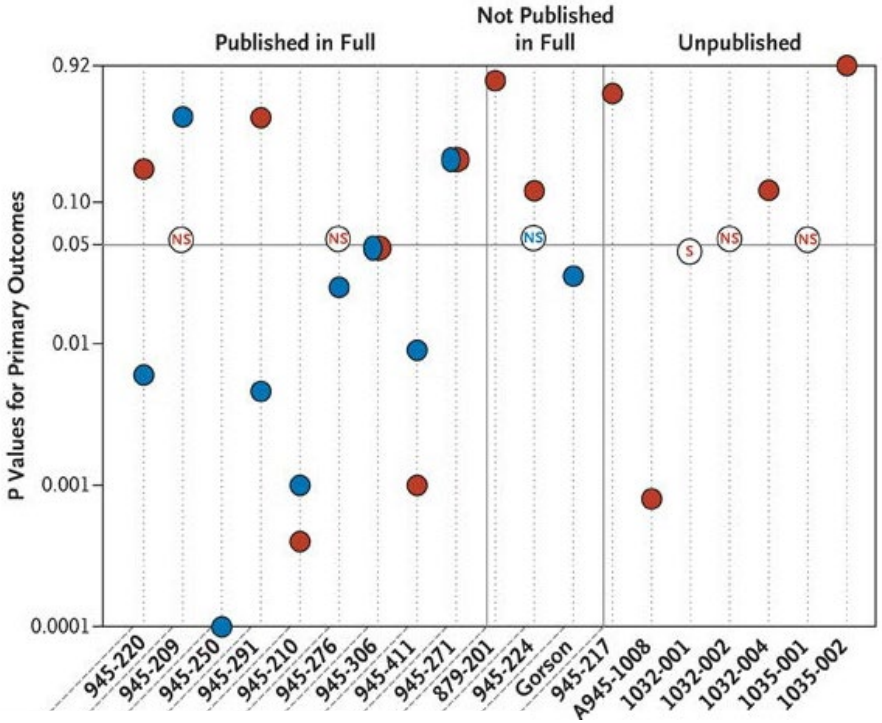
Mayo-Wilson, 2015. DOI: 10.1186/s13643-015-0134-z OA

Gabapentin: results for “primary” outcomes differ between sources



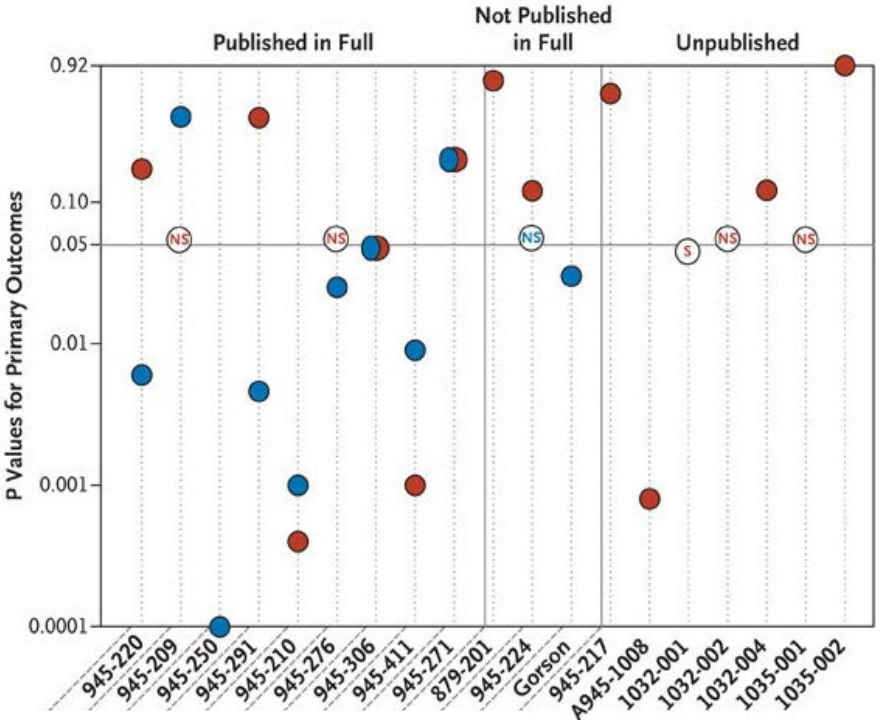
Vedula, 2009. DOI: 10.1056/NEJMsa0906126

