MINI REVIEW

Biases in reporting of adverse effects in clinical trials, and potential impact on safety assessments in systematic reviews and therapy guidelines

Tone Westergren[1] | Sigrid Narum[2,3] | Marianne Klemp[4]

[1]Regional Medicines Information & Pharmacovigilance Centre (RELIS), Department of Pharmacology, Oslo University Hospital HF, Oslo, Norway

[2]Centre for Psychopharmacology, Diakonhjemmet Hospital, Oslo, Norway

[3]Drugs and Therapeutics Committee, Department of Pharmacology, Oslo University Hospital, Oslo, Norway

[4]Department of Pharmacology, Institute of Clinical Medicine, University of Oslo, Oslo, Norway

Correspondence
Tone Westergren, Drugs and Therapeutics Committee, Department of Pharmacology, Oslo University Hospital HF, PO Box 4950 Nydalen, 0424 Oslo, Norway.
Email: twesterg@ous-hf.no

Abstract

Background: Clinical trials are an important source of adverse effects data, including analyses in systematic reviews and recommendations in therapy guidelines. Trial publication bias may have profound effects on safety perceptions. This MiniReview presents and discusses biases in reporting of safety data in clinical trials and the implications for systematic reviews and guidelines.

Objectives: The objectives of this work are to analyse risk of gastrointestinal bleeding in systemic corticosteroid trials and to assess adverse effects reporting in a fluoxetine trial in depression (Treatment for Adolescents With Depression Study [TADS]) and descriptions of adverse effects in adolescent depression therapy guidelines.

Methods: We performed literature reviews and descriptive analyse of clinical trials with corticosteroids, and publications from the TADS trial. Risk of gastrointestinal bleeding from corticosteroids was analysed by meta-analysis.

Findings: Gastrointestinal bleeding definitions varied considerably between trials. The incidence was significantly increased in hospitalized, but not in ambulant, patients compared to placebo. We identified several biases concerning TADS safety reporting, including severity thresholds and nonpublication of most adverse effects data beyond the initial 12 weeks. Therapy guidelines on adolescent depression mentioned suicidality risk, but many failed to mention other adverse effects.

Conclusions: We identified several pitfalls in adverse effects reporting in clinical trials. These include heterogeneous disease definitions, reporting thresholds, and incomplete reporting. Trial bias may have great impact on risk assessments in systematic reviews and meta-analyses.

KEYWORDS
adverse drug reactions, antidepressants, corticosteroids, randomized controlled trials, systematic reviews, therapy guidelines

Parts of this MiniReview have been published previously in the PhD thesis of Dr. Tone Westergren.

1 | INTRODUCTION

Despite the popular belief that clinical trials will identify risks associated with new medications, flaws and biases in reporting adverse effects in clinical trial publications have been identified, in particular for antidepressant medications.[1–9] In several cases, reassessments of data have resulted in considerably increased risk estimates for serious adverse effects.[3,4,10]

In this MiniReview, we give an overview of biases and weaknesses in the reporting of adverse effects in clinical trials, exemplified by our own research. We discuss how trial protocols and performance may influence identification and reporting of adverse effects. Trial publications are a major source of adverse event data and are essential for systematic reviews on drug safety.[11,12] Biases in trial publications will be continued along the evidence chain and have an impact on risk–benefit beliefs. For systematic reviews and therapy guidelines, there are no recommendations on the extent of safety information.

1.1 | Identification and reporting of adverse effects in clinical trials

Randomized clinical trials (RCTs) are central to drug development, as randomization is assumed to offer a nonbiased comparison of treatment and control groups for both efficacy and risk. Consequently, RCTs have a high ranking in the hierarchy of evidence.[5] There are, however, several sources of bias throughout the complicated process of performing an RCT. These include study design, patient monitoring, data analysis, and manuscript preparation (Table 1).

1.1.1 | Planning and collection of data

Most clinical trials are performed to assess treatment efficacy. Incidence of adverse effects is generally not a primary outcome, and trials are not powered to detect adverse effects unless very common.[13,14] A priori definitions of adverse effects is unrealistic due to the large number of possible adverse reactions. Consequently, few adverse effects are predefined with regard to severity criteria, diagnosis, identification, and classification.[5,6,13] Procedures for safety monitoring may vary considerably between trials and include general or specific physical examinations, laboratory tests, questionnaires, patient interviews, checklists, or application of diverse scoring tools.[15] Safety findings will largely depend on procedures for identifying adverse effects.

