BMJ Open

Restoring invisible and abandoned trials of gabapentin for neuropathic pain: a clinical and methodological investigation

Evan Mayo-Wilson ,[1] Xiwei Chen,[1] Riaz Qureshi,[2] Stephanie Dickinson,[1] Lilian Golzarri-Arroyo,[1] Hwanhee Hong,[3] Carsten Görg,[4] Tianjing Li[5]

To cite: Mayo-Wilson E, Chen X, Qureshi R, *et al*. Restoring invisible and abandoned trials of gabapentin for neuropathic pain: a clinical and methodological investigation. *BMJ Open* 2021;**11**:e047785. doi:10.1136/bmjopen-2020-047785

► Prepublication history and supplemental material for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2020-047785).

Received 09 December 2020
Accepted 26 May 2021

For numbered affiliations see end of article.

Correspondence to
Dr Evan Mayo-Wilson; emayowil@iu.edu

ABSTRACT

Introduction Gabapentin (Neurontin) is prescribed widely for conditions for which it has not been approved by regulators, including certain neuropathic pain conditions. There is limited evidence that gabapentin is safe and effective for the treatment of neuropathic pain. Published trial reports, and systematic reviews based on published trial reports, mislead patients and providers because information about gabapentin's harms has been published only partly. We confirmed that trials conducted by the drug developer have been abandoned, and we plan to conduct a restoration with support from the Restoring Invisible and Abandoned Trials Support Centre (https://restoringtrials.org/).

Methods and analysis In this study, we will analyse and report the harms that were observed in six trials of gabapentin, which have not been reported publicly (eg, in journal articles). We will use clinical study reports and individual participant data to identify and report the harms observed in each individual trial and to summarise the harms observed across all six trials. We will report all adverse events observed in the included trials by sharing deidentified data and summary tables on the Open Science Framework (https://osf.io/w8puv/). Additionally, we will produce a summary report that describes differences between the randomised groups in each trial and across trials for prespecified harms outcomes.

Ethics and dissemination We will use secondary data. This study was determined to be exempt from Institutional Review Board (IRB) review (protocol #1910607198).

Strengths and limitations of this study

- Gabapentin is prescribed widely and thus of ongoing clinical interest.
- Gabapentin's developer has no plans to publish complete information about its harms, so restoration is needed to complete the published record.
- This study will report previously undisclosed harms in six trials of gabapentin for neuropathic pain.
- Unpublished data to be used in this study will provide a more comprehensive account of the drug's effects compared with previous journal articles and systematic reviews limited to published evidence.

INTRODUCTION

Gabapentin (Neurontin) was approved in 1993 by the US Food and Drug Administration (FDA) for the treatment of epilepsy. It was later approved for the treatment of postherpetic neuralgia. To encourage doctors to prescribe gabapentin for many types of neuropathic pain, including 'off-label' indications other than those approved by FDA, the manufacturer published favourable clinical trial results in medical journals.[1] In 2004, the manufacturer pleaded guilty to civil and criminal charges related to illegal marketing and paid US$430 million to the US Department of Justice.[2] Pfizer, the manufacturer that ultimately acquired the drug, paid US$325 million in 2014 to settle plaintiffs claims that it defrauded patients and benefit providers in this manner.[3]

Gabapentin continues to be prescribed widely,[4] including for the treatment of pain, perhaps because published trial reports and systematic reviews based on published trial reports continue to mislead patients and providers about its benefits and harms. In published journal articles, undisclosed changes to primary outcomes and methods of analysis contributed to overestimating gabapentin's potential benefits.[5] Such changes were possible because the manufacturer conducted multiple analyses of primary and secondary outcomes and reported only a subset of the results.[6 7] Publications also include very little information about harms ('adverse events' (AEs)) that occurred in clinical trials,[8–10] which patients with chronic pain say are critically important to their decisions about whether to take drugs to treat pain.[11] More information about benefits and harms can be found in unpublished clinical study reports (CSRs), which the manufacturer

Table 1 Glossary of terms related to adverse events and sources

Term	Definition used in our study
Clinical study report (CSR)	A special document type originating with drug and device manufacturers for submitting information to regulators (eg, US Food and Drug Administration (FDA), European Medicines Agency). CSRs usually follow intentional guidelines for structure and content, and they often contain detailed summaries of trial design (including the trial protocol and statistical analysis plan) and results.[82]
Individual patient data (IPD)	A record of variables collected for each participant in a clinical trial (eg, clinical characteristics, scores on tests and questionnaires), usually stored in a database file.
Adverse event (AE)	The International Conference on Harmonisation defines an 'AE' as 'any untoward medical occurrence in a patient or clinical investigation subject administered a pharmaceutical product and which does not necessarily have to have a causal relationship with this treatment."[83] The US FDA and other regulators use this definition.[84 85]
Coding Symbols for a Thesaurus of Adverse Reaction Terms (COSTART)	The Coding Symbols for a Thesaurus of Adverse Reaction Terms (COSTART) is a terminology developed and used by the FDA for coding, filing and retrieving of post-marketing adverse drug/biologic experience reports.[33]COSTART was replaced by Medical Dictionary for Regulatory Activities in the late 1990s.[34]
Preferred term (COSTART symbol)	A 20-character code used to identify events using the COSTART system.[33]
Mid-level systems	According to the COSTART manual, a mid-level system is a 'mid-level pathophysiologic classification of COSTART for purposes of categorising and retrieving information based on disease associations.' 'This section is hierarchical in arrangement, allowing one to be very general or more specific and is a convenient strategy for searching for drug-induced disease'.[33]
Body systems	According to the COSTART manual, 'Essentially anatomic, this body system classification is sometimes the basis of search strategy. The classification is hierarchical in nature.'[33]
Serious	The US FDA and other regulators consider AEs as 'serious' when they lead to or prolong hospitalisation, cause death or disrupt normal life functions.[84 85] 'Serious' is not synonymous with 'severe,' the latter being a descriptive characteristic rather than a regulatory classification.
Time under observation	The length of time (eg, days) during which a participant could have reported harms to study investigators, which we estimated for this study using multiple data sources.

Definitions adapted from previous reports of the MUDS study.[7–10 16]
MUDS, Multiple Data Sources for Meta-Analysis.

called 'Research Reports', and individual patient data (IPD) that became available during litigation (table 1). Despite the availability of these data, information about harms has not been incorporated in reviews and guidelines about gabapentin, such as a highly cited Cochrane review,[12 13] that continue to be used to promote the use of gabapentin.[14]

We confirmed that trials conducted by the manufacturer have been abandoned.[15] We contacted Pfizer in 2015 to share the published protocol for the Multiple Data Sources for Meta-Analysis (MUDS) study, a methodological study about the use of different data sources for systematic reviews and meta-analysis.[16] We subsequently exchanged emails and spoke by telephone with Pfizer representatives. In 2016, Pfizer confirmed by telephone that it has no plans to publish additional data or analyses from trials of gabapentin for neuropathic pain. In 2017, we received an unexpected email from a Pfizer contractor in response to our earlier request for information about gabapentin trials; we replied, but Pfizer did not respond (online supplemental file 1). Thus, we issued a 'call to action' to produce a complete account of the harms that were observed but not reported in six important clinical trials,[17] for which we later received support from the Restoring Invisible and Abandoned Trials Support Centre.[18]

Objectives

Our objective in this study is to report the harms that were observed in six trials of gabapentin, which have not been reported publicly (eg, in journal articles). We will use CSRs and IPD to identify and report the harms observed in each individual trial and to summarise the harms observed across all six trials.

