Final Research Report: Integrating Multiple Data Sources for Meta-analysis to Improve Patient-centered Outcomes Research

Kay Dickersin, PhD1; Evan Mayo-Wilson, MPA, DPhil1; Tianjing Li1

1Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

For the MUDS Investigators

PCORI Project ID: ME-1303-5785
HSRproj ID: 20143578

To cite this document, please use:
Table of Contents

THE MUDS INVESTIGATORS ........................................................................................................................ 4

ABSTRACT .......................................................................................................................................................... 6

1. BACKGROUND .............................................................................................................................................. 9
   1.1. Objectives and overview ........................................................................................................................... 9
   1.2. Patient-centered outcomes ....................................................................................................................... 9
   1.3. Reporting biases ....................................................................................................................................... 10
   1.4. Multiple data sources ............................................................................................................................... 10

2. METHODS ................................................................................................................................................... 11
   2.1. Selection of case studies ........................................................................................................................... 11
      2.1.1. Gabapentin for neuropathic pain.................................................................................................................. 11
      2.1.2. Quetiapine for bipolar depression ............................................................................................................. 11
   2.2. Methods for identifying patient-centered outcomes ........................................................................... 12
      2.2.1. Eligibility criteria for a trial’s inclusion in our study ............................................................................ 12
      2.2.2. Search methods for identification of trials ............................................................................................ 13
      2.2.3. Selection of trials ......................................................................................................................................... 14
      2.2.4. Data collection ............................................................................................................................................. 14
   2.3. The effect of multiple data sources on meta-analyses ........................................................................ 12
      2.3.1. Series of meta-analyses ......................................................................................................................... 16
   2.4. Analysis .................................................................................................................................................. 15
      2.4.1. Assessment of trial quality .................................................................................................................... 15
      2.4.2. Series of meta-analyses ......................................................................................................................... 16

3. RESULTS ...................................................................................................................................................... 17
   3.1. Patient-centered outcomes identified through existing sources and stakeholder
       coinvestigators ................................................................................................................................................. 17
      3.1.1. Core outcome sets .................................................................................................................................... 17
      3.1.2. Social media ........................................................................................................................................... 17
      3.1.3. DRUGDEX ............................................................................................................................................... 17
      3.1.4. Food and Drug Administration prescribing information ........................................................................ 18
      3.1.5. Patient and stakeholder investigators ................................................................................................. 18
   3.2. The effects of adding and replacing data from multiple data sources on the results of meta-
       analyses ......................................................................................................................................................... 18
      3.2.1. Results of the search ............................................................................................................................ 18
      3.2.2. Completeness of information ............................................................................................................... 22
THE MUDS INVESTIGATORS

Lorenzo Bertizzolo  
Department of Epidemiology  
Johns Hopkins University Bloomberg School of Public Health

Joseph K. Canner  
Department of Surgery  
Johns Hopkins University School of Medicine

Terrie Cowley  
The TMJ Association, Ltd.

Kay Dickersin, PhD (Principal Investigator)  
Department of Epidemiology  
Johns Hopkins Bloomberg School of Public Health

Peter Doshi  
Pharmaceutical Health Services Research  
University of Maryland School of Pharmacy

Jeffrey Ehmsen  
Department of Neurology-Neuromuscular Medicine  
Johns Hopkins University School of Medicine

Nicole Fusco  
Department of Epidemiology  
Johns Hopkins University Bloomberg School of Public Health

Gillian Gresham  
Department of Epidemiology  
Johns Hopkins University Bloomberg School of Public Health

Nan Guo  
Department of Epidemiology  
Johns Hopkins University Bloomberg School of Public Health

Jennifer A Haythornthwaite  
Department of Psychiatry & Behavioral Sciences  
Johns Hopkins University School of Medicine

James Heyward  
Department of Epidemiology  
Johns Hopkins University Bloomberg School of Public Health

Hwanhee Hong  
Department of Mental Health  
Johns Hopkins University Bloomberg School of Public Health

Tianjing Li  
Department of Epidemiology
ABSTRACT

Background: International standards include recommendations that systematic reviews be comprehensive, but time and resources may render it impractical to search for and extract data from all possible sources of information. For example, many data sources exist for randomized controlled trials (RCTs), a number of which are not publicly available or are difficult or impossible to access. Searching nonpublic sources of RCTs may improve the impact of systematic reviews by identifying information that was recorded but not included in public sources, thus reducing the need for additional RCTs and improving research efficiency.

Objectives: To determine whether multiple data sources about RCTs affected systematic reviews and meta-analyses of patient-centered outcomes (PCOs) research (e.g., trial quality assessment, pooled effect estimates) and to produce open access guidance about using multiple data sources, for producers of systematic reviews of PCOs research.

Methods: We conducted a methods study related to the conduct of systematic reviews using 2 case studies: (1) gabapentin for neuropathic pain and (2) quetiapine for bipolar depression. Applying comprehensive searching methods as if we were conducting 2 systematic reviews, we attempted to identify all data sources for eligible RCTs. We extracted data and compared the information about trial design and trial quality (i.e., risk of bias) using the multiple data sources. Using these multiple data sources, we extracted information about prespecified outcome domains, including the definition of each outcome (i.e., the outcome domain, measure, metric, method of aggregation, and time-point) and the associated results (i.e., numerical estimates of treatment effectiveness). Then, we compared the results of meta-analyses using multiple data sources by conducting multiple meta-analyses in which we systematically added and replaced data from various sources. We compared the outcomes that matter most to patients with the outcomes that were examined in clinical trials.

Results: Most clinical trials in our case studies were associated with multiple data sources, including public sources (e.g., journal articles, conference abstracts, trial registrations, and Food and Drug Administration (FDA) reviews) and nonpublic sources (e.g., clinical study reports and individual patient data). We found 21 gabapentin RCTs (74 reports, 6 individual participant data) and 7 quetiapine RCTs (50 reports, 1 individual participant data); we found nonpublic sources for 6 of 21 (29%) gabapentin and 4 of 7 (57%) quetiapine RCTs. We identified literally hundreds of PCOs reported in the sources we found that could be included in our meta-analyses. We surmised that there were many opportunities for selective outcome reporting by trialists and systematic reviewers. The process of identifying all sources of information—and extracting and analyzing data—required considerable time and skill. Nonpublic sources were especially difficult to identify. Clinical
study reports were time consuming to extract, and individual participant data required time and skill to prepare the information for analysis.

Data sources differed in completeness. Most RCTs (18/21 [86%] and 6/7 [86%], respectively) were reported in journal articles, which often presented unclear information related to trial quality. When nonpublic sources were available for RCTs, clinical study reports contained mostly information about trial design and trial quality, and clinical study reports and individual participant data contained the most results. In these case studies, individual participant data were not accompanied by any metadata (e.g., codebooks or description of trial methods).

For a single RCT reported in a single source, 1 outcome domain (e.g., pain intensity) could be associated with multiple outcome definitions (e.g., using multiple measures) and results. Thus, multiple results for an outcome could be presented selectively, even when an outcome domain was prespecified. Multiple data sources associated with a single RCT sometimes contained contradicting information about trial design.

In the series of meta-analyses using results for a single outcome domain, measure, and time-point for each case study, the effect of gabapentin on pain intensity ranged from SMD = –0.46 (95% CI –0.64, –0.28) to SMD = –0.31 (95% CI –0.47, –0.15), and the effect of quetiapine on depression ranged from SMD = –0.42 (95% CI –0.62, –0.22) to SMD = –0.44 (95% CI –0.63, –0.26). The range of results might have been wider had we included other outcome measures and time-points in our meta-analyses.

Clinical study reports and individual participant data contained hundreds of harms (adverse events) that were not included in public sources. Further analyses would be required to determine if conclusions about the relative benefits and harms of treatment would be affected by a systematic synthesis of harms included only in nonpublic sources.

**Discussion:** There is tremendous variation in the information available across multiple data sources from individual trials. Reported estimates from individual trials were vulnerable to selective reporting at both the trial and the systematic review levels.

Although an “open science” environment is a positive move forward, it will not solve all the problems that we identified with how trial information is currently made available. For example, an open trial data set for the purpose of making research reproducible allows for the investigation of new research questions, systematic reviews, and individual participant data meta-analysis; however, the steps required to share a data set and the steps required to utilize an open data set are resource intensive and demand skills that are not taught routinely to trialists or systematic reviewers, such as creating, storing, accessing, synthesizing, and analyzing information from multiple data sources. Furthermore, patients and stakeholders are often unfamiliar with the
wide variety of existing data sources and what can be gained from each of them. Thus, current proposals to expand data sharing require careful and community-wide decision making about implementation.

Although guidance for systematic reviewers suggests searching extensively for information sources and including all data sources found, this approach is neither practical nor possible for many systematic reviews. Our research provides practical insights into a complex area that needs further research and discussion.
1. BACKGROUND

1.1. Objectives and overview

Our objectives were to determine whether multiple data sources about randomized controlled trials (RCTs) affected systematic reviews and meta-analyses of patient-centered outcomes (PCOs) research and to produce open access guidance about using multiple data sources for producers of systematic reviews of PCOs research.

To meet these objectives, we conducted a methods study related to the conduct of systematic reviews using 2 case studies: (1) gabapentin for neuropathic pain and (2) quetiapine for bipolar depression. We did not conduct systematic reviews; instead, we examined whether and how the results of systematic reviews might be affected by how trial data are made available, and how meta-analyses might be affected by those who use data from multiple sources for a single trial. Although we investigated methods used in conducting systematic reviews and the effect of those methods on meta-analyses, we did not consider our work to comprise multiple systematic reviews because we did not evaluate the benefits and harms of the interventions per se. Our findings may be of greatest interest to trialists, systematic reviewers, journal editors, and guideline developers.