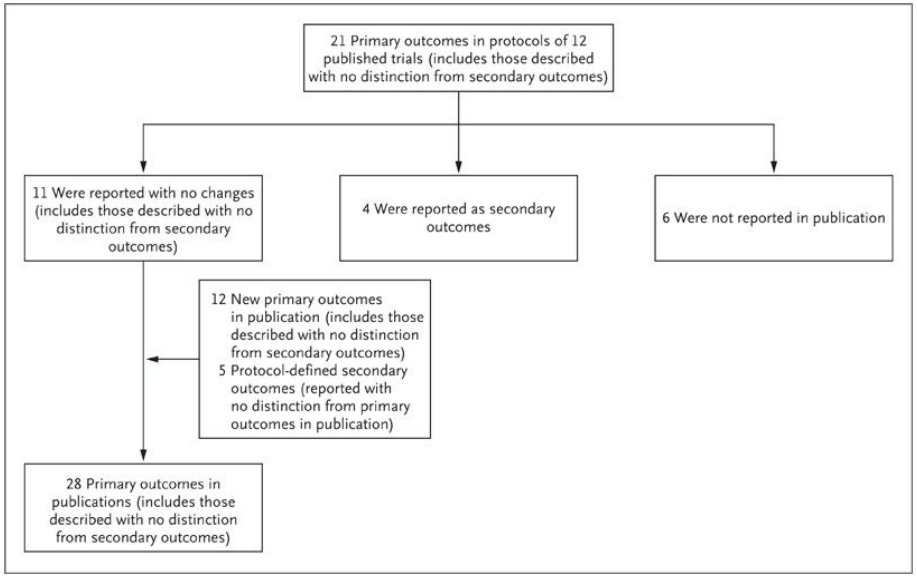
Gabapentin: results for “primary” outcomes differ between sources



Vedula, 2009. DOI: 10.1056/NEJMsa0906126

Gabapentin: results for “primary” outcomes differ between sources

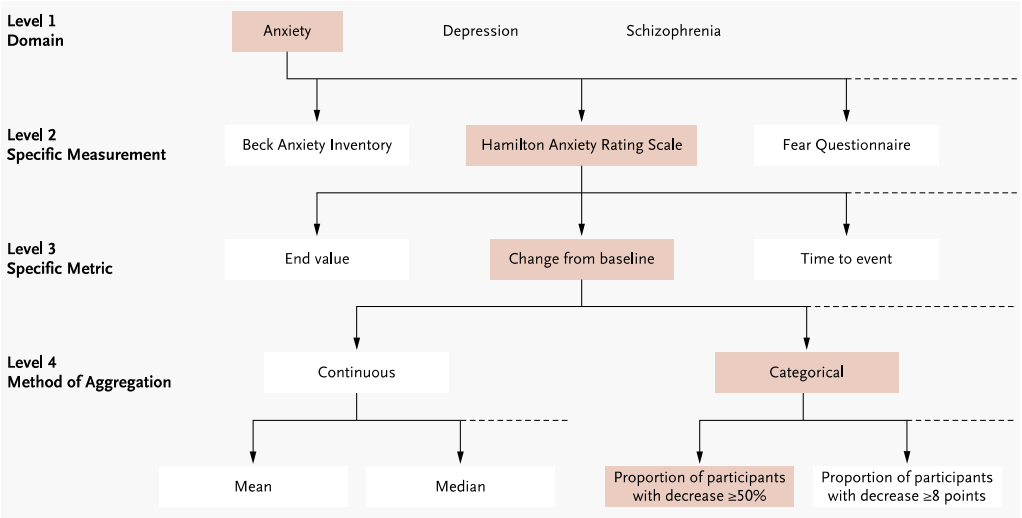




Vedula, 2009. DOI: 10.1056/NEJMsa0906126

Outcomes are defined in many ways

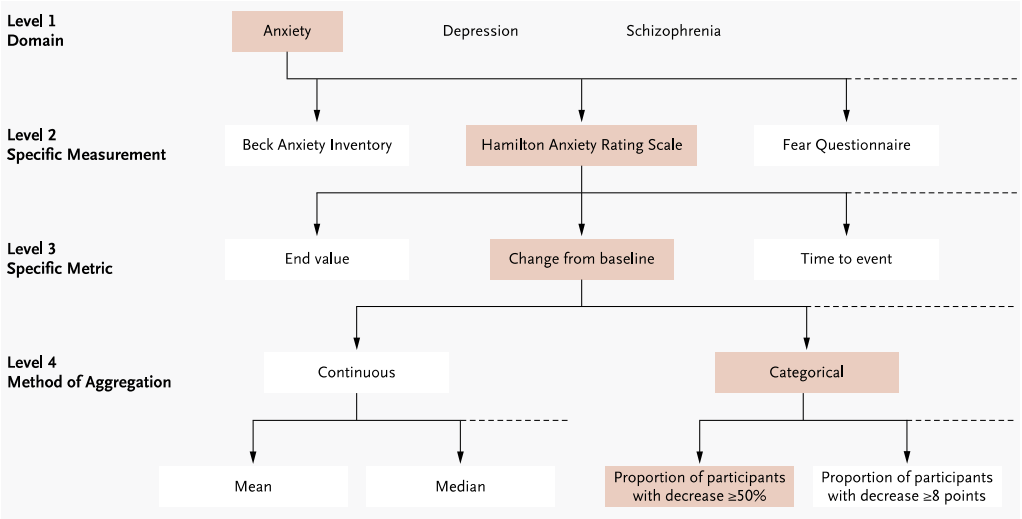
Elements of an outcome on ClinicalTrials.gov

