JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Cherry picking results: Effects on clinical trial reports and systematic reviews

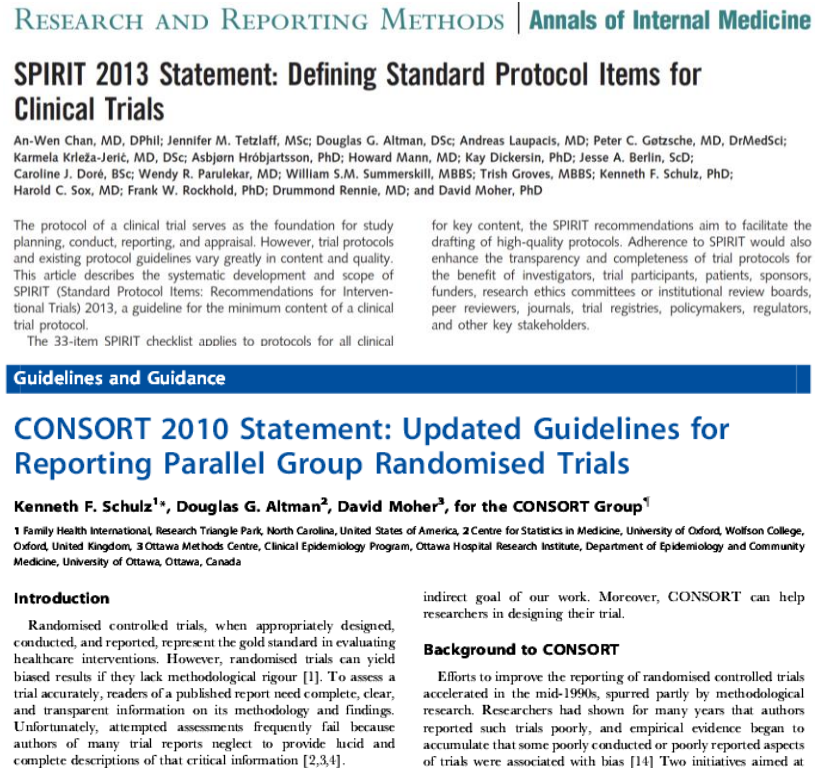
Evan Mayo-Wilson, MPA, DPhil

Center for Clinical Trials and Evidence Synthesis

“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

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.

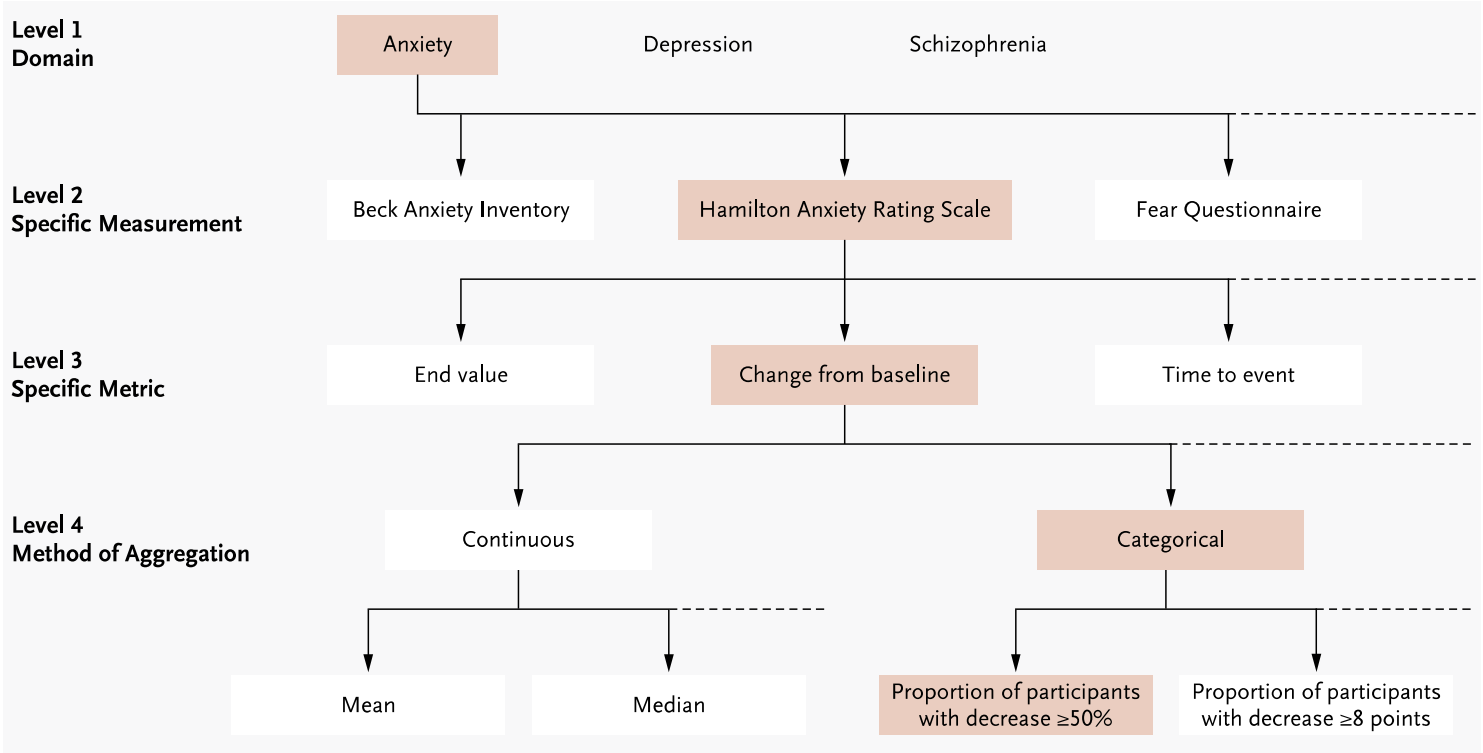
Item 6a: Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed.