1.2. Patient-centered outcomes

We aimed to examine whether we found additional PCOs (meaningful to patients) by looking through all reporting sources for each trial, not just a few sources. Systematic reviewers may rely on easy-to-find sources (e.g., public sources) for trial outcomes, and they may not list PCOs that could be available from other sources (e.g., nonpublic sources). Searching nonpublic sources may improve the impact of systematic reviews by identifying PCOs that were recorded but not included in public sources. Examining nonpublic data sources for PCOs could reduce the need for additional RCTs and improve the efficiency of PCOs research. As far as we are aware, this possibility has never been addressed in research.
1.3. Reporting biases

Reporting biases threaten the reliability of RCTs and systematic reviews. Failure to report trials at all and selective reporting of design, methods, or results represents an ethical breach with research participants who believed that by participating in a trial they were contributing to knowledge. Reporting bias means that information from a trial is not transmitted to the public and thus also constitutes wasted information.\(^1\) When failure to report trials (or failure to report trials completely) is not random, it biases the trial findings and what is generally believed to be true.\(^2\)

1.4. Multiple data sources

To decrease the threat of reporting biases, current best practices for the conduct of systematic reviews include a comprehensive search for trial information.\(^3,4\) Standards for such a search have focused on public sources of information about trials and include going beyond journal articles and searching the “gray” literature such as conference abstracts, FDA reviews, and trial registries. This is mainly because studies of reporting bias have shown that not all information about trial design and quality can be found in journals, and publication of trial results in a journal is associated with positive trial results.\(^5,6\)

In addition, studies have shown that additional information about a trial can be found in non-public sources.\(^7,9\) Although systematic reviewers may wish to search for and include all relevant and reliable data in systematic reviews, comprehensive searching adds to the time and resources required to complete systematic reviews. In addition, information from the various sources may not always agree,\(^8\) and systematic reviewers must decide which source is correct.\(^10\) The value of searching for non-public information is unclear, partly because the best methods for searching are unknown.

Assuming that comprehensive searching for trial information provides additional data, questions remain about which public and non-public sources should be searched for and used in systematic reviews. Empirically grounded guidance is needed to help systematic reviewers make choices about the use of data from multiple sources. Increasingly, the move to open science is making non-public sources available.\(^11\) Even if all that is currently hidden becomes public, we would still need guidance to address the gains and challenges from accessing multiple data sources for a single trial.
2. METHODS

We published a protocol describing detailed methods for our study. What follows is a summary of our methods.

2.1. Selection of case studies

To evaluate the effect of using data from multiple sources in systematic reviews and meta-analyses, we examined 2 case studies regarding the effectiveness and safety of (1) gabapentin (Neurontin®) for neuropathic pain in adults and (2) quetiapine (Seroquel®) for the treatment of depression in adults with bipolar disorder following a published protocol. We used methods consistent with the National Academy of Medicine (formerly the Institute of Medicine) standards for conducting systematic reviews.

We selected these cases for several reasons. First, gabapentin and quetiapine are used commonly for these respective conditions, so the included studies are clinically important. Second, pain and depression are associated with several patient-reported outcomes. Methodologically, these cases also allow us to consider different, typical situations that systematic reviewers might face with respect to data from multiple sources.

2.1.1. Gabapentin for neuropathic pain

Neuropathic pain occurs when there is damage to the peripheral nervous system, which is the array of nerves that transmit information from the central nervous system (brain and spinal cord) to other parts of the body. Between 3% and 10% of the population may be living with neuropathic pain, which affects 30% to 50% of people with diabetes in particular. Other consequences include loss of sleep, reduced quality of life, and high health care costs.

The FDA approved gabapentin for the treatment of epilepsy in 1993, and it approved gabapentin for the treatment of postherpetic neuralgia (i.e., residual pain in people who have had shingles) in 2002. Gabapentin is used off label for a variety of symptoms, including for neuropathic and other types of pain.

2.1.2. Quetiapine for bipolar depression

Bipolar disorder is characterized by episodes of depression and at least 1 episode of mania or hypomania and has a lifetime prevalence of 1% to 4%. Work, family life, and social life are impaired significantly by depressive episodes. People with bipolar disorder are at increased risk of suicide compared with the general population and compared with people who have other mental health problems.

Quetiapine is an antipsychotic medication originally used for the treatment of acute psychotic episodes. Quetiapine is currently recommended as a first-line choice for treatment of acute bipolar depression.
although it is associated with outcomes that are undesirable to patients, including daytime sleepiness, cognitive impairment, loss of libido, and rapid weight gain. Quetiapine and other antipsychotics are also associated with serious adverse events, including cardiac and metabolic effects, and extrapyramidal symptoms.23,24

2.2. Methods for identifying patient-centered outcomes

A goal of this project was to examine what matters most to patients who take gabapentin for neuropathic pain and to patients who take quetiapine for bipolar depression. We used existing resources and consulted patient coinvestigators. We compared the list of outcomes that matter most to patients with the outcomes that were examined in clinical trials.

We identified PCOs by reviewing existing information sources, and we worked with our patient coinvestigators and clinical experts to identify those outcomes they consider patient centered. Specifically, we examined the following:

1. **Core outcome sets (COSs)** using the Core Outcome Measures in Effectiveness Trials database25 (http://www.comet-initiative.org) and the James Lind Alliance (http://www.lindalliance.org) in April and May 2014, respectively. We asked coinvestigators and colleagues to identify COSs related to neuropathic pain and bipolar disorder.

2. A **social media website** (http://www.patientslikeme.com/) in April 2014. PatientsLikeMe allows patients to enter information about their medical history and interventions they have used.

3. **DRUGDEX** (http://micromedex.com/) in January 2015. DRUGDEX is a compendium health care providers use to make treatment decisions. The Centers for Medicare and Medicaid may use DRUGDEX ratings to make reimbursement decisions related to medically accepted treatments. DRUGDEX contains a list, by organ system, of each drug’s adverse effects.

4. **FDA prescribing information** (“package inserts”) accessed through the Drugs@FDA website in April 2014

After examining existing sources, we asked patient and stakeholder partners to identify outcomes they thought should be included in each case study. We discussed the outcomes identified through all sources, and we selected those outcomes that patient and stakeholder partners deemed most important. We then generated a list of outcomes reported in clinical trials for each case study, and we examined whether PCOs were included in sources of clinical trials.

2.3. The effect of multiple data sources on meta-analyses

2.3.1. Eligibility criteria for a trial’s inclusion in our study

Our published protocol includes detailed inclusion and exclusion criteria for each case study.12 Table 1 describes the eligibility criteria for participants, interventions, outcomes, comparisons, and time-points.
Table 1: Eligibility criteria for the case studies

<table>
<thead>
<tr>
<th>Gabapentin</th>
<th>Quetiapine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participants</strong></td>
<td><strong>Included:</strong>&lt;br&gt;• Adults (≥ 18 years)&lt;br&gt;• Neuropathic pain&lt;br&gt;• Not hospitalized</td>
</tr>
<tr>
<td><strong>Interventions</strong></td>
<td><strong>Included:</strong>&lt;br&gt;• Oral gabapentin&lt;br&gt;• ≥ 300mg per day&lt;br&gt;• Regular or extended release</td>
</tr>
<tr>
<td><strong>Comparisons</strong></td>
<td><strong>Included:</strong>&lt;br&gt;• No intervention&lt;br&gt;• Pill placebo&lt;br&gt;• Treatment(s) common to both groups (i.e., gabapentin was the only difference between groups)</td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td><strong>Five prespecified outcome domains:</strong>&lt;br&gt;• Mood&lt;br&gt;• Pain intensity&lt;br&gt;• Pain interference&lt;br&gt;• Quality of life&lt;br&gt;• Sleep disturbance</td>
</tr>
<tr>
<td><strong>Time-points</strong></td>
<td>• 8 weeks (4-13 weeks)&lt;br&gt;• 18 weeks (14-22 weeks)&lt;br&gt;• 27 weeks (23-31 weeks)</td>
</tr>
</tbody>
</table>

2.3.2. Search methods for identification of trials

We conducted electronic and additional searches to identify trials as described in our study protocol.12 We searched for the following types of sources, which are listed by their approximate level of expected detail (i.e., less detail to more detail):

1. Trial registrations without results posted on trial registries (e.g., www.ClinicalTrials.gov)
2. Short reports (e.g., conference abstracts, posters)
3. Trial registrations without with summary data posted on trial registries
4. Peer reviewed journal articles
5. Dissertations (e.g., master’s or doctoral theses)
6. Unpublished manuscripts and reports (e.g., reports to funders, clinical study reports)
7. Information from regulators (i.e., FDA medical and statistical reviews)
8. Individual patient data

2.3.3. Selection of trials

Two investigators independently screened titles and abstracts to determine which were eligible for inclusion in the review, as described in our study protocol.12 We independently screened full text reports of all potentially relevant trials and resolved disagreements through consensus and by asking a third investigator as needed.