This study was determined to be exempt from IRB review (IRB Protocol #: 1910607198).

METHODS AND ANALYSIS

Patient and public involvement

This study will use data that were previously collected by the developer and manufacturer of gabapentin. Members of our team contributed to data harmonisation as part of the MUDS study; patients were involved in the design,

conduct, reporting and dissemination of the MUDS study. Patients and the public will not be involved in this additional investigation.

Outcomes

We will report all AEs observed in the included trials by sharing deidentified data and summary tables on the Open Science Framework (https://osf.io/w8puv/).

Additionally, we will attempt to produce a summary report that describes differences between the randomised groups in each trial and across trials for the following outcomes:

1. Proportion of participants who experienced one or more AEs.
2. Proportion of participants who experienced one or more serious AEs.
3. Rate of one or more AEs.
4. Rate of one or more serious AEs.
5. Proportion of participants who discontinued their assigned intervention because of AEs.
6. Proportion of participants who discontinued their assigned intervention for any reason.
7. Time to discontinuation because of AEs.
8. Time to discontinuation for any reason.
9. Proportion of participants who experienced specific AEs at the level of preferred term, mid-level system, and body system.
10. Rate of specific AEs at the level of preferred term, mid-level system and body system.
11. Time to specific AEs at the level of preferred term, mid-level system and body system.

Eligible trials

This is an additional study using data from the MUDS study, which examined two drugs, gabapentin and quetiapine. The MUDS investigators selected drugs for which they had access to both public and non-public sources of information; they searched for public and non-public information, and they requested additional information from the manufacturers, as described previously.[7 16 19]

This study focuses on gabapentin only. The MUDS study included 21 parallel randomised clinical trials comparing gabapentin with placebo for neuropathic pain in adults. Crossover studies were excluded. Six trials included in the MUDS study were conducted by the manufacturer and are included in this study (table 2); for these trials, we have access to CSRs and IPD[20–27] as well as public data sources, such as journal articles.[28–31] Pfizer-released information about these studies during litigation in which Kay Dickersin, principal investigator of the MUDS study, served as an expert witness. During litigation, Pfizer provided to plaintiff's attorneys: a list of trials of gabapentin, internal company documents (Inferential Analysis Plans, Research Reports and memos), and Microsoft Access databases containing IPD. In response to requests for additional information and meta-data (eg, codebooks), Pfizer confirmed that materials not released during litigation either never existed or were lost (online supplemental file 1).

Data collection and management

Obtaining aggregate data

From each report of each eligible trial in the MUDS study, two investigators independently extracted data using the open access Systematic Review Data Repository (http://srdr.ahrq.gov/) and resolved discrepancies by consensus and through discussion with a third reviewer if necessary. The MUDS investigators shared the statistical code and

Table 2 Eligible placebo controlled trials

Study protocol number	Pain condition(s) included	No of participants assigned to gabapentin and placebo	Daily dose in each gabapentin group
945-210[20]	Diabetic neuropathic pain	165	Maximum (target) dose 3600 mg
945-224[21]	Diabetic neuropathic pain	325	Fixed doses 600, 1200 and 2400 mg/day
945-400-211[22–24]	Postherpetic neuralgia	229	Fixed dose 3600 mg/day
945-430-295[25]	Postherpetic neuralgia	334	Fixed doses 1800 and 2400 mg/day
945-430-306[26]	Neuropathic pain (Allowed: complex regional pain syndrome or reflex sympathetic dystrophy; pain because of traumatic injury; diabetic peripheral neuropathy; phantom limb pain or pain following amputation of limbs; postherpetic neuralgia; radicular pain or radiculopathy associated with spinal stenosis; stroke)	305	Maximum (target) dose 2400 mg
A945-1008[27]	Diabetic peripheral neuropathy	389	Maximum (target) dose 3600 mg

Table 3 Variables for each participant in the MUDS database

Variable name	Description
study_id	Study identification number as assigned by the manufacturer.
patient_id	Participant identification number as assigned by the manufacturer.
sex	Participant sex, harmonised to 'female' or 'male' by the MUDS investigators.
treatment	Treatment group allocation, including dose, as reported by the manufacturer.
treatment_dic	Treatment group allocation, recoded by the MUDS team as either 'placebo' or 'gabapentin'.
pain_d01 pain_d02 … pain_d98	Daily pain score. Each morning on arising, participants wrote down a number to rate their pain during the previous 24 hours on an 11-point Likert scale ranging from 0 (no pain) to 10 (worst possible pain).
sleep_d01 sleep_d02 … sleep_d98	Daily sleep score. In five of six trials,[20–25 27]each morning on arising, participants wrote down a number to rate how pain interfered with sleep on an 11-point Likert scale. The scale ranged from 0 (did not interfere with sleep) to 10 (completely interferes with sleep). The daily sleep score was not collected in one trial.[26]

MUDS, Multiple Data Sources for Meta-Analysis.

datasets from the MUDS study on the Dryad repository.[32] The MUDS investigators shared partially redacted CSRs on the Drug Industry Documents Archive,[20–26] which are complete except for appendices containing identifying information (eg, patient initials, date of birth, exact dates of medical examinations), which the MUDS investigators did not have resources to redact and recode.

Obtaining IPD

The MUDS study used CSRs and partially deidentified IPD that were provided as Microsoft Access Databases without codebooks to Professor Kay Dickersin for her expert witness report in litigation against Pfizer. The MUDS investigators developed codebooks and harmonised the databases by comparing the databases with case report forms (which show how and when data were recorded) and statistical analysis plans (which show how data were coded and analysed) to identify the variables contained in the databases.

In this study, we will use a subset of the IPD database that was harmonised by the MUDS investigators. The MUDS database includes two types of tables, which include information about participants and AEs, respectively (tables 3 and 4).