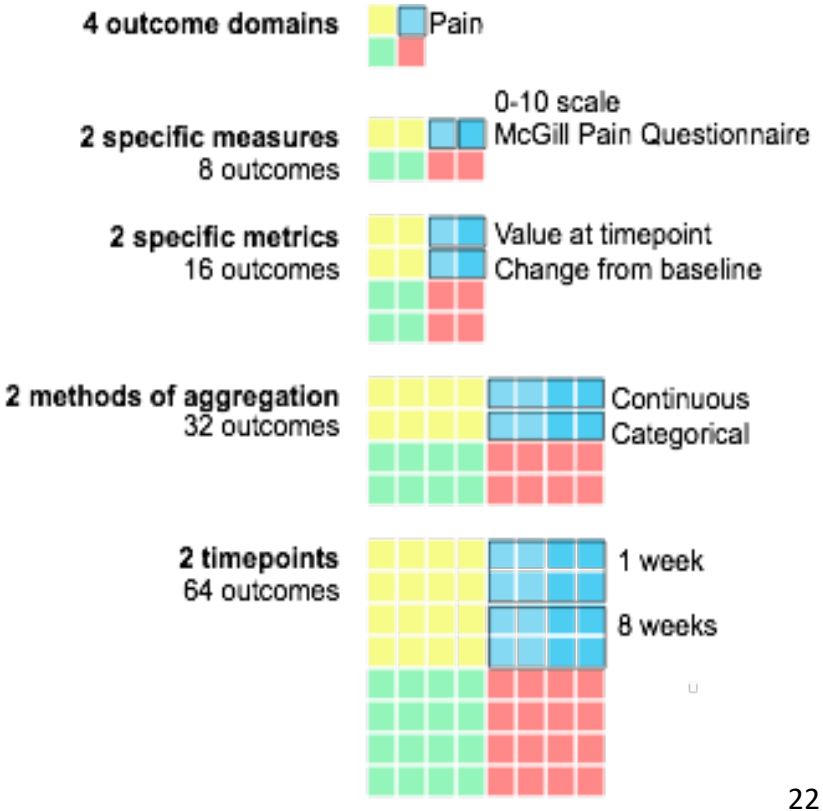


Zarin, 2011. DOI: 10.1056/NEJMsa1012065

Outcomes are defined in many ways

Elements of an outcome on ClinicalTrials.gov





Zarin, 2011. DOI: 10.1056/NEJMsa1012065
Mayo-Wilson, 2017. DOI: 10.1016/j.jclinepi.2017.05.007

Multiple analyses lead to *multiple results for the same outcome*

Analysis population

Participants eligible to be included in the analysis (e.g., people who took one dose, everyone randomized)

Handling missing data

Methods to account for missing data, including missing items and missing cases (e.g., multiple imputation, last observation carried forward)

Methods of analysis

Statistical methods, including analysis model, procedures (e.g., transformations, adjustments), and covariates included in the analysis

Mayo-Wilson, 2017. DOI: 10.1016/j.jclinepi.2017.05.007

Multiple Data Sources (MUDS) Study Design

- Two case studies:
  - Gabapentin for neuropathic pain
  - Quetiapine for bipolar depression

- Participants & investigators masked

- Placebo-controlled, parallel RCTs

- Comprehensive searches for published and unpublished data

Mayo-Wilson, 2015. DOI: 10.1186/s13643-015-0134-z OA

Characteristics of eligible trials

	Gabapentin	Quetiapine
Number of trials	21	7
Dates of reports	1997 to 2013	2003 to 2014
No. public reports / No. all reports	68/74	46/50
Individual participant data (No. trials, % of total)	6 (29%)	1 (14%)
Trial characteristics (No. trials, % of total)		
Manufacturer-funded	14 (67%)	7 (100%)
≥3 groups	11 (52%)	4 (57%)
Multi-center	14 (67%)	6 (86%)
English language	20 (95%)	7 (100%)
Number of participants randomized (median, range)	150 (26 to 452)	526 (100 to 802)
Sources of data for each trial (No. trials, % of all trials)		
Only public	15 (71%)	3 (43%)
Only non-public	1 (5%)	0 (0%)
Both public & non-public	5 (24%)	4 (57%)

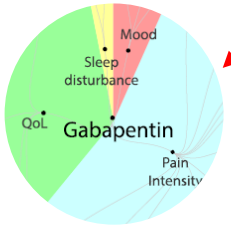
Mayo-Wilson, 2017. DOI: 10.1016/j.jclinepi.2017.05.007
Mayo-Wilson, 2017. DOI: 10.1016/j.jclinepi.2017.07.014

How many outcomes are there in clinical trials?