Reporting will be subject to individual judgement, inaccuracies, and spontaneity if definitions are missing, probably with increased risk in multicentre, multinational trials with greater numbers of investigators. In addition, risks may be affected by dosages, treatment duration, and use of concomitant medications, such as use of additional antidepressants in antidepressant trials. Patient selection and exclusion criteria may reduce the risk of adverse effects in trial patients compared to real-life populations.[5,14]

In trials where adverse effects have been defined as a primary or secondary outcome, and randomization, blinding, patient monitoring, data collection, events, assessment methods, and analyses are described and fully reported, the risk of bias and errors will clearly be reduced.

Even for approved indications and patient groups, safety data may be surprisingly limited, as illustrated by selective serotonin reuptake inhibitors (SSRIs) for depression in children and adolescents. For the most extensively studied SSRI, fluoxetine, the pivotal trial is the "Treatment for Adolescents With Depression Study (TADS)" where 216 of 439 patients received fluoxetine.[16,17]

1.1.2 | Assessment and data registration

The chain of events, from a patient experience of an adverse effect until the event has been registered in the study files, offer opportunities for individual judgements. Basically, events must be acknowledged as possible adverse effects and not dismissed as chance findings or caused by the condition being treated (e.g., suicidality and depression). Reanalysis of primary, individual patient data from a paroxetine and imipramine trial caused new cases of serious adverse effects to be identified.[4] For antidepressants, new criteria and reassessment of suicidality cases resulted in identification of new cases and elimination of previously reported cases.[18] For the antidiabetic drug rosiglitazone, reanalysis of individual patient data, as opposed to summary level data, identified additional cases of myocardial infarction, and the drug was withdrawn several years after marketing.[10]

Events (symptoms, verbal descriptions, laboratory findings) must be translated into diagnostic codes, using complex medical vocabularies. Interpretations and judgements during this process may give rise to variations and misclassifications and have been found to vary between individuals.[19] Coding omissions, where not all events are coded, have been identified through analysis of individual patient data.[4] Choice of classification terms may lessen the apparent severity of an event, as exemplified

TABLE 1 Aspects of adverse effects reporting in assessed trials

Trial phase	Examples of variables	Corticosteroid trials	TADS
Planning	Patient group Inclusion criteria Exclusion criteria Study duration	Different diseases and all ages Different study durations Exclusion criterion ongoing or previous peptic ulcer applied in some studies Varying study duration	Multiple exclusion criteria, primarily psychiatric diseases Double blind 12 weeks Duration max. 88 weeks.
	Medication and dose Control group	Different corticosteroids and doses	Adjunctive treatment allowed
	Specified adverse effects All adverse effects	Not specified in many studies	Adverse event criteria threshold
Data collection	General or specific examinations Laboratory tests Questionnaires or interviews Doctor or patient reporting	Differences in monitoring for adverse effects	Interview setting with parents and assessments by investigators
	Criteria Severity thresholds	Heterogeneity in definitions and severity thresholds of gastrointestinal bleeding	Severity threshold Varying criteria for mania diagnosing
Assessments	Grouping of adverse effects Classification system Translation to medical codes	Different terminology	Ambiguous terminology for some symptoms Not considering symptoms from discontinuation
	Criteria for causality Assessment of causality	Not always described	Suicidality assessment described in study manual. Reanalysis of causality on suicidal events. Variations in assessment of mania
	Intention-to-treat Per protocol As treated	Zero events in several studies	ITT analysis possibly unsuitable due to supplemental therapy ITT underestimating true AE risk?
Publication	Most frequent adverse effects Most serious adverse effects Adverse effects in specific organs	Specific information missing in many studies	Several publications Focus on suicidality Not published all adverse effects for entire trial time
	Absolute numbers Relative numbers General statements	General statements in some studies Both absolute and relative numbers reported	Varying between adverse effects, publications and time periods. Suicidality: Absolute numbers, mean scores, score changes, proportion of patients over threshold values

by using the coding term "emotional lability" in cases of suicidal behaviour.[4] Trials of checkpoint inhibitors in cancer treatment used 24 different terms for the adverse effect of colitis, and trials of orlistat in weight reduction treatment used 11 different terms for describing diarrhoea.[20,21] Other examples include difficulties in classifying cases of self-harm for antidepressants.[19]

In some cases, adverse effects that originally were recorded quantitatively, for example, liver enzyme levels or symptom scores, are converted to dichotomous values (criteria fulfilled or not).[6,14] Consequently, the threshold value for qualifying for an adverse event will have a large impact on the number of cases.

