IBTN 2020 GLOBAL

ibtn
international
behavioural
trials network

Outcomes: do's and dont's

Evan Mayo-Wilson, MPA, DPhil
Associate Professor
Department of Epidemiology and Biostatistics
Indiana University School of Public Health-Bloomington

May 28, 2020

Sources of support

American Public Health Association (*American Journal of Public Health*)

European Social Research Council (ESRC)

Arnold Ventures (Laura and John Arnold Foundation)

National Institutes of Health (NIH)

National Institute of Justice

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

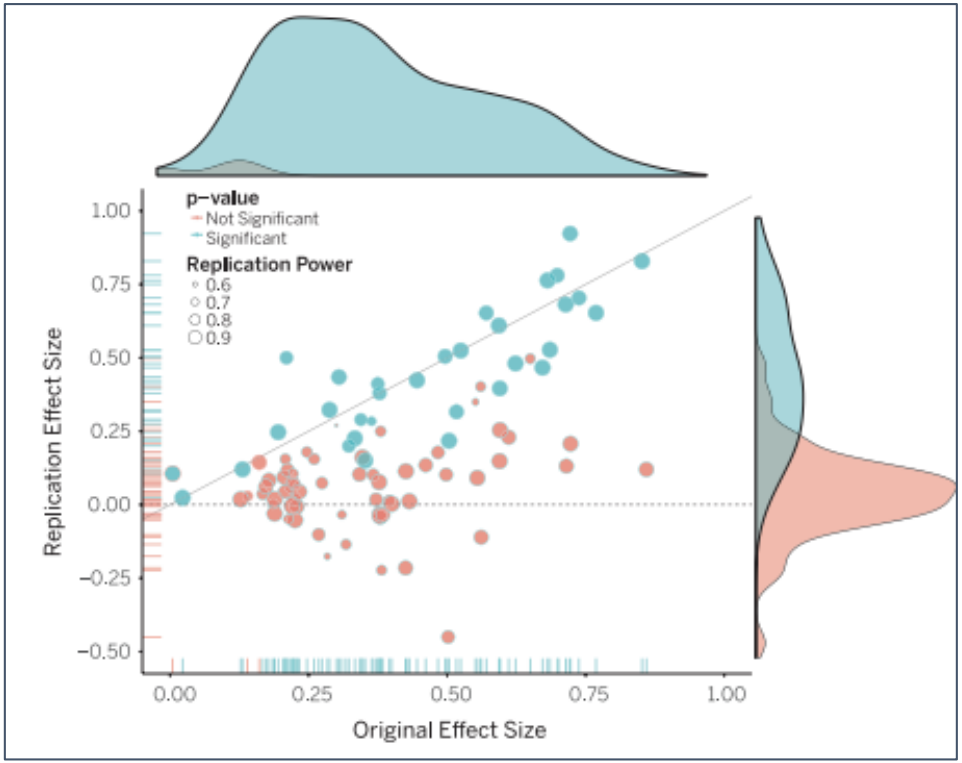
Swedish Board of Health and Welfare (Socialstyrelsen)

Patient Centered Outcomes Research Institute (PCORI)

Robert Wood Johnson Foundation

U.S. Food and Drug Administration (FDA)

Replication project



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

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 overinterpretation 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 remains a troubling frequency of published reports that claim a significant re-

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

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.

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 $Pr(E\|H_0) \leq .05$ (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

“Publication strategy” for Neurontin (gabapentin)

Earnings limited for epilepsy

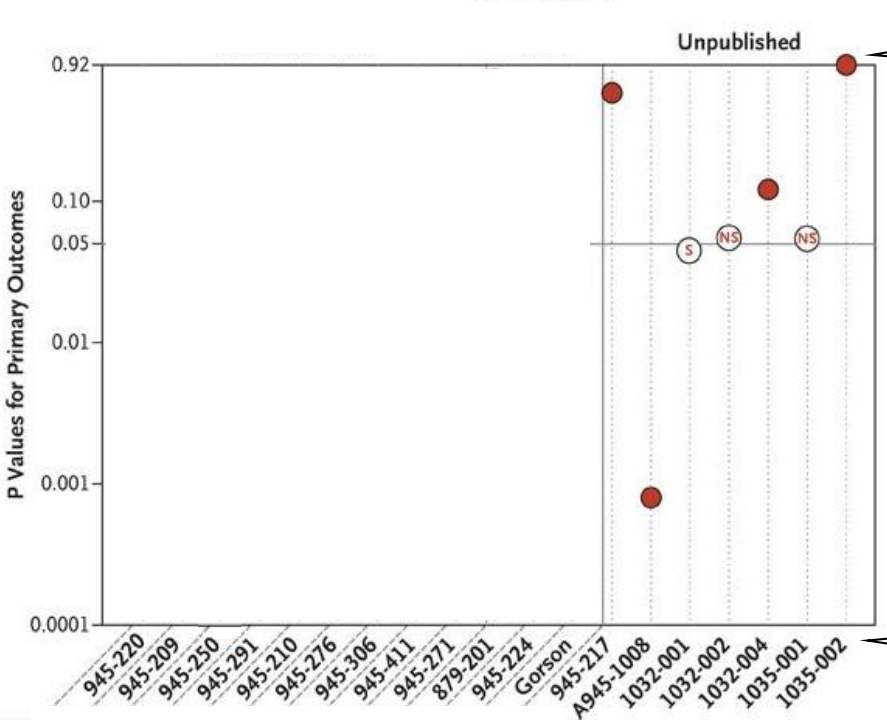
Marketing assessment

- Bipolar disorder
- Migraine
- Neuropathic pain
- Nociceptive pain
- Social anxiety disorder

Used “publication strategy” rather than “indication strategy”

Vedula, et al., 2009. DOI: 10.1056/NEJMsa0906126
Vedula, et al., 2012. DOI: 10.1186/1745-6215-13-136

Results for “primary” outcomes differ between sources

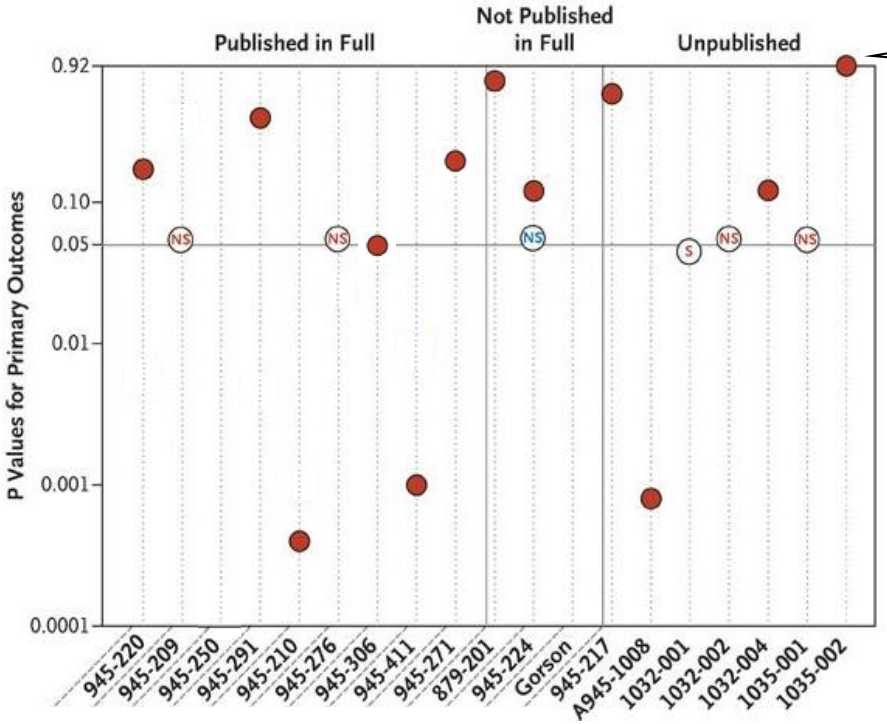


Primary outcome in unpublished research report (red)

Trial number

Vedula, 2009. DOI: 10.1056/NEJMsa0906126

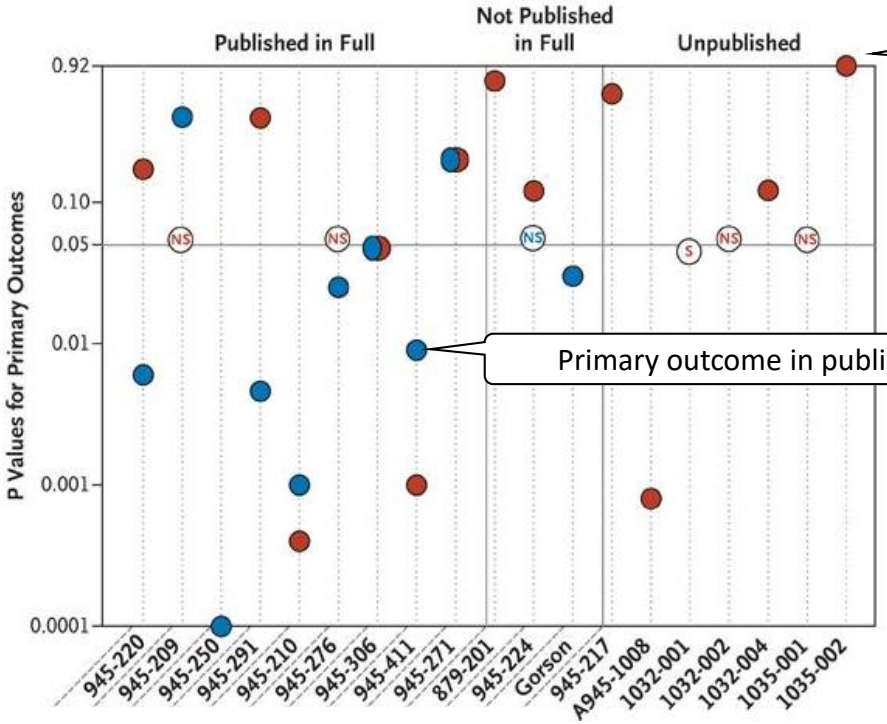
Results for “primary” outcomes differ between sources



Primary outcome in unpublished research report (red)

