Reporting of adverse effects in clinical trials, systematic reviews, and guidelines

How events are lost along the evidence chain

Tone Westergren

Regional Medicines Information and Pharmacovigilance Centre South-East (RELIS Sør-Øst), Department of Pharmacology, Oslo University Hospital

Department of Pharmacology, Institute of Clinical Medicine, Faculty of Medicine, University of Oslo

UiO : Faculty of Medicine
University of Oslo

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo

ISBN 978-82-8377-926-4

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

CONTENT

CONTENT 3
Acknowledgements 5
List of Papers 7
Sammendrag 8
Abbreviations 10
Summary 11
1. Introduction 13
1.1 Background 13
1.2 Adverse effects 14
1.2.1 History 14
1.2.2 Terminology 15
1.3 Adverse effects reporting in randomized clinical trials 17
1.3.1 Gastrointestinal bleeding risk in publications of corticosteroid trials 22
1.3.2 Adverse effects in publications of antidepressant trials in children and adolescents 22
1.4 Information on adverse effects in systematic reviews 23
1.4.1 Systematic reviews on corticosteroid-induced gastrointestinal bleeding 24
1.5 Information on adverse effects in clinical practice guidelines 25
2. Aim of the thesis 27
3. Materials and methods 28
3.1 Methods of data extraction and evaluation 28
3.1.1 Study designs 28
3.1.2 Literature search 28
3.1.3 Data extraction and assessment 29
3.2 Statistical methods 30
3.3 Ethical considerations and approvals 31
4. Summary of results 32
5. Discussion 35
5.1 Discussion of the main findings 35

5.1.1 Aim 1. Risk of gastrointestinal bleeding or perforation due to systemic corticosteroid treatment (Paper I). 35
5.1.2 Aim 2. Descriptions of gastrointestinal bleeding or perforation in corticosteroid trials, and assessment of the CONSORT Harms checklist as a quality tool (Paper II). 38
5.1.3 Aim 3. Identification of TADS trial publications, and assessment of adverse effects reporting (Paper III). 39
5.1.4 Aim 4. Adverse effects information in therapy guidelines on treatment of depression in children and adolescents (Paper IV). 43
5.1.5 Impact of the findings 44
5.2 Methodological considerations 46
5.2.1 Analysis of gastrointestinal bleeding risk associated with corticosteroid use (Paper I) 46
5.2.2 Analysis of descriptions of gastrointestinal bleeding or perforation in corticosteroid trials, and assessment of the CONSORT Harms checklist as a quality tool (Paper II) 48
5.2.3 Identification of TADS trial publications and assessment of adverse effects reporting (Paper III). 49
5.2.4 Analysis of clinical therapy guidelines on depression in children and adolescents (Paper IV) 49
6. Conclusions 51
7. Implications and future research 52
8. References 54
9. Errata 75
10.Papers I-IV 77

Acknowledgements

The work presented in this thesis has been inspired by, and carried out at, the Regional Medicines Information and Pharmacovigilance Centre in South-East Norway (RELIS Sør-Øst). Over the years, we have received numerous questions about the risk of adverse effects of medications, and have observed repeatedly that interpretation and assessment can be difficult. This fuelled my interest in the question of how adverse effects are identified and reported. I am deeply grateful for the opportunity to engage in this project.

I would like to extend great thanks to my main supervisor, Marianne Klemp, and my co-supervisor Sigrid Narum, for their enthusiasm and encouragement over several years. We have examined numerous study reports together and share many of the same concerns and interests. You have always been available for discussions and advice. Our meetings have been a source of inspiration, and given me new ideas and new perspectives.

I would also like to express my gratitude to all my present and former colleagues at RELIS Sør-Øst. Deep thanks to Ingunn Marie Randulff Nielsen, Anne Katrine Eek, Hanne Stenberg-Nilsen, Hilde Nordén, Cecilie Sogn Nergård, Randi Myhr, Helle Tjønn Lindland, Hilde Fjeld, Gro Cecilie Havnen, Karen Astrid Boldingh Debernard, Ellen Hilde Julsrud, Henning Gustav Økland, Karin Mi Ran Holmberg Dahle, Ellen Marie Egebæk Flatebø, Renate Brenden, and Mai-Linh Huynh Do, for your inspiration and continuing support. The ideas that formed the basis of this project were generated through our team discussions and reflections. You have made it possible for me to take time off for research and writing, and you have shouldered the additional workload in my periods of absence.

Thank you to my present and former colleagues in the RELIS leader group; Jan Schjøtt, Harald Langaas, Terje Nilsen and Trude Giverhaug, for your support.

My leader at the Department of Pharmacology, Head of Department Berit Muan, made it possible for me to take on a doctoral thesis, by allowing me to engage in research and to adjust my other duties accordingly. You have given me the trust and scope to take on this research project, for which I am deeply grateful and extend my heartfelt thanks.

