

Opportunities for selective reporting of harms in randomized clinical trials: Selection criteria for nonsystematic adverse events

Evan Mayo-Wilson (✉ evan.mayo-wilson@jhu.edu)

Johns Hopkins University Bloomberg School of Public Health <https://orcid.org/0000-0001-6126-2459>

Nicole Fusco

Johns Hopkins University Bloomberg School of Public Health

Hwanhee Hong

Duke University

Tianjing Li

Johns Hopkins University Bloomberg School of Public Health

Joseph K. Canner

Johns Hopkins University School of Medicine

Kay Dickersin

Johns Hopkins University Bloomberg School of Public Health

Research Article

Keywords: Harms, adverse events, clinical trials, reporting bias, selective outcome reporting, data sharing, trial registration

Posted Date: February 5th, 2019

DOI: <https://doi.org/10.21203/rs.2.268/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published on September 5th, 2019. See the published version at <https://doi.org/10.1186/s13063-019-3581-3>.

Abstract

Background: Adverse events (AEs) in randomized clinical trials may be reported in multiple sources. Different methods for reporting adverse events across trials, or across sources for a single trial, may produce inconsistent and confusing information about the adverse events associated with interventions

Methods: We sought to compare the methods authors use to decide which AEs to include in a particular source (i.e., “selection criteria”) and to determine how selection criteria could impact the AEs reported.

We compared sources (e.g., journal articles, clinical study reports [CSRs]) of trials for two drug-indications: gabapentin for neuropathic pain and quetiapine for bipolar depression. We identified selection criteria and assessed how criteria affected AE reporting.

Results: We identified 21 gabapentin trials and 7 quetiapine trials. All CSRs (6 gabapentin, 2 quetiapine) reported all AEs without applying selection criteria; by comparison, no other source reported all AEs, and 15/68 (22%) gabapentin sources and 19/48 (40%) quetiapine sources reported using selection criteria. Selection criteria greatly affected the number of AEs that would be reported. For example, 67/316 (21%) AEs in one quetiapine trial met the criterion “occurring in $\geq 2\%$ of participants in any treatment group,” while only 5/316 (2%) AEs met the criterion, “occurring in $\geq 10\%$ of quetiapine-treated patients and twice as frequent in the quetiapine group as the placebo group.”

Conclusions: Selection criteria for reporting AEs vary across trials and across sources for individual trials. If investigators do not pre-specify selection criteria, they might “cherry-pick” AEs based on study results. Even if investigators pre-specify selection criteria, selective reporting of AEs will produce biased meta-analyses and clinical practice guidelines. Data about all AEs identified in clinical trials should be publicly available; however, sharing data will not solve all the problems we identified in this study.

Keywords: Harms, adverse events, clinical trials, reporting bias, selective outcome reporting, data sharing, trial registration

Background

The U.S. Food and Drug Administration (FDA) and other regulators may approve drugs and biologics for marketing when the results of randomized clinical trials indicate that their potential benefits outweigh their potential adverse events (AEs). In clinical trials, AEs can be collected *non-systematically* when participants report them spontaneously to investigators or in response to open-ended questions such as “have you noticed any symptoms since your last examination?” (Box 1) [1-4]. Specific AEs can also be collected *systematically* by assessing their presence or absence using the same method for all participants in a trial (e.g., clinical examination, questionnaires, and medical instruments [1, 2]. Previous studies have described non-systematic AEs as “passively collected” and systematic AEs as “proactively collected” [1, 2]. AEs are categorized as “serious” when they lead to hospitalization or death, or disrupt normal life functions [4, 5].

Doctors, patients, and policy makers rely on systematic reviews and clinical practice guidelines to make medical decisions. Syntheses of clinical trial findings should include all available evidence; however, they are often based on AE information reported in public sources, such as journal articles (Box 1) [6]. Many systematic reviews that plan to synthesize AEs ultimately do not address AEs [7]. By contrast, regulators have access to non-public sources of trial information, including clinical study reports (CSRs) and individual patient datasets (IPD). These previously non-public sources are becoming available for some trials [8-10]; however, the results of many trials are available only in public sources [11, 12], which are often incomplete [12-21]. Reporting bias, which is the selective reporting of research results, occurs when reporting is influenced by the nature of results (e.g., the direction, magnitude, statistical significance of results). Reporting bias may lead to overestimating potential benefits of an intervention [12-17] and underestimating potential AEs [19, 22-29].

There is little evidence about the methods trialists use to select AEs for reporting, and there is little evidence about whether those methods contribute to outcome reporting bias. Previous guidance for reporting adverse events has discouraged vague statements about adverse events [33], and the CONSORT Extension for Harms discourages authors from reporting AEs above frequency thresholds (e.g., >10%) [34]. Nonetheless, there is little evidence about how different trials, and different sources for a single trial, report AEs. The objectives of this study were to (1) compare selection criteria for reporting non-systematic AEs (i.e., the methods authors use to decide which AEs to include in a particular source as illustrated in Box 2); (2) examine how different selection criteria could affect AE reporting; and (3) assess how different selection criteria could impact meta-analyses of AEs.

Methods

This analysis is a sub-study of the Multiple Data Sources (MUDS) study. The published protocol and amendments [21 ,35] describe the study methods. We searched for public and non-public sources (see definitions in Box 1) and requested additional non-public sources from the companies that manufacture gabapentin and quetiapine [19 ,35 ,36].

Eligible trials and sources

Eligible studies were parallel randomized clinical trials that compared either gabapentin for neuropathic pain or quetiapine for bipolar depression with placebo. The MUDS study included 21 gabapentin trials (80 sources, including 6 IPD) and 7 quetiapine trials (52 sources, including 2 IPD). Because IPD did describe methods for analyzing data, and because the CSRs included aggregate results that were consistent with the IPD, we excluded IPD from this sub-study. Thus, we included 68 public sources and 6 non-public sources (all CSRs) for gabapentin trials, and we included 46 public sources and 4 non-public sources (2 CSRs, 2 CSR-synopses) for quetiapine trials. All studies were reported in one or more public sources, except one gabapentin trial that was not reported in any public source.

Data extraction

From each source of each eligible trial, two investigators independently extracted data using the open access Systematic Review Data Repository (SRDR; <http://srd.ahrq.gov/>) and resolved differences by discussion. We extracted results (i.e. the number or proportion of participants who experienced AEs) that were reported for each trial; we did not extract results of analyses that pooled multiple trials.

We classified each AE as “systematic” if its presence or absence was recorded for every participant and assessed using specific measurement tools (e.g., questionnaires, checklists, laboratory tests, or clinical examinations); we classified all other AEs as “non-systematic” (Box 1). For each non-systematic AE, we extracted the name of the event (e.g., “dizziness”, “headache”) and the numerical results closest to 8 and 18 weeks (e.g., proportion of participants in each group who experienced the AE at 6 weeks). We extracted the names of non-systematic AEs even when numerical results were not reported; for example, if a source reported that “the most common AEs were dizziness and headache,” we extracted that dizziness and headache were reported and we entered the numerical results as missing data [19].