2.3.4. Data collection

We extracted information about trial design as described in our study protocol.12 We recorded all outcomes within the prespecified outcome domains, including the 5 elements described in Table 2.26,27

2.3.4.1. For sources of aggregate data

For trials associated with multiple sources (e.g., journal articles, clinical study reports), we extracted data from each source, including trial design, trial quality (i.e., risk of bias), outcomes, and results. We developed data collection forms and entered data in the Systematic Review Data Repository (SRDR).28,29

From single sources about more than 1 trial, we extracted information about each individual trial; we did not extract results from multiple trials when the data had already been combined (i.e., we were unable to separate data for an individual trial from the combined data).

We extracted results from figures and tables that included numerical data (e.g., counts and means included in graphs), including statistical significance (e.g., p-value above or below 0.05). We did not meta-analyze information that was presented only graphically. We extracted text from sources describing results as “statistically significant” or “not significant.”

Table 1. The 5 Elements of an Outcome

<table>
<thead>
<tr>
<th>Element</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain</td>
<td>Title27 or concept30 to describe 1 or more outcomes</td>
</tr>
<tr>
<td>Specific measure</td>
<td>Instrument used to assess the outcome domain, including (1) the name of the instrument or questionnaire and (2) the total score or the subscales that will be analyzed</td>
</tr>
<tr>
<td>Specific metric</td>
<td>Unit of measurement (e.g., change from baseline, value at a time)</td>
</tr>
<tr>
<td>Method of aggregation</td>
<td>The primary statistical procedure for estimating the treatment effect (e.g., mean, median), including definition of cutoff or pooling for</td>
</tr>
</tbody>
</table>
categorical variables (e.g., proportion with 50% decrease, proportion much or very much improved)

The procedure for estimating the treatment effect, including (1) if the outcome will be treated as a continuous, categorical, or time-to-event variable, and (2) for continuous variables, a measure of central tendency (e.g., mean value); for categorical and time-to-event data variables, proportion with an event and, if relevant, the specific cutoff or categories compared

<table>
<thead>
<tr>
<th>Time-point</th>
<th>The period of follow-up, including (1) when outcome measurements will be obtained (e.g., weeks or months since randomization) and (2) which of the outcome measurements will be analyzed</th>
</tr>
</thead>
</table>

Two investigators extracted data independently and resolved discrepancies by consensus and through discussion with a third investigator if necessary. We used Stata 1431 to reconcile the records and to produce a final data set that we will share publicly through SRDR at the end of the study.

2.3.4.2. For individual patient data

We received individual participant data for gabapentin trials from Pfizer as Microsoft Access database files. When codebooks were not available, we compared the individual participant data with case report forms (which often showed how and when data were recorded) and statistical analysis plans (which often showed how data were coded and analyzed) to identify the variables contained in the individual participant data.

We did not receive any data sets about quetiapine from AstraZeneca. In 1 quetiapine clinical study report, we identified an appendix with individual participant data, which we extracted using ABBYY FineReader version 11.32

2.4. Analysis

2.4.1. Assessment of trial quality

To compare the completeness of sources with respect to trial quality, we assessed each source using the Cochrane Collaboration Risk of Bias Tool3 as described in our study protocol.12 Two investigators independently rated each source for risk of bias (low, unclear, or high risk of bias) related to sequence generation, allocation concealment, masking of participants, masking of outcome assessors, masking of providers, and missing data. Investigators resolved discrepancies by consensus and through discussion with a third investigator if necessary.
2.4.2. Series of meta-analyses

For each case study, we conducted a series of meta-analyses to explore the impact of multiple data sources on the overall results. We analyzed these results by adding or replacing data as follows:

1. Including only results from publications in peer reviewed journals
2. Adding results from trial registries to Step 1
3. Adding results from short reports to Step 2
4. Adding or replacing results (from Step 3) using results obtained from regulators (e.g., FDA reviews)
5. Adding or replacing results (from Step 4) using summary results obtained from the authors or manufacturers (e.g., clinical study reports)
6. Replacing the best available aggregate results for all trials (Step 5) with individual patient data where available

For example, if there was no source for a trial in Steps 1 to 4 (i.e., no journal article, trial registration, conference abstract, or FDA review) but there was a clinical study report, we added the results from the clinical study report in Step 5. If there was a journal article, a FDA review, and a clinical study report, we replaced the results from the journal article with the results from the FDA review in Step 4, and we replaced the results from the FDA review with data from the clinical study report in Step 5.

For each outcome in each source, we determined if results included enough information to include them in meta-analyses. We considered results meta-analyzable if they included a point estimate and measure of precision (e.g., standard error, confidence interval). We meta-analyzed only data for which there were sufficient information in a source to calculate the treatment effect; for example, we did not impute precision if it was not reported.

For each case study, we combined results for the most common measure using the standardized mean difference (SMD) for the difference between groups. In meta-analyses that included both aggregate data and individual participant data, we analyzed individual participant data to obtain an aggregate result for each trial and then we combined the aggregate results across all trials.33-36

We conducted all meta-analyses using random effects, and we report all SMDs with 95% confidence intervals (CIs).37-39 We prespecified that we would use random effects because we anticipated that eligible trials might include differences (e.g., dosage, participant characteristics) that would violate the assumptions required for the fixed effect model. We compared information using the Dersimonian-Laird method, a very common method of synthesis, assuming the random effects model. We have no reason to think that differences in meta-analytic effect estimates using multiple data sources would be limited to whether we used 1 statistical model or another to perform the meta-analysis.
Additional analyses—including other outcome measures, metrics, and methods of aggregation—are included in manuscripts submitted for publication.

3. RESULTS

Manuscripts describing our analyses and findings are currently under review. What follows is a summary of our findings.

3.1. Patient-centered outcomes identified through existing sources and stakeholder coinvestigators

As our protocol describes, we selected outcomes that matter most to patients, using existing sources and in collaboration with patient and stakeholder partners.

3.1.1. Core outcome sets

We identified a core outcome set for pain: the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) consensus recommendations. IMMPACT recommends that the following 6 core domains be assessed in clinical trials of chronic pain: pain, physical functioning, emotional functioning, participant ratings of global improvement, symptoms and adverse events, and participant disposition.

We did not identify a core outcome set for bipolar depression. From the James Lind Alliance, we learned that a working group had been formed to identify priorities for bipolar disorder in conjunction with the Oxford Patients Active in Research project; no relevant outputs had been produced at the time of our study.

3.1.2. Social media

Information on social media could not be used to identify PCOs. PatientsLikeMe reported the percentage of patients who rated each intervention’s subjective effectiveness and side effects. Although patients could list specific outcomes related to effectiveness (e.g., quality of life) or side effects (e.g., dizziness), PatientsLikeMe does not include a function that allows patients to rate those specific side effects. PatientsLikeMe reports information about the 6 most commonly rated side effects for each medication.

3.1.3. DRUGDEX

DRUGDEX entries for gabapentin and quetiapine included information about the use of the drugs for multiple conditions. We identified information about the outcome domains pain intensity and depression, respectively; however, we found little information about other potential benefits of treatment for either on-label or off-
label uses of the drugs. DRUGDEX did not always state whether harms were associated with all conditions for which a drug might be used or with only specific conditions.

3.1.4. Food and Drug Administration prescribing information

Prescribing information included the outcome domains pain intensity for trials of gabapentin and depression for trials of quetiapine. Prescribing information contained little information about other potential benefits. For gabapentin, prescribing information did not include any information about potential benefits for off-label uses (i.e., types of neuropathic pain for which gabapentin is not approved by FDA).

Prescribing information included many potential harms, and it did not always state whether information about harms was applicable to any condition or to only specific conditions.

3.1.5. Patient and stakeholder investigators

Patient and stakeholder partners identified outcomes they thought should be included in each review. We created a questionnaire with a list of potential outcomes, and patients rated how much the outcomes mattered to them. We discussed the outcomes identified through all sources and selected those outcomes that patient and stakeholder partners thought were most patient centered.

3.2. The effects of adding and replacing data from multiple data sources on the results of meta-analyses

3.2.1. Results of the search

We identified 21 eligible trials of gabapentin (from 1997 to 2013) and 7 eligible trials of quetiapine (from 2003 to 2014). As Tables 3 and 4 show, most of these trials were associated with more than 1 source (71% and 100%, respectively). We identified 74 (gabapentin) and 50 (quetiapine) reports and extracted information from them. Journal articles and conference abstracts were the most common reports. We identified 1 gabapentin trial only through a conference abstract, and 1 gabapentin trial only on ClinicalTrials.gov (www.clinicaltrials.gov).

FDA reviews included information about 2 trials of gabapentin for postherpetic neuralgia. FDA reviews available online did not include trials of quetiapine for bipolar depression, which was added to the label in a supplemental new drug application.

We identified few registrations for trials of gabapentin on ClinicalTrials.gov and other registries, none of the trials conducted by the manufacturer (Pfizer). In contrast, all eligible trials of quetiapine were registered.
We received clinical study reports and individual participant data for 6 gabapentin trials as a result of the litigation we describe above\textsuperscript{7,41}; we received no additional information from Pfizer as a result of our requests. By searching online, we identified clinical study reports for 2 quetiapine trials that were released during litigation (http://psychrights.org/); 1 of the clinical study reports included individual participant data in an appendix. AstraZeneca declined our request for more information about quetiapine trials, and our correspondence with AstraZeneca has been published.\textsuperscript{42} Thus, in our study, all nonpublic sources were previously nonpublic, since they were available to us online from litigation sources (i.e., they were disclosed during patient lawsuits), and we received no additional sources from the trials' sponsors.