21 gabapentin trials

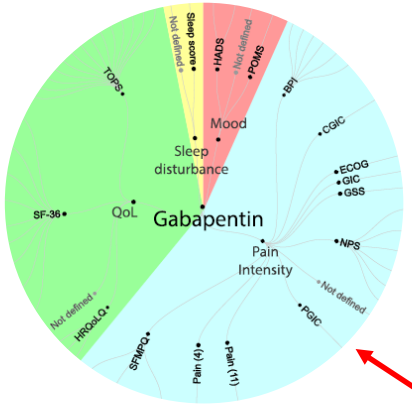
6 with non-public sources

4 Outcome domains



Mayo-Wilson, 2017. DOI: 10.1016/j.jclinepi.2017.05.007

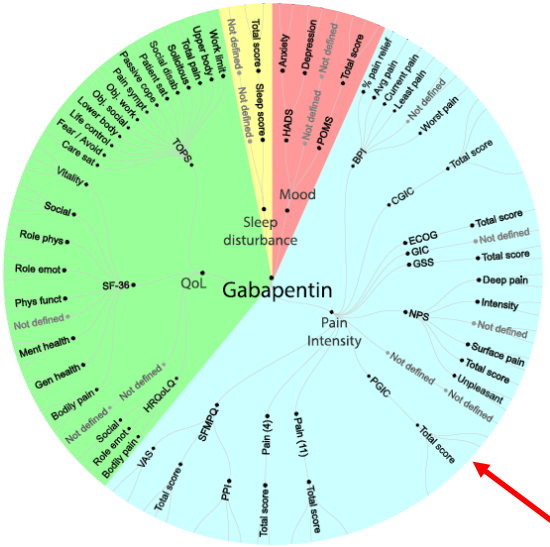
How many outcomes are there in clinical trials?



Multiple measures

Mayo-Wilson, 2017. DOI: 10.1016/j.jclinepi.2017.05.007

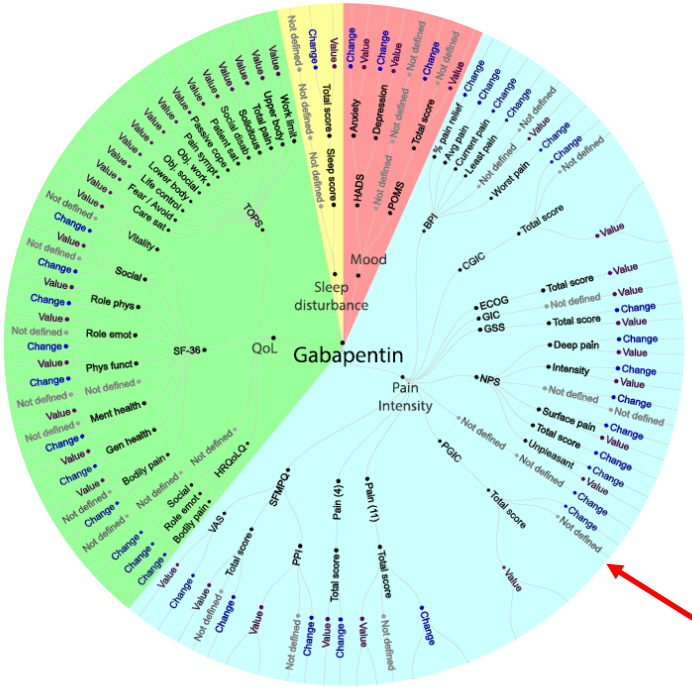
How many outcomes are there in clinical trials?



Multiple totals and subscales

Mayo-Wilson, 2017. DOI: 10.1016/j.jclinepi.2017.05.007

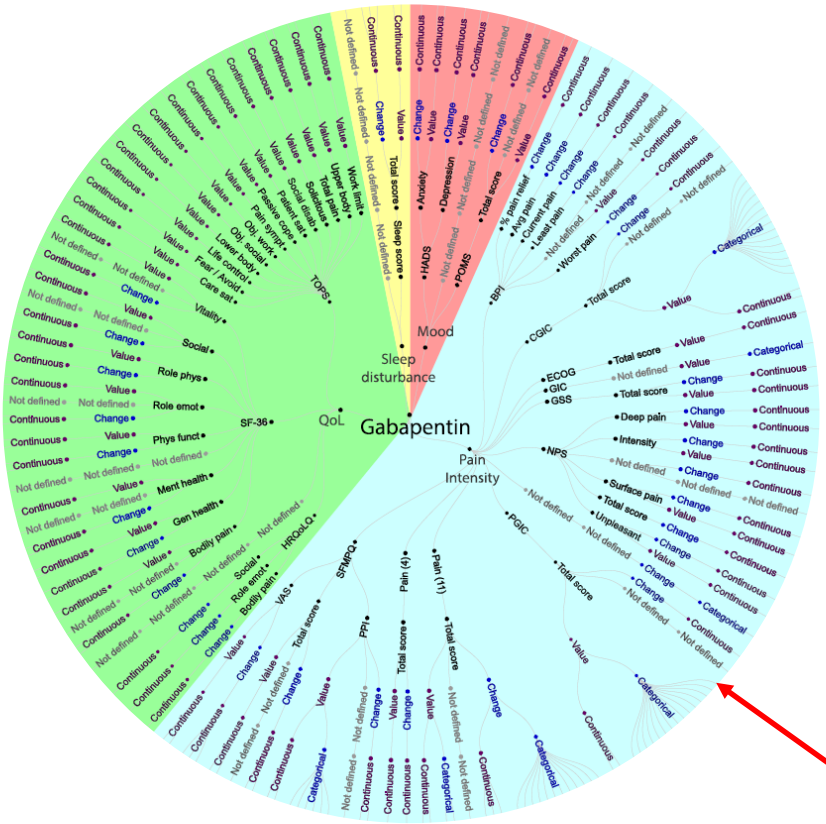
How many outcomes are there in clinical trials?