Vedula, 2009. DOI: 10.1056/NEJMsa0906126

Results for “primary” outcomes differ between sources

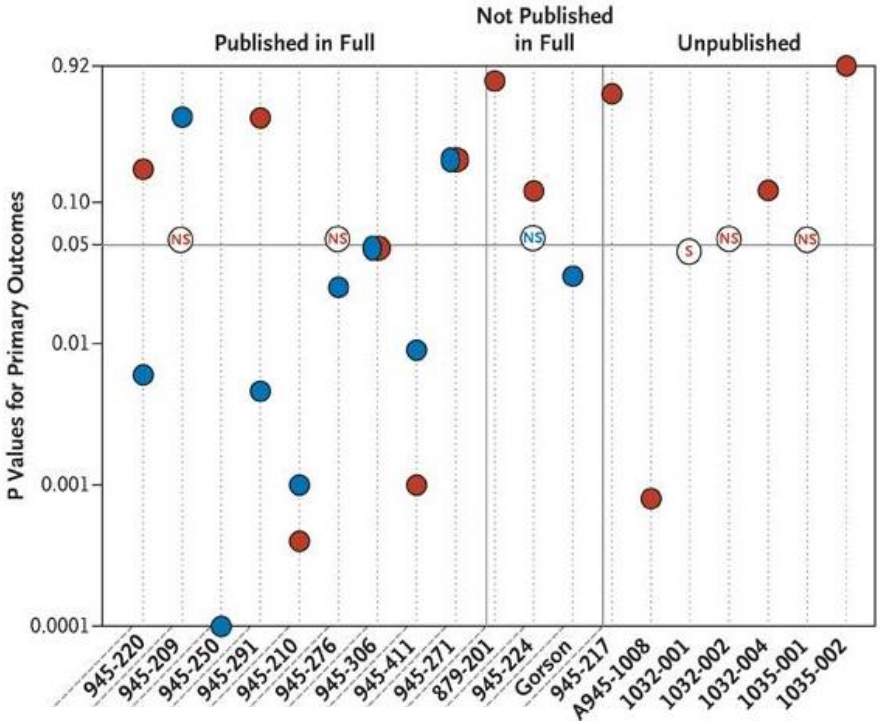


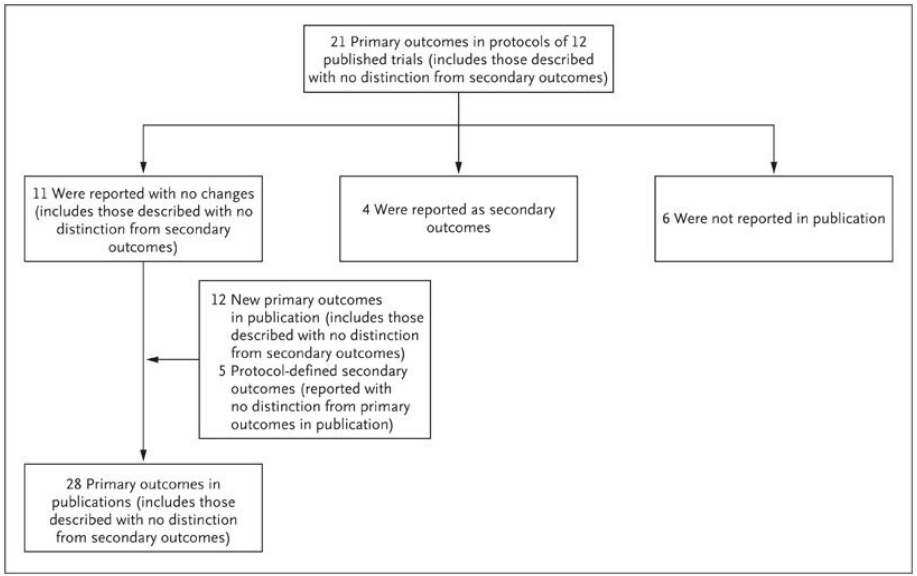
Primary outcome in unpublished research report (red)

Primary outcome in published journal article (blue)

Vedula, 2009. DOI: 10.1056/NEJMsa0906126

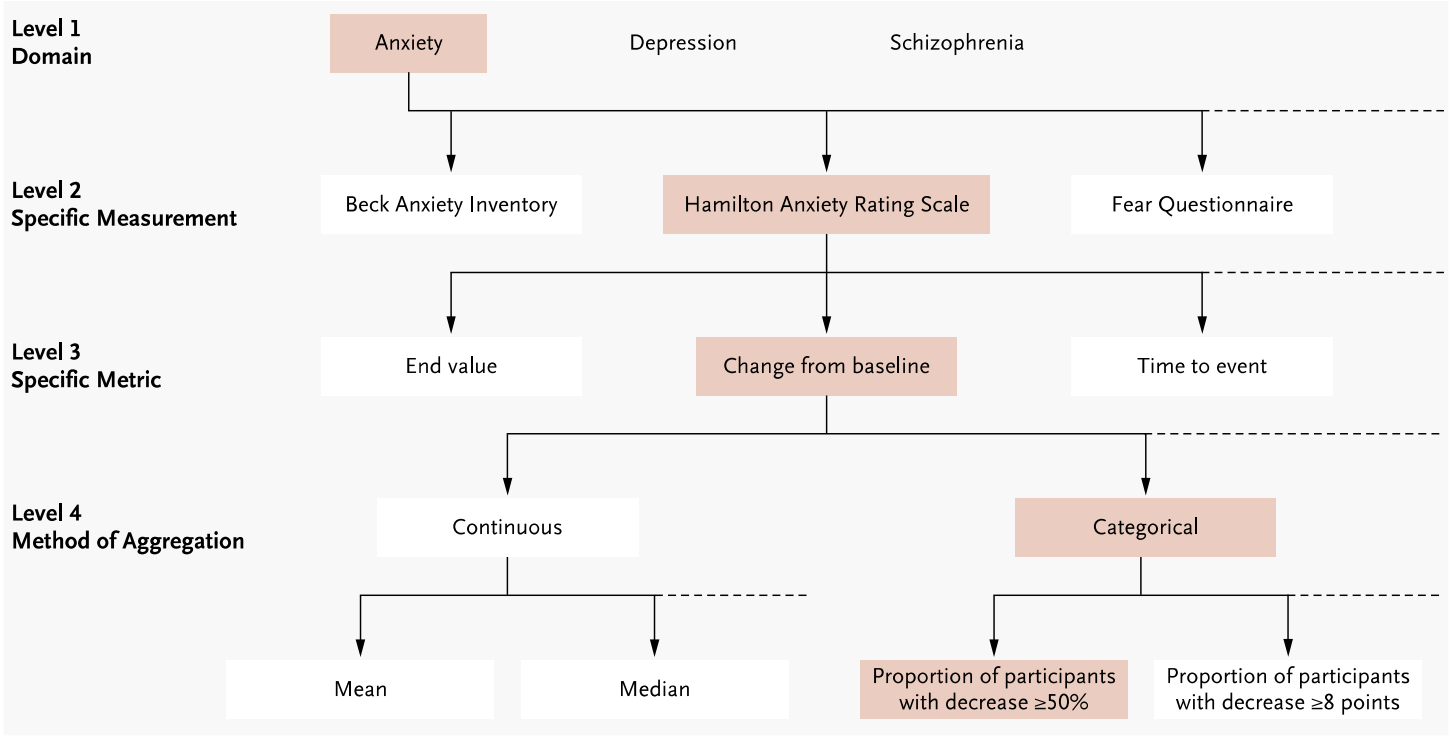
Results for “primary” outcomes differ between sources





Vedula, 2009. DOI: 10.1056/NEJMsa0906126

Elements of an outcome



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

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, *et al.,* 2017. DOI: 10.1016/j.jclinepi.2017.05.007

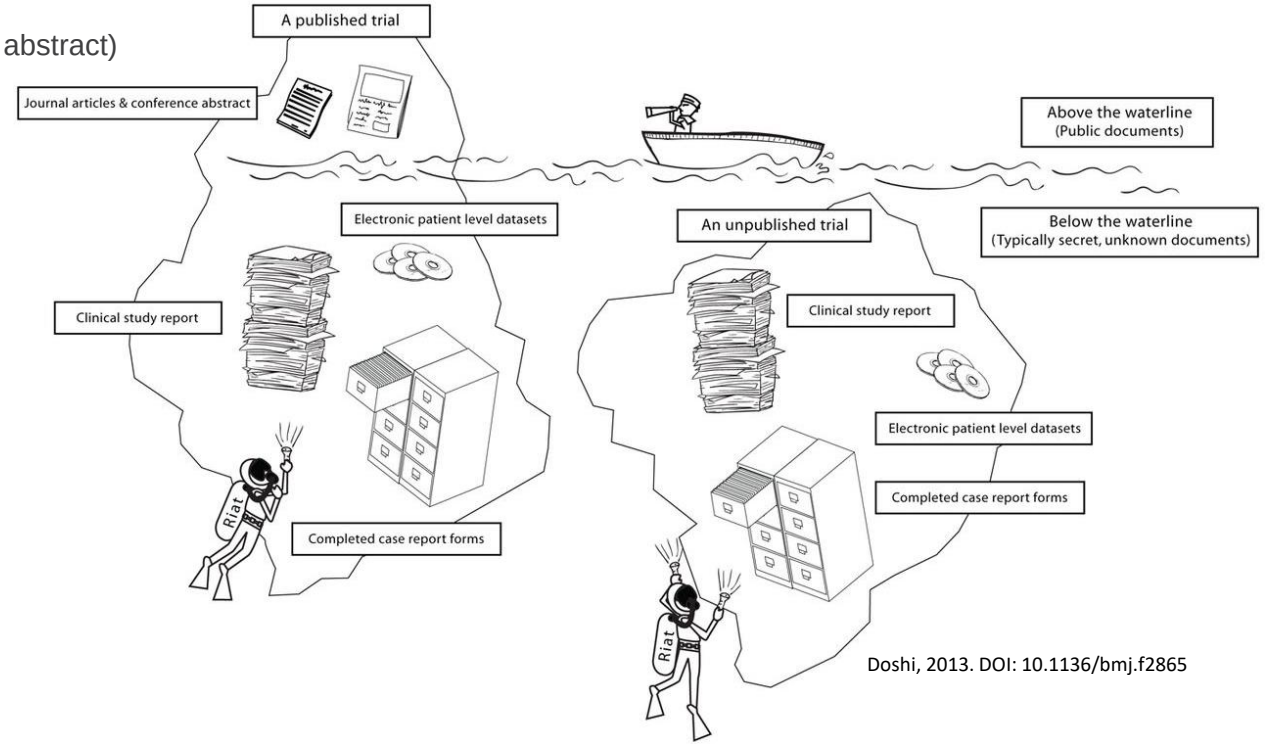
Multiple data sources

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

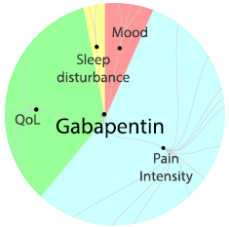
- Unpublished manuscript
- Individual participant data (IPD)
- Grant proposal
- Study protocol
- Case report form
- Memos and emails