From each source, we extracted the methods that authors said they used to select AEs for inclusion in the source, which we refer to as “selection criteria” or we recorded that the selection criteria were not reported (Box 2). Although we found evidence that potential benefits and systematic AEs are reported selectively [20], we did not find that systematic AEs were reported based on selection criteria as described in this sub-study. Thus, we describe selection criteria for non-systematic AEs only.

Comparing non-systematic AE selection criteria sources and trials

For each trial, we compared the selection criteria for non-systematic AEs reported across sources (see examples in Box 2). We recorded three pre-specified components of non-systematic AE selection criteria:

Numerical threshold: a cutoff for reporting the number or proportion of participants who experienced a non-systematic AE (e.g., $\geq 5\%$ of participants).

Participant group: the group(s) in a trial that must have experienced an AE (e.g., participants in the test intervention group, all participants in the trial).

Difference between groups: a cutoff for reporting a difference in the number or proportion of participants who experienced an AE, comparing one participant group with another (e.g., more frequent in the test intervention group compared with the placebo group).

Applying AE selection criteria to data in individual trials

To assess how different selection criteria might affect which and how many non-systematic AEs would be reported, we combined each of the numerical thresholds we observed (N=5) with each of the participant group criteria (N=3) and difference between groups criteria we observed (N=3) to create 45 combined selection criteria.

We observed some selection criteria in both gabapentin and quetiapine trials, and we observed other selection criteria in gabapentin trials only or in quetiapine trials only. Although we analyzed only numerical thresholds, participant groups, and differences between groups that we observed in eligible trials, we never observed some combined selection criteria in any trial.

Using all non-systematic AEs reported in each CSR, we then applied each of 45 combined selection criteria to the data found in CSRs for six gabapentin trials and two quetiapine trials. We estimated the number of AEs that would have been reported according to each combined criterion.

Statistical methods

We calculated descriptive statistics (i.e., counts) using Stata 14 [37].

Results

AE selection criteria in public and non-public sources

We identified selection criteria only in public sources and in CSR-synopses. CSRs, which included trial protocols, did not report using selection criteria and did not specify selection criteria to use in other sources [32]. Based on the information in CSRs, including trial protocols, CSRs appeared to include information about all observed AEs.

In public sources and CSR-synopses, we identified:

Five different numerical thresholds, including four in gabapentin sources (1%, 2%, 3%, and 5%) and two in quetiapine sources (5% and 10%);

Three different participant groups; all three participant groups were included in both gabapentin and quetiapine sources. These were (1) participants in any single intervention group, (2) participants receiving the test intervention (i.e., gabapentin or quetiapine), and (3) all trial participants (i.e., frequency in the combined intervention groups); and

Three different frequency thresholds, including two in gabapentin sources (no difference in frequency, frequency “higher” in the gabapentin group compared with the placebo group) and two in quetiapine sources (no difference in frequency, frequency “at least twice as high” in the quetiapine group).

We observed 9 combined selection criteria in gabapentin trials and 4 combined selection criteria in quetiapine trials, all in public sources and CSR-synopses (Figure 1). We found that all of the selection criteria that we observed included numerical

thresholds (albeit different numerical thresholds). Selection criteria did not always include information about participant group or frequency thresholds; when they did, selection criteria often included different participant groups and frequency thresholds.

When multiple sources described the same trial(s), selection criteria sometimes varied across sources (Appendix 1). Sources about the same trial were sometimes discrepant; in 2/7 (29%) trials with multiple sources that reported selection criteria, we found different sources reported different AEs (Appendix 1). In one additional trial, we found a public source that did not report AEs found in the CSR that should have been reported in the public source according to the selection criteria described in the public source (Appendix 1).

Number of trials and sources providing AE selection criteria

Although we found no trial with a public source that included all non-systematic AEs (Figure 2), 47/74 (64%) gabapentin sources and 22/50 (44%) quetiapine sources reported some AE results. Of the sources that did report AE results, 15/47 (32%) gabapentin sources and 19/22 (86%) quetiapine sources reported the selection criteria used.

Selection criteria affect AE reporting

We examined the impact of selection criteria on reporting AEs using all trials for which we had a CSR (Figure 3). By applying the 45 selection criteria to the AEs reported in eight CSRs, we determined that selection criteria could have a meaningful impact on the number of different AEs that would be reported in other sources (Figure 3). For example, there were 316 different AEs described in the CSR for Calabrese 2004 [38]. While 126/316 (40%) AEs met the selection criterion “occurring in $\geq 1\%$ of participants in any treatment group,” only 5/316 (2%) AEs met the selection criterion “occurring in $\geq 10\%$ of quetiapine-treated patients and twice as frequent in the quetiapine group as the placebo group.” All public sources about Calabrese 2004 that reported selection criteria indicated that they described all AEs occurring in $\geq 10\%$ of quetiapine-treated patients and twice as frequently in the quetiapine group compared with the placebo group.

Discussion

In this study, we examined how reports of clinical trials use selection criteria for reporting non-systematic AEs. We found that public sources and CSR-synopses applied selection criteria while the CSRs reported all AEs; consequently, most AEs and serious AEs were not reported in public sources. In public sources, every combined selection criterion we found included a numerical threshold; however, numerical thresholds were not consistent across trials or across sources for individual trials. Some selection criteria also included requirements related to the participant group and the difference between groups.

We found no evidence in public sources (e.g., journal articles) or non-public sources (i.e., CSRs) that any trial used pre-specified selection criteria. We assume that public sources reporting a small number of AEs used selection criteria; however, many sources did not describe the selection criteria they used. Among public sources that did *not* describe selection criteria, we could not determine why the authors reported some non-systematic AEs and not other non-systematic AEs. Even when selection criteria were described, public sources did not explain why the authors applied those selection criteria (e.g., following a pre-specified statistical analysis plan).

When we applied various selection criteria to eligible trials and compared what would be reported, we observed meaningful differences in the number of AEs that would be reported. Trialists might have used selection criteria to identify the most important AEs; however, we found no evidence that patients or clinicians we involved in deciding which AEs would be reported in public sources. Instead, it appears that public sources included the most common AEs. Trialists could have performed analyses similar to ours in which they applied different selection criteria and then decided *post hoc* which selection criteria would allow them to report, or not report, particular AEs. In a previous study, we found that trialists did not “group” AEs for reporting [19], which would be another method to consolidate AEs for reporting; however, grouping AEs can also disguise important AEs by combining them with less important AEs.