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

Defining an outcome



Zarin, *et al.,* 2011. DOI: 10.1056/NEJMsa1012065

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

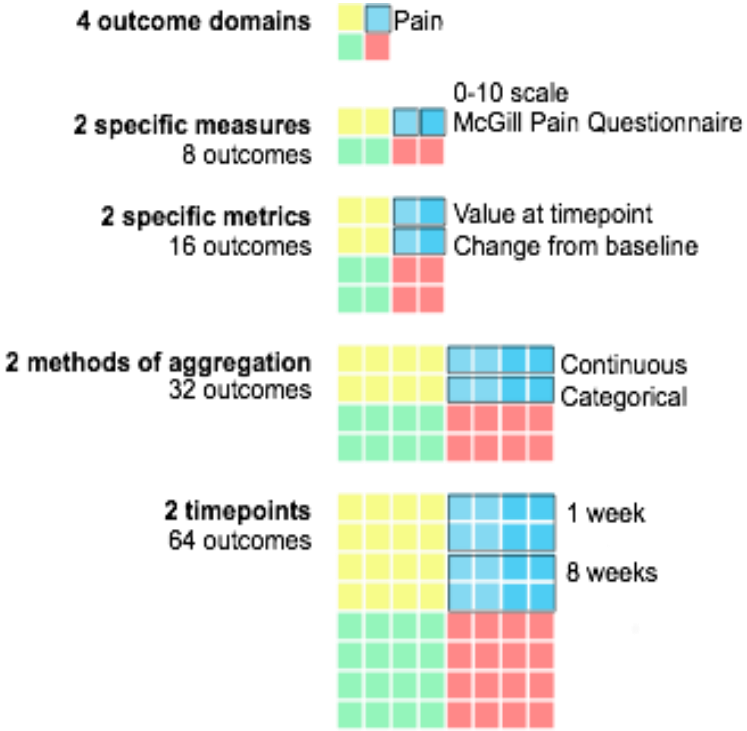
Methods to account for missing data, including missing items and missing cases (e.g., multiple imputation, last observation carried forward)