Individual harms and groups of harms

Multiple systems may be used to classify and analyse harms including Coding Symbols for a Thesaurus of Adverse Reaction Terms (COSTART), Medical Dictionary for Regulatory Activities (MedDRA), Systemized Nomenclature of Medicine (SNOMED), or Common Terminology Criteria for Adverse Events (CTCAE).[33–36] Such systems

Table 4 Variables for each AE in the MUDS database

Variable name	Description
COSTARTsymbol	For five of six trials,[20–26] COSTART symbol ('preferred term') for each AE, which the MUDS team matched to alphanumeric codes reported by the manufacturer. One trial[27] was not coded using COSTART, and we will match AE terms to COSTART symbols for this study.
COSTARTmid1	For five of six trials,[20–26] COSTART mid-level system for each AE as matched by the MUDS team. One trial[27] was not coded using COSTART, and we will match AE terms to COSTART symbols for this study.
COSTARTbodyA1	For five of six trials,[20–26] COSTART primary body system for AE as matched by MUDS team. One trial[27]was not coded using COSTART, and we will match AE terms to COSTART symbols for this study.
aetext	Text describing adverse events as reported by the manufacturer.
aestartday	Time of onset as reported by the manufacturer (days from start of medication to start of AE).
aeendday	Time of resolution as reported by the manufacturer (days from start of medication to end of AE).
recurrent	For five of six trials,[20–26]whether the AE recurred as reported by the manufacturer. Recurrence was not available in one of the six trial databases.[27]
severity	Severity of the AE as reported by the manufacturer (mild, moderate, severe).
serious	Whether AE was considered serious as reported by the manufacturer.
reltreat_dic	Whether the AE was judged as related to treatment or caused by treatment. Harmonised by the MUDS investigators as either 'yes'; 'no' or 'insufficient Information'.
action	Action taken following AE, as reported by the manufacturer ("None"; "Dose reduced"; "Dose interrupted"; "Discontinued"; or "Increased").

AE, adverse event; COSTART, Coding Symbols for a Thesaurus of Adverse Reaction Terms; MUDS, Multiple Data Sources for Meta-Analysis.

generally use a hierarchical structure with higher-order terms denoting the anatomic or physiological systems that are affected. Lower order terms typically denote the specific harms experienced. Because some harms are rare, grouping harms by anatomic or physiological systems can increase statistical power and increase the possibility of detecting drug-induced harms. Grouping can also disguise important harms by combining them with less important harms (eg, 'migraine' might be more severe than 'headache', but the distinction would be lost if combined under the higher order term 'headaches').

The COSTART was developed by FDA, and it was being used for regulatory trials when gabapentin was developed.[33] MedDRA replaced COSTART in the late 1990s.[34]

Using the COSTART system, analyses may be performed at the level of 'preferred term,' which is the lowest level for analysis in the hierarchy, and preferred terms may be grouped for analysis using the following higher levels of aggregation:

- Body systems.
- Body system subcategories.
- Mid-level system.
- Mid-level system subcategories.

In this study, we plan to assess harms at the level of preferred terms and grouped to mid-level and body systems. For preferred terms that could be matched to more than one body system, we will use the primary body system as recorded by the MUDS investigators.

Mapping individual patient data to the COSTART system

For five included trials, the IPD received from Pfizer included a description of each AE alongside a 5-character alphanumeric 'COSTART code'. After deduplication, the MUDS investigators identified 246 unique 5-character alphanumeric codes that were matched to COSTART preferred terms for analysis using the following methods:

1. In addition to the five-character alphanumeric code, the IPD for one trial[22–24] contained a field with the COSTART preferred term. The MUDS investigators used the IPD to match 120 (49%) of the 246 unique five-character alphanumeric codes to preferred terms.
2. IPD for two gabapentin trials[37 38] that were not eligible for the MUDS study contained both five-digit alphanumeric codes and COSTART preferred terms. The MUDS investigators used these IPD to validate the previously identified preferred terms and to match 33 (10%) additional alphanumeric codes to COSTART preferred terms.
3. An appendix in a CSR[39] included both five-digit alphanumeric codes and COSTART preferred terms. Using the software ABBYY FineReader,[40] the MUDS investigators extracted these data into a spreadsheet, which they used to match 19 (8%) additional five-character alphanumeric codes to COSTART preferred terms.
4. The MUDS investigators converted Index D of the COSTART Dictionary (Glossary of Included Terms) into a spreadsheet to match the remaining five-digit alphanumeric codes to COSTART preferred terms. This was accomplished for each five-character alphanumeric code by comparing the accompanying 'AE text' (ostensibly, what was written on the Case Report Form (CRF) to describe each AE) to the terms listed in the COSTART Glossary.[33] The COSTART Glossary includes approximately 6000 synonyms for COSTART preferred terms. This information was used to match 24 (10%) additional five-character alphanumeric codes to COSTART preferred terms.
5. Then the MUDS investigators matched 5 (2%) five-character alphanumeric codes and accompanying text to MedDRA, a more current AE coding system, and mapped them to the corresponding COSTART terms.
6. Finally, three clinicians worked in pairs to review the remaining 45 five-character alphanumeric codes, along with accompanying text and the COSTART Glossary,[33] and to propose appropriate preferred terms. Each of the three clinicians was given 30 five-character alphanumeric codes; the pairs compared their ratings and sought input from the third clinician to resolve any disagreements. The MUDS investigators successfully matched 34 (14%) five-character alphanumeric codes to preferred terms in this manner.
7. The remaining 11 (5%) alphanumeric codes could not be matched.

In one trial,[27] some AEs were recorded using terms that mapped to COSTART preferred terms while other terms did not map to COSTART. To analyse this trial and to compare it with the other trials in our study, we will attempt to map all terms to COSTART.

The first step in mapping these terms was performed automatically by the MUDS investigators in Stata (V.13) using the COSTART dictionary and the other included trials. This yielded a match for some entry terms, but many still require manual mapping. To manually map terms, two independent investigators will compare the AE text to COSTART preferred terms. Discrepancies will be reviewed by a third investigator and discussed, and a clinician will review any remaining unmapped terms.

To manually map terms, each of two independent investigators will use the BioPortal website (fifth Edition COSTART)[41] to compare AE text to COSTART preferred terms. BioPortal includes all entry terms and synonyms for preferred terms and also provides an intuitive presentation of the mappings to higher order terms. When mapping the AE text to preferred terms, we will extract the corresponding preferred term ('notation') and the 'prefLabel' (figure 1).

For all mapped preferred terms in trial A945-1008[27] that also appear in one of the other five trials, the previously mapped mid-level system and body system will be used in our study. For all preferred terms that are unique to A945-1008,[27] we will assign the mid-level system and body system terms that are more specific to the preferred term.