Multiple metrics

Mayo-Wilson, 2017. DOI: 10.1016/j.jclinepi.2017.05.007

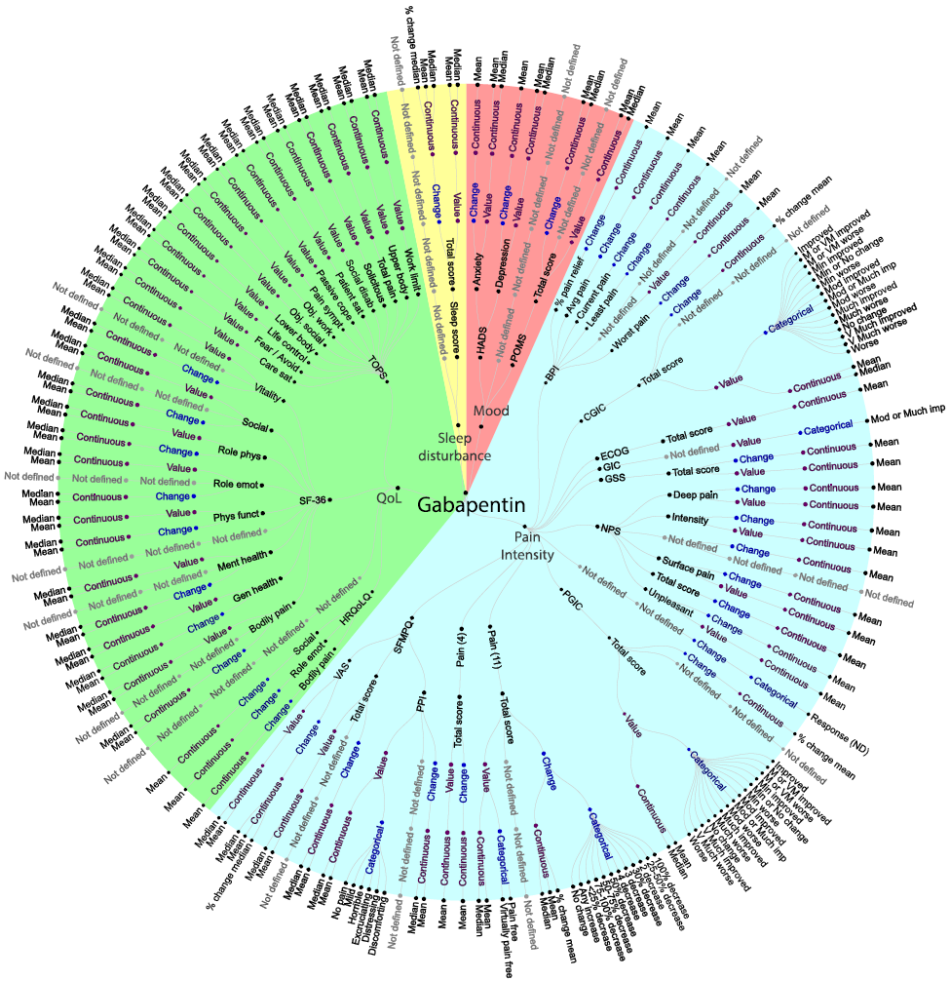
How many outcomes are there in clinical trials?



Multiple methods of aggregation

Mayo-Wilson, 2017. DOI: 10.1016/j.jclinepi.2017.05.007

How many outcomes are there in clinical trials?