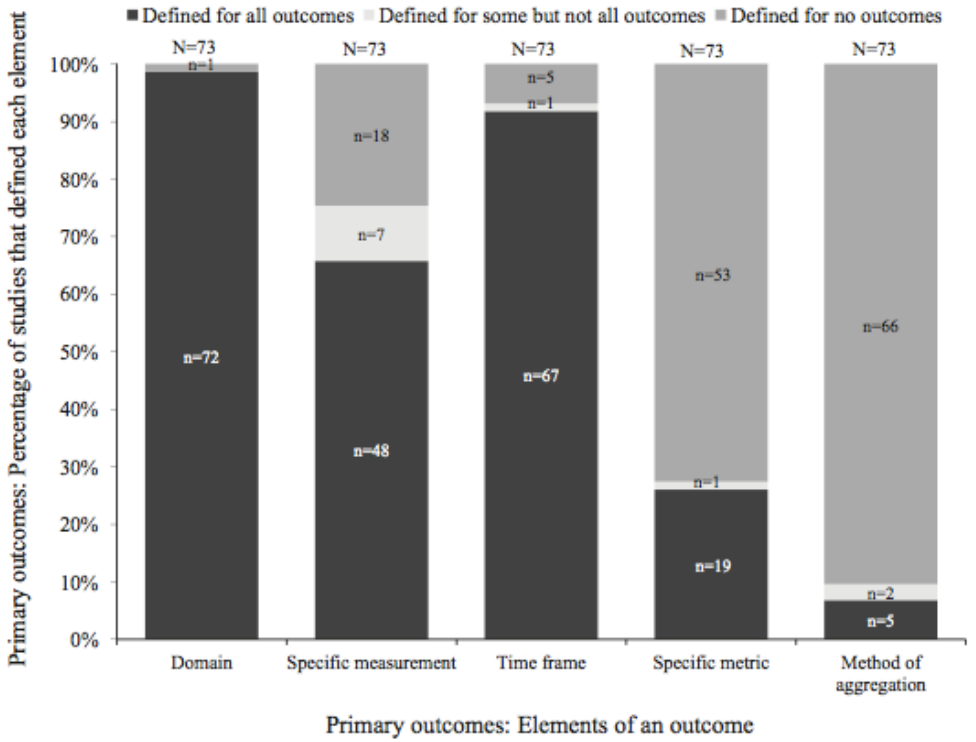
Methods of analysis

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

Outcomes are not defined in trial registers





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

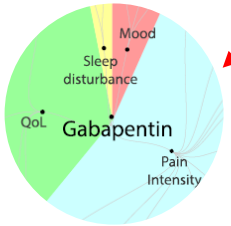
Cybulski, *et al.,* 2016. DOI: 10.1037/ccp0000115.

How much multiplicity is there in clinical trials?

21 trials

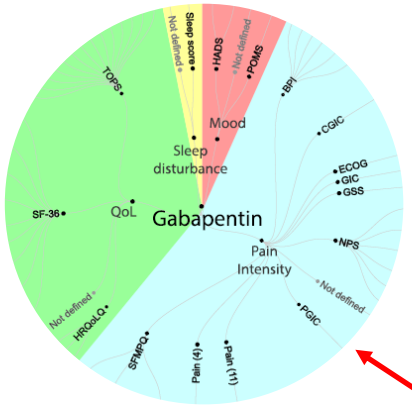
6 with non-public sources

4 Outcome domains



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

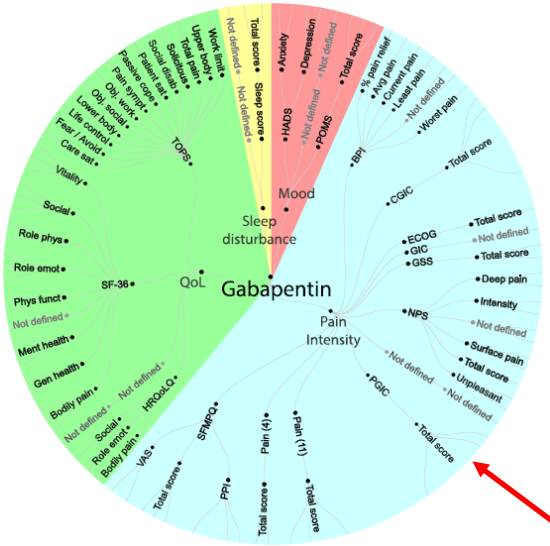
How much multiplicity is there in clinical trials?