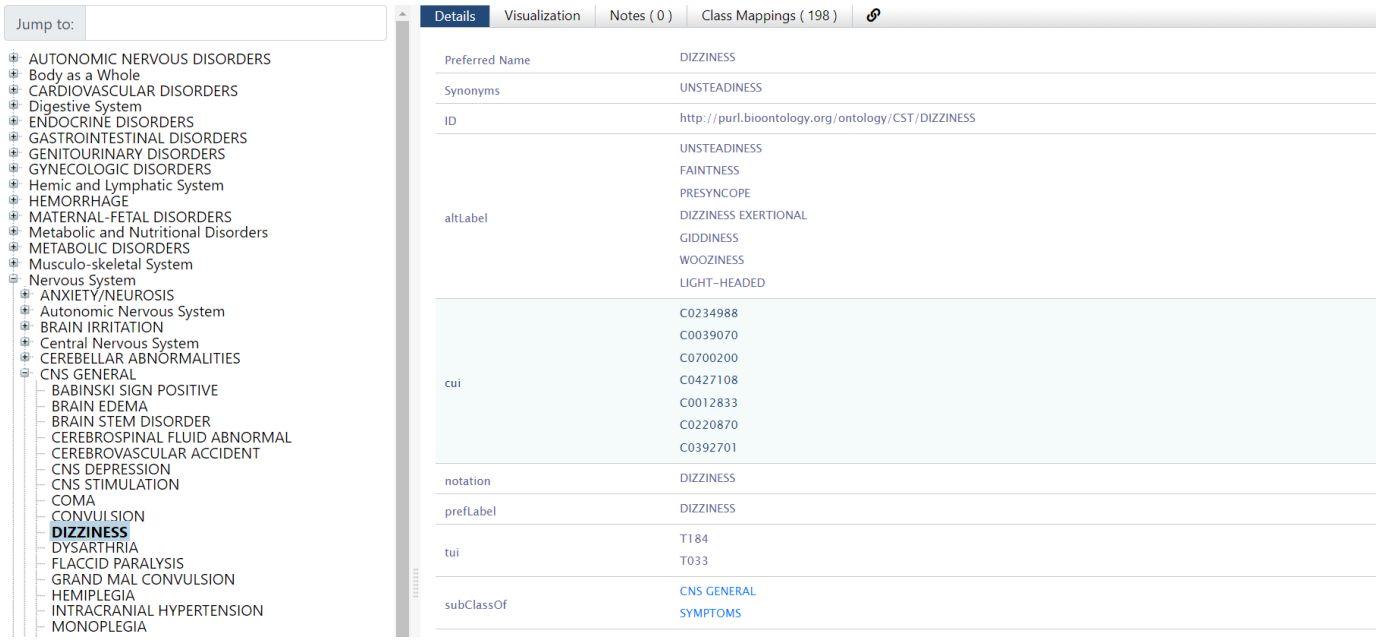

Figure 1 BioPortal COSTART dictionary result for 'Dizziness'. The BioPortal dictionary can be navigated manually on the left-hand side of the figure, or by searching for specific terms using the search function. To the right, the preferred term ('notation'), label ('prefLabel') and alternative labels ('altLabel') appear. The notation, prefLabel and altLabel are 'entry terms' that direct to the preferred term; for example, a physician could write 'light-headed' in their notes, which would be mapped to the preferred term 'Dizziness'. On BioPortal, entering an altLabel into the search bar will bring up the corresponding preferred term. The results on the right also include corresponding mid-level systems and body systems ('subClassOf'). Here, 'Dizziness' maps to 'CNS General' (mid-level) which falls under 'nervous system' (body system), as can be seen by looking on the left side of the page or by clicking on the link to CNS General (in blue). 'Dizziness' also maps to 'symptoms' (mid-level) which falls under 'non-specific disorders' (body system). COSTART, Coding Symbols for a Thesaurus of Adverse Reaction Terms

Methods of analysis in each individual trial

Time under observation for each participant

IPD received from the manufacturer did not appear to include a variable indicating the time during which each participant was under the observation. We will calculate time under the observation for each participant by checking the following sources of information. Where more than one value is available, we will choose the longest valid time (table 5).

For example, if a participant's pain and sleep records were available up to day 46 (then missing), the last recorded medication was taken on day 48, and the participant reported AEs on days 9, 42 and 62, we would estimate the time under observation as 62 days. We will check for potentially invalid values (eg, 620 days) before performing the analysis.

Study discontinuation (drop-out) status for each participant

IPD received as databases from the manufacturer did not include a variable indicating whether each participant had completed the study or discontinued. To compare dropout between groups, we will use information from the CSRs or calculated using the methods described below.

- One CSR[21] listed participant identifiers for participants who discontinued.
- Four CSRs[20 22–26] listed each participant's end-of-study status as 'Completed study/phase' or giving a reason for discontinuation (ie, 'AE', 'lack of efficacy', 'non-compliance' or 'other'); for these four trials, we will consider participants to have discontinued (dropped out) if their end-of-study status was not 'completed study/phase'.

Table 5 Variables used to calculate time under observation for each participant

Source	Variable name
CSR[22–24]	Last day in study
CSR[25 26]	Observation day
IPD	aestartday
IPD	aeendday
IPD	pain_d01 pain_d02 … pain_d98
IPD[20–25 27]	sleep_d01 sleep_d02 … sleep_d98
CSR[20 21]	Study day last double-blind medication
CSR[25 26]	Day of last dose of study drug

CSR, clinical study report; IPD, individual participant data.

- One trial[27] did not report end-of-study status in the CSR. Participants in this trial were assigned placebo for 1 week before random assignment to gabapentin or placebo for 14 weeks (98 days); thus, we will consider participants to have discontinued if their time under observation is less than the 98 days of treatment period.

Days with AEs

For each participant, we will calculate the number of days with AEs as the difference between the time of onset and time of resolution for each of the following levels: any AE, any serious AE, and specific AEs at level of preferred term, mid-level system and body system. Except for analyses at the level of preferred term, each day will be counted once. Thus, the number of days with AEs will not exceed the number of days of observation (eg, for each participant, a day with 'any AE' will be counted once whether the participant experienced one AE or multiple AEs on that day).

Consistency and data quality

To check the quality of the dataset, we will compare CSRs and IPD with regard to the time of onset of AEs, last medication time, time under observation and study period. We will flag any observations where the AE's time of onset is indicated to be at least 2 weeks (14 days) longer than the study period. These observations will be excluded from the primary analyses, but will be included in sensitivity analyses.

Risk of AE and risk of discontinuation

For the randomised participants, we will calculate the proportion (risk) of participants in each group reporting: any AE, any serious AE, each mid-level system, each body system and selected preferred terms. We will also calculate the proportion of randomised participants in each group who discontinued because of AEs and who discontinued for any reason. For these outcomes, we will report the differences between-groups as relative effects (eg, risk ratio (RR) or OR) and absolute effects (risk difference (RD)), including the corresponding 95% CIs.

Proportion of days with AE

For the randomised participants, we will calculate the proportion of days with any AE, any serious AE, selected preferred terms, each mid-level term and each body system. For these outcomes, we will report the ratio between groups.

The proportion of days with AE in each group is the total number of days with the AE divided by the total time under observation (eg, '5 days with headache per week'). Because randomised participants who do not take at least one dose of study medication will not contribute any person-time for this analysis, our planned analysis of the full intention-to-treat population will be identical to an analysis limited to the 'safety population'.

The proportion ratio is the proportion in the intervention group divided by the proportion in the placebo group, which we will express (eg, 'two times more days with headache').

Incidence rate

For the randomised participants, we will calculate the incidence rate of any AE, any serious AE, selected preferred terms, each mid-level term and each body system. For these outcomes, we will report the incidence rate ratio (IRR) and 95% CI to compare the differences between two treatment arms. We will consider the same models proposed for 'risk of AE and risk of discontinuation'.