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

Doshi, 2013. DOI: 10.1136/bmj.f2865

The problem of multiple outcome definitions



21 trials

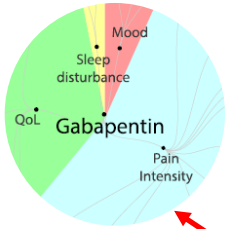
6 with non-public sources

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

The problem of multiple outcome definitions

21 trials

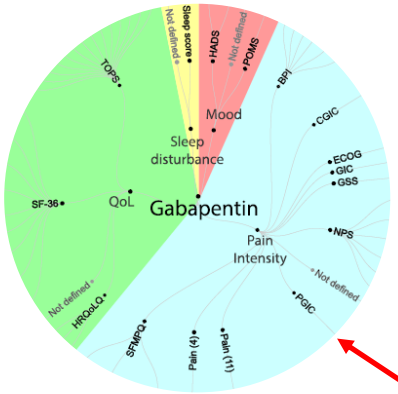
6 with non-public sources



4 Outcome domains

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

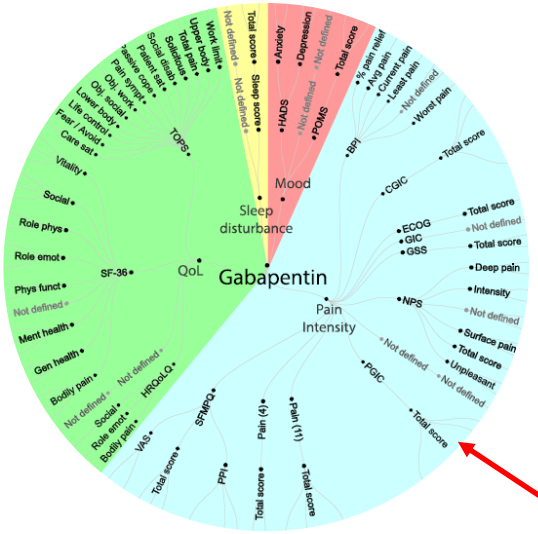
The problem of multiple outcome definitions



Multiple measures

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

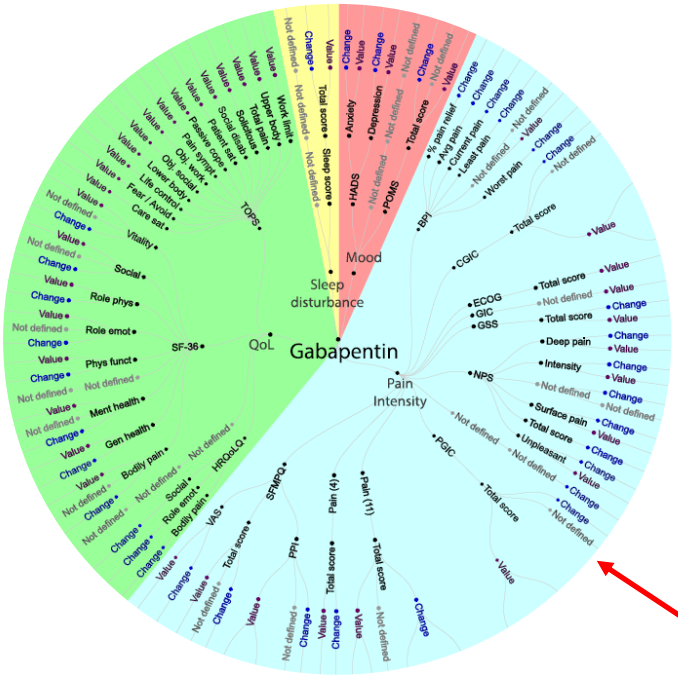
The problem of multiple outcome definitions



Multiple totals and subscales

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

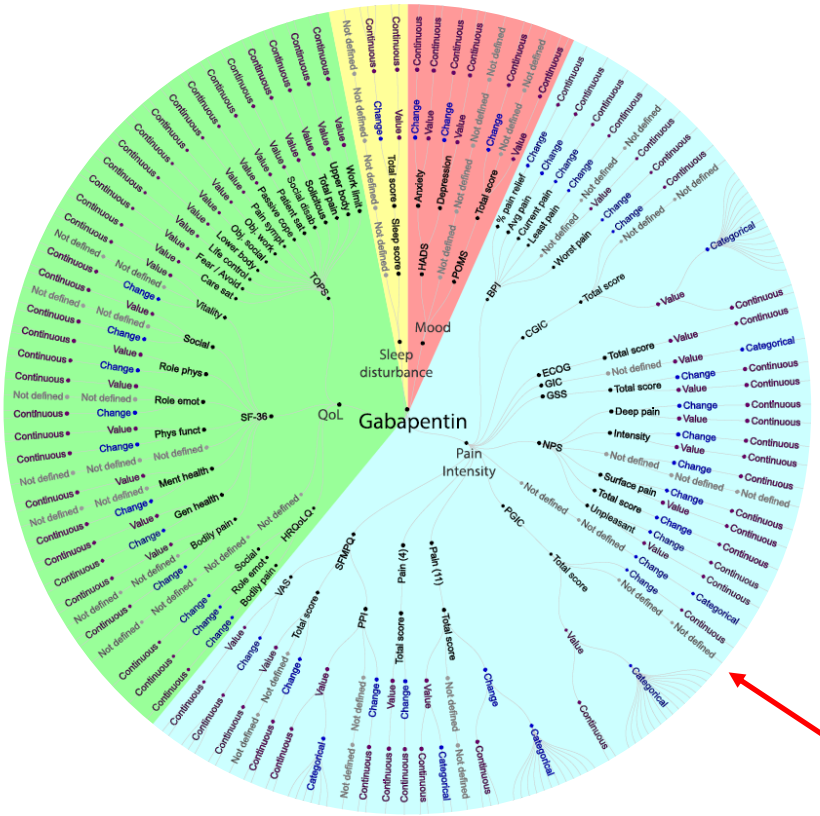
The problem of multiple outcome definitions



Multiple metrics

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

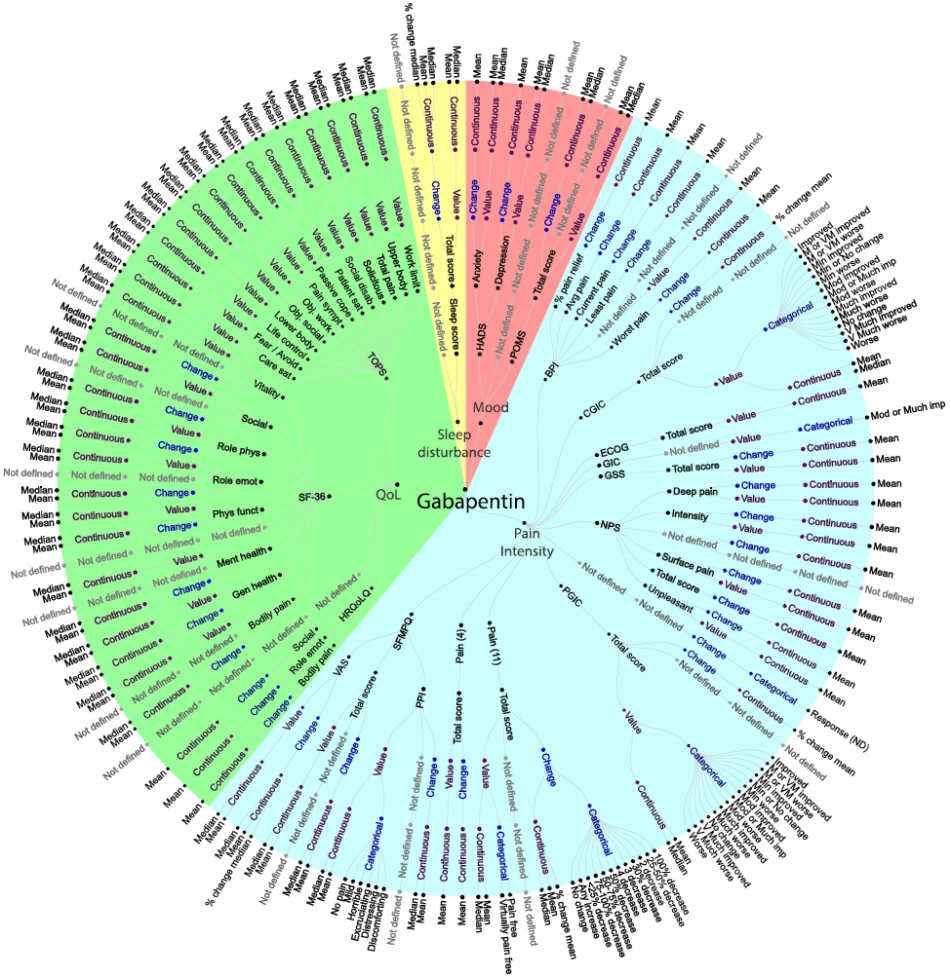
The problem of multiple outcome definitions