1.1.3 | Publication

To improve the reporting of adverse effects in clinical trial publications, the CONSORT recommendations (Consolidated Standards of Reporting Trials) were extended with more detailed recommendations for harms reporting (CONSORT Harms) in 2004 in order to improve quality and reduce risk of bias and errors.[22] Despite the CONSORT Harms recommendations, deviations to complete safety reporting have been identified. Many trials are not published, or safety data are not presented in full.[1,4,7,23,24] Some papers merely state that no major adverse effects were observed, fail to mention

serious adverse effects, or limit reporting to serious or severe reactions, adverse effects with incidence above a threshold value, or statistically significant risk increases.[1,4,6,7,21,25]

Reporting of adverse effects data is customarily summed up by organ systems, with varying levels of subspecifications. The level of grouping adverse effects for analyses is often not specified a priori,[26] and each alternative carry an inherent risk of errors. Grouping by major terms, for example, "gastrointestinal disorders," will increase group size and statistical power. However, grouping may easily combine highly disparate events with different aetiologies, and the increase in numbers may mask cases of rare but significant adverse reactions. Analysis by single, specific diagnoses will provide more information of certain risks but will be subject to errors and differences in classifications, as in "emotional lability" versus "suicidality," and may render categories too narrow to identify relevant risks.[18]

1.2 | Adverse effects in systematic reviews

Systematic reviews aim to answer specific research questions through comprehensive analyses of relevant, high-quality trials.[27] Most reviews focus on efficacy, and many have a limited mention of risks, or fail to address treatment risk altogether.[28–31] Recommendations for harms reporting in systematic reviews, the PRISMA Harms checklist, were published in 2016,[32] but the impact is still unknown.

Systematic reviews of risk are at risk of bias for many reasons. In contrast to high quality efficacy data, risk assessments are generally not based on primary outcomes in multiple trials, but on poorer quality heterogeneous data on secondary outcomes across trials, or spontaneous reports of varying quality. RCTs generally do not have the size and statistical power to identify or draw conclusions on potential risks. Trials may differ in their design, monitoring, and descriptions of risk, thereby affecting the quality of risk assessments.[29,30,33] Systematic reviews on identical topics may reach different conclusions, as illustrated by the question of corticosteroid-induced gastrointestinal bleeding, which has been debated for decades. Some researchers have found the risk to be significantly increased,[34] while others have not.[35,36] This may be the reason why databases and handbooks describe the risk association either as unlikely or weak[37] or as increased.[38]

Systematic reviews on risk do not necessarily address all relevant adverse effects, as illustrated by

SSRI safety reviews. Due to warnings on suicidality risk,[39] several reviews and meta-analyses have focused on suicidality,[40,41] without addressing other adverse effects.

1.3 | Adverse effects in clinical guidelines

Clinical therapy guidelines are expected to assess therapy benefits and risks, based on systematic literature reviews.[12,42] There is no standard for risk descriptions of adverse effects of different treatment options in clinical guidelines and there is little research on the subject. The quality instrument for evaluating therapy guidelines, the AGREE II tool, does not comment on level of risk information.[42] Clinical guidelines are known to focus on benefits and be liable to biases, including conflicts of interest.[43] It is not known to what extent safety concerns and adverse effects information is described in guidelines on antidepressant therapy in children and adolescents.

In our research, we aimed to assess publications of clinical trials with regard to reporting of adverse effects. The evaluations were performed in two model areas: Risk of gastrointestinal bleeding or perforation in trials of systemic corticosteroid therapy and descriptions of adverse effects from fluoxetine in a pivotal trial (the TADS study) in children and adolescents with depression. We also aimed to assess clinical therapy guidelines on depression in children and adolescents for their information on adverse effects.

2 | METHODS

For assessment of adverse effects reporting in clinical trials, we performed literature searches for trials of corticosteroids and publications arising from the TADS antidepressant trial, as described elsewhere.[44,45] Corticosteroid trial publications were analysed with regard to inclusion and exclusion criteria, extent and indication for corticosteroid use, concomitant medications, ambulant or hospital treatment, and definitions or criteria for gastrointestinal bleeding. Risk of gastrointestinal bleeding or perforation was analysed through meta-analysis.[44] Fulfilment of criteria for harms reporting[22] were analysed quantitatively.[26]

TADS trial publications were analysed descriptively with regard to information about adverse effects.