In summary, we found multiple sources for most trials, the largest contributors to multiple reports being journal articles and conference abstracts.
<table>
<thead>
<tr>
<th>Trial Name</th>
<th>Reports</th>
<th>Journal (1 Trial)</th>
<th>Journal (&gt; 1 Trial)</th>
<th>Public Sources</th>
<th>Other Source</th>
<th>Trial Registry</th>
<th>FDA Review</th>
<th>Nonpublic Sources</th>
<th>All Sources (Total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arai 2010</td>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Backonja 1998</td>
<td></td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>Caraceni 2004</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Hahn 2004</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Hui 2010</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Irving 2009</td>
<td></td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Milenkovic 2009</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Mishra 2012</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>NCT00475904</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Rice 2001</td>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>Rowbotham 1998</td>
<td></td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>Sandercock 2012</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Sang 2013</td>
<td></td>
<td>1</td>
<td>4</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>Serpell 2002</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Simpson 2001</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Study 945-224</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Study A945-1008</td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Tamez Pérez 2000</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Wallace 2010</td>
<td></td>
<td>1</td>
<td>4</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>Yildirim 2003</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Zepeda Vazquez 2001</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Trials (total)</strong></td>
<td></td>
<td><strong>17</strong></td>
<td><strong>7</strong></td>
<td><strong>10</strong></td>
<td><strong>9</strong></td>
<td><strong>5</strong></td>
<td><strong>2</strong></td>
<td><strong>6</strong></td>
<td><strong>21</strong></td>
</tr>
<tr>
<td><strong>Unique sources (total)</strong></td>
<td></td>
<td><strong>20</strong></td>
<td><strong>6</strong></td>
<td><strong>20</strong></td>
<td><strong>15</strong></td>
<td><strong>5</strong></td>
<td><strong>2</strong></td>
<td><strong>6</strong></td>
<td><strong>80</strong></td>
</tr>
</tbody>
</table>

Note: Journal (1 Trial) = journal article about a single included trial; Journal (> 1 Trial) = journal article about 2 or more included trials; Other Source = letters, summaries of journal articles, press releases, brief reports in trade publications (e.g., not peer reviewed); Trial Registry = trial registration on ClinicalTrials.gov or another registry; FDA Review = US Food and Drug Administration medical or statistical review; CSR = clinical study report; IPD = individual patient data. Each cell represents the number of sources of a particularly type (e.g., journal articles) for a given trial.
<table>
<thead>
<tr>
<th>Trial Name Reports</th>
<th>Journal (1 Trial)</th>
<th>Journal (&gt; 1 Trial)</th>
<th>Conference Abstract</th>
<th>Other Source</th>
<th>CSR-S</th>
<th>Trial Registry</th>
<th>CSR</th>
<th>IPD</th>
<th>All Sources (Total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calabrese 2004117-142</td>
<td>4</td>
<td>5</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Gao 2014143,144</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Li 2014145,146</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>McElroy 2010125-129,133,141,147-151</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Suppes 2010152-154</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Thase 2006125-133,138,141,155-160</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>Young 2008125-129,133,141,148,161-166</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td><strong>Trials (total)</strong></td>
<td><strong>6</strong></td>
<td><strong>4</strong></td>
<td><strong>6</strong></td>
<td><strong>4</strong></td>
<td><strong>2</strong></td>
<td><strong>7</strong></td>
<td><strong>2</strong></td>
<td><strong>1</strong></td>
<td><strong>7</strong></td>
</tr>
<tr>
<td><strong>Unique sources (total)</strong></td>
<td><strong>10</strong></td>
<td><strong>5</strong></td>
<td><strong>19</strong></td>
<td><strong>5</strong></td>
<td><strong>2</strong></td>
<td><strong>7</strong></td>
<td><strong>2</strong></td>
<td><strong>1</strong></td>
<td><strong>51</strong></td>
</tr>
</tbody>
</table>

Note: Journal (1 Trial) = journal article about a single included trial; Journal (> 1 Trial) = journal article about 2 or more included trials; Other Source = letters, summaries of journal articles, press releases, brief reports in trade publications (e.g., not peer reviewed); CSR-S = clinical study report synopsis; Trial Registry = trial registration on ClinicalTrials.gov or another registry; CSR = clinical study report; IPD = Individual patient data. Each cell represents the number of sources of a particularly type (e.g., journal articles) for a given trial.
3.2.2. Completeness of information

Trial design (e.g., participant, intervention characteristics) was best described in clinical study reports and journal articles. For example, conference abstracts were incomplete and sometimes contradicted other sources. Five trial registrations submitted to ClinicalTrials.gov by AstraZeneca described a single intervention, “quetiapine fumarate,” while other sources indicated that 2 groups received quetiapine at different doses (i.e., 300mg and 600mg).

Trial quality (defined as our assessment of the risk of bias for each trial) was often unclear in the sources we identified (see Tables 5 and 6). When we located enough information to assess risk of bias (usually in a journal article or clinical study report), that information almost always led from an unclear to a low risk of bias rating. The only exception was information about missing data, where identification of additional information led from an unclear to a high risk of bias rating for trials with large amounts of missing data.

Statistical information about many outcomes described in journal articles and short reports did not include sufficient detail for results to be included in meta-analysis. In contrast, data in clinical study reports were usually complete and could almost always be used for analysis.

For these reasons, clinical study reports—when they were available—were the most complete sources of information about methods and results.
Table 5. Trial Quality (Risk of Bias) for Sources of Gabapentin for Neuropathic Pain Trials

| Trial Name          | Total Reports | Sequence Generation | Allocation Concealment | Masking Participants | Masking Providers | Masking Outcome Assessment | Missing Data for Primary Outcome |
|---------------------|---------------|---------------------|------------------------|----------------------|-------------------|-----------------------------|---------------------------------
| Arai 2010           | 2             | 0 1 1               | 0 2 0                  | 0 2 0                | 0 2 0             | 0 2 0                       | 0 2 0                            |
| Backonja 1998       | 14            | 0 13 1              | 0 13 1                 | 0 12 2               | 0 14 0            | 0 14 0                      | 0 12 2                            |
| Caraceni 2004       | 3             | 0 3 0               | 0 2 1                  | 0 3 0                | 0 3 0             | 0 3 0                       | 0 2 1                            |
| Hahn 2004           | 1             | 0 1 0               | 0 0 1                  | 0 0 1                | 0 1 0             | 0 1 0                       | 0 1 0                            |
| Hui 2010            | 2             | 0 1 1               | 0 2 0                  | 0 1 1                | 0 2 0             | 0 2 0                       | 0 2 0                            |
| Irving 2009         | 2             | 0 1 1               | 0 2 0                  | 0 1 1                | 0 2 0             | 0 2 0                       | 0 1 1                            |
| Milenkovic 2009     | 1             | 0 1 0               | 0 1 0                  | 0 1 0                | 0 1 0             | 0 1 0                       | 0 1 0                            |
| Mishra 2012         | 1             | 0 0 1               | 0 1 0                  | 0 1 0                | 0 1 0             | 0 1 0                       | 0 1 0                            |
| NCT00475904         | 1             | 0 1 0               | 0 1 0                  | 0 1 0                | 0 1 0             | 0 1 0                       | 0 1 0                            |
| Rice 2001           | 13            | 0 11 2              | 0 11 2                 | 0 12 1               | 0 13 0            | 0 13 0                      | 0 7 5                            |
| Rowbotham           | 13            | 0 12 1              | 0 13 0                 | 0 12 1               | 0 13 0            | 0 13 0                      | 0 11 1                            |
| Sandercock 2012     | 3             | 1 2 0               | 0 3 0                  | 0 3 0                | 0 3 0             | 0 3 0                       | 0 1 1                            |
| Sang 2013           | 17            | 0 15 2              | 0 17 0                 | 0 16 1               | 0 17 0            | 0 17 0                      | 0 13 2                            |
| Serpell 2002        | 10            | 0 9 1               | 0 10 0                 | 0 10 0               | 0 10 0            | 0 10 0                      | 0 7 2                            |
| Simpson 2001        | 2             | 0 2 0               | 0 2 0                  | 0 2 0                | 0 2 0             | 0 2 0                       | 0 2 0                            |
| Study 945-224       | 6             | 0 6 0               | 0 5 1                  | 0 5 1                | 0 5 1             | 0 6 0                       | 0 5 1                            |
| Study A945-1008     | 1             | 0 0 1               | 0 1 0                  | 0 1 0                | 0 1 0             | 0 1 0                       | 0 0 1                            |
| Tamez Pérez 2000    | 2             | 0 2 0               | 0 2 0                  | 0 2 0                | 0 2 0             | 0 2 0                       | 0 2 0                            |
| Wallace 2010        | 16            | 0 15 1              | 0 15 1                 | 0 16 0               | 0 16 0            | 0 16 0                      | 0 13 1                            |
| Yildirim 2003       | 1             | 0 1 0               | 0 1 0                  | 0 1 0                | 0 1 0             | 0 1 0                       | 1 0 0                            |
| Zepeda Vazquez 2001 | 1             | 0 1 0               | 0 1 0                  | 0 1 0                | 0 1 0             | 0 1 0                       | 0 1 0                            |

Note: For each included trial, this table shows the number of sources (excluding individual participant data) at high (H), unclear (U), and low (L) risk of bias for each item. NA = not applicable because the source did not report outcomes.
Table 6. Trial Quality (Risk of Bias) by Source of Quetiapine for Bipolar Depression Trials