At the Department of Pharmacology, I would also like to thank advisor Marianne Spalder Larsen for her ready support. You have eased my administrative duties by your insight and rapid problem solving.

My family has been a source of love and support throughout the project. Your immediate response was one of great enthusiasm, which warmed my heart. My deepest thanks go to my husband Torkild, our children Katrine and Håkon, and partners James and Anette, for your backing, understanding, and encouragement.

List of Papers

Paper I

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

Paper II

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

Paper III

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

Paper IV

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

Sammendrag

Bakgrunn

Kunnskap om et legemiddels bivirkningsprofil er viktig for å kunne vurdere nytten av behandling i forhold til forventet risiko. Kliniske utprøvninger er en grunnleggende metode for å få informasjon om nytte og bivirkninger av et legemiddel, og resultater fra kliniske studier ligger til grunn for kunnskapsoppsummeringer og behandlingsretningslinjer. Flere studier har påvist mangelfull rapportering av bivirkninger når resultater fra kliniske studier publiseres i tidsskrifter. Det er økende fokus på svakheter i bivirkningsrapportering og på skjevrapportering ved forskning og publisering av studieresultater.

Kunnskapsoppsummeringer basert på publiserte data fra kliniske studier har i flere tilfeller ført til motstridende konklusjoner og anbefalinger. En årsak til dette kan være metodologiske begrensninger i det underliggende materialet.

Det har ikke vært undersøkt tidligere hvordan bivirkningsdata fra kliniske studier gjengis og beskrives i behandlingsretningslinjer.

I denne avhandlingen har jeg undersøkt forskjellige aspekter ved identifikasjon og beskrivelse av bivirkninger i kliniske studier, fra studiedesign og metode, til publisering og gjengivelse av bivirkningsdata i systematiske oversikter og behandlingsretningslinjer.

Hensikt

Hensikten med avhandlingen var å kvantifisere risiko for gastrointestinal blødning og perforasjon ved systemisk behandling med kortikosteroider, og å undersøke omfang og kvalitet av bivirkningsrapportering i kliniske studier, med utgangspunkt i to eksempler: Risiko for gastrointestinal blødning eller perforasjon i en gruppe kliniske studier av systemiske kortikosteroider og beskrivelse av bivirkningsprofilen for det antidepressive legemidlet fluoksetin hos barn og ungdom i en enkelt, sentral studie ("TADS-studien").

Vi ønsket også å kvantifisere risikoen for kortikosteroid-indusert gastrointestinal blødning og å undersøke hvordan bivirkninger beskrives i retningslinjer for behandling av depresjon hos barn og ungdom.

Metoder

Publikasjoner og retningslinjer ble identifisert gjennom litteratursøk.

Kliniske studier av bruk av systemiske kortikosteroider ble analysert kvantitativt, ved hjelp av metaanalyse av risiko for gastrointestinal blødning eller perforasjon. Beskrivelser av gastrointestinale blødninger og oppfyllelse av kvalitetskriterier for bivirkningsrapportering ved publikasjon av kliniske studier med kortikosteroider, beskrivelse av bivirkninger i publikasjoner fra TADS-studien av fluoksetin til barn og ungdom og beskrivelser av bivirkninger i retningslinjer for behandling av depresjon hos barn og ungdom, ble analysert deskriptivt.

Resultater og konklusjon

Vi fant at behandling med kortikosteroider var forbundet med økt risiko for gastrointestinal blødning eller perforasjon. Risikoøkningen var statistisk signifikant for pasienter innlagt på sykehus, men ikke for ambulante pasienter.

Analyse av publikasjoner om kliniske utprøvninger av kortikosteroider viste at det var store variasjoner i hvordan gastrointestinal blødning eller perforasjon var definert og kartlagt. Kvalitetskriterier for presentasjon av bivirkningsdata ved publisering av kliniske studier er ikke laget for, eller egnet til, å kvalitetsvurdere artikler og mange av kvalitetskriteriene er ikke entydige.

Analyse av bivirkningsinformasjon i publikasjoner fra TADS-studien viste ufullstendig rapportering av bivirkninger utover de første 12 ukene av studien og en betydelig risiko for skjevhet i kartleggingen av bivirkninger.

Vi fant en stor variasjon i omfang og presentasjon av bivirkningsinformasjon i retningslinjer for behandling av depresjon hos barn og ungdom. Alle retningslinjene nevnte selvmordsrisiko, men de fleste retningslinjene omtalte ikke risiko for somatiske bivirkninger.

De påviste begrensningene ved publikasjon av bivirkningsdata fra kliniske utprøvninger har potensielt stor betydning for utarbeidelse av systematiske oversikter og behandlingsretningslinjer når det gjelder informasjon om bivirkninger.