Therapy guidelines on treatment of depression in children and adolescents were identified through literature searches in PubMed, EMBASE, guideline collections,

and manual searches, as described elsewhere.[46] Presentation of adverse effect data and risk profiles were analysed descriptively.

3 | RESULTS

3.1 | Descriptions of gastrointestinal bleeding or perforation in corticosteroid trials

To assess how clinical trials had addressed the question of gastrointestinal bleeding from corticosteroids, we analysed the reporting of this adverse reaction in 159 published clinical trials of corticosteroid therapy.[44] Most trials addressed treatment efficacy. Monitoring procedures, and definitions of gastrointestinal bleeding or perforation, varied considerably between trials, from general, unspecific examinations to blood tests and faecal analysis. Overall, 37 terms had been used to describe gastrointestinal bleeding. Criteria varied greatly with regard to disease severity, ranging from "guaiac-positive aspirate" to "hematemesis" or "melena requiring transfusion." The variations and potential biases identified in the corticosteroid trials are described in Table 1. Overall, we found an increased risk of gastrointestinal bleeding or perforation of 40% (OR 1.43, 95% CI 1.22 to 1.66) in patients treated with systemic corticosteroids. The risk increase was statistically significant for hospitalized, and presumably sicker, patients. Few cases were reported for ambulant patients.

Assessment of publications according to the CONSORT Harms criteria showed that many studies did not state an intention of identifying adverse effects or describe plans for risk presentation or analysis.[26] We found the criteria ambiguous and unsuitable for retrospective trial assessment for quality, as publications that do not fulfil all criteria might still present relevant safety data despite low assessment scores.

3.2 | Descriptions of adverse effects in the TADS trial

To assess how a single, pivotal antidepressant trial monitored patients and reported on safety results, we identified and analysed TADS trial publications. In the TADS trial, patients underwent several interviews and assessments through screening tools, which are described in the trial protocol.[47] These include Affective Disorder Screening (ADS), Clinical Global Impressions (CGI), and Children's Global Assessment Scale (CGAS), with guidelines for rating major depression, suicidality, mania, and

overall functioning, but do not provide definitions or specific monitoring of other potential adverse effects. In addition, patients and parents were asked about any new health problems at assessments, but the trial protocol specified that only adverse events that fulfilled threshold criteria would be recorded in an adverse event form. For adverse events that were recorded in an adverse events form, the therapist would indicate an opinion with regard to causality.

We identified 48 publications with data from the TADS trial,[45] of which eight presented adverse effects data to some extent. All eight publications described cases of suicidal behaviour. The TADS study was performed as a randomized, double-blinded, placebo-controlled trial for the first 12 weeks, but many patients continued on open, noncontrolled treatment for up to 88 weeks. Data on adverse effects other than suicidality were only published for the controlled, double-blinded 12-week trial phase I. The variations and potential biases identified in the TADS trial are described in Table 1. We did not identify any publications describing other psychiatric, or somatic, adverse effects occurring after the first 12 weeks. Comparison of adverse effects reported in the TADS trial and the list of fluoxetine adverse effects in the Summary of Product Characteristics (SmPC)[48] showed that many well-known risks were not mentioned in publications from the TADS trial.

3.3 | Descriptions of adverse effects in guidelines on antidepressants in children and adolescents

Nineteen clinical guidelines on treatment of depression in children and adolescents were assessed and found to vary considerably with regard to risk descriptions.[46] All guidelines mentioned risk of suicidality, and many described other psychiatric reactions, but most guidelines failed to describe somatic adverse effects.

4 | DISCUSSION

4.1 | Reporting of adverse effects in clinical trials

4.1.1 | Definitions of adverse effects

The lack of clear and uniform definitions of what constitutes an adverse effect, and the large variations in monitoring methods, will result in the summing up of different entities. The lack of common definitions and

monitoring methods for adverse effects is likely to be valid across all medical areas and is potentially a large source of bias.[5,13] As exemplified by the diagnosis of gastrointestinal bleeding, monitoring for occult faecal blood must be expected to identify more cases than counting patients needing blood transfusion.