<table>
<thead>
<tr>
<th>Trial Name</th>
<th>Reports</th>
<th>Total Sources</th>
<th>Sequence Generation</th>
<th>Allocation Concealment</th>
<th>Masking Participants</th>
<th>Masking Providers</th>
<th>Masking outcome Assessment</th>
<th>Missing Data for Primary Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calabrese 2004</td>
<td>137-142</td>
<td>26</td>
<td>H 22 4 H U L 0 25 1</td>
<td>H U L 0 25 1</td>
<td>H U L 0 25 1</td>
<td>H U L NA 0 22 1 3</td>
<td>H U L NA 5 14 0 7</td>
<td></td>
</tr>
<tr>
<td>Gao 2014</td>
<td>143,144</td>
<td>2</td>
<td>H 2 0 H U L 0 1 1</td>
<td>H U L 0 1 1</td>
<td>H U L 0 1 1</td>
<td>H U L NA 0 2 0 0 0 1 1 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Li 2014</td>
<td>145,146</td>
<td>2</td>
<td>H 2 0 H U L 0 1 1</td>
<td>H U L 0 1 1</td>
<td>H U L 0 1 1</td>
<td>H U L NA 0 2 0 0 0 1 1 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>McElroy 2010</td>
<td>125-129,133,141,147-151</td>
<td>12</td>
<td>H 0 10 2 H U L 0 11 1</td>
<td>H U L 0 12 0</td>
<td>H U L NA 0 12 0 0 2 7 0 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suppes 2010</td>
<td>152-154</td>
<td>3</td>
<td>H 0 3 0 H U L 0 3 0</td>
<td>H U L 0 3 0</td>
<td>H U L NA 0 3 0 0 1 2 0 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thase 2006</td>
<td>125-133,138,141,155-160</td>
<td>17</td>
<td>H 0 14 3 H U L 0 17 0</td>
<td>H U L 0 17 0</td>
<td>H U L NA 0 17 0 4 10 0 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young 2008</td>
<td>125-129,133,141,148,161-166</td>
<td>14</td>
<td>H 0 12 2 H U L 0 13 1</td>
<td>H U L 0 13 1</td>
<td>H U L NA 0 14 0 0 2 9 0 3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: For each included trial, this table shows the number of sources (excluding individual participant data) at high (H), unclear (U), and low (L) risk of bias for each item. NA = not applicable because the source did not report outcomes.
3.2.3. Locating patient-centered outcomes in multiple data sources

Across all sources for all gabapentin trials, we located 4 of 5 prespecified outcome domains as described in Section “3.1. Patient-centered outcomes identified through existing sources and stakeholder coinvestigators” (Tables 7 and 8). Across all sources for all quetiapine trials, we identified 7 of 8 prespecified domains. Quetiapine sources included about 6 (median) outcome domains, and they were also highly variable (range 2 to 6 domains per source). To our knowledge, none of the trials measured the prespecified outcome domains for pain interference (gabapentin) or psychiatric hospitalization (quetiapine).

Conference abstracts and other short reports included little useful information about PCOs. Journal articles contained more PCOs than did other public sources, although clinical study reports were consistently the best source of aggregate information about PCOs. No outcome measures appeared in public sources that did not also appear in corresponding clinical study reports; however, we identified methods of aggregation in public sources (i.e., journal articles and an FDA review) for gabapentin trials that were not included in clinical study reports.

In addition to the prespecified outcomes related to treatment benefits, clinical study reports and individual participant data described hundreds of harms that were not included in public sources.167
Table 7. Number of Prespecified Outcome Domains in Each Source of Gabapentin for Neuropathic Pain Trials

<table>
<thead>
<tr>
<th>Trial Name Reports</th>
<th>Journal (1 Trial)</th>
<th>Journal (&gt; 1 Trial)</th>
<th>Conference Abstract</th>
<th>Other Source</th>
<th>Trial Registry</th>
<th>FDA Review</th>
<th>CSR</th>
<th>IPD</th>
<th>All Sources (Total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arai 201043,44</td>
<td>0</td>
<td>NA</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>Backonja 199845-58</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>NA</td>
<td>NA</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Caraceni 200459-61</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>0</td>
</tr>
<tr>
<td>Hahn 200452</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2</td>
</tr>
<tr>
<td>Hui 201063,64</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>Irving 200965,66</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2</td>
</tr>
<tr>
<td>Milenkovic 200957</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>Mishra 201258</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>NCT0047590469</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>Rice 200149-53,70,77</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>NA</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Rowbotham 199849-53,71,73,76-81</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>NA</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Sandercock 201282-84</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
<td>2</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2</td>
</tr>
<tr>
<td>Sang 201385-101</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2</td>
</tr>
<tr>
<td>Serpell 200249-53,71,102-105</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Simpson 2001106,107</td>
<td>4</td>
<td>NA</td>
<td>3</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>4</td>
</tr>
<tr>
<td>Study 945-22449-53,108</td>
<td>NA</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Study A945-1006109</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Tamez Pérez 2000110,111</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>Wallace 201088-100,112-114</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2</td>
</tr>
<tr>
<td>Yildirim 20033315</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>Zepeda Vazquez 20013316</td>
<td>1</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4</strong></td>
<td><strong>4</strong></td>
<td><strong>4</strong></td>
<td><strong>4</strong></td>
<td><strong>2</strong></td>
<td><strong>4</strong></td>
<td><strong>4</strong></td>
<td><strong>4</strong></td>
<td><strong>4</strong></td>
</tr>
</tbody>
</table>

Note: **Journal (1 Trial)** = journal article about a single included trial; **Journal (> 1 Trial)** = journal article about 2 or more included trials; **Other Source** = letters, summaries of journal articles, press releases, brief reports in trade publications (e.g., not peer reviewed); **Trial Registry** = trial registration on ClinicalTrials.gov or another registry; **FDA Review** = US Food and Drug Administration medical or statistical review; **CSR** = clinical study report; **IPD** = individual patient data. For each trial, we recorded the number of prespecified domains (pain intensity, pain interference, sleep disturbance, mood, and quality of life) that could be extracted from each type of source and from all sources about that trial. None of the reports described pain interference.
Table 8. Number of Prespecified Outcome Domains in Each Source of Quetiapine for Bipolar Depression Pain Trials

<table>
<thead>
<tr>
<th>Trial Name</th>
<th>Reports</th>
<th>Journal (1 Trial)</th>
<th>Journal (&gt;1 Trial)</th>
<th>Public Sources</th>
<th>Nonpublic Sources</th>
<th>All Sources (Total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calabrese 2004</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>NA</td>
<td>6</td>
</tr>
<tr>
<td>Li 2014</td>
<td>NA</td>
<td>NA</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>McElroy 2010</td>
<td>6</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Suppes 2010</td>
<td>3</td>
<td>NA</td>
<td>1</td>
<td>NA</td>
<td>1</td>
<td>NA</td>
</tr>
<tr>
<td>Thase 2006</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td>Young 2008</td>
<td>6</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>7</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

Note: Journal (1 Trial) = journal article about a single included trial; Journal (>1 Trial) = journal article about 2 or more included trials; Other Source = letters, summaries of journal articles, press releases, brief reports in trade publications (e.g., not peer reviewed); CSR-S = clinical study report synopsis; Trial Registry = trial registration on ClinicalTrials.gov or another registry; CSR = clinical study report; IPD = individual patient data. For each trial, we recorded the number of prespecified domains (depression, functioning, quality of life, anxiety, sleep, weight, psychiatric hospitalization, and suicide) that could be extracted from each type of source and from all sources about that trial. None of the reports described psychiatric hospitalization.
3.2.4. Multiple outcome definitions

As noted earlier, it is also possible for a single prespecified outcome domain to be assessed in multiple ways (i.e., by varying the measure, metric, and method of aggregation). For this reason, an outcome might appear to be reported as described in the protocol when in fact there were multiple opportunities for changing the associated results (Figure 1).

Figure 1. The Number of Outcomes in a Trial Is a Function of the Number of Definitions of Each of the 5 Elements

Note: In this hypothetical example, the number of treatment effects for a single trial is the product of definitions for each element. In a trial with 4 outcome domains, introducing 2 definitions for each of the other elements will result in 64 different outcomes.

For each case study, we identified hundreds of outcomes for the prespecified domains (Figure 2). By varying 4 elements, gabapentin RCTs reported 214 outcomes and quetiapine RCTs reported 81 outcomes. For pain intensity and depression, the most common outcome domains for gabapentin and quetiapine RCTs, we found 119 and 44 outcomes, respectively. Multiple methods of analysis also contributed to multiple results for the same defined outcome. For these reasons, many meta-analyses were possible for each outcome domain using the same included trials.167
Figure 2. The Number of Outcomes Observed Increased as a Consequence of Multiple Definitions of the Elements