Abbreviations

AGREE II	Appraisal of Guidelines for Research & Evaluation II
Cochrane	The Cochrane Collaboration
CONSORT	Consolidated Standards of Reporting Trials
EMA	European Medicines Agency
FDA	U.S. Food and Drug Administration
GI	Gastrointestinal
GRADE	Grading of Recommendations Assessments, Development, and Evaluation
ITT	Intention-to-treat
MedDRA	Medical Dictionary for Regulatory Activities
NIMH	National Institute of Mental Health
NSAIDs	Non-steroidal anti-inflammatory drugs
OR	Odds ratio
RCT	Randomized controlled trial
RELIS	Regional Medicines Information Centres
RR	Relative risk (Risk ratio)
SmPC	Summary of Product Characteristics
SSRI	Selective serotonin reuptake inhibitors
TADS	Treatment of Adolescents with Depression Study
WHO	World Health Organization

Summary

Background

Knowledge about adverse effects of a medication is essential to assess treatment benefits versus risks. Clinical trials are a major source of data for efficacy and safety assessments, systematic reviews, and treatment guidelines. Several researchers have identified shortcomings in reporting of adverse effects in journal publications of clinical trials, and there is increasing focus on research- and publication bias. Use of published adverse effects data from clinical trials in systematic reviews have resulted in conflicting risk conclusions on identical topics, possibly due to methodological limitations. It is not known how adverse effects data from clinical trials are reproduced and described in clinical therapy guidelines.

In this thesis, I address aspects of identification and dissemination of adverse effects data, from study design and methods of clinical trials of pharmacological therapies, through publication, to presentation of the data in systematic reviews and guidelines.

Aims

The aim was to quantify the risk of corticosteroid-induced gastrointestinal bleeding, and to examine the extent and quality of adverse effects reporting in publications of clinical trials, as exemplified in two model areas; the risk of gastrointestinal bleeding or perforation in a group of clinical trials of corticosteroids, and the adverse effect profile of the SSRI antidepressant fluoxetine in a single trial in children and adolescents (the TADS study). We also aimed to assess whether quality criteria for publication of harms are fulfilled, and how adverse effects are cited and presented in clinical therapy guidelines on treatment of depression in children and adolescents.

Methods

Papers and guidelines were identified through literature searches. Quantitative analyses were performed for the corticosteroid trials, by meta-analysis of gastrointestinal bleeding risk. Descriptive analyses were performed with regard to reports of gastrointestinal bleeding in the corticosteroid trials and quality criteria fulfilment for their adverse effects reporting, extent of publication of adverse effects data from a trial of fluoxetine in children and

adolescents, and descriptions of adverse effects in therapy guidelines for treatment of depression in children and adolescents.

Results and conclusions

Use of corticosteroids was associated with increased risk of gastrointestinal bleeding. The risk increase was statistically significant for hospitalized patients only.

Analysis of the corticosteroid trial publications showed considerable variations in definitions and monitoring for gastrointestinal bleeding or perforation. We also found that the current quality criteria for reporting of harms in trial publications are ambiguous and not well suited for quality assessments of adverse effects reporting in published studies.

We found incomplete reporting of adverse effects from the TADS trial of fluoxetine beyond 12 weeks of treatment, and considerable risk of bias in the published data.

We also found considerable variation in the extent of adverse effects information in clinical guidelines on depression in children and adolescents. All guidelines mentioned risk of suicidality, but somatic adverse effects were not described in most treatment guidelines.

The observed shortcomings in publication of adverse effects data from clinical trials have considerable implications for the performance of systematic reviews and development of treatment guidelines with regard to safety information.

1. Introduction

1.1 Background

Medications are exogenous substances introduced into the body with the intention to cure or alleviate illness by actions upon therapeutic targets in the body. The risk of adverse effects is inherently linked to pharmacological treatment of disease, as pharmacological actions of medications are often less specific than could be wished. Almost all medications have been shown to cause a wide range of adverse effects, with unintended and unwanted consequences for health. Knowledge of a medication's adverse effects profile, through reliable study assessments and comprehensive safety reporting, is essential when assessing expected benefit and potential safety issues, especially if treatments differ little with regard to benefits (1).

At the time of marketing of a new product, the adverse effects profile is not fully known, due to the limited patient data from clinical trials. In addition, reporting of adverse effects in clinical trials has been found to be suboptimal in study publications (1-8). Data on risk will accumulate during a product's life cycle, increasing the numbers of adverse effects in product monographs over time (9-11), changing risk-benefit assessments, and causing market withdrawals (12-14). In the last two decades, we have seen a growing awareness of research bias associated with study design, conflict of interest, publication issues, and the consequences with regard to risk and benefit assessments (15, 16). A major discussion in later years has concerned the adverse effects profile of antidepressants, where several researchers have identified shortcomings in published adverse effects data (4, 17-19). The discussions have focused on suicidality risk in children and adolescents, with references to regulatory warnings from European and American authorities (20, 21). Other, and probably more common adverse effects, has received less attention.