Time-to-event (survival analysis) for harms and discontinuation

For participants who took at least one dose of study medication (and whose time under observation is therefore greater than 0 days), we will use survival analyses to investigate differences in: any AE, any serious AE, each mid-level AE, each body system AE, discontinuation because of AEs and discontinuation for any reason, and selected preferred terms. For these outcomes, we will calculate the differences in median time-to-event, and we will report the HR and its 95% CI. Although we expect survival analyses to complement analyses based on risk, discontinuation is a competing risk for the reporting of AEs to trial investigators, so it is possible that RRs and HRs would differ in magnitude or direction. We will fit a stratified Cox regression model with random effects.[42] The proportional hazard assumption will be tested using Grambsch and Therneau test and Schoenfeld residuals. If the assumption is not valid, we will consider alternative models such as reporting time-varying HR or cumulative Cox regression.[43]

We will calculate time to event as the difference between randomisation and the day on which the AE started. For participants reporting the same AE more than once, we will use the time-to-first occurrence and we will exclude future occurrences from this analysis. Participants who discontinue the study will be censored from the time of study discontinuation. We will include the total time under the observation for participants who complete the study without reporting any AEs, or the AE of interest, as appropriate.

Statistical significance (p values and CIs)

For the effect estimates that we calculate, we will also calculate p values and CIs. Although p values and other methods of assessing statistical significance should not be used for null hypothesis testing with AEs, they can be helpful in aiding interpretation of results.[44–47] While it is important to show some measure of uncertainty surrounding an estimate to provide a sense of the strength of the evidence for an association–especially as effect estimates for AEs can be high when events are rare—it is also important that they not be overinterpreted as proof or lack of proof of associations. Thus, we will not consider these to be null hypothesis tests and will not interpret 'significant' values as evidence of causal relationships.

Instead, we will interpret these values together with the effect estimates and number of events.

Many non-systematically assessed harms will be uncommon. Trials are rarely designed or powered to detect differences between groups in the occurrence of non-systematically assessed harms, unlike potential benefits. Uncommon events may produce unstable estimates wherein the bounds of the CI are unreasonable and hypothesis tests may be rendered invalid.[48] Moreover, hundreds of different non-systematically assessed harms might be reported, and statistical challenges are exacerbated by the problem of multiple testing, although some statistical methods have been developed to ameliorate these issues (eg, False Discovery Rate).[48–50]

Methods for data synthesis

Combining effects across studies

We will estimate effects by combining evidence across studies using RRs, ORs, RDs and HRs and corresponding 95% CIs for: any AE, any serious AE, each mid-level AE, each body system, selected preferred terms, discontinuation because of AEs and discontinuation for any reason.[42 51 52] We will conduct two-stage IPD meta-analyses. One-stage and two-stage meta-analyses will give almost identical results when models are not adjusted by baseline covariates, and we do not expect to adjust using baseline covariates because relatively little information is available from the included trials (see table 3).[53 54] Specifically, the first stage will aggregate IPD for each trial and each arm. For the second stage, we will consider various meta-analytic methods to fit rare binary outcomes.[55] We will fit traditional frequentist methods including Peto and Mantel-Haenszel and Bayesian hierarchical meta-analytic methods that incorporate between-study heterogeneity with random effects. This method is preferable to a frequentist approach because studies with 'zero' cells are not a problem for the Peto, Mantel-Haenszel and Bayesian methods. For various approaches to handling such 'zero' cells for other frequentist meta-analysis models are proposed by Sweeting *et al*.[56] For the proportion of days with AEs, we will fit Bayesian Poisson regression with random effects to estimate the proportion ratio.

To identify a subset of preferred terms for reporting and for further analysis, we will conduct an interim analysis using the combined events in the gabapentin groups and the combined events in the placebo groups to calculate the frequency of each preferred term in the gabapentin group, and the RR, OR and the HR for the gabapentin group compared with the placebo group. We will focus on preferred terms associated with the gabapentin group (eg, RR >0 or HR >1.0). We will explore various methods for selecting AEs for further investigation and we will describe the implications of selection criteria for our results.

Level of analysis issues

For multiarm trials in which participants were randomised to different doses of gabapentin,[21 25] we will combine the gabapentin groups into a single group and compare it with the placebo group, as recommended in the Cochrane Handbook.[57]

Assessment of heterogeneity

We will assess clinical and methodological heterogeneity using data coded by the MUDS investigators. We will interpret the results alongside the characteristics of studies, including risk of bias assessments, in tables and in a structured narrative.

To quantify statistical heterogeneity in results, we will calculate I^2 and perform Cochran's χ^2 test (ie, Q test), and we will visually examine the forest plots. An I^2 of over 75% will be considered as high heterogeneity, and we will adopt a p value of 0.10 as a threshold for statistical significance of Q test.[52] We will also report estimated SD of random effects from Bayesian random effects models.

Subgroup analysis

Because women tend to be smaller than men, we expect that women might experience more AEs than men for any given dose of a drug. We will explore differences between sex (men compared with women) for discontinuation because of AEs and for each body system. Other exploratory subgroup analyses may be considered depending on the findings.

Sensitivity analysis

Because two trials[21 25] randomised participants to multiple doses of gabapentin (eg, 600 mg/day, 2400 mg/day) or placebo, we will perform sensitivity analysis for selected outcomes by dose of gabapentin. For selected outcomes, we will 'split' randomly the shared placebo group into two or more groups with smaller sample size to account for the fact that the group has been used twice or more, and include each pairwise comparison separately in the meta-analyses across trials.[57] For example, in the three-arm trial of gabapentin 1800 mg, 2400 mg and placebo, we could compare 1800 mg gabapentin with half of the placebo group and 2400 mg gabapentin with the other half of the placebo group.

Analysing and reporting preferred terms

To identify a subset of preferred terms for reporting and for further analysis, we will conduct interim analyses in which we calculate the frequency of each preferred term in the gabapentin group and the treatment effects comparing the gabapentin group with the placebo group. We will focus on preferred terms that are associated with gabapentin (eg, RD >0.0, RR >1.0, IRR>1.0, HR >1.0). Of those, we will explore various methods for selecting AEs for further investigation and we will describe the implications of selection criteria for our results.

In each individual trial, we will attempt to calculate RRs, ORs, RDs, IRRs, and HRs for the selected preferred terms. If these results cannot be calculated in a given trial, we will report descriptive results (eg, no events were observed in any group). To synthesise the results across trials, we will calculate RRs, RDs and HRs for the preferred terms

selected for further analysis (see the 'Combining effects across studies' section).

STATISTICAL SOFTWARE

Analyses will be conducted, and figures will be drawn, in R statistical software (R V.4.0.3 and RStudio V.1.2.5001)[58] as needed.

ETHICS AND DISSEMINATION

We will use secondary data. This study was determined to be not human subjects research and thus exempt from IRB review (protocol #1910607198).