Note: The points nearest the center of the circle represent the prespecified outcome domains for which we found any information in eligible trials (i.e., 4 outcome domains for gabapentin and 7 outcome domains for quetiapine); these are represented by different colors. The branches represent variations in outcome definitions associated with those domains (i.e., different measures, metrics, methods of aggregation, or time-points). Each point around the outer edge of the circle thus represents a unique outcome described in 1 or more of the trial sources we identified. (A) **Specific measure (name of instrument)**: BPI = Brief Pain Inventory; CGIC = Clinician Global Impression of Change; ECOG = Eastern Cooperative Oncology Group; GIC = Global Impression of Change; GSS = Global Symptom Score; HADS = Hospital Anxiety and Depression Scale; HRQoL = Health Related Quality of Life Questionnaire; NPS = Neuropathic Pain Scale; Pain (11 pt) = pain score (11-point scale); Pain (4 pt) = pain score (4-point scale); PGIC = Patient Global Impression of Change; POMS = Profile of Mood States; SF-36 = Short Form 36 item; SFMPQ = Short Form McGill Pain Questionnaire; Sleep score =
sleep score (11-point scale); TOPS = Treatment Outcomes in Pain Survey. **Specific measure (total or subscale):** % relief = percentage pain relief; Ave. pain = average pain score; Care sat = health care satisfaction; Current pain = current pain score; Fear/Avoid = fear avoidance; Gen health = general health; Globe intensity = global intensity; Globe unpleas = global unpleasantness; Lower body = lower body functional limitations; Ment health = mental health; Obj. social = objective family/social disability; Obj. work = objective work disability; Pain sympt = pain symptom; Passive cope = passive coping; Patient sat = patient satisfaction with outcomes; Phys funct = physical functioning; PPI = present pain intensity; Role emot = role emotional; Role phys = role physical; Social disab = perceived family/social disability; Social funct = social functioning; Solicitous = solicitous responses; Total pain = total pain experience; Total score = total; Upper body = upper body functional limitations; VAS = visual analog scale; Work limit = work limitations. **Metric:** Change = change from baseline; Value = value at a time-point. **Method of aggregation:** % change mean = percentage change mean; % change median = percentage change median; 2 decrease = at least 2-point decrease; 3 decrease = at least 3-point decrease; 4 decrease = at least 4-point decrease; M or VM improved = much or very much improved; M or VM worse = much worse or very much worse; Min or No change = minimally improved or no change; Min improved = minimally improved; Min worse = minimally worse; Mod improved = moderately improved; Mod or much imp = moderately or much improved; Mod worse = moderately worse; No change = no change (0%); Pain free = totally pain free; Response (ND) = response (not defined); V much improved = very much improved; V much worse = very much worse. (B) **Specific measure (name of instrument):** BMI = Body Mass Index; CGI-I = Clinical Global Impression - Improvement; CGI-S = Clinical Global Impression - Severity; HAM-A = Hamilton Anxiety Rating Scale; HAM-D = Hamilton Depression Rating Scale; MADRS = Montgomery Asberg Depression Rating Scale; MADRS/YMRS = Montgomery Asberg Depression Rating Scale & Young Mania Rating Scale; MOS Cog = Medical Outcomes Study Cognitive Scale; PSQI = Pittsburgh Sleep Quality Assessment; Q-LES-Q = Quality of Life Enjoyment and Satisfaction Questionnaire; QIDS-SR-16 = Quick Inventory of Depressive Symptomology Self-Report 16; SDS = Sheehan Disability Scale; Ideation = suicidal ideation (self-reported). **Metric:** Change = change from baseline; Value = value at a time-point. **Method of aggregation:** “= 10” = less than or equal to 10; “= 12” = less than or equal to 12; “= 7” = less than or equal to 7; “= 8” = less than or equal to 8; 15% increase = at least 15% increase; 25% increase = at least 25% increase; 30% decrease = at least 30% decrease; 40% decrease = at least 40% decrease; 50% decrease = at least 50% decrease; 60% decrease = at least 60% decrease; 7% increase = at least 7% increase; 70% decrease = at least 70% decrease; Extremely ill = among the most extremely ill; M or VM improved = much or very much Improved; MADRS = 12, YMRS = 12 = MADRS less than or equal to 12 and YMRS less than or equal to 12; MADRS = 12, YMRS = 8 = MADRS less than or equal to 12 and YMRS less than or equal to 8; Mean % Max = mean percentage of the maximum; Min improved = minimally improved; Min worse = minimally worse; Normal or Borderline = normal, not at all ill or borderline ill; Normal, not ill = normal, not at all ill; Proportion = proportion of participants who experienced the event; Remission (ND) = remission (not defined); Response (ND) = response (not defined); V Much improved = very much improved; V Much worse = very much worse.
3.2.5. Impact of adding or replacing multiple data sources on pooled effect sizes and precision

For each case study, we conducted a series of meta-analyses in which we added or replaced data from each of the multiple sources for a single outcome domain, measure, and time-point to evaluate the effect of adding or replacing data. In the presence of reporting bias, whereby negative or null results are reported in nonpublic sources but not in public sources, we would expect the addition of nonpublic sources to change the average effects estimated in meta-analyses. This change would lead to a meta-analytic effect estimate closer to the null.

Not all trials could contribute data to a meta-analysis (Tables 9 and 10). Of the 21 gabapentin and 7 quetiapine trials, 9 (43%) and 4 (57%), respectively, could not be included in an analysis of pain intensity or depression using continuous data (e.g., the mean) because no source for those trials included meta-analyzable data (see “2.4.2. Series of meta-analyses”).

When restricted to the most common outcome defined as the same outcome domain, measure, and time-point, results in both case studies were largely consistent across sources. For gabapentin, including data from previously nonpublic sources (i.e., clinical study reports and individual participant data) reduced the magnitude of the effect estimate; adding additional sources did not affect the statistical significance of the treatment effect. In our series of meta-analyses, the effect of gabapentin on pain intensity ranged from SMD = −0.46 (95% CI −0.64, −0.28) to SMD = −0.31 (95% CI −0.47, −0.15). For quetiapine, there were no meaningful differences as a result of adding and replacing data from multiple sources. The effect of quetiapine on depression ranged from SMD = −0.42 (95% CI −0.62, −0.22) to SMD = −0.44 (95% CI −0.63, −0.26).
Table 9. Series of Meta-analyses and the Effect of Adding or Replacing Data Sources on the Pooled Estimate of Pain Intensity in Trials of Gabapentin for Neuropathic Pain (SMD and 95% CI)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Backonja 1998</td>
<td>45-58</td>
<td>−0.57 (−0.88, −0.25)</td>
<td>−0.57 (−0.88, −0.25)</td>
<td>−0.57 (−0.88, −0.25)</td>
<td>−0.56 (−0.87, −0.24)</td>
<td>−0.53 (−0.84, −0.22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irving 2009</td>
<td>65,66</td>
<td>−0.37 (−0.71, −0.04)</td>
<td>−0.37 (−0.71, −0.04)</td>
<td>−0.37 (−0.71, −0.04)</td>
<td>−0.37 (−0.71, −0.04)</td>
<td>−0.37 (−0.71, −0.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mishra 2012</td>
<td>48</td>
<td>−0.44 (−0.95, 0.06)</td>
<td>−0.44 (−0.95, 0.06)</td>
<td>−0.44 (−0.95, 0.06)</td>
<td>−0.44 (−0.95, 0.06)</td>
<td>−0.44 (−0.95, 0.06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCT00475904</td>
<td>69</td>
<td>No sources</td>
<td>0.32 (0.04, 0.60)</td>
<td>0.32 (0.04, 0.60)</td>
<td>0.32 (0.04, 0.60)</td>
<td>0.32 (0.04, 0.60)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rice 2001</td>
<td>53,70,77</td>
<td>−0.64 (−0.92, −0.36)</td>
<td>−0.64 (−0.92, −0.36)</td>
<td>−0.64 (−0.92, −0.36)</td>
<td>−0.63 (−0.86, −0.36)</td>
<td>−0.56 (−0.80, −0.33)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rowbotham 1998</td>
<td>55,71,76</td>
<td>−0.86 (−1.13, 0.59)</td>
<td>−0.86 (−1.13, 0.59)</td>
<td>−0.86 (−1.13, 0.59)</td>
<td>−0.86 (−1.13, 0.59)</td>
<td>−0.83 (−1.11, 0.56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sandercock 2012</td>
<td>82,84</td>
<td>−0.49 (−0.84, −0.15)</td>
<td>−0.49 (−0.84, −0.15)</td>
<td>−0.49 (−0.84, −0.15)</td>
<td>−0.49 (−0.84, −0.15)</td>
<td>−0.49 (−0.84, −0.15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sang 2013</td>
<td>85-101</td>
<td>−0.22 (−0.40, 0.03)</td>
<td>−0.22 (−0.40, 0.03)</td>
<td>−0.22 (−0.40, 0.03)</td>
<td>−0.22 (−0.40, 0.03)</td>
<td>−0.22 (−0.40, 0.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serpell 2002</td>
<td>59,71,102,105</td>
<td>No sources</td>
<td>No sources</td>
<td>No sources</td>
<td>No sources</td>
<td>−0.22 (−0.45, 0.01)</td>
<td>−0.20 (−0.43, 0.03)</td>
<td></td>
</tr>
<tr>
<td>Study 945-224</td>
<td>53,108</td>
<td>No sources</td>
<td>No sources</td>
<td>No sources</td>
<td>No sources</td>
<td>−0.09 (−0.35, 0.17)</td>
<td>−0.06 (−0.32, 0.19)</td>
<td></td>
</tr>
<tr>
<td>Study A945-1008</td>
<td>109</td>
<td>No sources</td>
<td>No sources</td>
<td>No sources</td>
<td>No sources</td>
<td>−0.43 (−0.66, 0.20)</td>
<td>−0.26 (−0.46, 0.06)</td>
<td></td>
</tr>
<tr>
<td>Wallace 2010</td>
<td>108,112,114</td>
<td>−0.20 (−0.40, 0.01)</td>
<td>−0.20 (−0.40, 0.01)</td>
<td>−0.20 (−0.4, 0.01)</td>
<td>−0.20 (−0.40, 0.01)</td>
<td>−0.20 (−0.40, 0.01)</td>
<td>−0.20 (−0.40, 0.01)</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Number (No.) of trials</strong></td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>12</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>No. of participants</strong></td>
<td>1805</td>
<td>2019</td>
<td>2019</td>
<td>2019</td>
<td>3074</td>
<td>3144</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Meta-analytic effect size (SMD)</strong></td>
<td>−0.46 (−0.64, −0.02)</td>
<td>−0.38 (−0.60, 0.15)</td>
<td>−0.38 (−0.60, 0.15)</td>
<td>−0.34 (−0.51, −0.02)</td>
<td>−0.31 (−0.47, −0.02)</td>
<td>−0.31 (−0.47, −0.02)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>r²</strong></td>
<td>68.75</td>
<td>82.56</td>
<td>82.56</td>
<td>82.56</td>
<td>79.00</td>
<td>76.94</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Note: Data for individual trials are provided in each row. In the columns, we first estimated the effect sizes using only journal articles; we then estimated the effect by including other sources to the analysis (i.e., adding data from new sources or replacing the data from sources appearing in columns the left). When we used IPD, we first obtained aggregated data and then combined data from all sources, as described in the Methods. For each included study, we calculated the standardized mean difference (SMD) and 95% CI for the intervention compared with placebo. We combined the results (“Total”) by conducting a meta-analysis using random effects. Effect sizes are given for all trials that reported pain intensity (measured using the average daily pain score on an 11-point scale) in at least 1 source.
Table 10. Series of Meta-analyses and the Effect of Adding or Replacing Data Sources on the Pooled Estimate of Depression in Trials of Quetiapine Bipolar Depression (SMD and 95% CI)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>-0.63 (–0.84, –0.41)</td>
<td>-0.63 (–0.84, –0.41)</td>
<td>-0.63 (–0.84, –0.41)</td>
<td>-0.63 (–0.84, –0.41)</td>
<td>-0.59 (–0.78, –0.41)</td>
<td>-0.59 (–0.78, –0.41)</td>
</tr>
<tr>
<td>Calabrese 2004127-142</td>
<td></td>
<td>-0.37 (–0.56, –0.17)</td>
<td>-0.37 (–0.56, –0.17)</td>
<td>-0.37 (–0.56, –0.17)</td>
<td>-0.37 (–0.56, –0.17)</td>
<td>-0.46 (–0.66, –0.27)</td>
<td>-0.46 (–0.66, –0.27)</td>
</tr>
<tr>
<td>Thase 2006125-133,138,141,155-160</td>
<td></td>
<td>-0.27 (–0.46, –0.08)</td>
<td>-0.27 (–0.46, –0.08)</td>
<td>-0.27 (–0.46, –0.08)</td>
<td>-0.27 (–0.46, –0.08)</td>
<td>-0.27 (–0.46, –0.08)</td>
<td>-0.27 (–0.46, –0.08)</td>
</tr>
<tr>
<td>Young 2008127-129,133,141,148,161-166</td>
<td></td>
<td>-0.42 (–0.62, –0.22)</td>
<td>-0.42 (–0.62, –0.22)</td>
<td>-0.42 (–0.62, –0.22)</td>
<td>-0.42 (–0.62, –0.22)</td>
<td>-0.44 (–0.63, –0.26)</td>
<td>-0.44 (–0.62, –0.26)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Number (No.) of trials</td>
<td></td>
<td>1453</td>
<td>1453</td>
<td>1453</td>
<td>1453</td>
<td>1625</td>
<td>1629</td>
</tr>
<tr>
<td>No. of participants</td>
<td></td>
<td>1453</td>
<td>1453</td>
<td>1453</td>
<td>1453</td>
<td>1625</td>
<td>1629</td>
</tr>
<tr>
<td>Meta-analytic effect size (SMD)</td>
<td></td>
<td>-0.42 (–0.62, –0.22)</td>
<td>-0.42 (–0.62, –0.22)</td>
<td>-0.42 (–0.62, –0.22)</td>
<td>-0.42 (–0.62, –0.22)</td>
<td>-0.44 (–0.63, –0.26)</td>
<td>-0.44 (–0.62, –0.26)</td>
</tr>
<tr>
<td>$I^2$</td>
<td></td>
<td>66.45</td>
<td>66.45</td>
<td>66.45</td>
<td>66.45</td>
<td>63.70</td>
<td>63.15</td>
</tr>
</tbody>
</table>