214 outcomes

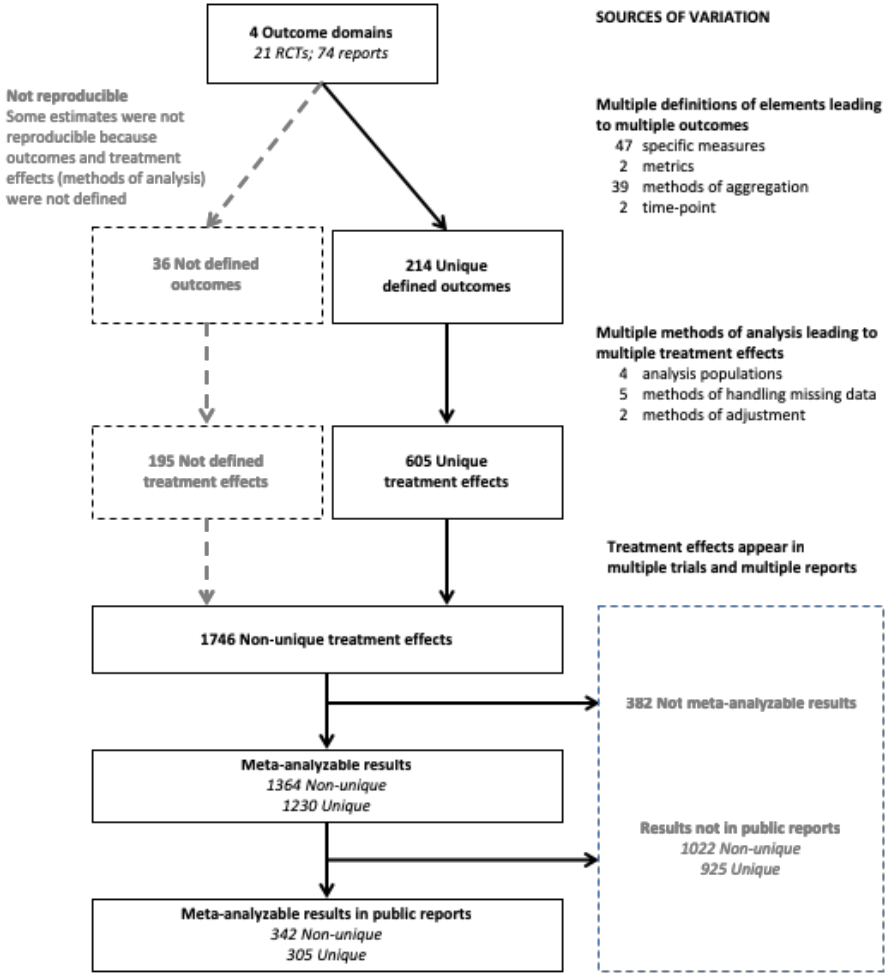
1230 results

305 (25%) publicly reported

More hidden…

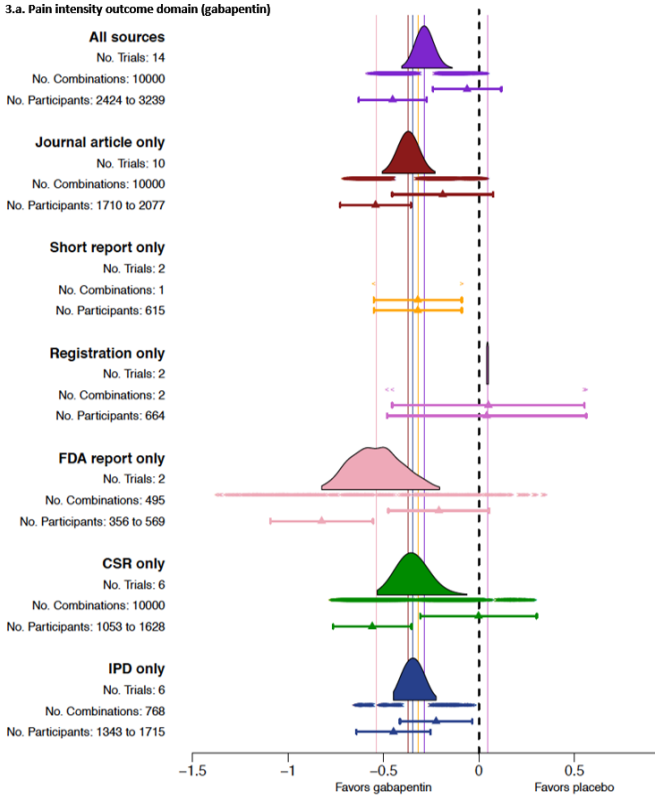
Mayo-Wilson, 2017. DOI: 10.1016/j.jclinepi.2017.05.007

Multiple outcomes and analyses in trials of gabapentin for neuropathic pain

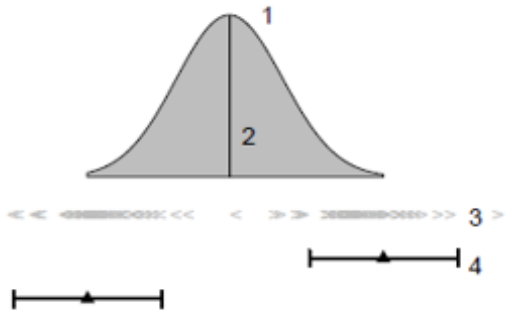


Mayo-Wilson, 2017. DOI: 10.1016/j.jclinepi.2017.05.007

Consequences of multiplicity for systematic reviews



Key



Item 1: Histogram showing the distribution of means (SMDs) from meta-analyses using one continuous effect estimate per study (selected at random)

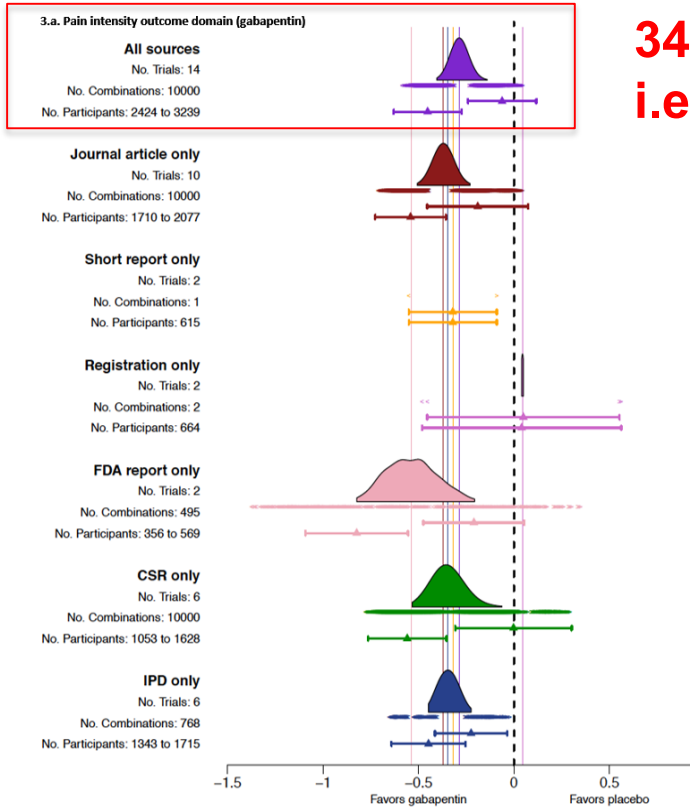
Item 2: Average of the mean effects (SMDs)

Item 3: 95% confidence interval (CI) corresponding to the mean effects (SMDs) in the histogram, including lower (<) and upper (>) limits.

