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

"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.

---

| 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 H0 with Pr(E|H0) ≤ .05 (3) | Number of Research Reports that Fail to Reject H0 (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 (1960) | 118                                 | 94                                                    | 91                                                       | 2                                               | 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
Results for “primary” outcomes differ between sources

Vedula, 2009. DOI: 10.1056/NEJMsa0906126
Results for “primary” outcomes differ between sources

Vedula, 2009. DOI: 10.1056/NEJMsa0906126
Results for “primary” outcomes differ between sources

Vedula, 2009. DOI: 10.1056/NEJMsa0906126
Results for “primary” outcomes differ between sources

28 Primary outcomes in publications (includes those described with no distinction from secondary outcomes)

12 New primary outcomes in publication (includes those described with no distinction from secondary outcomes)

5 Protocol-defined secondary outcomes (reported with no distinction from primary outcomes in publication)

4 Were reported as secondary outcomes

6 Were not reported in publication

21 Primary outcomes in protocols of 12 published trials (includes those described with no distinction from secondary outcomes)

11 Were reported with no changes (includes those described with no distinction from secondary outcomes)

Vedula, 2009. DOI: 10.1056/NEJMsa0906126
Elements of an outcome

Level 1
Domain
- Anxiety
- Depression
- Schizophrenia

Level 2
Specific Measurement
- Beck Anxiety Inventory
- Hamilton Anxiety Rating Scale
- Fear Questionnaire

Level 3
Specific Metric
- End value
- Change from baseline
- Time to event

Level 4
Method of Aggregation
- Continuous
  - Mean
  - Median
- Categorical
  - Proportion of participants with decrease ≥50%
  - Proportion of participants with decrease ≥8 points

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

<table>
<thead>
<tr>
<th>Analysis population</th>
<th>Handling missing data</th>
<th>Methods of analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants eligible to be included in the analysis (e.g., people who took one dose, everyone randomized)</td>
<td>Methods to account for missing data, including missing items and missing cases (e.g., multiple imputation, last observation carried forward)</td>
<td>Statistical methods, including analysis model, procedures (e.g., transformations, adjustments), and covariates included in the analysis</td>
</tr>
</tbody>
</table>

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
The problem of multiple outcome definitions

Multiple measures
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
The problem of multiple outcome definitions

Mayo-Wilson, et al., 2017. DOI: 10.1016/j.jclinepi.2017.05.007
The problem of multiple outcome definitions

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

21 trials

214 outcomes

1230 results

305 (25%) publicly reported

More hidden...
Consequences of multiplicity for meta-analysis

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

Mayo-Wilson, et al., 2017. DOI: 10.1016/j.jclinepi.2017.07.014
Consequences of multiplicity for meta-analysis

- Wide distribution of possible effects
- Largest possible
  - Big effect, “significant”
- Smallest possible
  - Small effect, “not significant”

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**

- Measured for all participants in the same way
- Selected *a priori*
- Analyzed and reported using pre-specified methods

**NON-SYSTEMATIC HARMs**

- Reported spontaneously by patients
- Reported based on the results
- 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. Analysis of Covariance With Last Observation Carried Forward Analysis, Full Analysis Set

- Aripiprazole lauroxil 441 mg (n = 196)
- Aripiprazole lauroxil 882 mg (n = 204)
- Placebo (n = 196)

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

- Aripiprazole lauroxil 441 mg (n = 196)
- Aripiprazole lauroxil 882 mg (n = 204)
- Placebo (n = 196)

*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 **harm**s of Aristada (aripiprazole)

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

<table>
<thead>
<tr>
<th>Preferred Term (%)</th>
<th>Aripiprazole Lauroxil</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>441 mg (n = 207)</td>
<td>882 mg (n = 208)</td>
</tr>
<tr>
<td>Any TEAE</td>
<td>58.9</td>
<td>57.2</td>
</tr>
<tr>
<td>Insomnia</td>
<td>9.7</td>
<td>12.0</td>
</tr>
<tr>
<td>Akathisia</td>
<td>11.6</td>
<td>11.5</td>
</tr>
<tr>
<td>Headache</td>
<td>8.2</td>
<td>8.7</td>
</tr>
<tr>
<td>Anxiety</td>
<td>2.9</td>
<td>5.3</td>
</tr>
<tr>
<td>Injection site pain</td>
<td>3.4</td>
<td>4.8</td>
</tr>
<tr>
<td>Toothache</td>
<td>2.4</td>
<td>3.8</td>
</tr>
<tr>
<td>Nausea</td>
<td>2.9</td>
<td>3.4</td>
</tr>
<tr>
<td>Constipation</td>
<td>2.9</td>
<td>2.4</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>2.4</td>
<td>2.4</td>
</tr>
<tr>
<td>Weight increase</td>
<td>2.9</td>
<td>2.4</td>
</tr>
<tr>
<td>Neck pain</td>
<td>1.0</td>
<td>2.4</td>
</tr>
<tr>
<td>Sedation</td>
<td>1.9</td>
<td>2.4</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>5.8</td>
<td>2.4</td>
</tr>
<tr>
<td>Restlessness</td>
<td>2.9</td>
<td>1.9</td>
</tr>
<tr>
<td>Blood CPK increase</td>
<td>4.3</td>
<td>1.4</td>
</tr>
</tbody>
</table>

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
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).

DOSEAGE 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 5 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).

DOSEAGE FORMS AND STRENGTHS
For extended-release injectable suspension: 441 mg, 662 mg or 882 mg single-use pre-filled syringes (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.
Non-systematic AEs are reported inconsistently across sources.

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.
Most non-systematic harms never mentioned publicly

Most non-systematic harms never mentioned publicly
Supporting Investigators

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

Butterfly diagram

Account characteristics
Policies
Procedures
Computer systems
Staff

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

---

### 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.

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

---


---

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

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.
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

Grant, et al., 2018. DOI: 10.1186/s13063-018-2735-z
Statistical analysis plan (SAP)

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)
Publishing protocols & statistical analysis plans

Study protocol

Criteria

Study protocol articles will only be considered for proposed or ongoing trials that have not completed participant recruitment at the time of submission. Trials advises that study protocols are submitted well before recruitment completes. Please confirm the

Trial Protocol

These manuscripts are documents that describe the organization and plan for a randomized clinical trial, including the trial's objective(s), design, methodology, all outcomes to be measured, and statistical analysis plan. All trial protocol manuscripts must include a copy of the trial protocol including the complete statistical analysis plan (see Protocols). All clinical trials that have begun randomization must be registered at an appropriate online public registry (see Trial Registration requirements).
Multiple Data Sources (MUDS) Investigators

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

**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

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

**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

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

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