DISCUSSION

To make informed decisions about health interventions, patients and other stakeholders need accurate and complete information about both benefits and harms. Syntheses of clinical trial findings should include all available evidence; however, they are often only based on information reported in public sources, such as journal articles,[59–61] which are often incomplete.[5–7 61–72] Compared with journal articles, CSRs and IPD contain much more information about harms observed in clinical trials.[73–75] Moreover, a comprehensive analysis of the harms associated with any commonly prescribed drug would include observational evidence, which might apply to relatively larger and more heterogeneous populations over longer periods of time.

The methods used to assess and to report harms in clinical trials contribute to challenges for interpreting and synthesising trial evidence, and many systematic reviews that plan to synthesise harms ultimately do not address them.[76] While AEs can be assessed systematically in clinical trials[9]—using methods like those used to assess potential benefits[6 7]—AEs are often assessed non-systematically in response to open-ended questions such as 'have you noticed any symptoms since your last visit?' or by spontaneous reporting. Evidence syntheses (eg, systematic reviews, clinical practice guidelines) could help identify rare AEs if all observed AEs were available for all trials[77]; however, rare AEs cannot be identified when clinical trials report only those AEs occurring above certain thresholds. Just as selectively reporting potential benefits based on quantitative results leads to biased meta-analyses[78–81] selection criteria for reporting AEs will lead to biased overall estimates. At the same time, reporting hundreds of events might overwhelm patients and clinicians with information that does not help them make decisions about whether and how to use medicines.

This restoration will address both problems by publishing the complete AEs observed in these trials and producing a clinically informative summary following a prespecified Statistical Analysis Plan (SAP).[1 7] Moreover, it will advance methods for analysing and reporting AEs in clinical trials. It is a limitation that this study will be include only a subset of known gabapentin trials, and it is a limitation that the system used to categorise and analyse AEs in these trials (ie, COSTART) is no longer in regular use (ie, it has been replaced by MedDRA). We hope this project will facilitate future guidance for treating neuropathic pain, and it will help patients and clinicians make informed decisions about the use of gabapentin.

Author affiliations
[1]Department of Epidemiology and Biostatistics, Indiana University School of Public Health-Bloomington, Bloomington, Indiana, USA
[2]Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA
[3]Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina, USA
[4]Department of Biostatistics and Informatics, University of Colorado School of Public Health, Aurora, Colorado, USA
[5]Department of Ophthalmology, University of Colorado, Denver, Colorado, USA

Acknowledgements We are grateful to the MUDS investigators who prepared much of the data used in this analysis. Kay Dickersin was principal investigator of the MUDS study. The MUDS investigators also included Lorenzo Bertizzolo, Joseph K. Canner, Theresa (Terrie) Cowley, Peter Doshi, Jeffrey Ehmsen, Nicole Fusco, Gillian Gresham, Nan Guo, Jennifer Haythornthwaite, James Heyward, Susan Hutfless, Jennifer Payne, Diana Pham, Lori Rosman, Elizabeth A. Stuart, Catalina Suarez-Cuero, Elizabeth Tolbert, Claire Twose and Swaroop Vedula.

Contributors EM-W, TL, SD and RQ obtained funding. EM-W, XC and RQ wrote the first draft of the manuscript. HH, SD, LG-A, CG and TL contributed to the refinement of the study methods and critical revision of the manuscript. All authors read and approved the final version of the manuscript.

Funding EM-W, XC, SD, LG-A and TL receive support for this study through a grant from the RIAT Support Centre at the University of Maryland, which is supported by Arnold Ventures (Laura and John Arnold Foundation).

Competing interests EM-W and TL previously received support for research about gabapentin from the Patient Centred Outcomes Research Institute (PCORI).

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iD
Evan Mayo-Wilson http://orcid.org/0000-0001-6126-2459

REFERENCES

1 Vedula SS, Goldman PS, Rona IJ, *et al*. Implementation of a publication strategy in the context of reporting biases. A case study based on new documents from Neurontin litigation. *Trials* 2012;13:136.
2 Lenzer J. Pfizer pleads guilty, but drug sales continue to SOAR. *BMJ* 2004;328:1217.
3 Pfizer to pay $325 million in Neurontin settlement. Reuters, 2014. Available: https://www.reuters.com/article/us-pfizer-neurontin-settlement-idUSKBN0ED1IS20140602 [Accessed 27 May 2020].