Multiple methods of aggregation

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

The problem of multiple outcome definitions



21 trials

214 outcomes

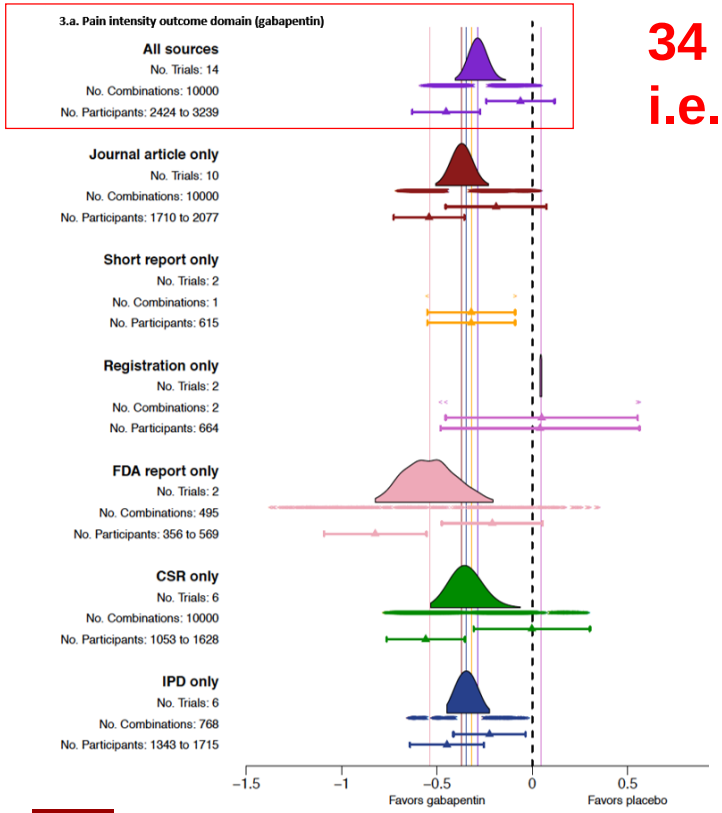
1230 results

305 (25%) publicly reported

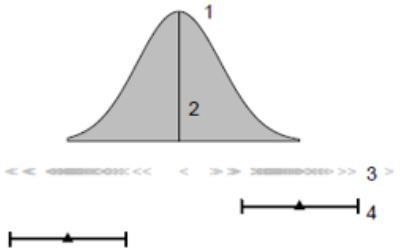
More hidden…

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

Consequences of multiplicity for meta-analysis



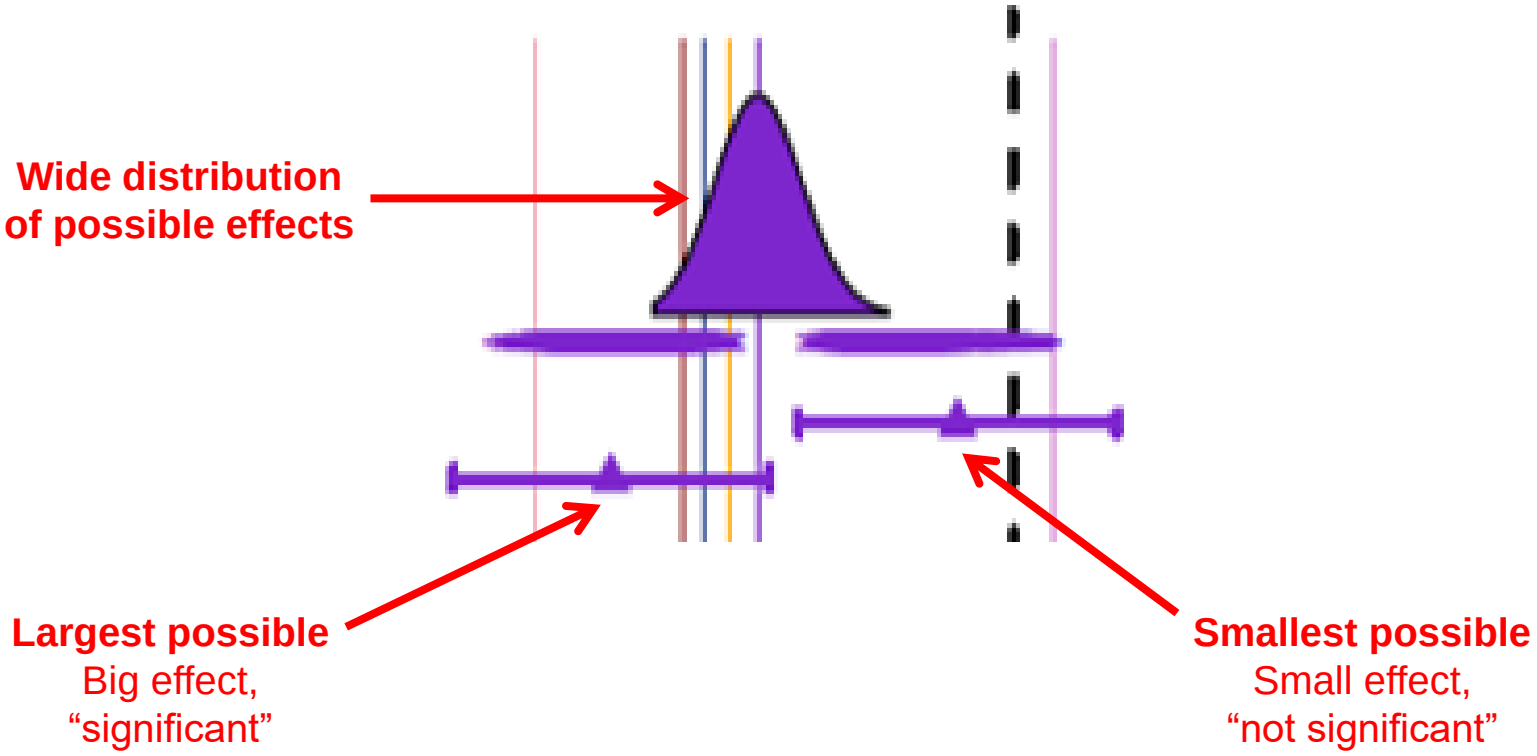
34 trillion possible meta-analyses of "pain"
i.e., combinations of *the same trials*



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, *et al.,* 2017. DOI: 10.1016/j.jclinepi.2017.07.014

Consequences of multiplicity for meta-analysis



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

Registration in clinical psychology

25 journals (Instructions to Authors)

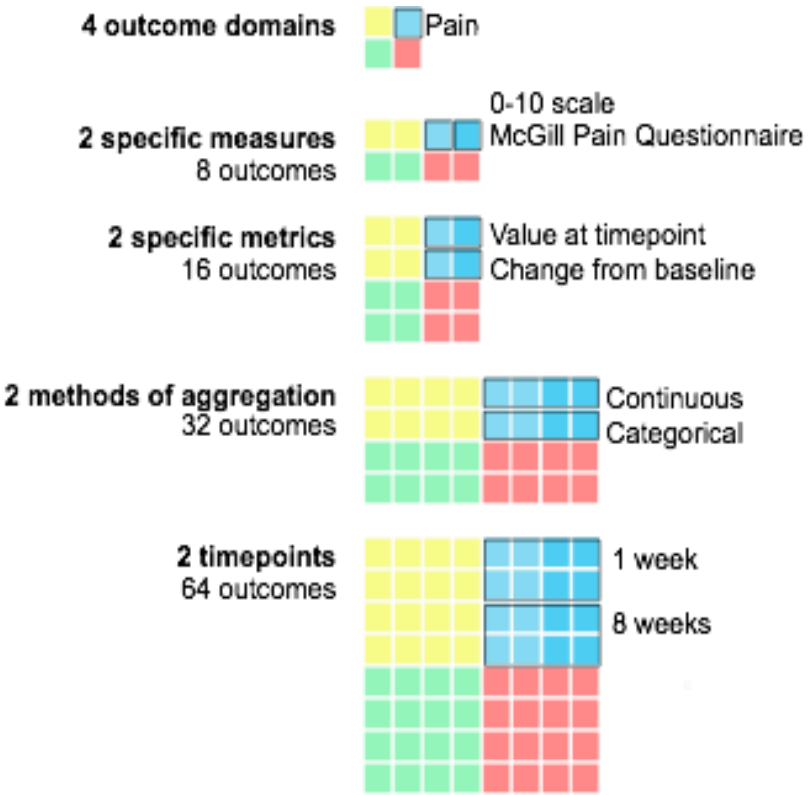
- 14 (56%) reference reporting guidelines
- 3 (12%) require registration

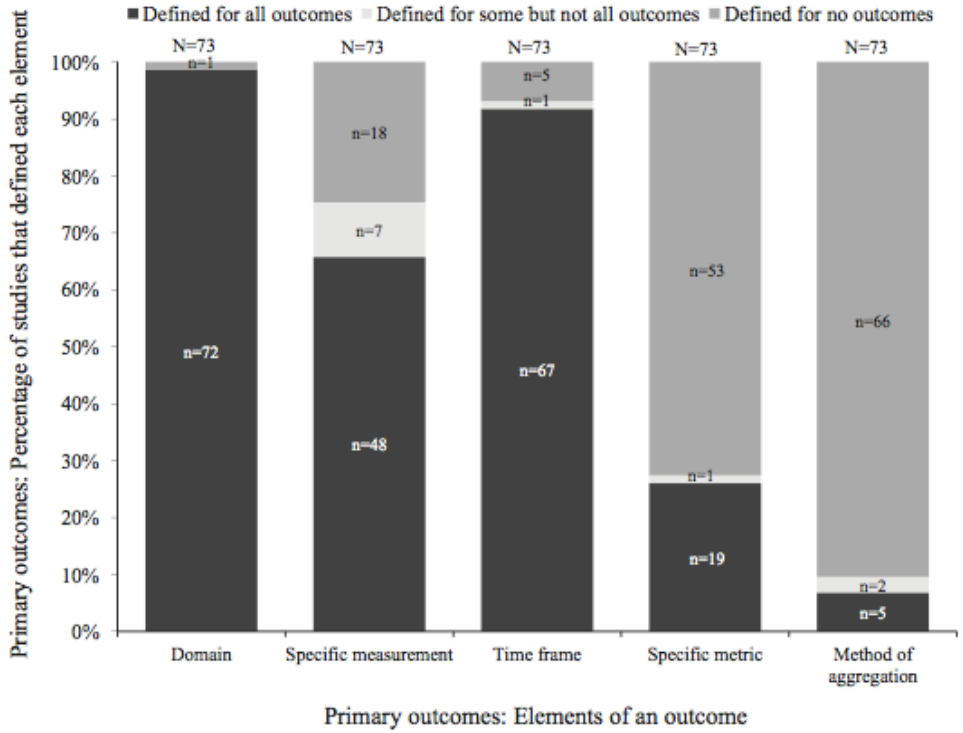
170 RCTs

- 38 (22%) reported registration status
- 68 (40%) registered

Cybulski *et al.*, 2016. DOI: 10.1037/ccp0000115

Outcomes not fully defined





Mayo-Wilson, *et al.,* 2017. DOI: 10.1016/j.jclinepi.2017.05.007
Cybulski, *et al.,* 2016. DOI: 10.1037/ccp0000115

Assessing benefits and harms

BENEFITS AND SYSTEMATIC HARMS	NON-SYSTEMATIC HARMS
Measured for all participants in the same way	Reported spontaneously by patients
Selected *a priori*	Reported based on the results
Analyzed and reported using pre-specified methods	Methods for analysis / reporting often unclear

Assessing potential benefits of Aristada (aripiprazole)

Figure 2. Mean Change From Baseline to Each Assessment for the Positive and Negative Syndrome Scale (PANSS) Total Score Over 85 Days[a]

A. Analysis of Covariance With Last Observation Carried Forward Analysis, Full Analysis Set

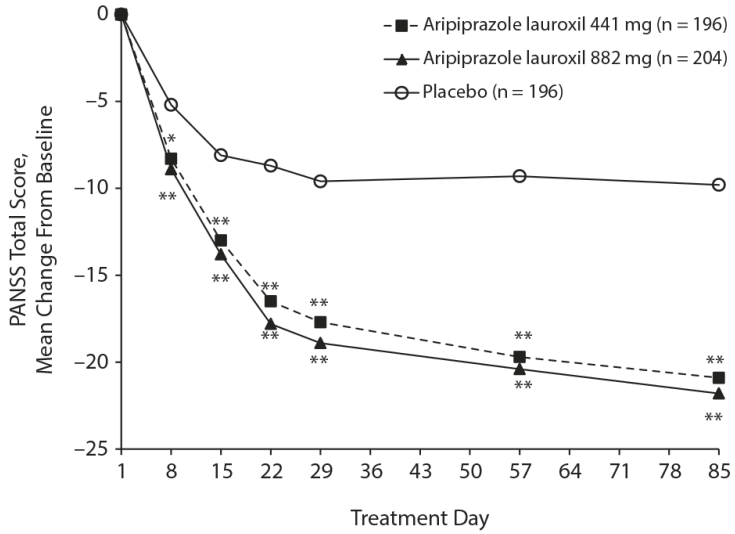
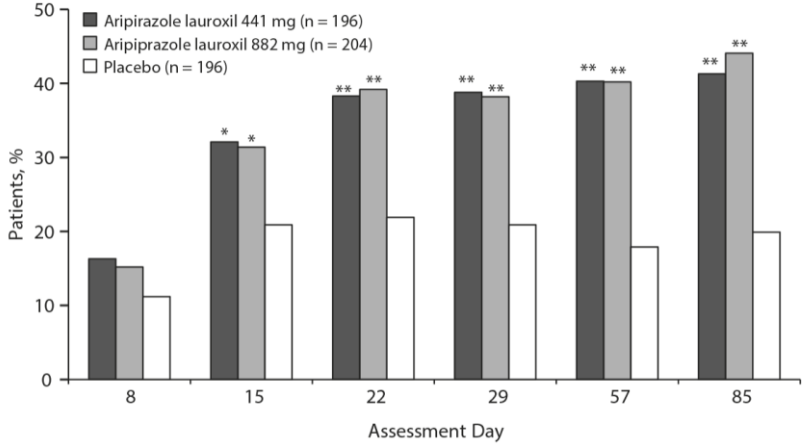


Figure 3. Proportion of Patients Reporting Ratings of Very Much or Much Improved on the Clinical Global Impressions-Improvement Scale[a]



[a]Proportion of patients with very much improved or much improved in full analysis set at each assessment time points. *P* values are for the aripiprazole lauroxil 441-mg and 882-mg dose groups versus placebo. Logistic regression model adjusting for study region. Missing values were imputed with no improvement.