Systematic reviews, and the treatment guidelines that build on them, are largely based on data from randomized controlled clinical trials (22, 23). Consequently, shortcomings in adverse effects descriptions in publications of clinical trials have a potentially large impact on presentation of risks in systematic reviews and clinical guidelines. Systematic reviews on

the same topic may reach different conclusions, as is the case for previous reviews on gastrointestinal bleeding risk due to corticosteroid treatment.

Presentation of risks in clinical therapy guidelines has been little studied, and it is not known to what extent they reflect existing knowledge of adverse effect profiles. The adverse effects information in therapeutic guidelines for antidepressant treatment in children and adolescents has never been assessed, despite the potential impact on treatment and the increasing use of antidepressants in this age group (24).

1.2 Adverse effects

1.2.1 History

Developments in medication safety surveillance have largely been triggered by severe injuries or disasters. Reports on toxic reactions to treatment were published in medical journals as early as 1814, in connection with arsenic and mercury treatments (25). Then, as now, interpretation of events and questions of causality were key issues. Events that later were acknowledged as adverse effects were initially attributed to the underlying illness or the healing process. In 1848, a death caused by recently introduced chloroform anaesthesia led to widespread concern and an enquiry initiated by The Lancet journal. The findings, published in 1893, were based on methods that closely resemble present standards of adverse effects assessments (26). Legal safety regulations by medicinal authorities were not established until much later. In 1938, the United States Congress passed legislation requiring that new drugs must be shown to be safe before marketing (27). This was triggered by a scandal in which 107 patients died, due to use of the toxic solvent diethylene glycol in a sulphanilamide elixir.

The first published randomised clinical trial in 1948 (28), examining the effect of streptomycin in pulmonary tuberculosis (29) included observations on toxic reactions. In 1963, Bradford Hill stated that controlled trials might rapidly identify side-effects of a treatment, while acknowledging that trials would not be likely to identify rare effects (30).

The relationship between exposure to a pharmacological agent and risk of adverse effects began to be discussed extensively in the 1950s and 1960s, in relation to the increasing number of pharmaceutical products coming onto the market (31, 32). The question of

thalidomide teratogenicity was first raised in a letter to The Lancet in 1961 (33). The ensuing disaster, in which thousands of infants were born with deformed limbs, led to improvements in safety studies, strengthened safety legislation (34), establishment of the UK Yellow card scheme for spontaneous reporting of adverse effects (1964) (35) and the WHO Programme for International Drug Monitoring (1968) (36). In Norway, the national adverse effects committee was established in 1970 (37).

At the present time, extensive national and international regulations and safety surveillance systems are in place. This has reduced, but not eliminated, unacceptable safety issues. Within the last two decades, examples of market withdrawals after considerable patient exposures include the COX-2-inhibitor rofecoxib and the antidiabetic rosiglitazone, both due to cardiovascular safety issues. Rofecoxib was withdrawn in 2004 after several years on the market and probably thousands of patients with severe drug-induced cardiovascular disease (38, 39). Rosiglitazone was withdrawn in Europe in 2010 (40, 41).

1.2.2 Terminology

There is no uniform terminology to describe adverse effects of medications. In the literature, terms such as adverse effects, adverse reactions, adverse drug reactions, side effects, adverse events, drug toxicity, adverse outcomes, undesirable effect, risks or treatment harms are used. Definitions may vary, and sometimes include medication errors or events that probably are unrelated to medication use (42).

The European Medicines Agency (EMA) defines an **adverse reaction** (synonymous with adverse drug reaction and adverse effect) as a response to a medicinal product which is noxious and unintended, and have a suspected causal relationship between a medicinal product and an occurrence, including occupational exposure, off-label use, misuse, abuse and medication errors (43). The World Health Organization (WHO) defines an **adverse effect** as a negative or harmful patient outcome that seems to be associated with treatment, including there being no effect at all. An **adverse event** is any negative or harmful occurrence that takes place during treatment, that may or may not be associated with a medicine (44). A **serious adverse reaction** is a reaction that is life-threatening or fatal, requires new or extended hospitalisation, results in persistent or significant disability or incapacity, or is a congenital anomaly or birth defect (45).

Pharmacovigilance is the science and activities relating to the detection, assessment, understanding and prevention of adverse effects or any other possible drug-related problems (46). To achieve a complete picture of risk within the shortest possible time, pharmacovigilance systems allow for inclusion of data from clinical trials, case reports, spontaneous reports, population databases, and epidemiological studies, as illustrated in Figure 1. As a result, adverse effects data are highly variable with regard to quality.

A **clinical trial** is an experiment to compare the effects of two or more healthcare interventions (47)**.** Most **randomised controlled trials (RCTs)** assess treatment **efficacy** by comparing two or more treatment options in a controlled, experimental setting, in contrast to effectiveness, which is associated with use in ordinary, everyday circumstances.