Note: Data for individual trials are provided in each row. In the columns, we first estimated the effect sizes using only journal articles; we then estimated the effect by including other sources to the analysis (i.e., adding data from new sources or replacing the data from sources appearing in columns the left). When we used IPD, we first obtained aggregated data and then combined data from all sources, as described in the Methods. For each included trial, we calculated the standardized mean difference (SMD) and 95% CI for the intervention compared with placebo. We combined the results (“Total”) by conducting a meta-analysis using random effects. Effect sizes are given for all trials that reported depression measured using the Montgomery Åsberg Depression Rating in at least 1 source.
3.2.6. Resources required to use multiple data sources in systematic reviews and meta-analyses

Available guidance for performing systematic reviews (e.g., from the Cochrane Collaboration and the National Academy of Medicine [formerly the Institute of Medicine]) suggests that reviewers should use all data sources for clinical trials.\textsuperscript{3,4} Identifying and extracting data from multiple data sources was resource intensive and required many skills, and our findings suggest that existing guidance could be improved to help systematic reviewers use multiple data sources efficiently (Table 11).
<table>
<thead>
<tr>
<th>Findings</th>
<th>Recommendation</th>
<th>Audience</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FINDING 1: Accessing, Obtaining, and Analyzing Multiple Sources Is Resource Intensive and Tedious, and Requires Many Skills.</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Most trials lead to multiple sources. | • Users of research should be able to identify clinical trials.  
• PCOs can be found in multiple sources about trials. | • Clinicians, patients  
• Guideline developers, patients, systematic reviewers |
| Multiple sources are needed to understand what was planned and what actually occurred in clinical trials. | • To prevent bias and reduce waste, trials should use common PCOs (core outcome sets).  
• All information about a clinical trial should be stored in a single public source.  
• Trialists and manufacturers should make information consistent across sources. | • Funders, trialists  
• Funders, trialists  
• Trialists |
| Some sources (e.g., trial protocols) are rarely available. | • Protocols submitted to IRBs should be publicly available.  
• IRBs should monitor compliance with registration and reporting requirements. | • Funders, trialists  
• Funders, trialists |
| Some trials are not reported in public sources (e.g., journal articles), and nonpublic sources are available for many trials that are reported. | • All trials should be registered in advance.  
• Results from all trials should be shared with sufficient information to understand the variables and to reproduce the analyses.  
• To reduce waste, trials should use common data elements.  
• Individual patient data are useful for some applications but not all. | • Trialists  
• Funders, trialists  
• Funders, trialists  
• Systematic reviewers |
| **FINDING 2: Information Varies Tremendously Across Sources.** | | |
| Sources differ in completeness. | • Public sources should be available for all trials with sufficient information to understand their design, quality, and results.  
• Outcomes data should be reported with sufficient information to include results in meta-analyses. | • Funders, guideline developers, trialists  
• Trialists |
| Sources include different information. | • All sources should include complete descriptions of all outcomes examined, including all elements of each outcome.  
• Users should expect variation across sources and prioritize variables/sources in advance. | • Trialists  
• Guideline developers, systematic reviewers |
| Trials and sources included results that could not be compared as well as incomplete information. | • To be usable, databases should employ common fields.  
• Shared data should be accompanied by codebooks and other metadata that allow examination of the data. | • Trialists |
| Nonpublic data sources included additional information. | • Trialists should anticipate requests for data and plan to share from the time a trial is initiated.  
• Systems are needed to archive and share new research, and to share “legacy” trials related to drugs in current use. | • Funders, trialists  
• Guideline developers, funders |
In general, we were able to extract information from conference abstracts quickly; journal articles took several hours to extract.

Clinical study reports, which included thousands of pages of appendices and hundreds of outcomes and analyses, required dozens of hours. Two extractors each spent at least 5 8-hour days extracting data from each clinical study report.

Because we did not receive codebooks for individual participant data, each data set was a unique puzzle we had to solve even before we could synthesize the data. Members of our team spent months developing codebooks and creating a common data format to combine data across trials. Even with this investment, there were some variables we could not identify with certainty.