*$P < .05$.

**$P < .001$.

Meltzer, 2015. DOI: 10.4088/JCP.14m09741

Assessing potential harms of Aristada (aripiprazole)

Table 2. Treatment-Emergent Adverse Events (TEAEs) Occurring in ≥ 2% of Aripiprazole Lauroxil–Treated Patients, Safety Population

	Aripiprazole Lauroxil		
Preferred Term (%)	441 mg (n = 207)	882 mg (n = 208)	Placebo (n = 207)
Any TEAE	58.9	57.2	62.3
Insomnia	9.7	12.0	11.6
Akathisia	11.6	11.5	4.3
Headache	8.2	8.7	8.2
Anxiety	2.9	5.3	6.8
Injection site pain	3.4	4.8	1.9
Toothache	2.4	3.8	0.5
Nausea	2.9	3.4	1.9
Constipation	2.9	2.4	3.9
Diarrhea	2.4	2.4	3.4
Weight increase	2.9	2.4	0.5
Neck pain	1.0	2.4	1.4
Sedation	1.9	2.4	1.4
Schizophrenia	5.8	2.4	10.6
Restlessness	2.9	1.9	1.9
Blood CPK increase	4.3	1.4	0.5

Abbreviation: CPK = creatinine phosphokinase.

"Treatment-emergent adverse events occurring in ≥ 2% of patients in the aripiprazole lauroxil treatment groups are reported in Table 2. The most common TEAEs occurring in > 5% of patients in the aripiprazole lauroxil groups were insomnia, akathisia, headache, and anxiety. Akathisia was the only TEAE with an incidence of ≥5% in each aripiprazole lauroxil group that was at least twice the rate of placebo (11.6%, 11.5%, and 4.3%). The majority (> 75%) of all akathisia episodes occurred before the second injection, generally within the first 3 weeks, when the patients in the aripiprazole lauroxil groups were also receiving oral aripiprazole. There were 3 cases of akathisia that occurred after the second injection in the aripiprazole lauroxil 441-mg group and 1 case in the placebo group. No cases of akathisia occurred in the aripiprazole lauroxil 882-mg group beyond 1 month after the first injection."

Meltzer, 2015. DOI: 10.4088/JCP.14m09741

Prescribing information ("drug label")

HIGHLIGHTS OF PRESCRIBING INFORMATION
These highlights do not include all the information needed to use ARISTADA™ safely and effectively. See full prescribing information for ARISTADA™.

ARISTADA™ (aripiprazole lauroxil) extended-release injectable suspension, for intramuscular use
Initial U.S. Approval: 2015

WARNING: INCREASED MORTALITY IN ELDERLY PATIENTS WITH DEMENTIA-RELATED PSYCHOSIS
See full prescribing information for complete boxed warning.

- **Elderly patients with dementia-related psychosis treated with antipsychotic drugs are at an increased risk of death. (5.1)**
- **ARISTADA is not approved for the treatment of patients with dementia-related psychosis. (5.1)**

INDICATIONS AND USAGE

ARISTADA is an atypical antipsychotic indicated for the treatment of schizophrenia (1).

DOSAGE AND ADMINISTRATION

- To be administered by intramuscular injection in the deltoid (441 mg dose only) or gluteal (441 mg, 662 mg or 882 mg) muscle by a healthcare professional (2.1).
- For patients naïve to aripiprazole, establish tolerability with oral aripiprazole prior to initiating treatment with ARISTADA (2.1).
- ARISTADA can be initiated at a dose of 441 mg, 662 mg or 882 mg administered monthly or 882 mg dose every 6 weeks (2.1).
- In conjunction with the first ARISTADA injection, administer treatment with oral aripiprazole for 21 consecutive days (2.1).
- Dosing regimen adjustments may be required for missed doses (2.2).
- Dose adjustments are required for 1) known CYP2D6 poor metabolizers and 2) for patients taking CYP3A4 inhibitors, CYP2D6 inhibitors, or CYP3A4 inducers for more than 2 weeks (2.4).

DOSAGE FORMS AND STRENGTHS

For extended-release injectable suspension: 441 mg, 662 mg or 882 mg single-use pre-filled syringe (3)

CONTRAINDICATIONS

Known hypersensitivity to aripiprazole (4)

WARNINGS AND PRECAUTIONS

- *Cerebrovascular Adverse Reactions in Elderly Patients with Dementia-Related Psychosis:* Increased incidence of cerebrovascular adverse reactions (e.g., stroke, transient ischemia attack, including fatalities) (5.2).
- *Neuroleptic Malignant Syndrome*: Manage with immediate discontinuation and close monitoring (5.3).
- *Tardive Dyskinesia*: Discontinue if clinically appropriate (5.4).
- *Metabolic Changes*: Monitor for hyperglycemia, dyslipidemia, and weight gain (5.5).
- *Orthostatic Hypotension*: Monitor heart rate and blood pressure and warn patients with known cardiovascular or cerebrovascular disease, and risk of dehydration or syncope (5.6).
- *Leukopenia, Neutropenia, and Agranulocytosis*: Perform complete blood counts in patients with a history of a clinically significant low white blood cell (WBC) count. Consider discontinuation if clinically significant decline in WBC in the absence of other causative factors (5.7).
- *Seizures*: Use cautiously in patients with a history of seizures or with conditions that lower the seizure threshold (5.8).
- *Potential for Cognitive and Motor Impairment*: Use caution when operating machinery (5.9).

ADVERSE REACTIONS

Most commonly observed adverse reaction with ARISTADA (incidence ≥5% and at least twice that for placebo) was akathisia (6.1).

To report SUSPECTED ADVERSE REACTIONS, contact Alkermes, Inc. at 1-866-274-7823 or FDA at 1-800-FDA-1088 or *www.fda.gov/medwatch*.

USE IN SPECIFIC POPULATIONS

- *Pregnancy:* May cause extrapyramidal and/or withdrawal symptoms in neonates in women exposed during the third trimester of pregnancy (8.1).

See 17 for PATIENT COUNSELING INFORMATION and Medication Guide.

https://www.accessdata.fda.gov/drugsatfda_docs/label/2015/207533s000lbl.pdf

Non-systematic AEs are reported inconsistently across sources

ARISTADA (aripiprazole laurixil)
(air-is-TAH-dah)
Alkermes Inc.
Approval date: October 5, 2015

1) Snapshot

Table 3. Adverse Reactions that Occurred in 2% or more of ARISTADA-Treated Patients and at Greater Incidence than in the Placebo-Treated Patients

2) Prescribing information

ADVERSE REACTIONS

Most commonly observed adverse reaction with ARISTADA (incidence ≥5% and at least twice that for placebo) was akathisia (6.1).

3) Trial registration (NCT01469039)

Frequency Threshold

Threshold above which other adverse events are reported	5%

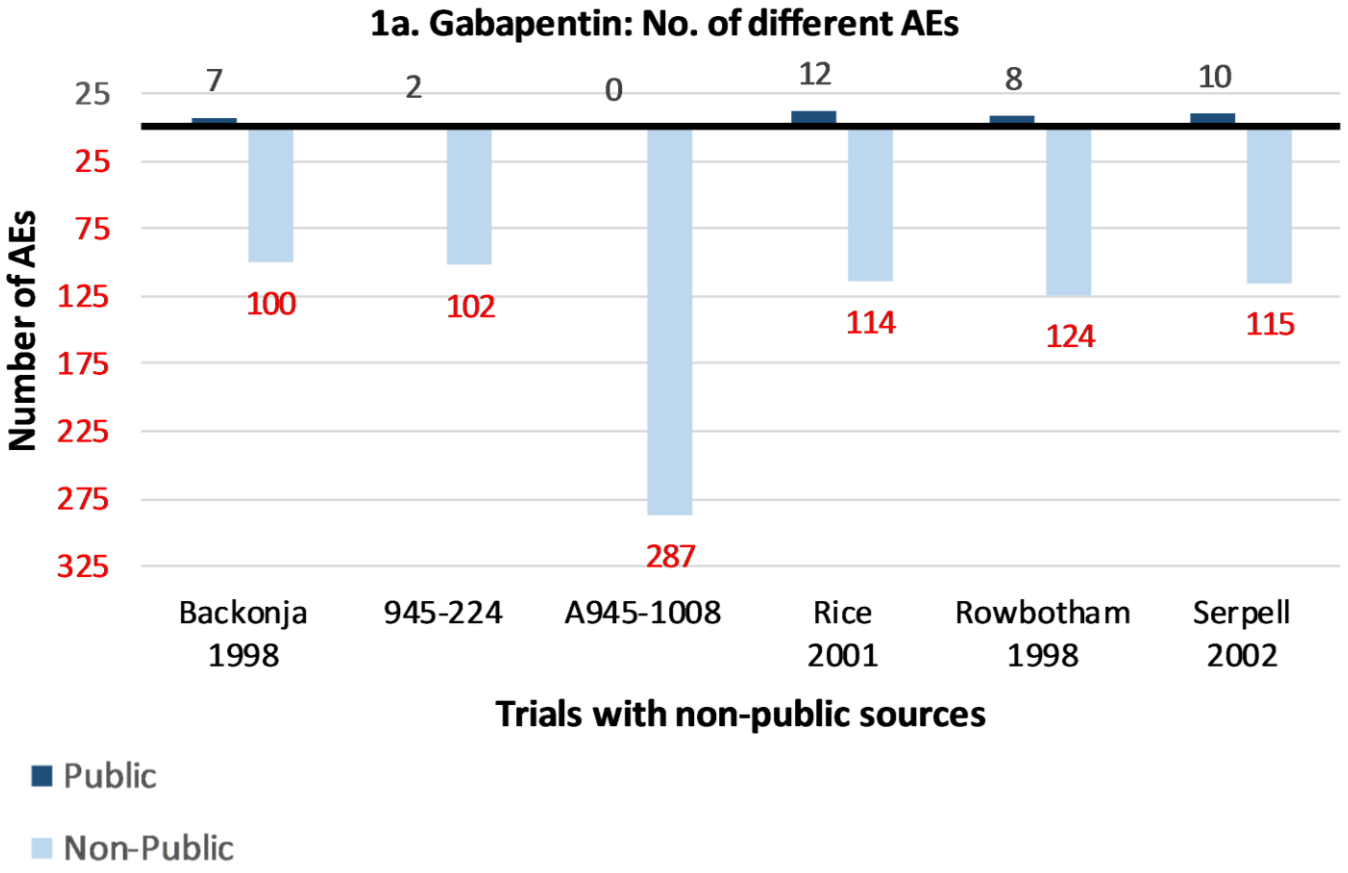
4) Journal article (Meltzer *et al.*, 2016)