Pharmacovigilance reporting systems (also called **spontaneous reporting systems)** rely on healthcare professionals and patients to identify and report any suspected adverse effects from medicines (44).

Published case reports and **spontaneous reports** concern individual patients. They have an important role in identifying safety issue signals, generating risk hypotheses, or market withdrawals (12), however, they are anecdotal by nature and do not provide definite proof of causality (48, 49) or frequency estimates (50, 51).

Epidemiological studies analyse observations in real-life populations, and are suitable for risk comparisons in exposed and unexposed groups. They are, however, non-randomized, with a risk of bias due to poor control of variables, such as degree of exposure and underlying risk factors (52). **Pharmacoepidemiological studies** focus on use and effects of medications in population groups, especially risks of adverse effects, due to the discovery of considerable health injuries of medications (53). They draw on a variety of data sources, including health records, claims databases, and health registries (54, 55).

Safety surveillance data are included in **Periodic Safety Update Reports (PSURs)**, which are conducted by the European Medicines Agency and updated by Marketing Authorisation Holders (pharmaceutical companies) (56, 57). Safety information is provided in product monographs, **Summaries of Product Characteristics (SmPCs)** (58), which often lack information on severity and timing, and can vary between countries (59-61).

In **causality assessments** of adverse effects, the aspects of the case and quality of the documentation are assessed in a structured way (62). The variables and structure of causality assessments were described by Bradford Hill in 1965 (63) and later developed further (62, 64, 65). Determination of causality is a complex process, due to confounders, variations in individual judgments, prior expectations, level of patient information, and insight into other possible causes (66).

Figure 1. Piecing together adverse effect profiles.

Graphics adapted from original at Pixabay, https://pixabay.com/vectors/jigsaw-puzzle-jigsaw-puzzle-piece-303503/, free download for commercial use.

1.3 Adverse effects reporting in randomized clinical trials

Due to their central role in drug development, randomized clinical trials (RCTs) are cornerstones in the process of providing data on benefits and harms of pharmacological therapies. In the hierarchy of evidence, RCTs are generally classified above case series and observational studies (15, 67), although shortcomings and biases may cause evidence from trials to be downgraded (68). In theory, clinical trials are expected to identify and report treatment harms. In reality, this is not always the case. Clinical trials involve decisions, activities, and individuals that may affect the outcome. Several researchers have identified risks of bias and shortcomings related to adverse effects reporting in clinical trials (1, 15, 49, 69-72). Biases can be caused by several factors, such as choice of study design, monitoring

parameters, handling of patient data, and publication thresholds. The main factors affecting adverse effects outcomes detection are illustrated in Table 1.

Planning phase

The main focus of a clinical trial is to establish whether the treatment has an effect on the condition being treated. In most studies, study power is determined by expected differences in efficacy. Adverse effects are, at best, defined as secondary outcomes. As a consequence, studies in general will not have sufficient power to detect differences in adverse effect profiles, or to detect other than the most common adverse effects (50, 73-77). Treatment periods may be short, and participants selected according to rigid inclusion and exclusion criteria, with increased likelihood of showing effect, and less risk of adverse effects than patients with multiple comorbidities and medications (15, 77).

Adverse effects may cover a range of symptoms and occur in several organ systems. As a consequence, possible adverse effects are often not predefined with regard to diagnostic criteria or procedures, unless the trial aims to identify specific risks. Trial protocols often include statements about safety monitoring, without specifying if any particular adverse effects will be looked for. Consequently, criteria for identifying and classifying adverse effects are often lacking (15, 50).

Adherence to trial protocol treatment and use of concomitant medications may have a large impact on adverse effect risk. Protocol noncompliance, with lower medication use than intended, will reduce the exposure and presumably the occurrence of drug-induced adverse effects. Use of concomitant medications may increase the risk of adverse effects (non-steroidal anti-inflammatory drugs, NSAIDs, in addition to corticosteroids), reduce the risk (gastric ulcer prophylaxis), or use of similar medications to the one is being tested (additional antidepressants in SSRI trials).