Multiple measures

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

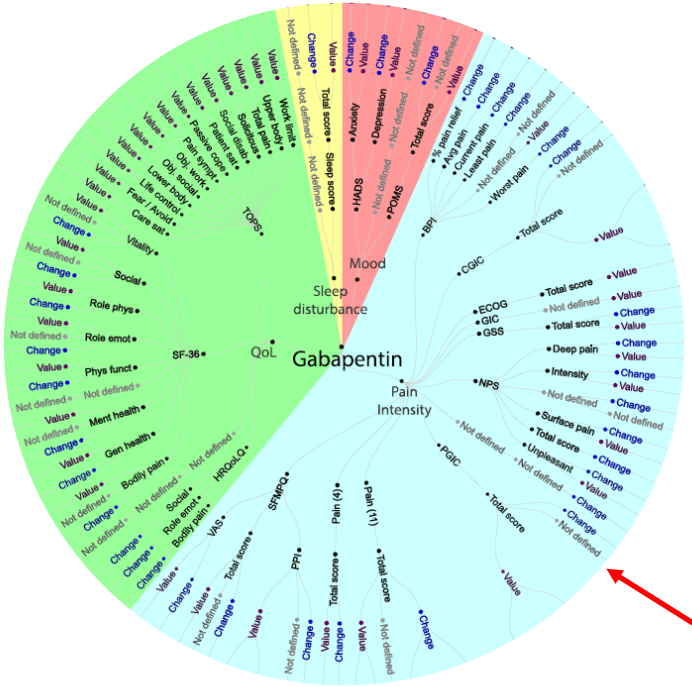
How much multiplicity is there in clinical trials?



Multiple totals and subscales

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

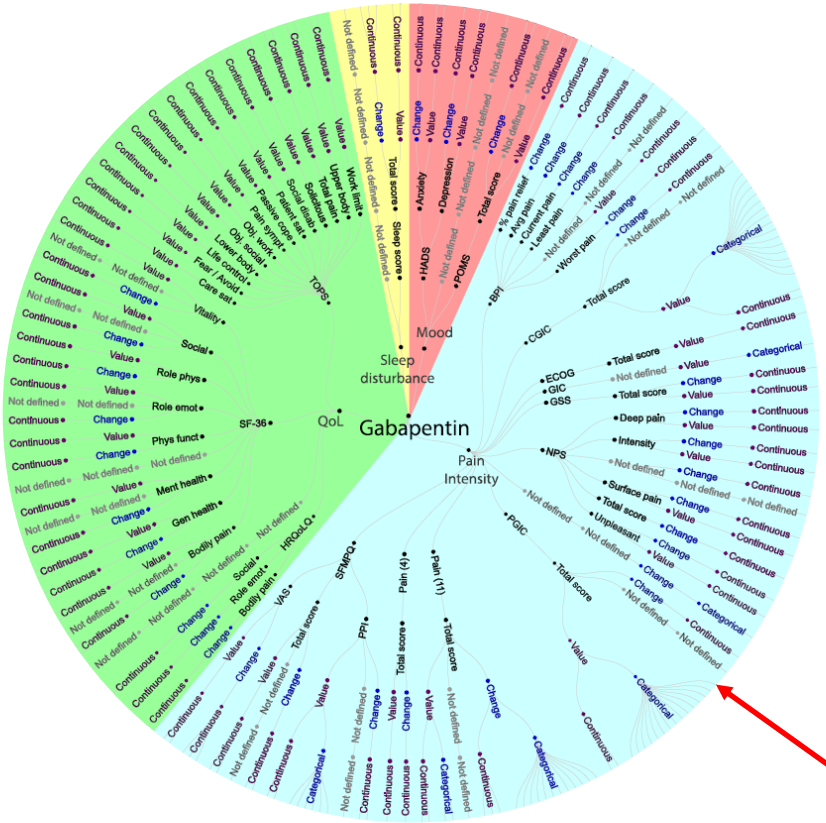
How much multiplicity is there in clinical trials?