Treatment-emergent adverse events occurring in ≥2% of patients in the aripiprazole lauroxil treatment groups are reported in Table 2. The most common TEAEs occurring in >5% of patients in the aripiprazole lauroxil groups were insomnia, akathisia, headache, and anxiety. Akathisia was the only TEAE with an incidence of ≥5% in each aripiprazole lauroxil group that was at least twice the rate of placebo (11.6%, 11.5%, and 4.3%). The majority (>75%) of all akathisia episodes occurred before the second injection,

Table 2. Treatment-Emergent Adverse Events (TEAEs) Occurring in ≥2% of Aripiprazole Lauroxil–Treated Patients, Safety Population

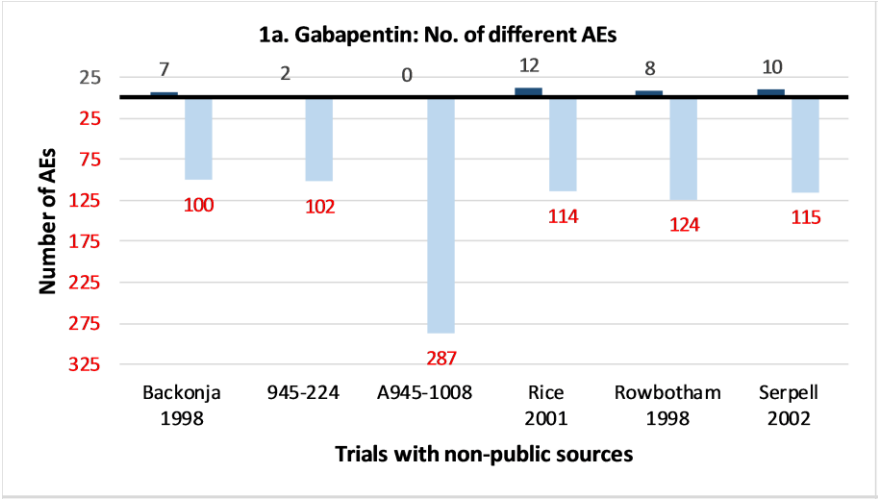
	Aripiprazole Lauroxil		
Preferred Term (%)	441 mg (n=207)	882 mg (n=208)	Placebo (n=207)
Any TEAE	58.9	57.2	62.3
Insomnia	9.7	12.0	11.6
Akathisia	11.6	11.5	4.3

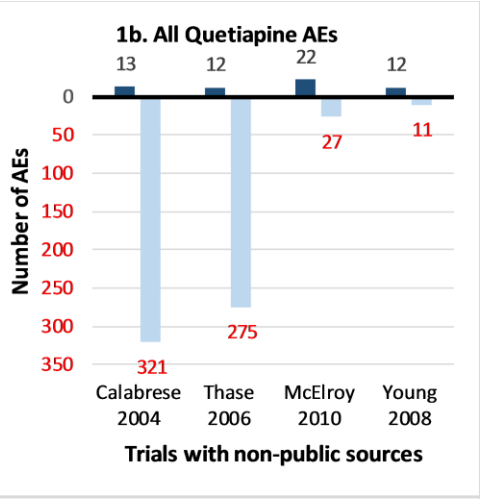
Most non-systematic harms never mentioned publicly

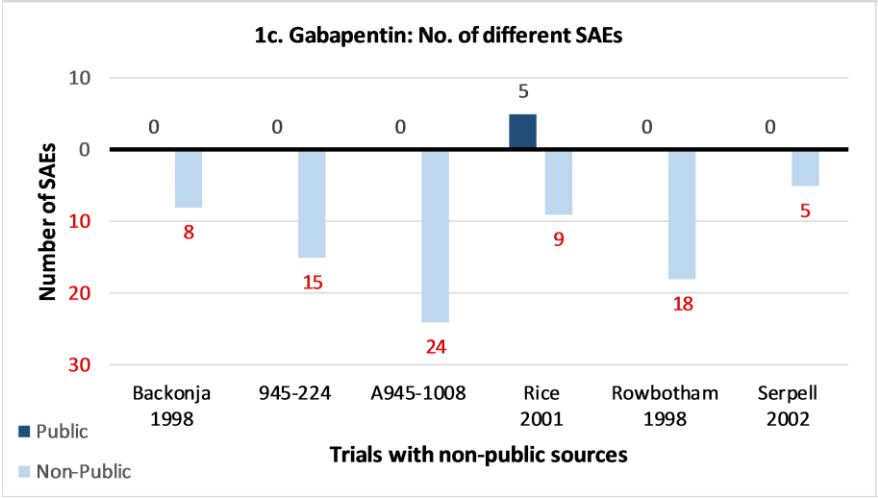


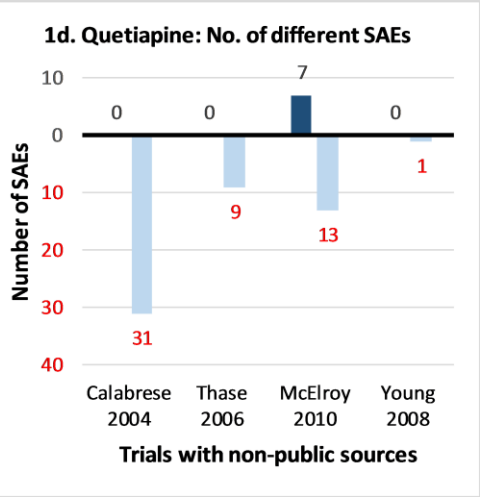
Mayo-Wilson, *et al.* (2019) DOI: 10.1016/j.jclinepi.2019.04.020

Most non-systematic harms never mentioned publicly



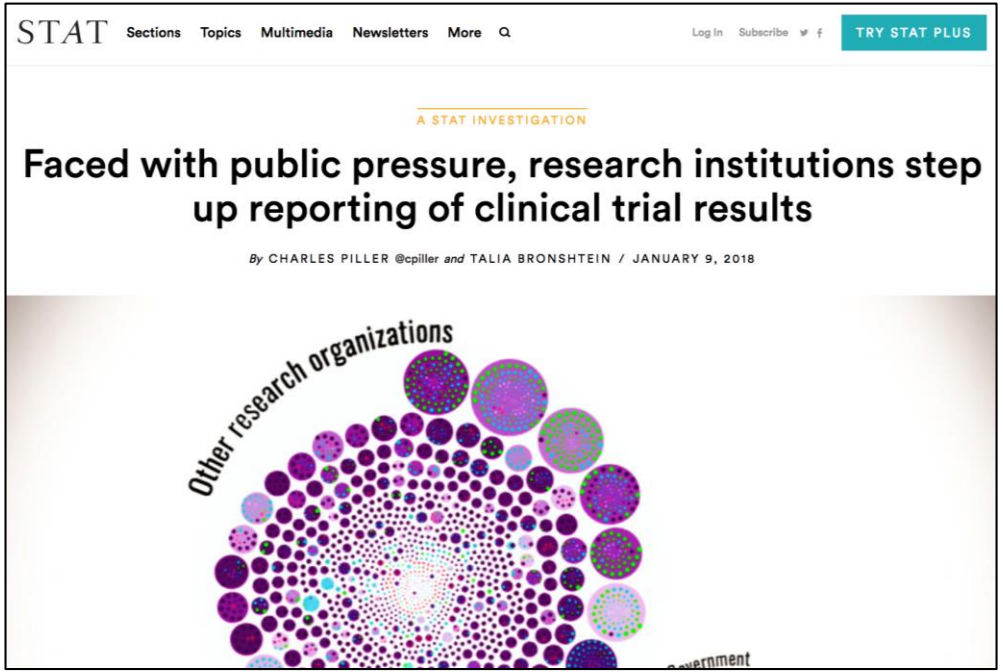






Mayo-Wilson, *et al.* (2019) DOI: 10.1016/j.jclinepi.2019.04.020

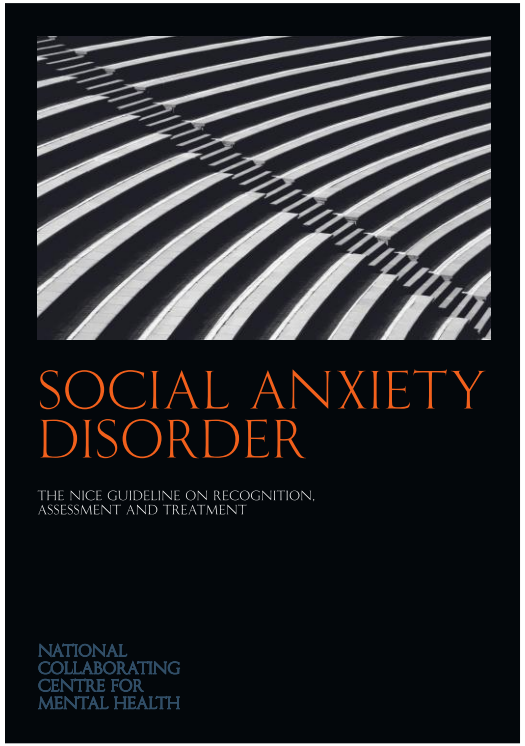
Supporting Investigators



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

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

Interventions for social anxiety disorder: Outcome measures



Anxiety Disorders Interview Schedule (ADIS-IV)

Brief Social Phobia Scale

Clinical Global Impression (CGI): Severity

Fear of Negative Evaluation Scale

Fear Questionnaire (FQ): Social Phobia

Liebowitz Social Anxiety Scale (64 of 101 trials)

Social Avoidance and Distress Scale (SADS)

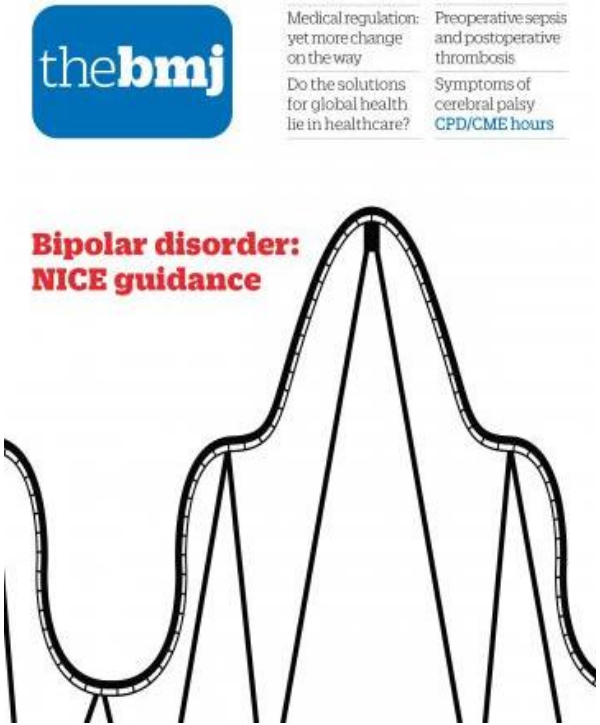
Social Interaction Anxiety Scale (SIAS)

Social Phobia Scale (SPS)

Social Phobia Anxiety Inventory

Pilling et al., 2013. DOI: 10.1136/bmj.f2541
Mayo-Wilson, et al., 2014. DOI: DOI: 10.1016/S2215-0366(14)70329-3