Item 4: The smallest and largest possible treatment effect from a meta-analysis (with associated 95% CI) calculated by selecting the most extreme results from any report about each included trial.

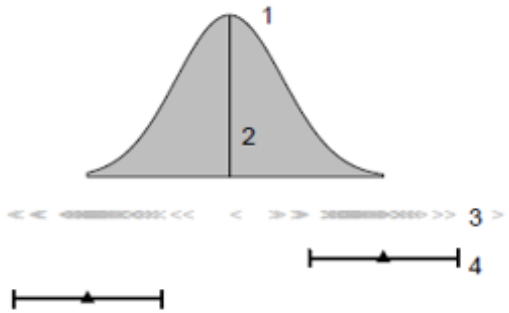
Mayo-Wilson, 2017. DOI: 10.1016/j.jclinepi.2017.07.014

Consequences of multiplicity for systematic reviews



34 trillion possible meta-analyses of "pain" domain i.e., combinations of *the same trials*

Key



Item 1: Histogram showing the distribution of means (SMDs) from meta-analyses using one continuous effect estimate per study (selected at random)

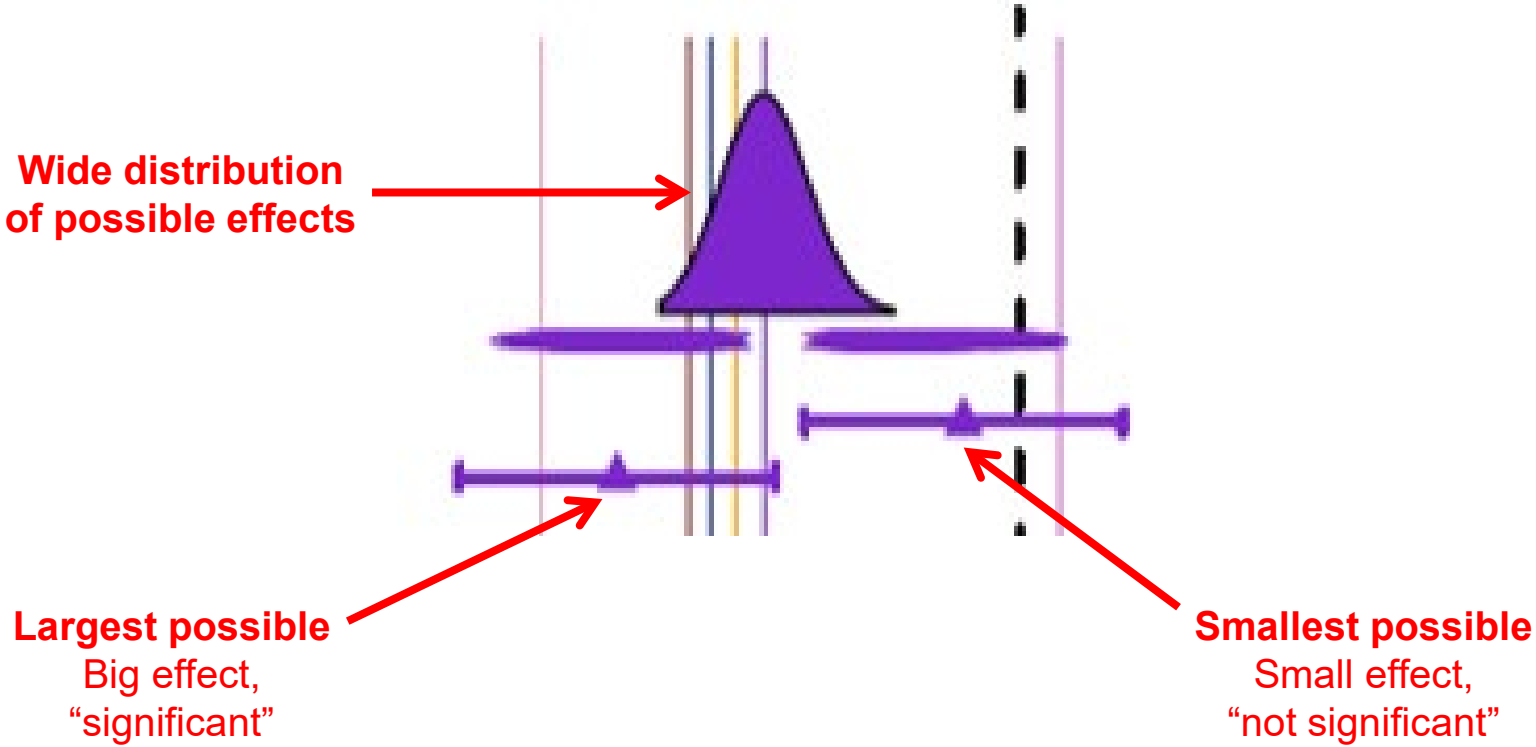
Item 2: Average of the mean effects (SMDs)

Item 3: 95% confidence interval (CI) corresponding to the mean effects (SMDs) in the histogram, including lower (<) and upper (>) limits.