Multiple metrics

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

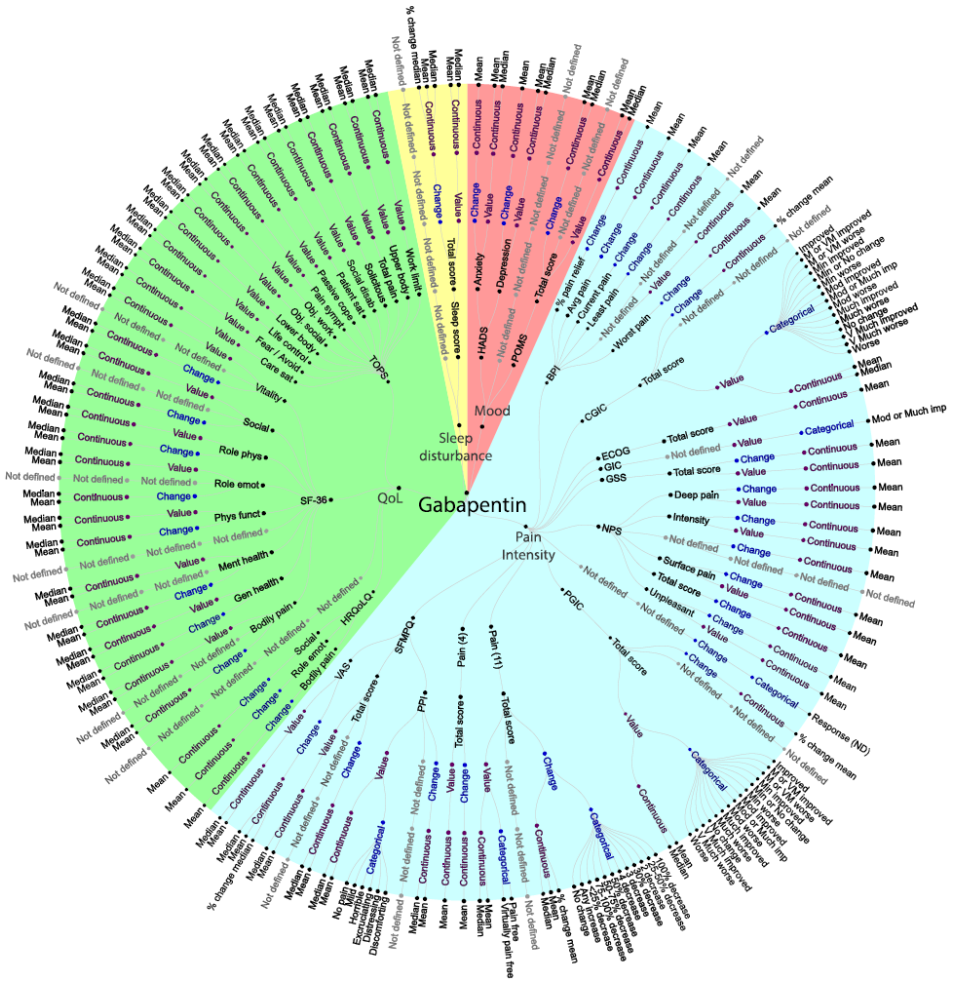
How much multiplicity is there in clinical trials?



Multiple methods of aggregation

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

How much multiplicity is there in clinical trials?