4.1.2 | Limitations, thresholds, and filters

The use of reporting thresholds will have a profound effect on the number of cases registered as an adverse effect. The TADS trial illustrates this point, as the trial protocol describes thresholds and limitations on adverse effects reporting that must have influenced results. A major threshold that has received little attention in the literature citing the trial was the fact that adverse events would not be recorded unless they caused a clinically significant interference with functioning, required medical attention, or caused a need to take medication.[17]

4.1.3 | Salami publications

The 48 publications from the TADS trial illustrate the problem of salami publications,[49] that is, publications of several papers from a single trial. The sheer number of publications may easily give readers the impression of extensive research, while in reality the published adverse effects data cover only 216 patients treated with fluoxetine for 12 weeks.

4.1.4 | Statistical analysis of adverse outcomes

Choice of statistical methodology may influence risk estimates in clinical trials. In many trials, reporting is descriptive, giving the number of cases in treatment and control groups (incidence rate). For statistical analyses, recommendations state that adverse effects should be analysed as intention-to-treat (ITT).[22] Potential biases, and erroneous risk estimates inherent to ITT analyses of adverse effect data, have been little discussed. If many patients leave the study early, or receive additional treatment that differ from their assigned group, ITT analysis will tend to reduce group differences.[14] The adverse effects incidence will be underestimated if the denominator includes nonadherent patients, as was the case in the TADS trial. Conversely, risk estimates patients will exclude events in patients who withdrew from treatment due to adverse events.

4.1.5 | What is missing in harms reporting?

As described in Table 1, risks of bias may arise from a number of causes during performance and publication of clinical trials. For adverse effects, these include nonpublication, skewed presentations, and lack of statistical strength. Bias may be expected if study design is suboptimal, patient monitoring not clearly defined, outcomes diffuse, and reporting selective. For readers, absence of relevant data may be difficult to notice unless an article is subjected to close scrutiny. Publication of safety data according to the CONSORT Harms recommendations[22] will potentially improve descriptions of trial performance with regard to any specific adverse reactions being addressed, monitoring methods, disease definitions, data collection, withdrawals, and risk analysis. It will not, however, reduce bias due to heterogeneity in patient groups, disease definitions, or monitoring methods.

Adverse effects are usually presented as number of cases, or percent incidence, for each adverse effect, by organ systems. There are few descriptions of event severity,[5] latency and duration of adverse effects,[5,21] and overall impact on patients. Obviously, a light and passing headache will differ from a debilitating, enduring headache in terms of acceptability. Furthermore, the present descriptive methods do not assess benefit and harms together for each patient, and current methods do not assess skewedness in distribution of benefits and risks within the patient groups.

There is currently great interest in the possibilities of in-depth analysis in individual patient data from trials, but issues regarding data availability and methodology are still pending.

4.2 | Adverse effects in systematic reviews

Systematic reviews will include heterogeneous data, given the differences in included trials. A systematic review addressing adverse effects of a specific medication may appear to be conclusive, but biases and limitations in the underlying data may not be taken sufficiently into account by researchers and readers. There are no clear recommendations as to when heterogeneity is large enough to preclude performance of systematic reviews.[5] The Cochrane Handbook for performance of systematic reviews discusses several sources of heterogeneity and states that "review authors must recognize the possibility of poor case definition, inadequate monitoring and incomplete reporting when synthesizing data."[50] This raises the question of whether systematic reviews can be more misleading than valid. Heterogeneity may be the

reason for different conclusions in systematic reviews. In the case of corticosteroid-induced gastrointestinal bleeding, systematic reviews, including our own, have included different trial publications, used different inclusion criteria and limitations, and included trials with different monitoring and definitions of gastrointestinal bleeding. Even if review authors describe selection criteria and biases, the take-home message from a systematic review will probably be limited to the main findings for most readers.

4.3 | Descriptions of adverse effects in therapy guidelines

Our assessment of therapy guidelines on depression in children and adolescents showed that many guidelines mentioned only selected adverse effects and failed to mention several common risks. Analysis of therapy guidelines is challenging, partly due to identification and collection difficulties. Any professional body or organization is free to develop guidelines, and they are in many cases published locally, outside peer-reviewed journals. The extent of adverse effects descriptions in therapy guidelines has been little studied and is, at present, to be decided by guideline authors, with no clear expectations from society or readers as to the level of information.