Evidence syntheses (e.g., systematic reviews, clinical practice guidelines) could help identify rare AEs if all observed AEs were available for all trials [39]; however, rare AEs cannot be identified when clinical trials report only those AEs occurring above numerical thresholds. Just as selectively reporting potential benefits based on quantitative results leads to biased meta-analyses [40-43] selective reporting of AEs based on trial results will necessarily lead to biased overall estimates. Selection criteria make it impossible for systematic reviewers and meta-analysts to accurately assess all AEs caused by medical interventions.

We found that no public source reported all AEs for any trial. Only serious AEs and AEs occurring in more than 5% of participants are required to be reported in www.ClinicalTrials.gov according to the Food and Drug Administration Amendments Act of 2007 (FDAAA) [1, 44]. Doctors and patients often rely on post-marketing surveillance studies to identify rare AEs, including serious AEs, yet these studies lack comparison groups that are present in clinical trials. It might be possible to detect rare AEs in clinical trials, and to calculate between-group differences, if investigators would not use numerical thresholds to determine which AEs to report. For common AEs, such as headache, investigators must ensure that data are collected consistently across intervention groups so that proportions in each group can be compared. This is difficult when AEs are collected non-systematically. Although systematically collecting AEs would result in more trustworthy and usable information about effects between groups [19, 20], it is not always possible, or even desirable, to anticipate which AEs patients will experience.

Authors have previously discouraged the use of selection criteria [33, 34]; however, trials would have to report hundreds or thousands of AEs if selection criteria were not used. Making CSRs and equivalent reports for non-industry trials public, and grouping AEs for analysis and reporting, would partially address the problem of reporting many AEs in journal articles and other public sources. A complete solution to the problems we have identified is not obvious.

Multiple sources of public AE information from trials (e.g., journal articles, FDA reviews), which may be written by different authors for different purposes, lead to inconsistent information for patients and doctors [45]. For example, FDA reviewers have access to CSRs that report all AEs that occurred in each trial. FDA reviewers apply clinical and statistical judgment when deciding what to report in medical and statistical reviews about new drugs and biologics. FDA and manufacturers also decide what to include in prescribing information (drug “labels”) written for patients and doctors. By comparison, other stakeholders get their information about interventions from a variety of sources, and those sources may use different selection criteria for reporting. U.S. law requires that investigators report on www.ClinicalTrials.gov both serious AEs and non-serious AEs that occur in more than 5% of participants in applicable clinical trials [44]. Individual authors and journal editors can decide what to report in journal articles, which vary tremendously. Different selection criteria across reports of clinical trials, including multiple reports of the same trial, lead to inconsistent and confusing information for stakeholders; consistent standards for reporting AEs, and open access to trial information (e.g., CSRs) could help.

Because current methods for selecting AEs lead to biased reporting that will necessarily produce incorrect effect estimates in systematic reviews and clinical practice guidelines, journals could require that trialists report all AEs on www.ClinicalTrials.gov or other registries. When it is not feasible to report all AEs in a given source (e.g., a conference abstract), the source could direct readers to additional information in a registry such as www.ClinicalTrials.gov. FDA and other regulators could make all AEs publicly available. If all AEs are not made available by regulators or by trialists, systematic reviewers and meta-analysts should interpret results with extreme caution and explain the limitations of using only publicly available data.

There is a pressing need to make clinical trial data available to the public, especially CSRs and IPD; however, sharing CSRs and IPD will not solve all problems identified in our research. First, reports describing hundreds of AEs might overwhelm doctors and patients by including “too much” information [46]. Many AEs reported in CSRs and IPD are not intervention-related; selection criteria might have been used to help decision-makers identify AEs that are caused by medical products. Sharing lengthy and confusing reports and datasets could increase the appearance of transparency while actually

disguising important information. Second, reanalyzing clinical trial data is time consuming and therefore expensive, and reanalysis should not be necessary to identify AEs. Most decision-makers want trustworthy *summaries* of clinical trials.

To avoid reporting bias, and to avoid overwhelming patients and doctors with information, trials could prespecify which AEs will be collected and reported. Core outcome sets for AEs could also improve the comparability of clinical trials. Core outcome sets could also identify those AEs that are most important to patients, not just the AEs that occurred most often [19,20,46].

It is a limitation that we included only two drugs-indications, and that we had non-public information for a minority of trials in each case study. The use of selection criteria, and the extent to which they are prespecified and consistent, might differ across companies and investigators. Nonetheless, the results of this methodological investigation highlight fundamental problems with methods currently used to report adverse events that occur in clinical trials.

Conclusions

Stakeholders require complete AE information to make informed health care decisions. Our findings show that systematic reviews and clinical practice guidelines are unlikely to provide accurate effect estimates for AEs because the use of selection criteria results in limited and necessarily biased information about AEs. It may be both difficult and undesirable to include all AEs in journal articles and other public reports, so standards are needed to determine which AEs to include in reports of clinical trials and which AEs to report in other public sources such as trial registers [47].

Abbreviations

AE	Adverse event
CSR	Clinical study report
IPD	Individual patient dataset
FDA	U.S. Food and Drugs Administration
FDAAA	Food and Drugs Administration Amendments Act of 2007
MUDS	Multiple Data Sources study
SRDR	Systematic Review Data Repository

Declarations

Ethics approval and consent to participate

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

Consent for publication

Not applicable

Availability of data and material

The datasets generated and analyzed during the current study, and statistical code, are available from the Dryad repository at <https://datadryad.org/resource/doi:10.5061/dryad.mp26fb1>.

Competing interests

Up to 2008, K.D. served as an unpaid expert witness for Greene LLP, the plaintiffs' lawyers in litigation against Pfizer that provided several gabapentin documents used for this study. Swaroop Vedula was paid by the plaintiffs' attorneys for research assistance provided to Kay Dickersin for her work as an expert witness. Greene LLP provided a fund for scholarly research on reporting biases that provided salary support for NF and EMW.

Funding

This work was supported by contract ME 1303 5785 from the Patient Centered Outcomes Research Institute (PCORI) and funds established for scholarly research on reporting biases at Johns Hopkins by Greene LLP. The funders were not involved in the design or conduct of the study, manuscript preparation, or the decision to submit the manuscript for publication.

Authors contributions

Concept and design: This study was completed as a sub-study of the Multiple Data Sources (MUDS) study, a larger study conceived of and designed as described in the published protocol. EMW drafted the overarching protocol, and NF contributed to the protocol and led the design of this sub-study.

Acquisition, analysis, or interpretation of data: KD, NF, TL, EMW contributed to drafting and finalizing the data extraction forms. KD, NF, EMW, and TL screened studies for inclusion. JC, NF, and EMW extracted data. JC, NF, HH, and EMW managed the data and prepared data for analysis. NF and HH analyzed the data.

Drafting of the manuscript: NF wrote an initial draft of the paper with EMW and KD. EMW and KD revised the manuscript substantially.

Critical revision: All authors reviewed, provided critical revisions, and approved the manuscript for publication.