214 outcomes

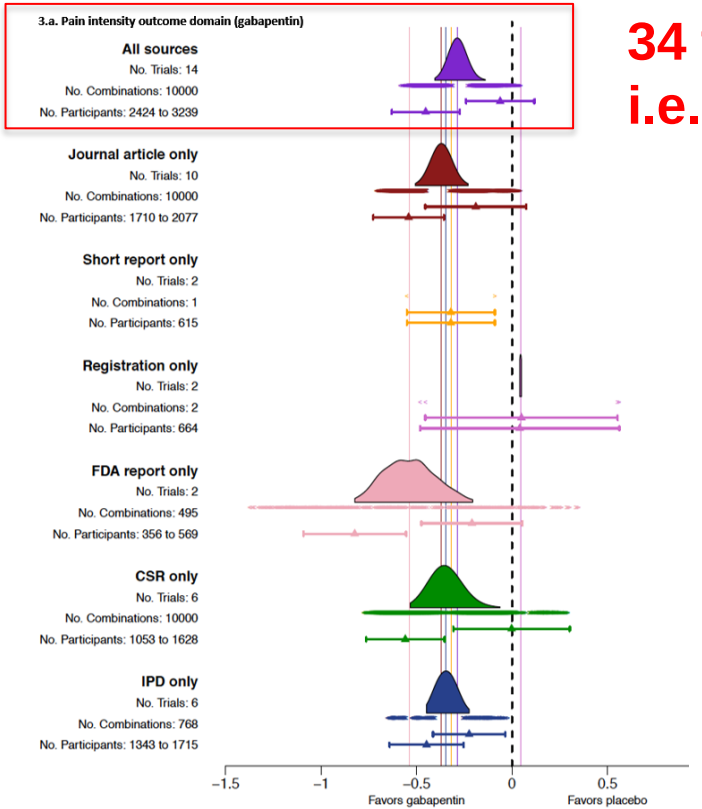
1230 results

305 (25%) publicly reported

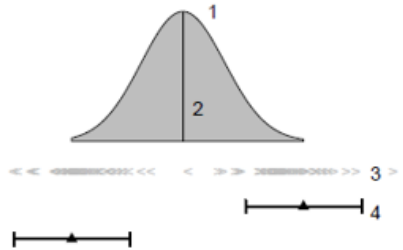
More hidden…

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

Consequences of multiplicity for clinical guidelines and practice



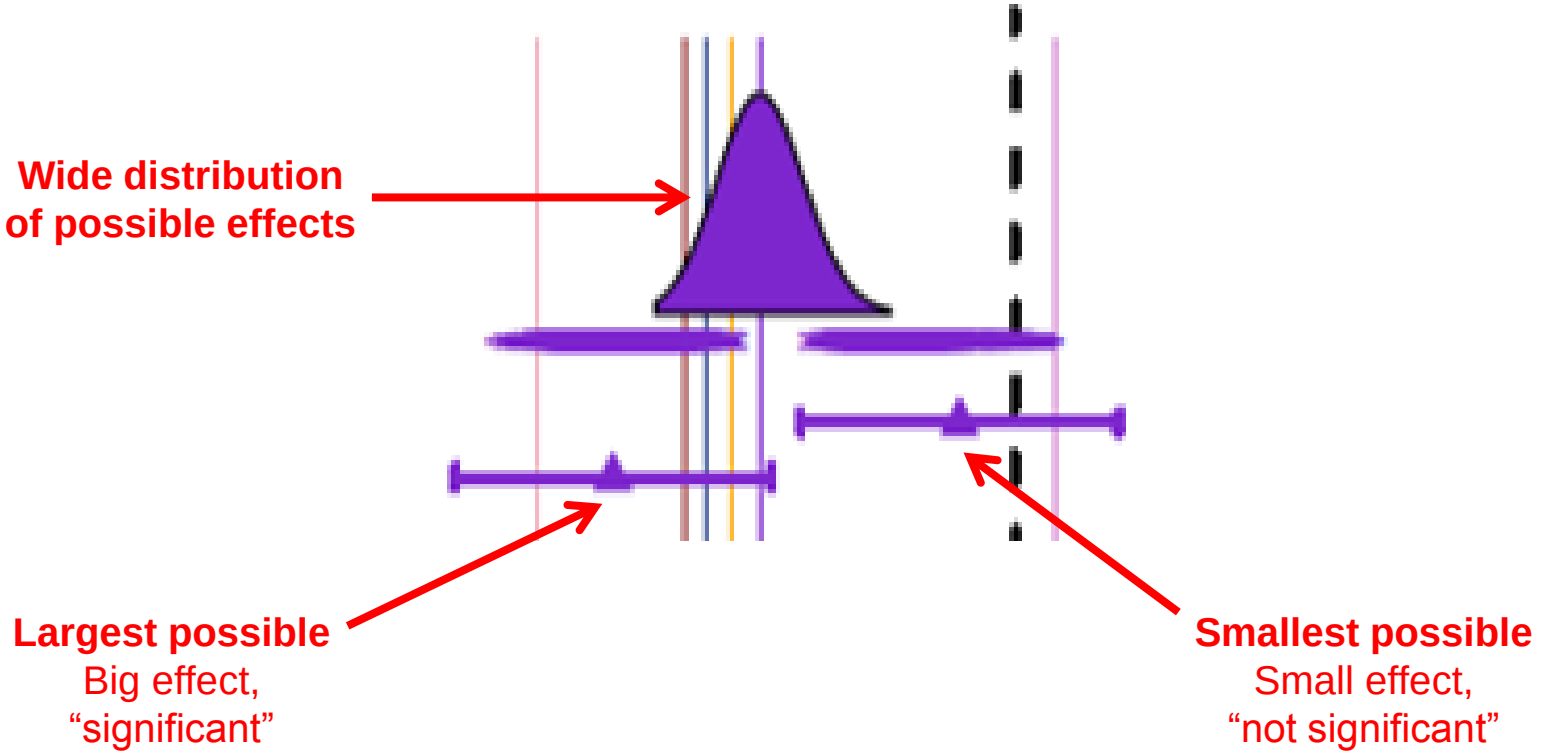
34 trillion possible meta-analyses of "pain"
i.e., combinations of *the same trials*