5 | CONCLUSION

Our findings confirm the results of other researchers and show considerable pitfalls in identification and reporting of adverse effects in clinical trials. Some biases are caused by methods choices including study design and patient monitoring, and others are caused by use of thresholds, filters, and selective reporting. In corticosteroid trials, the main issue was the highly heterogeneous definitions and diagnostic criteria for gastrointestinal bleeding, as well as heterogeneous patient groups with regard to underlying disease severity, risk factors, and concomitant medications. In the TADS trial, the main issue was the reporting threshold with regard to event severity, and failure to publish adverse effects data beyond the initial 12 weeks of treatment for other adverse effects than suicidality. Trial publication biases have potentially great implications for attempts to assess adverse effects risk in systematic reviews and meta-analyses based on trial publications.

ORCID
Tone Westergren https://orcid.org/0000-0002-4990-062X

REFERENCES

1. Ioannidis JP. Adverse events in randomized trials: neglected, restricted, distorted, and silenced. *Arch Intern Med.* 2009; 169(19):1737-1739. doi:10.1001/archinternmed.2009.313
2. Hughes S, Cohen D, Jaggi R. Differences in reporting serious adverse events in industry sponsored clinical trial registries and journal articles on antidepressant and antipsychotic drugs: a cross-sectional study. *BMJ Open.* 2014;4(7):e005535. doi:10.1136/bmjopen-2014-005535
3. Hogberg G, Antonuccio DO, Healy D. Suicidal risk from TADS study was higher than it first appeared. *Int J Risk Saf Med.* 2015;27(2):85-91. doi:10.3233/jrs-150645
4. Le Noury J, Nardo JM, Healy D, et al. Restoring study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. *BMJ.* 2015;351:h4320. doi:10.1136/bmj.h4320
5. Hammad TA, Pinheiro SP, Neyarapally GA. Secondary use of randomized controlled trials to evaluate drug safety: a review of methodological considerations. *Clin Trials.* 2011;8(5): 559-570. doi:10.1177/1740774511419165
6. Phillips R, Hazell L, Sauzet O, Cornelius V. Analysis and reporting of adverse events in randomised controlled trials: a review. *BMJ Open.* 2019;9(2):e024537. doi:10.1136/bmjopen-2018-024537
7. de Vries YA, Roest AM, Beijers L, Turner EH, de Jonge P. Bias in the reporting of harms in clinical trials of second-generation antidepressants for depression and anxiety: a meta-analysis. *Eur Neuropsychopharmacol.* 2016;26(11):1752-1759. doi:10.1016/j.euroneuro.2016.09.370
8. Hengartner MP. Methodological flaws, conflicts of interest, and scientific fallacies: implications for the evaluation of antidepressants' efficacy and harm. *Front Psych.* 2017;8:275. doi:10.3389/fpsyt.2017.00275
9. Hengartner MP, Ploderl M. Newer-generation antidepressants and suicide risk in randomized controlled trials: a re-analysis of the FDA database. *Psychother Psychosom.* 2019;88(4): 247-248. doi:10.1159/000501215
10. Wallach JD, Wang K, Zhang AD, et al. Updating insights into rosiglitazone and cardiovascular risk through shared data: individual patient and summary level meta-analyses. *BMJ.* 2020;368:I7078. doi:10.1136/bmj.l7078
11. *The Oxford Levels of Evidence v.2.1.* Oxford Centre for Evidence-Based Medicine, University of Oxford; 2011. Accessed 30.12.2020. https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence
12. Graham R, Mancher M, Wolman DM, Greenfield S, Steinberg E (Eds). *Clinical Practice Guidelines We Can Trust.* Institute of Medicine (US); 2011. Accessed April 19, 2021. https://www.ncbi.nlm.nih.gov/books/NBK209539/
13. Singh S, Loke YK. Drug safety assessment in clinical trials: methodological challenges and opportunities. *Trials.* 2012; 13(1):138. doi:10.1186/1745-6215-13-138

14. CIOMS Working Group VI. *Management of Safety Information From Clinical Trials*. Council for International Organizations of Medical Sciences (CIOMS); 2005. Accessed 14.01.2019. https://cioms.ch/wp-content/uploads/2017/01/Mgment_Safety_Info.pdf