Acknowledgements

We used the open access database, Systematic Review Data Repository (SRDR) and are grateful to Jens Jap, Bryant Smith, and Joseph Lau for making this possible without charge and for the support they provided.

The MUDS Investigators include: Lorenzo Bertizzolo, Joseph K. Canner, Terrie Cowley, Kay Dickersin, Peter Doshi, Jeffrey Ehmsen, Nicole Fusco, Gillian Gresham, Nan Guo, Jennifer A. Haythornthwaite, James Heyward, Hwanhee Hong, Tianjing Li, Evan Mayo-Wilson, Jennifer Payne, Diana Pham, Lori Rosman, Elizabeth Stuart, Catalina Suarez-Cuervo, Elizabeth Tolbert, Claire Twose, S. Swaroop Vedula. Swaroop Vedula and Peter Doshi contributed to the design of the MUDS study and to writing the grant application. Terrie Cowley, Gillian Gresham, James Heyward, Diana Pham, and Elizabeth Tolbert contributed to drafting and finalizing the data extraction forms. Lori Rosman and Claire Taylor designed and ran the electronic searches. Gillian Gresham, James Heyward, Susan Hutfless, and Swaroop Vedula screened studies for inclusion. Lorenzo Bertizzolo, Jeffrey Ehmsen, Gillian Gresham, James Heyward, Diana Pham, and Catalina Suarez-Cuervo extracted data.

References

1. Department of Health and Human Services. 42 CFR Part 11. Clinical trials registration and results information submission: Final rule. 2016. <https://www.federalregister.gov/documents/2016/09/21/2016-22129/clinical-trials-registration-and-results-information-submission>. Access date: May 12, 2018.
2. Zarin DA, Tse T, Williams RJ, et al. Trial reporting in ClinicalTrials.gov—the final rule. *The New England journal of medicine* 2016;375(20):1998-2004. doi: 10.1056/NEJMs1611785
3. Department of Health and Human Services. CFR Title 21, Section 312. Investigational New Drug Application. 2010

4. European Medicines Agency. Guideline on good pharmacovigilance practices (GVP): Annex I - Definitions (Rev 4), 2017. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2013/05/WC500143294.pdf . Access date: May 12, 2018.
5. Department of Health and Human Services. CFR Title 21, Section 312.32. IND safety reporting. 2010. <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRsearch.cfm?CFRPart=312>. Access date: May 12, 2018.
6. Golder S, Loke YK, Wright K, et al. Most systematic reviews of adverse effects did not include unpublished data. *Journal of Clinical Epidemiology* 2016;77:125-33. doi: 10.1016/S0889-8588(05)70309-9
7. Parsons R, Golder S, Watt I. More than one-third of systematic reviews did not fully report the adverse events outcome. *Journal of Clinical Epidemiology* 2019;108:95-101. doi: 10.1016/j.jclinepi.2018.12.007
8. Bonini S, Eichler H-G, Wathion N, et al. Transparency and the European Medicines Agency—sharing of clinical trial data. *The New England Journal of Medicine* 2014;371(26):2452-55.
9. Doshi P, Jefferson T. The first 2 years of the European Medicines Agency's policy on access to documents: secret no longer. *JAMA Intern Med* 2013;173(5):380-82.
10. U.S. Food and Drug Administration. FDA Commissioner Scott Gottlieb, M.D., on new steps FDA is taking to enhance transparency of clinical trial information to support innovation and scientific inquiry related to new drugs [press release]. January 16, 2018
11. Dickersin K, Min Y-I, Meinert CL. Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. *JAMA* 1992;267(3):374-78. doi: 10.1001/jama.1992.03480030052036
12. Dwan K, Gamble C, Williamson PR, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS One* 2013;8(7):e66844. doi: 10.1371/journal.pone.0066844
13. Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 2005;330(7494):753-56. doi: 10.1136/bmj.38356.424606.8F
14. Chan AW, Hrobjartsson A, Haahr MT, et al. Empirical evidence for selective reporting of outcomes in randomized trials - Comparison of Protocols to published articles. *JAMA* 2004;291(20):2457-65. doi: 10.1001/jama.291.20.2457
15. Chan AW, Krieza-Jeric K, Schmid I, et al. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *Can Med Assoc J* 2004;171(7):735-40. doi: 10.1503/cmaj.1041086
16. Dwan K, Altman DG, Clarke M, et al. Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. *PLoS Med* 2014;11(6):e1001666. doi: 10.1371/journal.pmed.1001666
17. Mayo-Wilson E, Fusco N, Li T, et al. Multiple outcomes and analyses in clinical trials create challenges for interpretation and research synthesis. *Journal of Clinical Epidemiology* 2017;86:39-50. doi: 10.1016/j.jclinepi.2017.05.007
18. Vedula SS, Bero L, Scherer RW, et al. Outcome reporting in industry-sponsored trials of gabapentin for off-label use. *The New England Journal of Medicine* 2009;361(20):1963-71. doi: 10.1056/NEJMsa0906126
19. Mayo-Wilson E, Fusco N, Li T, et al. Harms are assessed inconsistently and reported selectively: Non-systematic adverse events in clinical trials. *Submitted* 2018
20. Mayo-Wilson E, Fusco N, Li T, et al. Harms are assessed inconsistently and reported selectively: Systematic adverse events in clinical trials. *Submitted* 2018
21. Mayo-Wilson E, Li T, Fusco N, et al. Cherry-picking by trialists and meta-analysts can drive conclusions about intervention efficacy. *Journal of Clinical Epidemiology* 2017;91:95-110. doi: 10.1016/j.jclinepi.2017.07.014
22. Golder S, Loke YK, Wright K, et al. Reporting of adverse events in published and unpublished studies of health care interventions: a systematic review. *PLoS Med* 2016;13(9):e1002127. doi: 10.1371/journal.pmed.1002127
23. Halfpenny NJA, Quigley JM, Thompson JC, et al. Value and usability of unpublished data sources for systematic reviews and network meta-analyses. *Evid-Based Med* 2016;21(6):208-13. doi: 10.1136/ebmed-2016-110494