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 clinical guidelines and practice



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

Benefits and Harms

BENEFITS	HARMS
Measured systematically for all participants	Reported spontaneously by patients
Selected *a priori*	Reported based on the results

FDA Snapshot

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

DRUG TRIALS SNAPSHOT SUMMARY:

What is the drug for?

ARISTADA is a drug used for the treatment of schizophrenia.

Schizophrenia is a brain disorder characterized by hearing voices, believing other people are reading one's mind or controlling one's thoughts, and being suspicious or withdrawn.

How is this drug used?

ARISTADA is an injection given into the muscle of the arm or buttock by a healthcare provider once a month to once every 6 weeks.

What are the benefits of this drug?

ARISTADA improved symptoms of schizophrenia.

FDA Snapshot: Why measure & report these harms?

What are the possible side effects?

The table below summarizes adverse reactions in the Safety population defined as all patients who received at least one injection dose of study drug.

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

Adverse Reaction	Placebo N=207 (%)	ARISTADA	
		441 mg N=207 (%)	882 mg N=208 (%)
General disorders and administration site conditions			
Injection site pain	2	3	4
Investigations			
Increased weight	1	2	2
Increased blood creatine phosphokinase	0	2	1
Nervous system disorders			
Akathisia	4	11	11
Headache	3	3	5
Psychiatric disorders			
Insomnia	2	3	4
Restlessness	1	3	1

FDA Snapshot: Why measure & report these harms?

What are the possible side effects?

The table below summarizes adverse reactions in the Safety population defined as all patients who received at least one injection dose of study drug.

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

Adverse Reaction	Placebo N=207 (%)	ARISTADA	
		441 mg N=207 (%)	882 mg N=208 (%)
General disorders and administration site conditions			
Injection site pain	2	3	4
Investigations			
Increased weight	1	2	2
Increased blood creatine phosphokinase	0	2	1
Nervous system disorders			
Akathisia	4	11	11
Headache	3	3	5
Psychiatric disorders			
Insomnia	2	3	4
Restlessness	1	3	1

Reporting depends on results

FDA Snapshot: Why measure & report these harms?

What are the possible side effects?

The table below summarizes adverse reactions in the Safety population defined as all patients who received at least one injection dose of study drug.

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

Adverse Reaction	Placebo N=207 (%)	ARISTADA 441 mg N=207 (%)	ARISTADA 882 mg N=208 (%)
General disorders and administration site conditions			
Injection site pain	2	3	4
Investigations			
Increased weight	1	2	2
Increased blood creatine phosphokinase	0	2	1
Nervous system disorders			
Akathisia	4	11	11
Headache	3	3	5
Psychiatric disorders			
Insomnia	2	3	4
Restlessness	1	3	1

Systematic

Not Systematic

Inconsistent reporting

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

Specifying harms *a priori in* trials and core outcome sets

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, *et al*. 2005

Implications for research synthesis: Junk in, junk out

Table 3.—Most Frequently Reported Adverse Events*

Preferred Terms	Gabapentin (n = 84)	Placebo (n = 81)	*P* Value†
Dizziness	20 (23.8)	4 (4.9)	<.001
Somnolence	19 (22.6)	5 (6.2)	.004
Headache	9 (10.7)	3 (3.7)	.13
Diarrhea	9 (10.7)	7 (8.6)	.79
Confusion	7 (8.3)	1 (1.2)	.06
Nausea	7 (8.3)	4 (4.9)	.54

*Data are number (percentage).
†Data were calculated using the Fisher exact test.