15. Allen EN, Chandler CI, Mandimika N, Leisegang C, Barnes K. Eliciting adverse effects data from participants in clinical trials. *Cochrane Database Syst Rev*. 2018;2018(1):MR000039. doi:10.1002/14651858.MR000039.pub2

16. Walkup JT. Antidepressant efficacy for depression in children and adolescents: industry- and NIMH-funded studies. *Am J Psychiatry*. 2017;174(5):430-437. doi:10.1176/appi.ajp.2017.16091059

17. March JS, Silva S, Petrycki S, et al. Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for Adolescents With Depression Study (TADS) randomized controlled trial. *Jama*. 2004;292(7):807-820. doi:10.1001/jama.292.7.807

18. Posner K, Oquendo MA, Gould M, Stanley B, Davies M. Columbia Classification Algorithm of Suicide Assessment (C-CASA): classification of suicidal events in the FDA's pediatric suicidal risk analysis of antidepressants. *Am J Psychiatry*. 2007;164(7):1035-1043. doi:10.1176/ajp.2007.164.7.1035

19. Schroll JB, Maund E, Gotzsche PC. Challenges in coding adverse events in clinical trials: a systematic review. *PLoS ONE*. 2012;7(7):e41174. doi:10.1371/journal.pone.0041174

20. Yang A, Baxi S, Korenstein D. ClinicalTrials.gov for facilitating rapid understanding of potential harms of new drugs: the case of checkpoint inhibitors. *J Oncol Pract*. 2018;14(2):72-76. doi:10.1200/jop.2017.025114

21. Schroll JB, Penninga EI, Gotzsche PC. Assessment of adverse events in protocols, clinical study reports, and published papers of trials of orlistat: a document analysis. *PLoS Med*. 2016;13(8):e1002101. doi:10.1371/journal.pmed.1002101

22. Ioannidis JP, Evans SJ, Gøtzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med*. 2004;141(10):781-788. doi:10.7326/0003-4819-141-10-200411160-00009

23. Wang SM, Han C, Lee SJ, et al. Efficacy of antidepressants: bias in randomized clinical trials and related issues. *Expert Rev Clin Pharmacol*. 2018;11(1):15-25. doi:10.1080/17512433.2017.1377070

24. Hodkinson A, Kirkham JJ, Tudur-Smith C, Gamble C. Reporting of harms data in RCTs: a systematic review of empirical assessments against the CONSORT harms extension. *BMJ Open*. 2013;3(9):e003436. doi:10.1136/bmjopen-2013-003436

25. Mayo-Wilson E, Fusco N, Hong H, Li T, Canner JK, Dickersin K. Opportunities for selective reporting of harms in randomized clinical trials: selection criteria for non-systematic adverse events. *Trials*. 2019;20(1):553. doi:10.1186/s13063-019-3581-3

26. Westergren T, Narum S, Klemp M. Characterization of gastro-intestinal adverse effects reported in clinical studies of cortico-steroid therapy. *J Clin Epidemiol*. 2018;94:19-26. doi:10.1016/j.jclinepi.2017.10.018

27. Higgins J, Thomas J, Chandler J, et al. (Eds). Chapter I: Introduction. In: *Cochrane Handbook for Systematic Reviews of Interventions*; Version 6.3, 2022. Accessed April 13, 2022. https://training.cochrane.org/handbook/current/chapter-i

28. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). *J Clin Epidemiol*. 2011;64(4):407-415. doi:10.1016/j.jclinepi.2010.07.017

29. Zorzela L, Golder S, Liu Y, et al. Quality of reporting in systematic reviews of adverse events: systematic review. *BMJ*. 2014;348:f7668. doi:10.1136/bmj.f7668

30. Saini P, Loke YK, Gamble C, Altman DG, Williamson PR, Kirkham JJ. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ*. 2014;349(nov21 3):g6501. doi:10.1136/bmj.g6501

31. Yavchitz A, Ravaud P, Altman DG, et al. A new classification of spin in systematic reviews and meta-analyses was developed and ranked according to the severity. *J Clin Epidemiol*. 2016;75:56-65. doi:10.1016/j.jclinepi.2016.01.020

32. Zorzela L, Loke YK, Ioannidis JP, et al. PRISMA harms checklist: improving harms reporting in systematic reviews. *BMJ*. 2016;352:i157. doi:10.1136/bmj.i157