24. Hart B, Lundh A, Bero L. Effect of reporting bias on meta-analyses of drug trials: reanalysis of meta-analyses. *BMJ* 2012;344:d7202. doi: 10.1136/bmj.d7202
25. Hartung DM, Zarin DA, Guise J-M, et al. Reporting discrepancies between the ClinicalTrials.gov results database and peer-reviewed publications. *Ann Intern Med* 2014;160(7):477-83. doi: 10.7326/M13-0480
26. Jefferson T, Jones MA, Doshi P, et al. Neuraminidase inhibitors for preventing and treating influenza in healthy adults and children. *Cochrane Database Syst Rev* 2012;1 Art. No.: CD008965 doi: 10.1002/14651858.CD008965.pub4
27. Le Noury J, Nardo JM, Healy D, et al. Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. *BMJ* 2015;351:h4320. doi: 10.1136/bmj.h4320
28. Riveros C, Dechartres A, Perrodeau E, et al. Timing and completeness of trial results posted at ClinicalTrials.gov and published in journals. *PLoS Med* 2013;10(12):e1001566. doi: 10.1371/journal.pmed.1001566
29. Rodgers MA, Brown JV, Heirs MK, et al. Reporting of industry funded study outcome data: comparison of confidential and published data on the safety and effectiveness of rhBMP-2 for spinal fusion. *BMJ* 2013;346:f3981. doi: 10.1136/bmj.f3981
30. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline: Clinical Safety Data Management: Definitions and Standards for Expedited Reporting E2A. 1994.
http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E2A/Step4/E2A_Guideline.pdf; Access date: May 12, 2018.
31. Zarin DA, Tse T, Williams RJ, et al. The ClinicalTrials.gov results database—update and key issues. *The New England Journal of Medicine* 2011;364(9):852-60. doi: 10.1056/NEJMs1012065
32. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline: Structure and Content of Clinical Study Reports E3. 1995.
<http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/structure-and-content-of-clinical-study-reports.html>. Access date: May 12, 2018.
33. Lineberry N, Berlin JA, Mansi B, et al. Recommendations to improve adverse event reporting in clinical trial publications: a joint pharmaceutical industry/journal editor perspective. *BMJ* 2016;355:i5078.
34. Ioannidis JP, Evans SJ, Gøtzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004;141(10):781-88. doi: 10.7326/0003-4819-141-10-200411160-00009
35. Mayo-Wilson E, Hutfless S, Li T, et al. Integrating multiple data sources (MUDS) for meta-analysis to improve patient-centered outcomes research: a protocol for a systematic review. *Systematic Reviews* 2015;4:143 (Correction: 2018; 7:48). doi: 10.1186/s13643-015-0134-z
36. Mayo-Wilson E, Doshi P, Dickersin K. Are manufacturers sharing data as promised? *BMJ* 2015;351:h4169. doi: 10.1136/bmj.h4169
37. Stata Statistical Software: Release 14 [program]. College Station, TX, 2015.
38. AstraZeneca. A multicenter, double-blind, randomized, placebo-controlled, double-dummy trial of the use of quetiapine fumarate (SEROQUEL[Registered]) in the treatment of patients with bipolar depression. Study code: 5077US/0049, 2005.
39. Chalmers I. Proposal to outlaw the term “negative trial”. *BMJ* 1985;290(6473):1002.
40. Hutton J, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. *J Roy Stat Soc C-App* 2000;49(3):359-70. doi: 10.1111/1467-9876.00197
41. Kirkham JJ, Dwan KM, Altman DG, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365. doi: 10.1136/bmj.c365
42. Williamson PR, Gamble C. Identification and impact of outcome selection bias in meta-analysis. *Statistics in Medicine* 2005;24(10):1547-61. doi: 10.1002/sim.2025

43. Williamson PR, Gamble C, Altman DG, et al. Outcome selection bias in meta-analysis. *Statistical Methods in Medical Research* 2005;14(5):515-24. doi: 10.1191/0962280205sm4150a [published Online First: 2005/10/27]
44. 110th Congress of the United States of America. Food and Drug Administration Amendments Act, Section 801. 2007
45. Mayo-Wilson E, Li T, Fusco N, et al. Practical guidance for using multiple data sources in systematic reviews and meta-analyses (with examples from the MUDS study). *Res Synth Methods* 2018;9(1):2-12. doi: 10.1002/jrsm.1277
46. Mayo-Wilson E, Golozar A, Cowley T, et al. Methods to identify and prioritize patient-centered outcomes for use in comparative effectiveness research (in press). *Pilot and Feasibility Studies* 2018;4(95) doi: 10.1186/s40814-018-0284-6
47. Dickersin K, Mayo-Wilson E. Standards for design and measurement would make clinical research reproducible and usable. *P Natl Acad Sci USA* 2018;115(11):2590-94. doi: 10.1073/pnas.1708273114

Tables

Box 1. Terms related to adverse events (AEs)	
Adverse event (AE)	The International Conference on Harmonisation (ICH) defines an “adverse event” as “any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product and which does not necessarily have to have a causal relationship with this treatment” [30]. The U.S. Food and Drug Administration (FDA) and other regulators use this definition [4 ,5].
Non-systematic adverse events	According to The Final Rule [1 ,2], “‘non-systematic assessment’ relies on the spontaneous reporting of adverse events, such as unprompted self-reporting by participants”. Non-systematic adverse events may be collected by asking questions such as “Have you noticed any symptoms since your last examination?”.
Result	In the Multiple Data Source (MUDS) study, a “result” is a numerical contrast between a treatment and comparison arm (e.g., relative risk, mean difference).
Serious adverse events	The International Conference on Harmonisation (ICH) defines a “serious adverse event” as that which “results in death, is life-threatening, requires inpatient hospitalisation or prolongation of existing hospitalisation, results in persistent or significant disability/incapacity, or is a congenital anomaly/birth defect” [30]. FDA and other regulators use this definition [4 ,5].
Systematic adverse events	According to The Final Rule [1 ,2], “‘systematic assessment’ involves the use of a specific method of ascertaining the presence of an adverse event (e.g., the use of checklists, questionnaires, specific laboratory tests at regular intervals)”. Like a potential benefit of treatment, a systematic AE can be defined using five elements: (1) domain, (2) specific measurement, (3) specific metric, (4) method of aggregation, and (5) time-point [31]. For example, “proportion of participants with 50% change from baseline to 8 weeks on the Young Mania Rating Scale total score.”
Terms related to sources	
Clinical study report (CSR)	A comprehensive document, often created by a pharmaceutical manufacturer for submission to a regulator, detailing the design, methods, analyses, and results of a study. Appendices sometimes contain tables of individual patient data, also called “patient data listings”, and study protocols [32].
Clinical study report synopsis (CSR-synopsis)	A document that summarizes the information contained in a clinical study report. Clinical study report-synopses are much shorter than clinical study reports; the two clinical study report-synopses we examined were each 13 pages in length.
Individual patient data (IPD)	A table or database in which each record contains data for a single participant [32].
Non-public sources	In the MUDS study, non-public sources include individual patient data, clinical study reports, and clinical study report-synopses.
Public sources	In the MUDS study, public sources include journal articles, conference abstracts, commentaries, posters, trial registrations and associated results, and Medical Reviews and Statistical Reviews written by the FDA.

Box 2. Examples of selection criteria, with each component identified in bold

Two selection criteria:

Adverse events are reported if they occurred in $\geq 5\%$ of **participants in any intervention group**.

Adverse events are reported if they occurred in $\geq 2\%$ of **participants receiving gabapentin**.

Three selection criteria:

Adverse events are reported if they occurred in $\geq 2\%$ of **participants receiving gabapentin** and if they occurred **at least twice as frequently** in participants receiving gabapentin, compared with participants receiving placebo.