4 ClinCalc DrugStats. Medical expenditure panel survey (MEPS) 2008-2018. agency for healthcare research and quality (AHRQ), Rockville, MD. ClinCalc DrugStats database version 21.1. Available: https://clincalc.com/DrugStats/Top300Drugs.aspx [Accessed 27 May 2020].
5 Vedula SS, Bero L, Scherer RW, *et al*. Outcome reporting in industry-sponsored trials of gabapentin for off-label use. *N Engl J Med* 2009;361:1963–71.
6 Mayo-Wilson E, Fusco N, Li T, *et al*. Multiple outcomes and analyses in clinical trials create challenges for interpretation and research synthesis. *J Clin Epidemiol* 2017;86:39–50.
7 Mayo-Wilson E, Li T, Fusco N, *et al*. Cherry-picking by Trialists and meta-analysts can drive conclusions about intervention efficacy. *J Clin Epidemiol* 2017;91:95–110.
8 Mayo-Wilson E, Fusco N, Hong H, *et al*. Opportunities for selective reporting of harms in randomized clinical trials: selection criteria for non-systematic adverse events. *Trials* 2019;20:553.
9 Mayo-Wilson E, Fusco N, Li T, *et al*. Harms are assessed inconsistently and reported inadequately part 1: systematic adverse events. *J Clin Epidemiol* 2019;113:20–7.
10 Mayo-Wilson E, Fusco N, Li T, *et al*. Harms are assessed inconsistently and reported inadequately Part 2: nonsystematic adverse events. *J Clin Epidemiol* 2019;113:11–19.
11 Mayo-Wilson E, Golozar A, Cowley T, *et al*. Methods to identify and prioritize patient-centered outcomes for use in comparative effectiveness research. *Pilot Feasibility Stud* 2018;4:95.
12 Wiffen PJ, Derry S, Bell RF, *et al*. Gabapentin for chronic neuropathic pain in adults. *Cochrane Database Syst Rev* 2017;6:CD007938.
13 Moore A, Derry S, Wiffen P. Gabapentin for chronic neuropathic pain. *JAMA* 2018;319:818–9.
14 Mathieson S, Lin C-WC, Underwood M, *et al*. Pregabalin and gabapentin for pain. *BMJ* 2020;369:m1315.
15 Doshi P, Dickersin K, Healy D, *et al*. Restoring invisible and abandoned trials: a call for people to publish the findings. *BMJ* 2013;346:f2865.
16 Mayo-Wilson E, Hutfless S, Li T, *et al*. Integrating multiple data sources (MUDS) for meta-analysis to improve patient-centered outcomes research: a protocol. *Syst Rev* 2015;4:143.
17 Restoring misreported trials of gabapentin for neuropathic pain, 2020. Available: https://www.bmj.com/content/346/bmj.f2865/rr-14 [Accessed 27 May 2020].
18 Doshi P, Shamseer L, Jones M, *et al*. Restoring biomedical literature with RIAT. *BMJ* 2018;361:k1742.
19 Mayo-Wilson E, Doshi P, Dickersin K. Are manufacturers sharing data as promised? *BMJ* 2015;351:h4169.
20 Hes M, Koto E, LaMoreaux LK. *A double-blind placebo-controlled trial of gabapentin for treatment of painful diabetic peripheral neuropathy (protocol 945-210). Parke-Davis pharmaceutical research division of Warner-Lambert company*. Michigan, Ann Arbor, 1998.
21 Roder B, Maisonobe P. Research report: Rr 720-04130. A double-blind placebo-controlled trial with 3 doses of gabapentin for treatment of painful diabetic neuropathy. Available: https://www.industrydocuments.ucsf.edu/drug/docs/#id=fgjm0223 [Accessed 14 Oct 2020].
22 Magnus-Miller L, Podolnick P. Research report: Rr 995-00070. double-blind, randomized, placebo-controlled, parallel groups, multicenter trial to determine the efficacy and safety of Neurontin in subjects with peripheral neuropathy (post-herpetic neuralgia). Available: https://www.industrydocuments.ucsf.edu/drug/docs/#id=jjjm0223 [Accessed 14 Oct 2020].
23 Magnus-Miller L, Podolnick P. Research report: Rr 995-00070. double-blind, randomized, placebo-controlled, parallel groups, multicenter trial to determine the efficacy and safety of Neurontin in subjects with peripheral neuropathy (post-herpetic neuralgia). appendix E. Available: https://www.industrydocuments.ucsf.edu/drug/docs/#id=kjjm0223
24 Magnus-Miller L, Podolnick P. Research report: Rr 995-00070. double-blind, randomized, placebo-controlled, parallel groups, multicenter trial to determine the efficacy and safety of Neurontin in subjects with peripheral neuropathy (post-herpetic neuralgia). appendix E. Available: https://www.industrydocuments.ucsf.edu/drug/docs/#id=ljjm0223
25 Bibby SJ, Maton S, Rosen JP. A double blind placebo controlled trial of gabapentin for the treatment of post herpetic neuralgia.. Available: https://www.industrydocuments.ucsf.edu/drug/docs/#id=zfjm0223 [Accessed 14 Oct 2020].
26 Bibby SJ, Maton S, Wensley S. A double blind placebo controlled trial of gabapentin for the treatment of patients exhibiting symptoms of neuropathic pain. Available: https://www.industrydocuments.ucsf.edu/drug/docs/#id=tfjm0223 [Accessed 14 Oct 2020].
27 Pfizer. Final study report: gabapentin protocol A945-1008. A 15 week, randomized, double-blind, placebo-controlled, parallel-group, multi-center study of Neurontin (gabapentin) for efficacy and quality of life in patients with painful diabetic peripheral neuropathy. Available: https://www.industrydocuments.ucsf.edu/drug/docs/#id=nnjm0223 [Accessed 14 Oct 2020].
28 Serpell MG, Neuropathic Pain Study Group. Gabapentin in neuropathic pain syndromes: a randomised, double-blind, placebo-controlled trial. *Pain* 2002;99:557–66.
29 Rice ASC, Maton S, Postherpetic Neuralgia Study Group. Gabapentin in postherpetic neuralgia: a randomised, double blind, placebo controlled study. *Pain* 2001;94:215–24.
30 Rowbotham M, Harden N, Stacey B, *et al*. Gabapentin for the treatment of postherpetic neuralgia: a randomized controlled trial. *JAMA* 1998;280:1837–42.
31 Backonja M, Beydoun A, Edwards KR, *et al*. Gabapentin for the symptomatic treatment of painful neuropathy in patients with diabetes mellitus: a randomized controlled trial. *JAMA* 1998;280:1831–6.
32 Mayo-Wilson E, Li T, Fusco N. *Data from: Cherry-picking by Trialists and meta-analysts can drive conclusions about intervention efficacy*, 2018.
33 Office of Management and Operations. *COSTART: coding symbols for thesaurus of adverse reaction terms*. Fifth ed, 1995. https://bioportal.bioontology.org/ontologies/COSTART
34 MedDRA M. *Introductory guide MedDRA version 21.1*, 2018.
35 International Health Terminology Standards Development Organisation. Snomed. Available: http://www.snomed.org/ [Accessed 15 Sep 2020].
36 Common terminology criteria for adverse events (CTCAE). Available: https://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm [Accessed 15 Sep 2020].
37 Double-Blind, randomized, placebo-controlled, multicenter trial to determine the efficacy and safety of Neurontin (gabapentin) in migraine prophylaxis administered in doses divided three times a day (tid) (protocol 945-220). PARKE-DAVIS pharmaceutical research division of WARNER-LAMBERT company medical and scientific Affairs department Morris plains,New Jersey, 1999. Available: https://www.industrydocuments.ucsf.edu/drug/docs/#id=klgw0217
38 Double-Blind, randomized, placebo-controlled, multicenter trial to determine the efficacy and safety of Neurontin (gabapentin) in migraine prophylaxis (protocol 945-217). Parke-Davis pharmaceutical research division of Waner-Lambert company medical and scientific Affairs department Morris plains, new Jersey, 2000. Available: https://www.industrydocuments.ucsf.edu/drug/docs/#id=llgw0217
39 Pfizer global research & development ann arbor laboratories Ann Arbor, Michigan. *A randomized, open label trial to determine the relative efficacy and safety of a fixed dose of gabapentin versus optional titration to effect for the treatment of painful diabetic peripheral neuropathy (protocol 945-411*, 2002.
40 ABBYY FineReader PDF 15. ABBYY pdf.abbyy.com/finereader. Available: https://pdf.abbyy.com/
41 NCBO BioPortal. Coding Symbols for a Thesaurus of Adverse Reaction Terms - Summary. Available: https://bioportal.bioontology.org/ontologies/COSTART/?p=summary [Accessed 14 Oct 2020].
42 Smith CT, Williamson PR, Marson AG. Investigating heterogeneity in an individual patient data meta-analysis of time to event outcomes. *Stat Med* 2005;24:1307–19.
43 Gregson J, Sharples L, Stone GW, *et al*. Nonproportional Hazards for Time-to-Event Outcomes in Clinical Trials: JACC Review Topic of the week. *J Am Coll Cardiol* 2019;74:2102–12. doi:10.1016/j.jacc.2019.08.1034
44 US Food and Drug Administration. Manual of policies and procedures. good review practice: clinical review template (MAPP 6010.3 Rev. 1). Available: https://www.fda.gov/files/about%20fda/published/Good-Review-Practice--Clinical-Review-Template.pdf
45 ICH E9 statistical principles for clinical trials ICH Harmonised tripartite guideline. *Int Conf Harmon* 2020 http://www.ich.org/products/guidelines/efficacy/efficacy-single/article/statistical-principles-for-clinical-trials.html
46 Wasserstein RL, Lazar NA. The ASA Statement on *p* -Values: Context, Process, and Purpose. *Am Stat* 2016;70:129–33.
47 Crowe BJ, Xia HA, Berlin JA, *et al*. Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: a report of the safety planning, evaluation, and reporting team. *Clin Trials* 2009;6:430–40.
48 Ma H, Ke C, Jiang Q, *et al*. Statistical considerations on the evaluation of imbalances of adverse events in randomized clinical trials. *Ther Innov Regul Sci* 2015;49:957–65.
49 Mehrotra DV, Heyse JF. Use of the false discovery rate for evaluating clinical safety data. *Stat Methods Med Res* 2004;13:227–38.