3.3. Guidance for using multiple data sources in systematic reviews

As a final step in our investigation, we developed guidance for systematic reviewers (Table 12). For the open science movement to succeed, systematic reviewers need to make efficient use of evidence from clinical trials, including previously nonpublic information. Although it would have been desirable to identify a best source of information for clinical trials, we did not find a single source or combination of sources that provided complete information about trial design, trial quality, outcomes, and results. Our recommendations incorporate the main findings for each step in the review process.
### Table 12: Recommendations for Systematic Reviews Using Multiple Data Sources

<table>
<thead>
<tr>
<th>Task</th>
<th>Guidance for Systematic Reviewers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Selecting trials</strong></td>
<td>• Determine in advance if individual participant data, aggregate data (e.g., clinical study reports), or both are required to answer the research questions.</td>
</tr>
<tr>
<td></td>
<td>• Search electronic databases for all publicly available sources.</td>
</tr>
<tr>
<td></td>
<td>• Use registries (e.g., ClinicalTrials.gov), conference abstracts, and regulatory data (e.g., FDA, EMA) to identify unpublished trials.</td>
</tr>
<tr>
<td></td>
<td>• Search repositories (e.g., <a href="http://www.clinicalstudydatarequest.com">www.clinicalstudydatarequest.com</a>) for clinical study reports and individual participant data.</td>
</tr>
<tr>
<td></td>
<td>• For unpublished data, allow time to establish data use agreements as required for access to trial data; make alternative plans in case requests are ignored or refused.</td>
</tr>
<tr>
<td></td>
<td>• Contact trial authors and sponsors to confirm which sources are related to each trial.</td>
</tr>
<tr>
<td><strong>Assessing trial design and trial quality</strong></td>
<td>• Use clinical study reports to extract information about trial design and trial quality.</td>
</tr>
<tr>
<td></td>
<td>• In addition to clinical study reports, use journal articles and FDA reviews because they sometimes include information not contained in clinical study reports.</td>
</tr>
<tr>
<td></td>
<td>• Check trial registrations for information about prespecified methods and outcomes.</td>
</tr>
<tr>
<td></td>
<td>• Consider excluding conference abstracts when other sources are available for data extraction.</td>
</tr>
<tr>
<td><strong>Extracting results</strong></td>
<td>• Anticipate multiple definitions of outcomes; prespecify outcomes for extraction within each outcome domain.</td>
</tr>
<tr>
<td></td>
<td>• When clinical study reports are available, check journal articles and FDA reviews for additional analyses.</td>
</tr>
<tr>
<td></td>
<td>• Where clinical study reports are not available, use journal articles to assess effects on primary outcomes.</td>
</tr>
<tr>
<td><strong>Analyzing results</strong></td>
<td>• Consider excluding trials reported only in conference abstracts from meta-analyses.</td>
</tr>
<tr>
<td></td>
<td>• Be cautious in interpreting trials with only public data sources (e.g., conduct sensitivity analyses limited to trials with clinical study reports and individual participant data).</td>
</tr>
<tr>
<td></td>
<td>• Use individual participant data for investigating research questions not addressed in clinical study reports (e.g., subgroup analyses, potential mediators and moderators) and for older trials for which new statistical methods to account for missing data (e.g., multiple imputation) might affect the results.</td>
</tr>
<tr>
<td></td>
<td>• Allow time (12-18 months) to recode individual participant data for analysis and to merge databases.</td>
</tr>
</tbody>
</table>

Note: **EMA** = European Medicines Agency, **FDA** = US Food and Drug Administration.
4. DISCUSSION

4.1. Identifying patient-centered outcomes

Our efforts to identify PCOs revealed little formal research about what patients with neuropathic pain and bipolar disorder consider important outcomes. Better evidence is needed to determine what should be measured by clinical trials in these areas and to determine how those outcomes should be assessed (e.g., choice of specific measures and time-points).

4.2. The effects of adding and replacing data from multiple data sources on the results of meta-analyses

In both case studies, most trials were associated with multiple sources of information about their methods and results (mostly reports about the trials and some individual participant data). The most common sources, short reports (e.g., conference abstracts), were often the least informative about trial design, trial quality, outcomes, and results. Journal articles contained information that was consistent with nonpublic sources, but they did not include all of the information available from nonpublic sources. FDA reviews contained less information than hoped, largely because the drugs and indications we examined did not meet FDA’s criteria for inclusion in its public information. Trial registries were also less useful than we had hoped because trials supporting current uses of the drugs in our case studies were performed before legislation mandated trial registration and results reporting. The nonpublic sources we obtained were made public through litigation, an approach that has been used in previous studies to compare data sources.168

Because many sources were incomplete and inconsistent regarding trial design (e.g., number of groups), sometimes we found it difficult to understand whether sources were about the same trial or different trials.

Most sources for a single trial, viewed on their own, were unclear about whether there was risk of bias in a trial’s design (also referred to as trial quality). An advantage of identifying multiple data sources for each trial was that often information was provided in 1 source but not another; thus, when all sources were taken together, we were able discern more about the trial quality.

The effects of multiple data sources on meta-analysis for gabapentin were consistent with evidence that using FDA reviews and nonpublic sources may change the results of meta-analyses because additional trials can be included using FDA reviews and nonpublic sources and because results in these sources sometimes differ from results in journal articles and conference abstracts.169-172 Findings for quetiapine were consistent with evidence from other case studies that indicate the
results reported in multiple data sources sometimes agree.\textsuperscript{173-175} We are not aware of any other cases in which researchers have had access to this range and number of sources.

Although they sometimes identified trials not reported in other sources (e.g., trials not published in a journal article), short reports such as conference abstracts did not contribute new information about methods and results when other data sources were available.

Most of the outcomes in each trial we examined were associated with a few domains (i.e., they measured a few underlying concepts, such as pain intensity and depression), and these domains were assessed using a variety of measures, metrics, methods of aggregation, and time-points both within individual trials and across all eligible trials.\textsuperscript{167} For this reason, we believe it is possible for investigators to prespecify an outcome domain, assess it in multiple ways, and choose the most favorable result for public presentation.

In the 2 case studies we examined, inferences about effectiveness derived from meta-analyses were only slightly different when we added or replaced data following the process described above. However, some results associated with outcome definitions in each outcome domain were statistically significant and some were not; such differences might lead trial authors to report outcomes selectively or lead systematic reviewers to include outcomes selectively in their meta-analyses. Reporting biases related to differences in outcome measures, metrics, and methods of aggregation could lead decision makers to recommend or select 1 treatment over another.

Potential harms were viewed by our patient coinvestigators as among the most important outcomes. There was limited information about harms in publicly available sources. Further investigation of harmful outcomes as reported in public and nonpublic sources is warranted; findings could have important implications for PCOs research.

4.3 Implications for data sharing

Discussions about open access have largely focused on developing systems for archiving and sharing data from new trials. The drugs in our study are widely prescribed even though the trials were completed before trial registration and reporting requirements came into effect.\textsuperscript{176} Open access to trial data would make what are now nonpublic sources (e.g., clinical study reports, individual participant data) more widely available. Our findings highlight the importance of open access to legacy trial data (e.g., trials conducted before requirements to share data) because many older drugs enjoy current widespread use. Before data similar to those we examined are lost forever, the research community should develop a system for sharing legacy trials with future generations.
4.4 Limitations

We were not funded to analyze potential harms, so we do not know whether the relative benefits and harms of these interventions differ between public and nonpublic data sources. Further research is needed to explore the benefits of using clinical study reports and individual patient data meta-analysis for identifying the frequency of specific harms and of groups of harms.

While many studies have shown that selective reporting is a problem, until now only a few studies have explored the potential for systematic reviewers to select outcomes for meta-analysis,\(^{177-179}\) and none have used both public and nonpublic sources. Although our results were consistent with previous studies comparing various sources with trial registrations,\(^{180,181}\) conference abstracts,\(^{6,182}\) FDA reviews,\(^{174,183}\) journal articles,\(^{184}\) protocols,\(^{185-187}\) clinical study reports,\(^{8,188}\) and individual participant data,\(^{9,189,190}\) we do not know the extent of applicability of our results to other CER questions. These case studies included trials of relatively short duration, and there may be even greater benefits to using nonpublic sources for trials with multiple assessments over long periods of time.

5. CONCLUSIONS

For systematic reviewers and meta-analysts, using multiple data sources for the same trial is time consuming and comes with both rewards (e.g., more complete data) and challenges (e.g., the need to resolve differences among contradictory sources). We found no advantage in using conference abstracts except for identifying the existence of unregistered and otherwise unpublished trials. Trial registrations in these case studies were relatively old (e.g., we only found information in ClinicalTrials.gov, not other registries), and they provided less information than users might find in more recent registrations. Systematic reviewers should anticipate the vast amount of information that could be extracted from nonpublic sources and plan accordingly. Further research is needed to determine if the results of these case studies are representative of all trials and to determine how multiple data sources might affect meta-analyses of potential harms.

6. REFERENCES


51. Backonja MM, Mutisya E. Gabapentin demonstrates a nonlinear dose-response across five multicenter trials for neuropathic pain. Poster presented at: EFNS Annual Congress; October 26-29, 2002; Vienna, Austria.


71. Slawson D. How effective and safe is gabapentin (Neurontin) for reducing the pain of postherpetic neuralgia? Evidence-Based Practice. 2002;7(7):335.


150. AstraZeneca. An international, multi-centre, double-blind, randomised, parallel-group, placebo-controlled, phase III study of the efficacy and safety of quetiapine fumarate (Seroquel [Registered], single oral 300 mg or 600 mg dose) and paroxetine as monotherapy in adult patients with bipolar depression for 8 weeks and quetiapine in continuation treatment for 26 up to 52 weeks. Study code: D1447C00134. Synopsis 5 October 2007.
162. AstraZeneca. Multi-centre, double-blind, randomised, parallel-group, placebo-controlled, phase 3 study of the efficacy & safety of quetiapine fumarate & lithium as monotherapy in adult patients with bipolar


164. AstraZeneca. An international, multi-centre, double-blind, randomised, parallel-group, placebo-controlled, phase III study of the efficacy and safety of quetiapine fumarate (Seroquel, single oral 300 mg or 600 mg dose) and lithium as monotherapy in adult patients with bipolar depression for 8 weeks and quetiapine in continuation treatment for 26 up to 52 weeks. Study code: D1447C0001. Synopsis.


Copyright © 2018 Johns Hopkins University. All Rights Reserved

Disclaimer:

The [views, statements, opinions] presented in this report are solely the responsibility of the author(s) and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute® (PCORI®), its Board of Governors or Methodology Committee.

Acknowledgement:

Research reported in this report was [partially] funded through a Patient-Centered Outcomes Research Institute® (PCORI®) Award (#ME-1303-5785).