Adverse events are reported if they occur in $\geq 1\%$ of **all participants** and if they occur **more frequently** in participants receiving gabapentin, compared with participants receiving placebo.

Figures

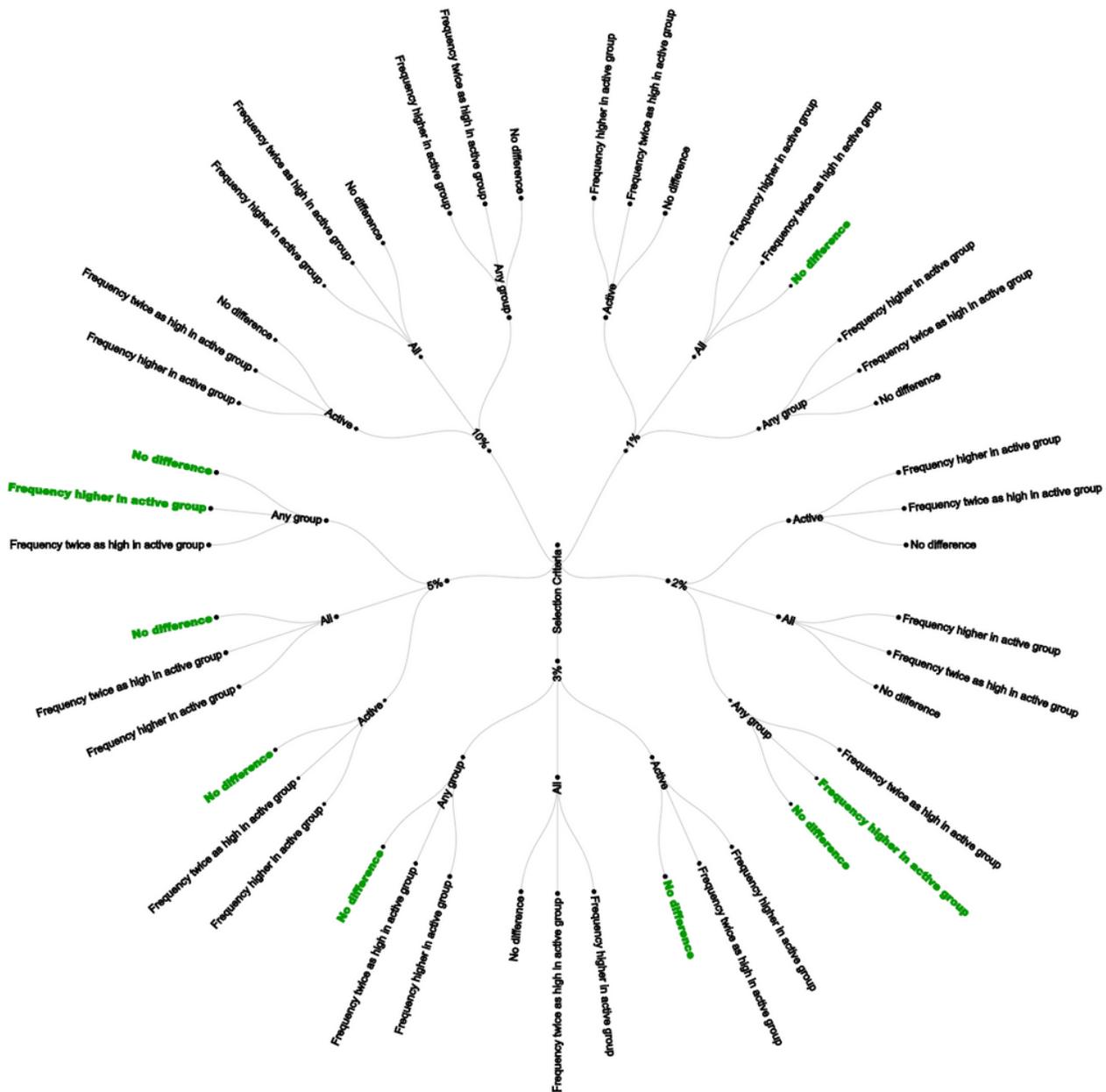


Figure 1

Figure 1A - Selection criteria in gabapentin trials. Observed components of selection criteria: This figure shows the selection criteria we applied in sources about gabapentin. The figures below show all possible combinations of selection criteria. Each of the three rings of the circle represents a different component of the selection criteria: numerical threshold, participant group, and difference in frequency threshold. Green text indicates selection criteria reported in public sources. All=All participants combined across groups; Any group=Participants in a particular intervention group; Active=Participants in the active intervention group (i.e., gabapentin or quetiapine).