Item 4: The smallest and largest possible treatment effect from a meta-analysis (with associated 95% CI) calculated by selecting the most extreme results from any report about each included trial.

Mayo-Wilson, 2017. DOI: 10.1016/j.jclinepi.2017.07.014

Consequences of multiplicity for systematic reviews



Mayo-Wilson, 2017. DOI: 10.1016/j.jclinepi.2017.07.014

Core outcome sets for clinical trials and systematic reviews

"minimum set of outcome measures that must be reported in all RCTs in a given health condition"

http://www.comet-initiative.org/about/overview
Boers, 2014. DOI: 10.1016/j.jclinepi.2013.11.013

Core outcome sets: IMMPACT recommended outcomes

IMMPACT (pain trials)

1) Pain

a. 11-point (0-10) rating of pain intensity
b. Usage of rescue analgesics
c. Categorical rating of pain intensity

2) Physical functioning (either one of two measures)

a. Multidimensional pain inventory interference scale
b. Brief Pain Inventory interference items

3) Emotional functioning (at least one one of two measures)

a. Beck Depression Inventory
b. Profile of Mood States

4) Participant ratings of global improvement and satisfaction with treatment

a. Patient Global Impression of Change

5) Symptoms and adverse events

a. Passive capture of spontaneously reported adverse events

6) Participant disposition

Dworkin, 2005. DOI: 10.1016/j.pain.2004.09.012

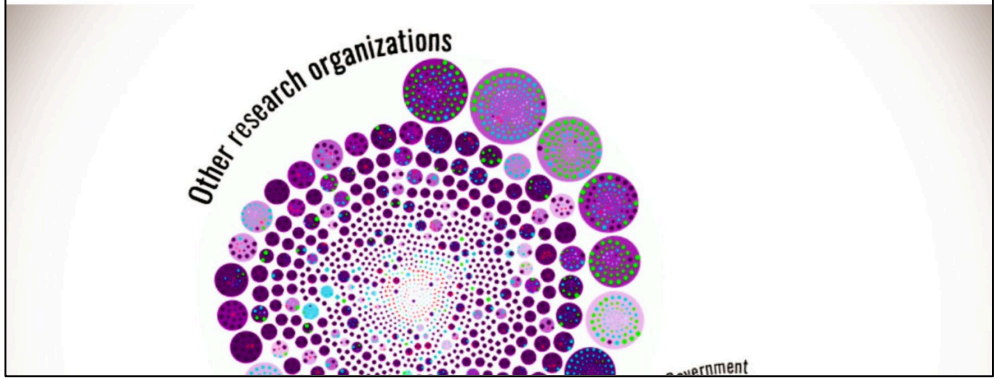
Supporting Investigators in registering and reporting results

STAT Sections Topics Multimedia Newsletters More Log In Subscribe TRY STAT PLUS

A STAT INVESTIGATION

Faced with public pressure, research institutions step up reporting of clinical trial results

By CHARLES PILLER @cpiller *and* TALIA BRONSHTEIN / JANUARY 9, 2018



RESEARCH ARTICLE Open Access

Clinical trial registration and reporting: a survey of academic organizations in the United States

Evan Mayo-Wilson[1*], James Heyward[1], Anthony Keyes[2], Jesse Reynolds[3], Sarah White[4], Nidhi Atri[5], G. Caleb Alexander[6], Audrey Omar[3], Daniel E. Ford[5] and on behalf of the National Clinical Trials Registration and Results Reporting Taskforce Survey Subcommittee

- Account characteristics
- Policies
- Procedures
- Computer systems
- Staff

Mayo-Wilson, 2018. DOI: 10.1186/s12916-018-1042-6

Supporting Investigators in registering and reporting results

Participants

- Invited 783 “University/Organization” accounts
- 366 (47%) partially or fully completed
- Large organizations most likely to participate

Results

- 43% had a trial registration policy
- 35% had a results reporting policy
- 19% used computer software to manage records
- Median staffing 8% of one full-time equivalent

Mayo-Wilson, 2018. DOI: 10.1186/s12916-018-1042-6

Conclusions

- Multiple outcomes, and multiple analyses, lead to many results
- Many results for an “outcome domain” create opportunities for cherry-picking
- To increase consistency, and to reduce research waste, trials and systematic reviews should use core outcome sets
- To prevent bias, outcomes and analysis plans should be registered completely and prospectively
- Institutions can support investigators in meeting registration and reporting requirements

Li, 2018. DOI: 10.1186/s13063-018-2888-9
Mayo-Wilson, 2018. DOI: 10.1002/jrsm.1277

Evan Mayo-Wilson, DPhil, MPA
Assistant Scientist
Center for Clinical Trials and Evidence Synthesis
Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health

Evan.Mayo-Wilson@jhu.edu

www.evanmayo-wilson.org