50 Berry SM, Berry DA. Accounting for multiplicities in assessing drug safety: a three-level hierarchical mixture model. *Biometrics* 2004;60:418–26.
51 Higgins JP, Li T, Deeks JJ. Chapter 6: Choosing effect measures and computing estimates of effect. In: Higgins JPT, Thomas J, Chandler J, *et al*, eds. *Cochrane Handbook for systematic reviews of interventions version 6.0 (updated July 2019)*, 2019. https://training.cochrane.org/handbook/current/chapter-06
52 Deeks JJ, Higgins JP, Altman DG. Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, *et al*, eds. *Cochrane Handbook for systematic reviews of interventions version 6.2 (updated February 2021)*, 2021. https://https://training.cochrane.org/handbook/current/chapter-10
53 Burke DL, Ensor J, Riley RD. Meta-Analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Stat Med* 2017;36:855–75.
54 Debray TPA, Moons KGM, van Valkenhoef G, van VG, *et al*. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. *Res Synth Methods* 2015;6:293–309.
55 Hong H, Wang C, Rosner GL. Meta-Analysis of rare adverse events in randomized clinical trials: Bayesian and frequentist methods. *Clin Trials* 2021;18:3–16.
56 Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? use and avoidance of continuity corrections in meta-analysis of sparse data. *Stat Med* 2004;23:1351–75.
57 Higgins JP, Eldridge S, Li T. Chapter 23: Including variants on randomized trials. In: Higgins JPT, Thomas J, Chandler J, *et al*, eds. *Cochrane Handbook for systematic reviews of interventions version 6.2 (updated February 2021)*, 2021. https://training.cochrane.org/handbook/current/chapter-23
58 R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2020. https://www.R-project.org/
59 Golder S, Loke YK, Wright K, *et al*. Most systematic reviews of adverse effects did not include unpublished data. *J Clin Epidemiol* 2016;77:125–33.
60 Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. follow-up of applications submitted to two institutional review boards. *JAMA* 1992;267:374–8.
61 Dwan K, Gamble C, Williamson PR, *et al*. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS One* 2013;8:e66844.
62 Chan A-W, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 2005;330:753.
63 Chan A-W, Hróbjartsson A, Haahr MT, *et al*. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457–65.
64 Chan A-W, Krleza-Jerić K, Schmid I, *et al*. Outcome reporting bias in randomized trials funded by the Canadian Institutes of health research. *CMAJ* 2004;171:735–40.
65 Dwan K, Altman DG, Clarke M, *et al*. Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. *PLoS Med* 2014;11:e1001666.
66 Breau RH, Gaboury I, Scales CD, *et al*. Reporting of harm in randomized controlled trials published in the urological literature. *J Urol* 2010;183:1693–7.
67 Carlesso LC, Macdermid JC, Santaguida LP. Standardization of adverse event terminology and reporting in orthopaedic physical therapy: application to the cervical spine. *J Orthop Sports Phys Ther* 2010;40:455–63.
68 de Vries TW, van Roon EN. Low quality of reporting adverse drug reactions in paediatric randomised controlled trials. *Arch Dis Child* 2010;95:1023–6.
69 Hochman M, McCormick D. Characteristics of published comparative effectiveness studies of medications. *JAMA* 2010;303:951–8.
70 Loke YK, Derry S. Reporting of adverse drug reactions in randomised controlled trials - a systematic survey. *BMC Clin Pharmacol* 2001;1:3.
71 Pitrou I, Boutron I, Ahmad N, *et al*. Reporting of safety results in published reports of randomized controlled trials. *Arch Intern Med* 2009;169:1756–61.
72 Pope A, Adams C, Paton C, *et al*. Assessment of adverse effects in clinical studies of antipsychotic medication: survey of methods used. *Br J Psychiatry* 2010;197:67–72.
73 Bonini S, Eichler H-G, Wathion N, *et al*. Transparency and the European Medicines Agency--sharing of clinical trial data. *N Engl J Med* 2014;371:2452–5.
74 Doshi P, Jefferson T. The first 2 years of the European medicines Agency's policy on access to documents: secret no longer. *JAMA Intern Med* 2013;173:380–2.
75 FDA. FDA commissioner Scott Gottlieb, M.D., on new steps FDA is taking to enhance transparency of clinical trial information to support innovation and scientific inquiry related to new drugs, 2018. Available: https://www.fda.gov/news-events/press-announcements/fda-commissioner-scott-gottlieb-md-new-steps-fda-taking-enhance-transparency-clinical-trial [Accessed 27 May 2020].
76 Parsons R, Golder S, Watt I. More than one-third of systematic reviews did not fully report the adverse events outcome. *J Clin Epidemiol* 2019;108:95–101.
77 Chalmers I. Proposal to outlaw the term "negative trial". *BMJ* 1985;290:1002.
78 Hutton JL, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. *J R Stat Soc Ser C Appl Stat* 2000;49:359–70.
79 Kirkham JJ, Dwan KM, Altman DG, *et al*. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365.
80 Williamson PR, Gamble C, Altman DG, *et al*. Outcome selection bias in meta-analysis. *Stat Methods Med Res* 2005;14:515–24.
81 Williamson PR, Gamble C. Identification and impact of outcome selection bias in meta-analysis. *Stat Med* 2005;24:1547–61.
82 Doshi P, Jefferson T. Clinical study reports of randomised controlled trials: an exploratory review of previously confidential industry reports. *BMJ Open* 2013;3:e002496.
83 Abraham J. International Conference On Harmonisation Of Technical Requirements For Registration Of Pharmaceuticals For Human Use. In: Tietje C, Brouder A, eds. *Handbook of transnational economic governance regimes*. Brill | Nijhoff, 2010: 1041–53.
84 CFR - Code of Federal Regulations Title 21, Section 312.32. IND safety reporting. Available: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=312.32 [Accessed 26 May 2020].
85 European Medicines Agency. Guideline on good pharmacovigilance practices (GVP) - Annex I - Definitions (Rev 4). Available: https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-good-pharmacovigilance-practices-annex-i-definitions-rev-4_en.pdf