Table 1. Implementation of clinical trials. Factors that may influence adverse effects outcomes		
Trial phase	**Protocol decisions**	**Examples of variables**
Planning	Trial design	Patient group Inclusion criteria Exclusion criteria Study duration
	Trial therapy	Medication and dose Control group
	Adverse effects focus	Specified adverse effects All adverse effects
Data collection	Patient monitoring and reporting	General or specific examinations Laboratory tests Questionnaires Interviews Doctor reporting Patient reporting Judgmental procedures
	Adverse effects definitions	Criteria Severity thresholds
Assessments	Translations and coding	Grouping of adverse effects Choice of classification system Translation of vernacular to medical codes
	Causality	Criteria for causality Assessment of causality
	Seriousness	Definition of serious adverse event (SAE) Criteria for assessing SAE Procedures for assessing SAEs
	Group analysis	Intention-to-treat Per protocol As treated
Publication	Inclusion of adverse effects	Most frequent adverse effects Most serious adverse effects Adverse effects in specific organ groups
	Risk of adverse effects	Absolute numbers Relative numbers General statements

Data collection phase

Definitions of adverse events and principles of severity grading are not described in many trial reports (2). Few trials include formal procedures to assess adverse effects, and methods of assessment are often not described (70). There is no consensus as to how study patients should be examined or how data should be collected. Methods for safety monitoring include

general, specific physical, or laboratory examinations, patient interviews with general or specific questions, use of checklists and scoring tools. Choice of method, and patient and investigator expectations, may have considerable impact on the number of adverse effects reported. Use of checklists, with suggestions of specific adverse effects, has been found to increase the number of reported adverse effects, compared to open-ended questions. Patients' desires to participate in a trial, for example if few other treatment options are available, may reduce their interest in reporting adverse effects (69, 78-80). Monitoring procedures may vary between clinical trials, making comparisons and comprehensive assessments difficult (2, 79).

Assessment phase

Transfer of data from case report forms into a medical vocabulary, and application of diagnostic codes on adverse effects, has been shown to cause misclassifications and errors due to different interpretations and evaluations of events. It may be difficult to distinguish between newly occurring disease as an adverse effect, and exacerbation of underlying disease with similar symptoms, as in the case of pneumonia or heart failure in COPD patients. Disease criteria, for example suicidality, may cause uncertainty and variability in classification of cases of self-harm (50, 81, 82). Terminology may vary between studies on the same subject, as in trials of checkpoint inhibitors where 24 terms were used to describe colitis (83). Schroll et al. describe how a considerable number of medical terms were used to code cases of diarrhoea and constipation in clinical trials of orlistat, making the overall risk more difficult to assess (84). Coding errors and differences between coding staff have been noted, probably due to difficulties in matching description of events to the complex medical terminology coding systems (85). In the case of rosiglitazone, mentioned in Section 1.2.1, researchers reanalysing the data after market withdrawal identified more cases of myocardial infarction and heart failure when assessing individual patients' data than in summary data from the same trials (86). Causality assessments require judgements by clinical investigators despite use of algorithms and scoring systems. In some trials, investigator judgment may have a potentially large impact on reporting of events and classifications of seriousness, as illustrated by a paroxetine trial where trial investigators originally had judged several serious adverse effects to be unrelated to treatment (19).

Interpretation of trial events as chance findings, unrelated to treatment, has in some cases led to marketing of therapeutic agents that were later withdrawn due to safety issues (16).

In many cases, adverse effects originally recorded on a continuous scale, such as liver enzyme levels, or psychiatric scores, are converted to binary outcomes (illness or not) (70). Binary classification criteria and thresholds for defining illness have the risk of identifying too few, or too many, events depending on the threshold value, and sensitivity related to degree of change may be lost (77).

Reporting of adverse effects is often limited to counting of cases (incidence rate) and comparison of number of cases between treatment and control groups (risk difference) (87). Analysis of adverse effects in study groups may be performed by comparing patient groups according to randomization (intention-to-treat, ITT), by adherence to treatment protocol throughout the trial (per protocol), or by actual treatment received (as treated) (88). According to current recommendations, ITT analysis should be preferred both for efficacy and harms data, to uphold the balance achieved by randomization (89).

Publication

The CONSORT statement (Consolidated Standards of Reporting Trials), a set of recommendations for reporting results from clinical trials, was introduced in the late 1990s (90), but with little focus on reporting of harms. An extension on harms reporting (CONSORT Harms) was published in 2004 (89) after identification of serious flaws in clinical trials (91-93). Ideally, all adverse effects suspected in a clinical trial should be reported and fully described in the resulting publication. In reality, many trials are not published, with the result that safety and efficacy data are unavailable for large numbers of patients (16, 94, 95).

Several studies have found incomplete reporting of adverse effects, despite implementation of the CONSORT Harms guidelines (1, 2, 6, 18, 71, 72, 96-104). In an analysis by the German Institute for Quality and Efficiency in Health Care, researchers noted extensive publication bias, including withholding of safety data (16). In some publications, no details on adverse effects are provided beyond a general statement that no major adverse events were observed (70, 102). Analyses of antidepressant and analgesic clinical trials have found that a large proportion of the published studies did not mention serious adverse effects (18, 98).

Publication selection of adverse effects by frequency and severity is common. Examples of such limitations are reporting of adverse effects occurring in more than a certain percentage of patients, reporting only the serious or severe cases (70, 71, 84, 102, 105-107), or limiting reporting to statistically significant differences in adverse effects despite lack of study power (1, 108).