Psychological interventions for bipolar depression: Outcome measures



- Bech–Rafaelsen Melancholia Scale (BRMS)
- Beck Depression Inventory (BDI)
- Center for Epidemiological Studies Depression Scale (CES-D)
- Goldberg Anxiety and Depression Scale (GADS)
- Hamilton Depression Rating Scale (HAM-D)
- Montgomery-Asberg Depression Rating Scale (MADRS)
- Bipolar Longitudinal Investigation of Problems (BLIP)
- Internal State Scale (ISS)
- Depression and Schedule for Affective Disorders and Schizophrenia, change version (SADS-C)

Kendall, et al. 2014. DOI: 10.1136/bmj.g5673
Oud, et al. 2016. DOI: 10.1192/bjp.bp.114.157123

Core outcome sets for benefits and harms

"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

Journal policies: TOP Guidelines

SCIENTIFIC STANDARDS

Promoting an open research culture

Author guidelines for journals could help to promote transparency, openness, and reproducibility

By **B. A. Nosek,* G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, C. D. Chambers, G. Chin, G. Christensen, M. Contestabile, A. Dafoe, E. Eich, J. Freese, R. Glennerster, D. Goroff, D. P. Green, B. Hesse, M. Humphreys, J. Ishiyama, D. Karlan, A. Kraut, A. Lupia, P. Mabry, T. Madon, N. Malhotra, E. Mayo-Wilson, M. McNutt, E. Miguel, E. Levy Paluck, U. Simonsohn, C. Soderberg, B. A. Spellman, J. Turitto, G. VandenBos, S. Vazire, E. J. Wagenmakers, R. Wilson, T. Yarkoni**

Summary of the eight standards and three levels of the TOP guidelines

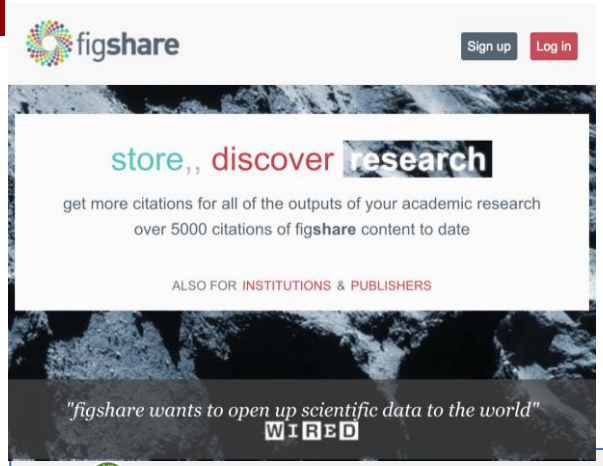
Levels 1 to 3 are increasingly stringent for each standard. Level 0 offers a comparison that does not meet the standard.

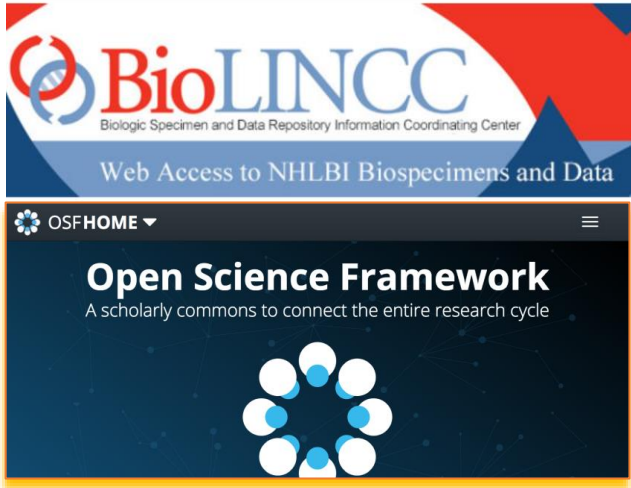
	LEVEL 0	LEVEL 1	LEVEL 2	LEVEL 3
Citation standards	Journal encourages citation of data, code, and materials—or says nothing.	Journal describes citation of data in guidelines to authors with clear rules and examples.	Article provides appropriate citation for data and materials used, consistent with journal's author guidelines.	Article is not published until appropriate citation for data and materials is provided that follows journal's author guidelines.
Data transparency	Journal encourages data sharing—or says nothing.	Article states whether data are available and, if so, where to access them.	Data must be posted to a trusted repository. Exceptions must be identified at article submission.	Data must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Analytic methods (code) transparency	Journal encourages code sharing—or says nothing.	Article states whether code is available and, if so, where to access them.	Code must be posted to a trusted repository. Exceptions must be identified at article submission.	Code must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.
Research materials transparency	Journal encourages materials sharing—or says nothing	Article states whether materials are available and, if so, where to access them.	Materials must be posted to a trusted repository. Exceptions must be identified at article submission.	Materials must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.

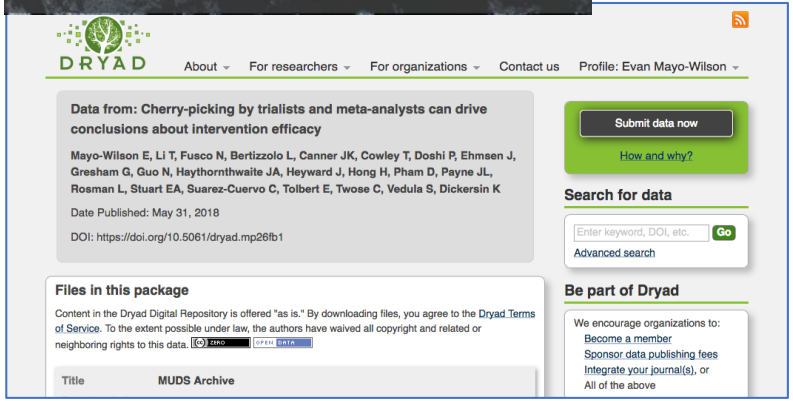
Nosek, *et al.*, 2015. DOI: 10.1126/science.aab2374

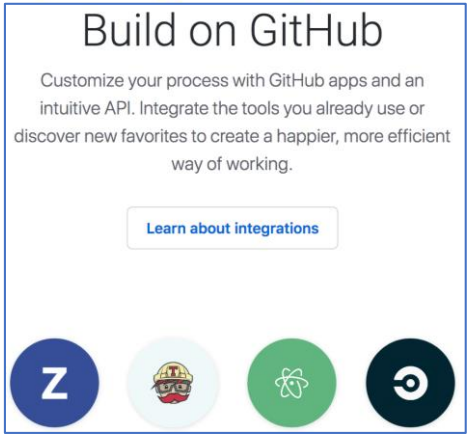
The TOP Statement Working Group 2018. DOI: 10.17605/OSF.IO/SM78T

Repositories for data and code

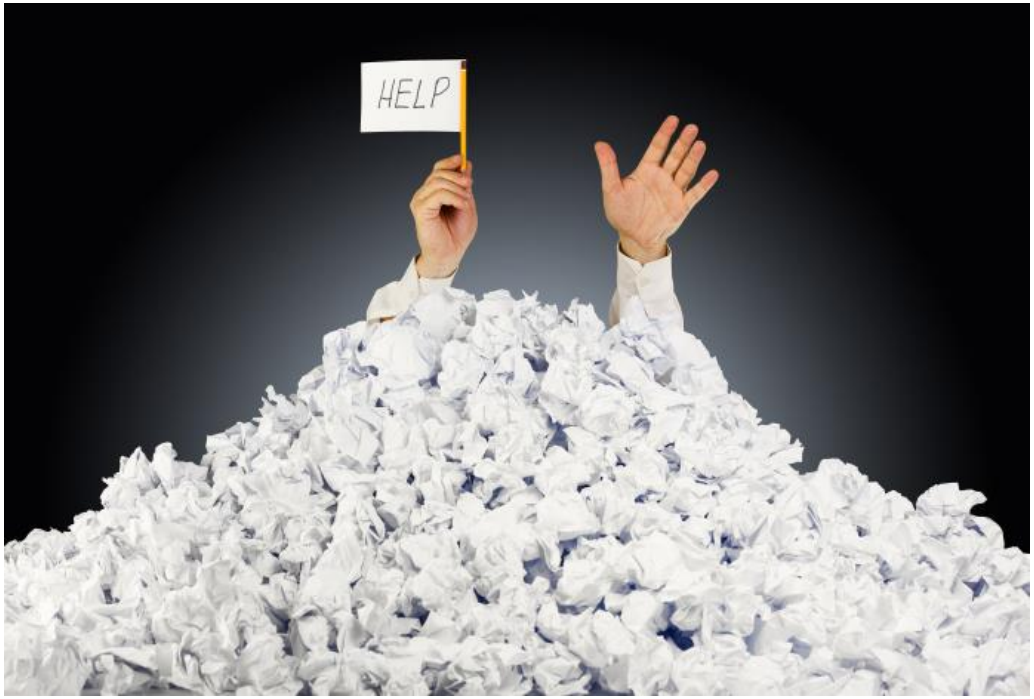








Summaries are important



Reanalyzing data is difficult and time consuming

Most people do not have the time, skills, or interest

Science depends of trustworthy *summaries* of research

Mayo-Wilson, *et al.*, 2017. DOI: 10.1002/jrsm.1277
Dickersin and Mayo-Wilson, 2017. DOI: 10.1073/pnas.1708273114
Hoffmann, *et al.*, 2017. DOI: 10.1136/bmj.j2782

Reporting guidelines minimize cherry picking

RESEARCH AND REPORTING METHODS | Annals of Internal Medicine

SPIRIT 2013 Statement: Defining Standard Protocol Items for Clinical Trials

An-Wen Chan, MD, DPhil; Jennifer M. Tetzlaff, MSc; Douglas G. Altman, DSc; Andreas Laupacis, MD; Peter C. Gøtzsche, MD, DrMedSci; Karmela Krleža-Jerić, MD, DSc; Asbjørn Hróbjartsson, PhD; Howard Mann, MD; Kay Dickersin, PhD; Jesse A. Berlin, ScD; Caroline J. Doré, BSc; Wendy R. Parulekar, MD; William S.M. Summerskill, MBBS; Trish Groves, MBBS; Kenneth F. Schulz, PhD; Harold C. Sox, MD; Frank W. Rockhold, PhD; Drummond Rennie, MD; and David Moher, PhD

The protocol of a clinical trial serves as the foundation for study planning, conduct, reporting, and appraisal. However, trial protocols and existing protocol guidelines vary greatly in content and quality. This article describes the systematic development and scope of SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) 2013, a guideline for the minimum content of a clinical trial protocol.

The 33-item SPIRIT checklist applies to protocols for all clinical

for key content, the SPIRIT recommendations aim to facilitate the drafting of high-quality protocols. Adherence to SPIRIT would also enhance the transparency and completeness of trial protocols for the benefit of investigators, trial participants, patients, sponsors, funders, research ethics committees or institutional review boards, peer reviewers, journals, trial registries, policymakers, regulators, and other key stakeholders.