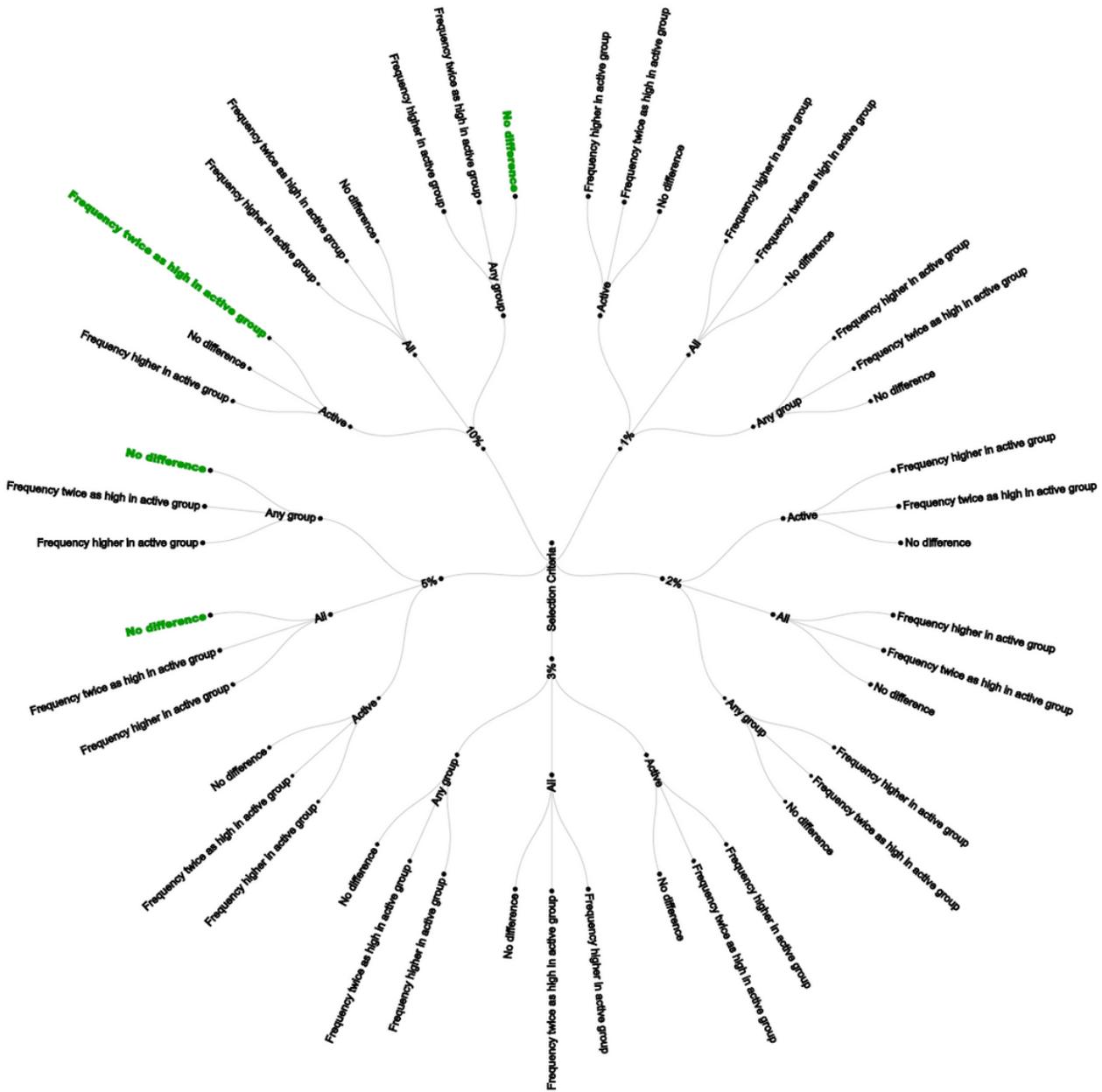


Figure 2

Figure 1B - Selection criteria in quetiapine trials. Observed components of selection criteria: This figure shows the selection criteria we applied in sources about quetiapine. The figures below show all possible combinations of selection criteria. Each of the three rings of the circle represents a different component of the selection criteria: numerical threshold, participant group, and difference in frequency threshold. Green text indicates selection criteria reported in public sources. All=All

participants combined across groups; Any group=Participants in a particular intervention group; Active=Participants in the active intervention group (i.e., gabapentin or quetiapine).

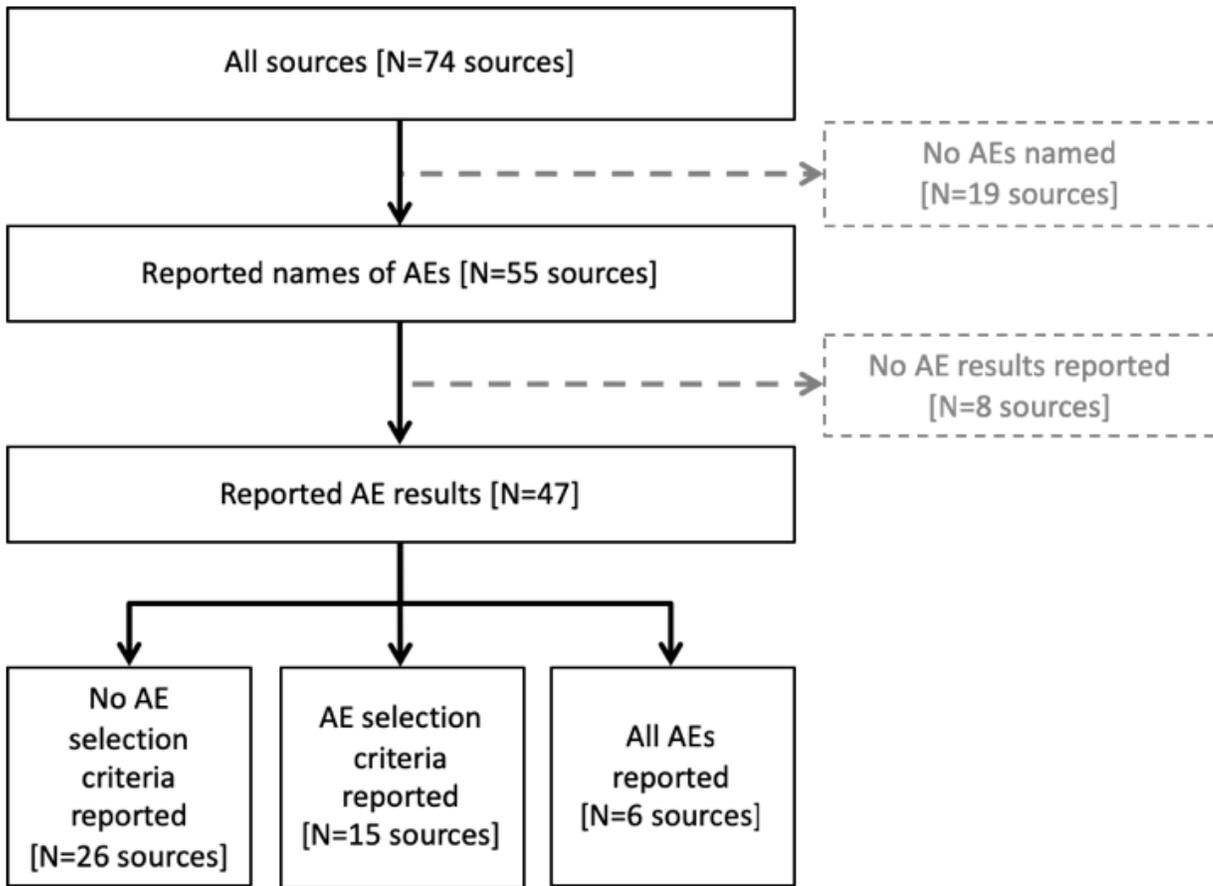


Figure 3

Figure 2A - Gabapentin. Legend: Six CSRs we found appeared to report all AEs. All other sources in this figure are public sources.