33. Hammad TA, Neyarapally GA, Pinheiro SP, Iyasu S, Rochester G, Dal Pan G. Reporting of meta-analyses of randomized controlled trials with a focus on drug safety: an empirical assessment. *Clin Trials*. 2013;10(3):389-397. doi:10.1177/1740774513479467

34. Messer J, Reitman D, Sacks HS, Smith H Jr, Chalmers TC. Association of adrenocorticosteroid therapy and peptic-ulcer disease. *N Engl J Med*. 1983;309(1):21-24. doi:10.1056/nejm198307073090105

35. Conn HO, Blitzer BL. Nonassociation of adrenocorticosteroid therapy and peptic ulcer. *N Engl J Med*. 1976;294(9):473-479. doi:10.1056/nejm197602262940905

36. Conn HO, Poynard T. Corticosteroids and peptic ulcer: meta-analysis of adverse events during steroid therapy. *J Intern Med*. 1994;236(6):619-632. doi:10.1111/j.1365-2796.1994.tb00855.x

37. Corticosteroids, general use. 2019. British National Formulary. Accessed June 10, 2020. https://www.medicinescomplete.com/#/content/bnf/_111415959?hspl=corticosteroids#content%2Fbnf%2F_111415959%23section_111415959-1

38. Major side effects of systemic glucocorticoids. 2020. UpToDate. Accessed May 20, 2020. https://www.uptodate.com/contents/major-side-effects-of-systemic-glucocorticoids/print?search=glucocorticoids%26source=search_result%26selectedTitle=1%7E150%26usage_type=default%26display_rank=1

39. European Medicines Agency finalises review of antidepressants in children and adolescents. 2005. European Medicines Agency. Accessed December 30, 2020. https://www.ema.europa.eu/en/news/european-medicines-agency-finalises-review-antidepressants-childrenand-adolescents

40. Hammad TA, Laughren T, Racoosin J. Suicidality in pediatric patients treated with antidepressant drugs. *Arch Gen Psychiatry*. 2006;63(3):332-339. doi:10.1001/archpsyc.63.3.332

41. Bridge JA, Iyengar S, Salary CB, et al. Clinical response and risk for reported suicidal ideation and suicide attempts in pediatric antidepressant treatment: a meta-analysis of randomized controlled trials. *Jama*. 2007;297(15):1683-1696. doi:10.1001/jama.297.15.1683

42. The AGREE II Instrument. The AGREE Next Steps Consortium. 2017. Accessed August 15, 2019. http://www.agreetrust.org

43. Khan R, Scaffidi MA, Rumman A, Grindal AW, Plener IS, Grover SC. Prevalence of financial conflicts of interest among authors of clinical guidelines related to high-revenue medications. *JAMA Intern Med*. 2018;178(12):1712-1715. doi:10.1001/jamainternmed.2018.5106

44. Narum S, Westergren T, Klemp M. Corticosteroids and risk of gastrointestinal bleeding: a systematic review and meta-analysis. *BMJ Open*. 2014;4(5):e004587. doi:10.1136/bmjopen-2013-004587

45. Westergren T, Narum S, Klemp M. Critical appraisal of adverse effects reporting in the 'Treatment for Adolescents With Depression Study (TADS)'. *BMJ Open*. 2019;9(3): e026089. doi:10.1136/bmjopen-2018-026089

46. Westergren T, Narum S, Klemp M. Adverse effects information in clinical guidelines on pharmacological treatment of depression in children and adolescents: a systematic review. *BMJ Open*. 2020;10(7):e036412. doi:10.1136/bmjopen-2019-036412

47. Treatment for adolescents with depression study (TADS) Clinician Assessment Manual. The TADS Team Duke University Medical Center. 2000. Accessed November 3, 2021. http://tads.dcri.org/tads-manuals

48. Prozac. Summary of Product Characteristics. 2018. Eli Lilly and Company Ltd. Accessed May 31, 2018. https://www.medicines.org.uk/emc/product/3768/smpc#PRODUCTINFO

49. Ebrahim S, Montoya L, Kamal el Din M, et al. Randomized trials are frequently fragmented in multiple secondary publications. *J Clin Epidemiol*. 2016;79:130-139. doi:10.1016/j.jclinepi.2016.05.016

50. Higgins J, Thomas J. Chapter 19: adverse effects. In: Chandler J, Cumpston M, Li T, Page MJ, Welch VA, eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 6.3*; 2022 Accessed April 13, 2022. https://training.cochrane.org/handbook/current/chapter-19