1.3.1 Gastrointestinal bleeding risk in publications of corticosteroid trials

The Norwegian Regional Medicines Information Centres (RELIS) have received repeated queries on gastrointestinal bleeding risk and need for gastroprotective treatment for patients on corticosteroid therapy. Previous systematic reviews and meta-analyses have arrived at conflicting conclusions with regard to risk of gastrointestinal bleeding during corticosteroid therapy, as discussed in Section 1.4.1. A comprehensive assessment of gastrointestinal adverse effects described in all included publications has not been done previously. The quality of the clinical trials that form the basis of the systematic reviews has not been analysed with regard to adverse effects reporting, and it has not been examined whether differences in patient monitoring, diagnostic criteria, or other parameters may have had an impact on the trials' identification of adverse effects. On this background, we decided that analysis of the gastrointestinal bleeding reports in the group of corticosteroid trials would be a suitable model system for a detailed assessment of the quality of monitoring and reporting of adverse effects. We also believed that the analysis could give insights relevant to monitoring and reporting of gastrointestinal bleeding in other trials and medical areas.

1.3.2 Adverse effects in publications of antidepressant trials in children and adolescents

Treatment with antidepressants in young people is increasing in many countries. In England, use of antidepressants doubled from 2005 to 2017 in young people aged 12-17 (109). In Norway, use of selective serotonin reuptake inhibitors (SSRIs) in the age group 10-19 years increased by 70% from 2005 to 2019 (110). Several prescribers have contacted RELIS with requests concerning SSRI safety profiles in young patients, in order to perform individual risk-benefit assessments. We identified several analyses of suicidality risk, but less detailed information on other, and presumably more common, adverse effects.

Suicidality risk has long been a major issue in safety assessments of antidepressants in children and adolescents. In 2004, the U.S. Food & Drug Administration issued a black box warning for antidepressant use in children and adolescents, due to increased risk of suicidality (21). The European Medicines Agency, EMA, published similar warnings in 2004 and 2005 (20, 111). Risk of suicidality was addressed in several reviews and meta-analyses during the 1990s (112-114).

Fluoxetine is one of the most commonly used SSRIs in children and adolescents, and the only SSRI approved for use in depression in children and adolescents in Norway. Fluoxetine has been examined in a limited number of clinical trials in this age group. The major study is the 'Treatment for Adolescents With Depression Study (TADS)' which included 439 patients (115) and is regarded as a high-quality study (116). TADS was initiated and publicly funded by the National Institute of Mental Health (NIMH) (117), which is part of the National Institutes of Health in the USA. The trial was coordinated by the Duke University Medical Center. An in-depth analysis of the adverse effects data from this pivotal clinical trial had not been performed previously. Researchers have found under-reporting of adverse effects, biases, and flawed data analyses in other antidepressant trials (17, 19, 106, 107, 118-122). We undertook a detailed analysis of the reporting of adverse effects in the TADS study, in order to describe the quality and weaknesses of the published data in a single, high-quality trial, and to assess the full range of adverse effects observed in the trial.

1.4 Information on adverse effects in systematic reviews

In a **systematic review,** the aim is to examine all relevant research pertaining to a specific question, by methods and assessments that are transparent and clearly stated and where data from all studies are analysed together (47). Due to the large numbers of published clinical trials, systematic assessments, as provided by systematic reviews and meta-analyses, provide an accessible overview of existing evidence (123, 124). In some systematic reviews, **meta-analyses** are used for statistical analysis of data from all studies combined, enabling group comparisons in larger data materials (125). Meta-analyses may include both RCTs and observational studies; however, inclusion is often limited to RCTs due to the less controlled settings and increased risk of confounding in observational studies (126).

Systematic reviews of RCTs are seen as producing high quality evidence both for benefit and harms (22), despite the fact that RCTs are not powered to identify harms (127). Treatment harms that have been identified through case reports or epidemiological studies are not always included (128-131). Systematic reviews and meta-analyses focus mostly on treatment effects (benefit) and have far less focus on harms (132, 133). In systematic reviews focusing on treatment harms as a primary outcome, study heterogeneity and poor reporting of adverse effects in underlying clinical trials may have a large impact on quality (128, 130, 134-136). Many systematic reviews do not mention adverse events, fail to include harms as an outcome of interest, or give an inadequate description of harms and the quality and risk of bias of included studies (129, 130, 134, 137, 138). Incorporation of data into systematic reviews carry the risk of strengthening the initial impression given by biased trial data (139), and the increasing number of systematic reviews and meta-analyses with focus on benefit provide misleading conclusions with regard to risk (140). In 2009, the PRISMA (Preferred Reporting Items for Systematic reviews and MetaAnalyses) statement was developed to improve reporting and transparency (141), with focus on systematic reviews of efficacy. An extension to the PRISMA statement, the PRISMA harms checklist, was published in 2016 to improve systematic reviews and meta-analyses on treatment harms (142).