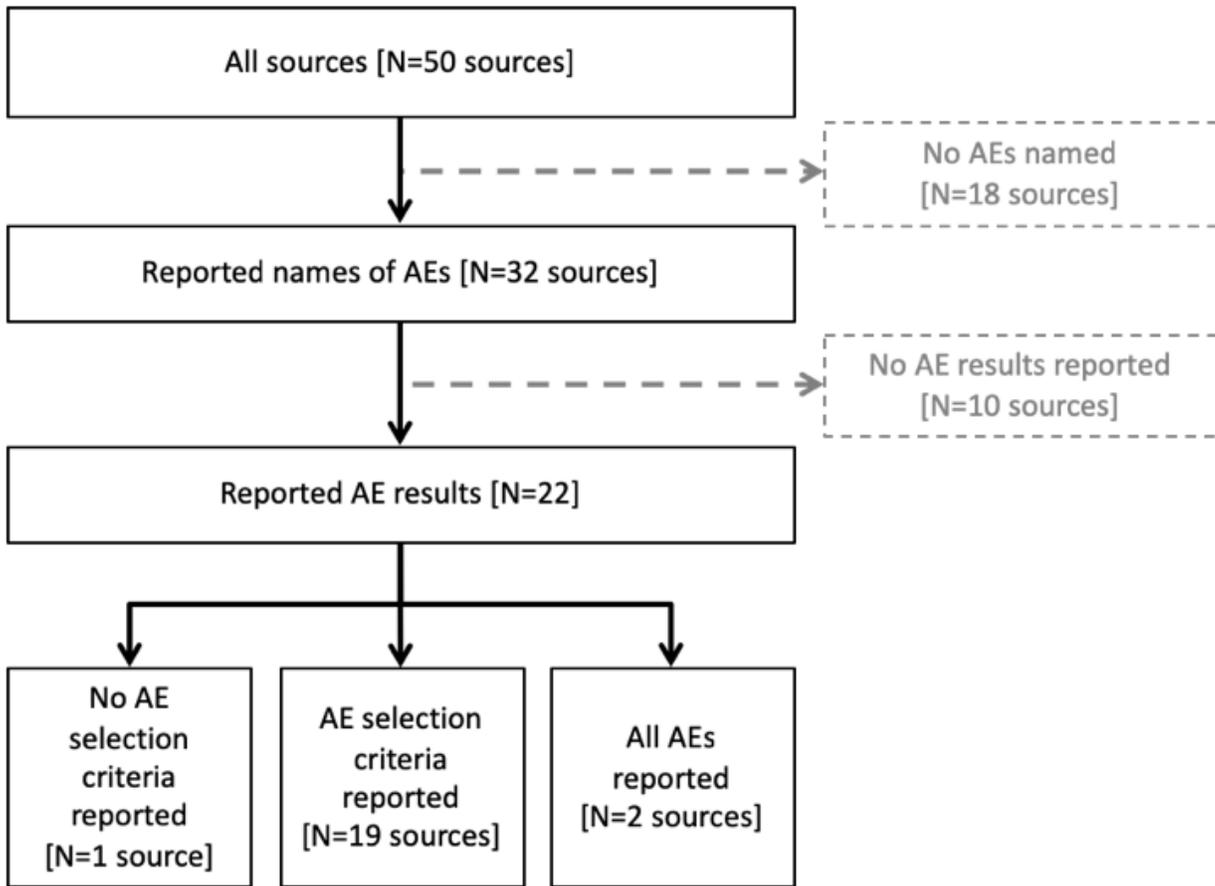


Figure 4

Figure 2B - Quetiapine. Legend: Two CSRs we found appeared to report all AEs. Two CSR-synopses reported AE results and AE selection criteria. All other sources in this figure are public sources.

Difference in frequency threshold	Percent of adverse events in each trial reported using different selection criteria											
	A	B	C	A	B	C	A	B	C	A	B	C
Numerical threshold and participant group	945-224 ¹ (N=91)			A945-1008 ¹ (N=198)			Backonja 1998 ¹ (N=96)			Rice 2001 ¹ (N=108)		
1% any treatment group	100	65	57	38	19	13	99	59	55	40	30	24
1% gabapentin/quetiapine treatment group	86	65	57	25	19	13	78	59	55	31	30	24
1% all participants combined across groups	23	19	11	30	18	12	47	23	19	27	21	16
2% any treatment group	29	22	14	17	12	8	39	23	19	26	20	16
2% gabapentin/quetiapine treatment group	25	22	14	14	12	8	29	23	19	22	20	16
2% all participants combined across groups	13	10	4	15	10	7	21	14	9	13	11	8
3% any treatment group	16	13	7	12	9	7	23	15	10	16	12	9
3% gabapentin/quetiapine treatment group	14	13	7	10	9	7	19	15	10	14	12	9
3% all participants combined across groups	8	7	4	8	6	4	19	11	7	9	7	6
5% any treatment group	9	7	4	4	4	3	13	9	7	9	7	6
5% gabapentin/quetiapine treatment group	7	7	4	4	4	3	9	9	7	7	7	6
5% all participants combined across groups	2	2	2	4	4	3	7	5	3	4	4	3
10% any treatment group	1	1	1	2	2	2	5	4	3	3	3	3
10% gabapentin/quetiapine treatment group	1	1	1	2	2	2	4	4	3	3	3	3
10% all participants combined across groups	0	0	0	2	2	2	2	2	2	2	2	2
	Rowbotham 1998 ² (N=122)			Serpell 2002 ¹ (N=111)			Calabrese 2004 ² (N=316)			Thase 2006 ² (N=273)		
1% any treatment group	39	23	20	50	32	22	40	27	21	45	27	21
1% gabapentin/quetiapine treatment group	31	23	20	42	32	22	37	27	21	37	27	21
1% all participants combined across groups	30	18	15	29	17	10	24	17	14	24	15	9
2% any treatment group	23	16	12	23	14	10	21	16	13	21	14	8
2% gabapentin/quetiapine treatment group	20	16	12	22	14	10	19	16	13	18	14	8
2% all participants combined across groups	16	10	7	18	10	5	12	9	7	12	10	5
3% any treatment group	13	10	7	19	12	8	10	8	6	11	8	5
3% gabapentin/quetiapine treatment group	12	10	7	16	12	8	10	8	6	10	8	5
3% all participants combined across groups	12	7	4	11	5	2	7	5	3	8	5	3
5% any treatment group	6	4	4	10	5	2	5	4	3	8	5	3
5% gabapentin/quetiapine treatment group	4	4	4	8	5	2	5	4	3	7	5	3
5% all participants combined across groups	6	3	3	8	5	2	4	3	2	5	3	2
10% any treatment group	2	2	2	4	2	2	3	2	2	3	2	2
10% gabapentin/quetiapine treatment group	2	2	2	2	2	2	2	2	2	3	2	2
10% all participants combined across groups	2	2	2	3	1	1	2	1	1	2	1	1

Figure 5

Figure 3 - Percentage of adverse events that would be reported using different selection criteria. Legend: We applied 45 different selection criteria to AEs in each of the eight trials for which we identified a clinical study report (CSR). To illustrate the potential variation in reported non-systematic AEs for each trial, we calculated the percentage of adverse events (AEs) that would be reported using each selection criterion, and we colored the figure using a heat map with green representing the most AEs and red representing the fewest AEs. Squares outlined in black represent the selection criteria used in at least one source about that trial (e.g., a source describing Serpell 2002 reported AEs that occurred in $\geq 5\%$ of gabapentin participants). Some trials did not have any sources that described the selection criteria (e.g., 945-224). A=No difference in frequency threshold; B=Higher frequency in gabapentin/quetiapine than in placebo; C=Frequency in gabapentin/quetiapine at least twice as high as placebo; 1Gabapentin trial; 2Quetiapine trial.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement1.docx](#)