Guidelines and Guidance

CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomised Trials

Kenneth F. Schulz[1]*, Douglas G. Altman[2], David Moher[3], for the CONSORT Group[¶]

1 Family Health International, Research Triangle Park, North Carolina, United States of America, 2 Centre for Statistics in Medicine, University of Oxford, Wolfson College, Oxford, United Kingdom, 3 Ottawa Methods Centre, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Canada

Introduction

Randomised controlled trials, when appropriately designed, conducted, and reported, represent the gold standard in evaluating healthcare interventions. However, randomised trials can yield biased results if they lack methodological rigour [1]. To assess a trial accurately, readers of a published report need complete, clear, and transparent information on its methodology and findings. Unfortunately, attempted assessments frequently fail because authors of many trial reports neglect to provide lucid and complete descriptions of that critical information [2,3,4].

indirect goal of our work. Moreover, CONSORT can help researchers in designing their trial.

Background to CONSORT

Efforts to improve the reporting of randomised controlled trials accelerated in the mid-1990s, spurred partly by methodological research. Researchers had shown for many years that authors reported such trials poorly, and empirical evidence began to accumulate that some poorly conducted or poorly reported aspects of trials were associated with bias [14] Two initiatives aimed at

Item 12: Primary, secondary, and other outcomes, including the specific measurement variable (e.g., systolic blood pressure), analysis metric (e.g., change from baseline, final value, time to event), method of aggregation (e.g., median, proportion), and time point for each outcome. Explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended.

Item 6a: Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed.

E&E

More than 70 outcomes were used in 196 RCTs of non-steroidal anti-inflammatory drugs for rheumatoid arthritis [108], and 640 different instruments had been used in 2000 trials in schizophrenia, of which 369 had been used only once [33]. Investigation of 149 of those 2000 trials

Chan, *et al.*, 2013. DOI: 10.7326/0003-4819-158-3-201302050-00583
Schulz, *et al.*, 2010. DOI: 10.1371/journal.pmed.1000251
Moher, *et al.*, 2010. DOI: 10.1136/bmj.c869

CONSORT Extension for Social and Psychological Interventions (CONSORT-SPI)

Incorporates other extensions

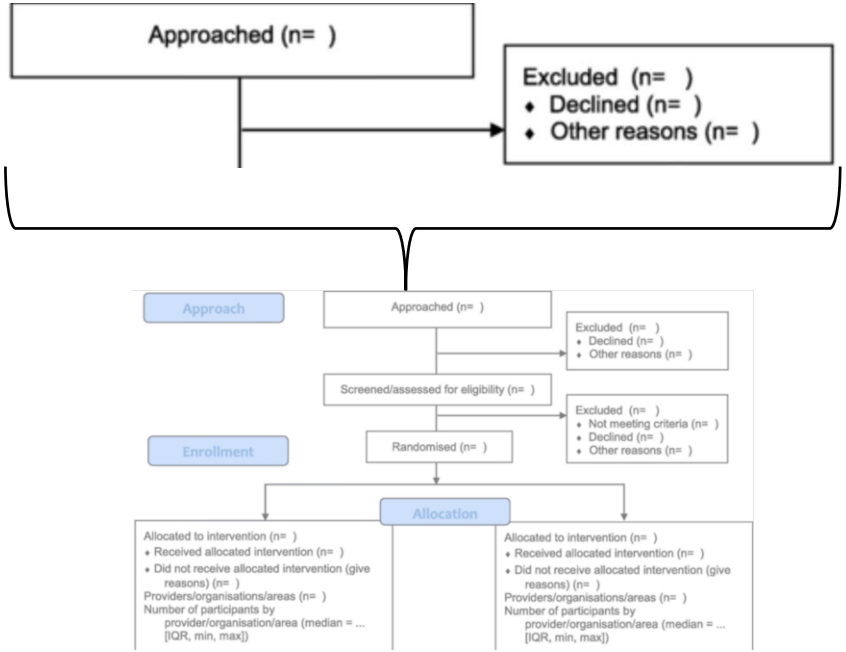
- CONSORT for Abstracts
- CONSORT for Cluster-Randomized Trials
- CONSORT for Non-pharmacologic Treatments
- TIDIER

Extends 9 of 25 items

- How the intervention might work
- Eligibility criteria for settings and providers
- Extent to which interventions delivered and taken-up
- Assignment of providers to intervention groups
- Involvement of the intervention developer in the study
- Stakeholder involvement in design, conduct, analysis
- Incentives offered

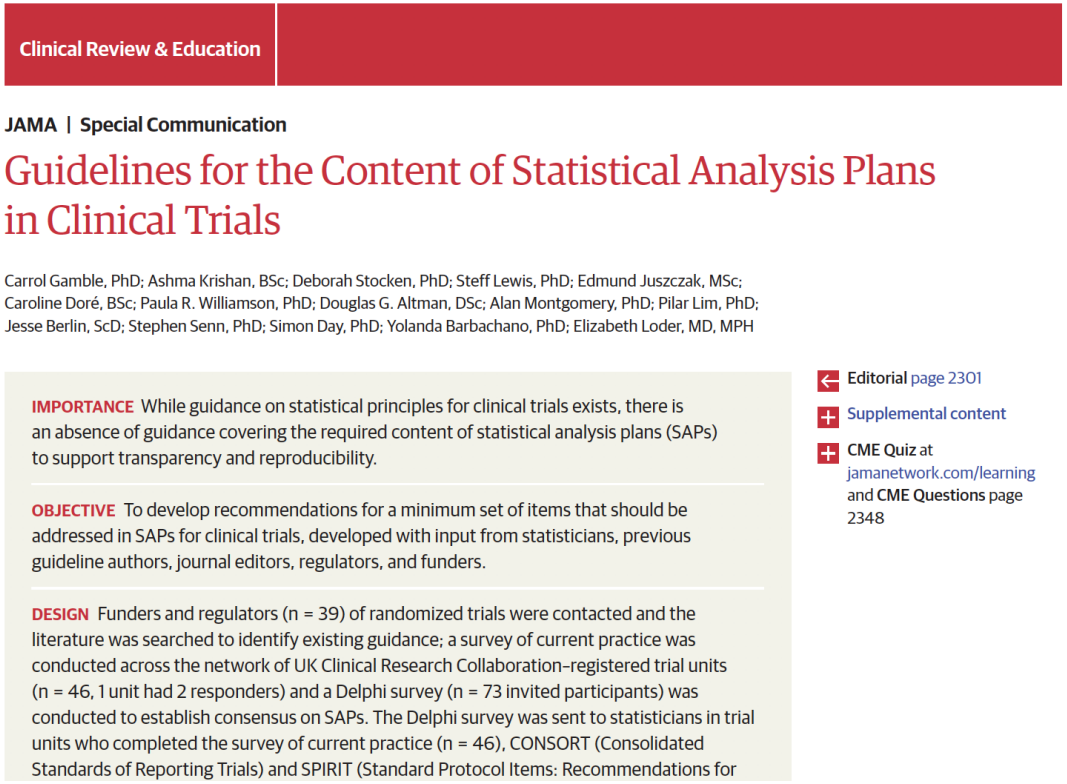
Some items not specific to SPI trials

- Methods for imputing missing data
- Availability of study materials and trial data



Montgomery, et al., 2018. DOI: 10.1186/s13063-018-2733-1
Grant, et al., 2018. DOI: 10.1186/s13063-018-2735-z

Statistical analysis plan (SAP)



Clinical Review & Education

JAMA | Special Communication

Guidelines for the Content of Statistical Analysis Plans in Clinical Trials

Carrol Gamble, PhD; Ashma Krishan, BSc; Deborah Stocken, PhD; Steff Lewis, PhD; Edmund Juszczak, MSc; Caroline Doré, BSc; Paula R. Williamson, PhD; Douglas G. Altman, DSc; Alan Montgomery, PhD; Pilar Lim, PhD; Jesse Berlin, ScD; Stephen Senn, PhD; Simon Day, PhD; Yolanda Barbachano, PhD; Elizabeth Loder, MD, MPH

IMPORTANCE While guidance on statistical principles for clinical trials exists, there is an absence of guidance covering the required content of statistical analysis plans (SAPs) to support transparency and reproducibility.

OBJECTIVE To develop recommendations for a minimum set of items that should be addressed in SAPs for clinical trials, developed with input from statisticians, previous guideline authors, journal editors, regulators, and funders.

DESIGN Funders and regulators (n = 39) of randomized trials were contacted and the literature was searched to identify existing guidance; a survey of current practice was conducted across the network of UK Clinical Research Collaboration–registered trial units (n = 46, 1 unit had 2 responders) and a Delphi survey (n = 73 invited participants) was conducted to establish consensus on SAPs. The Delphi survey was sent to statisticians in trial units who completed the survey of current practice (n = 46), CONSORT (Consolidated Standards of Reporting Trials) and SPIRIT (Standard Protocol Items: Recommendations for

Editorial page 2301

Supplemental content

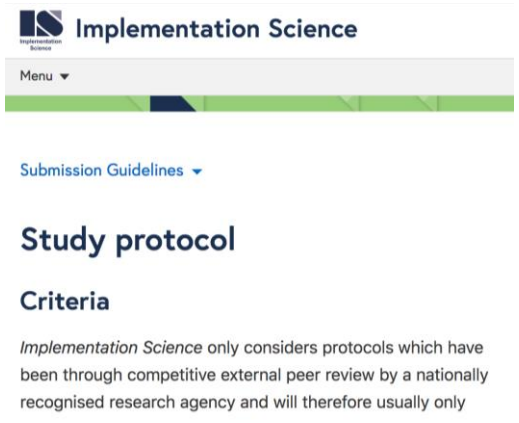
CME Quiz at jamanetwork.com/learning and CME Questions page 2348

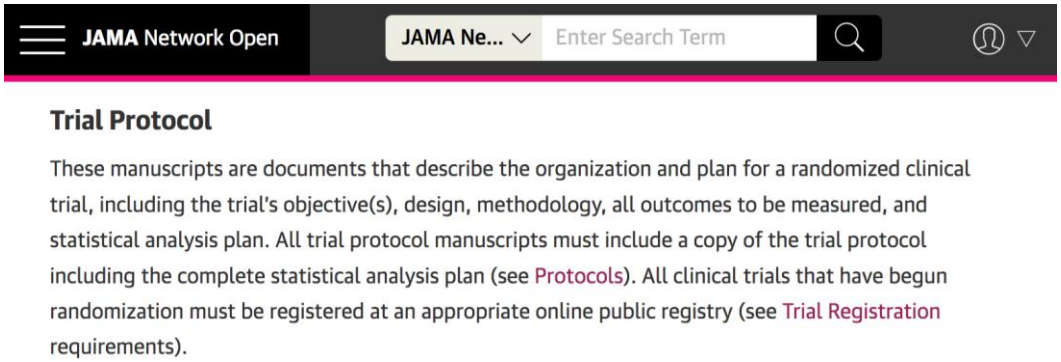
SAP can be part of the protocol or a separate document

May be finalized after study registration / start of enrollment (ideally before unmasking / preliminary analyses)

Gamble et al., 2017. DOI: 10.1001/jama.2017.18556

Publishing protocols & statistical analysis plans





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

Scan me

emayowil@iu.edu

@EvanMayoWilson

https://evanmayo-wilson.org/

INDIANA UNIVERSITY BLOOMINGTON

FULFILLING *the* PROMISE