1.4.1 Systematic reviews on corticosteroid-induced gastrointestinal bleeding

Despite the quality framework, systematic reviews on the same topic may reach different conclusions. The question of corticosteroid treatment and risk of gastrointestinal bleeding is a case in point. It has long been debated whether corticosteroid treatment increases the risk of gastrointestinal bleeding or perforation. Corticosteroids are used for several indications in a large number of patients, and an increased risk of gastrointestinal bleeding could have considerable health implications. Previous systematic reviews and meta-analyses have arrived at conflicting conclusions, with statements of no increased risk (143, 144), and a statistically significant increased risk (145), respectively. In 1976, Conn and Blitzer published an analysis of randomized, controlled studies with corticosteroids (143). They included 26 double-blinded trials of 3358 patients, and found a frequency of peptic ulcer of 1.4% in the corticosteroid group and 1.0% in the control group. The difference was not statistically significant, and the authors concluded that corticosteroids could not be shown to be associated with increased risk of peptic ulcer. In 1983, Messer et al. published an analysis of

peptic ulcer complications in 71 controlled corticosteroid trials with close to 6000 patients (145). Unlike Conn and Blitzer, Messer et al. found a statistically significant risk increase of diagnosed peptic ulcers, with 1.8% in the corticosteroid group and 0.8% in the control group. Inclusion of only double-blinded studies gave incidences of 2.6% and 1.5%, respectively (p=0.04). In 1994, Conn and Poynard published a meta-analysis of corticosteroid therapy complications, including peptic ulcer, in studies published from 1964 to 1982 (144). In this review of 93 studies with 6602 patients, the authors found a non-significant incidence of peptic ulcer, with 0.4% in the corticosteroid group and 0.3% in the placebo group. The conflicting conclusions in previous reviews are reflected in medicines information databases and other literature sources. While some sources state that corticosteroid treatment may increase the risk of ulcers and gastrointestinal bleeding, especially in combination with NSAIDs (146-148), others describe the association as weak (149, 150) or unlikely (151). Repeated queries to RELIS regarding ulcer prophylaxis in patients who are receiving corticosteroids indicated that differing conclusions were causing uncertainty among health practitioners, and that an updated, extensive, and conclusive review was needed.

1.5 Information on adverse effects in clinical practice guidelines

Clinical practice guidelines are statements that include an assessment of alternative care options, based on high-quality systematic reviews (23), including evidence and judgments regarding benefits and harms of treatment options (152, 153). To our knowledge, research on adverse effects information in clinical therapy guidelines has not been performed.

Clinical practice guidelines should rate the quality of the evidence and the strength of the recommendations according to the GRADE system (Grading of Recommendations Assessments, Development, and Evaluation) (154) that address transparency, conflict of interest, participants in guideline development, and quality of evidence. In GRADE, randomized trials are categorized as high-quality evidence, with options for downgrading according to risk of bias and other criteria (132, 155, 156).

To evaluate the quality of clinical guidelines, the AGREE II (Appraisal of Guidelines for Research & Evaluation II) instrument was proposed as an assessment tool in 2003 (157). AGREE II state that guidelines are expected to present both benefits and harms of an intervention, and addresses scope and purpose, stakeholder involvement, rigour of

development, clarity of presentation, applicability, and editorial independence (158). Only one AGREE II item concerns consideration of side effects and risks associated with treatment, and no level of risk communication or inclusion criteria for harms data have been defined.

Assessments of guideline quality have been performed in several medical areas, and many have found the overall quality to be poor or variable according to the AGREE II criteria (159-166). Guideline shortcomings and biases include conflicts of interest (167-171), and more focus on benefit than harms (168). As for systematic reviews and meta-analyses, there is a risk of compounding weaknesses and biases in underlying data when included in the evidence base of practice guidelines (23, 140, 172). The shortcomings have a potentially large impact on the presentation of risk profiles in guidelines and clinicians' subsequent perception of risks. We have not identified any specific assessments of adverse effects mentioned in guidelines.

2. Aim of the thesis

The overall aim of this thesis was to examine the reporting of adverse effects in publications of clinical trials in two model areas associated with considerable uncertainty, to analyse adverse effects data through a systematic review, and to assess how adverse effects of medications are presented in clinical therapy guidelines.

Aim 1. To assess and quantify the risk of gastrointestinal bleeding or perforation due to systemic corticosteroid treatment and to establish whether corticosteroid treatment is in fact associated with an increased risk (Paper I).

Aim 2. To assess whether studies of systemic corticosteroid treatment mentioned gastrointestinal bleeding or perforation as an adverse effect, to examine the terminology and diagnostic criteria applied in the trials, and to assess whether the publications adhered to international guidelines on harms reporting and whether the CONSORT Harms checklist is a suitable tool for evaluating adverse effects reporting in clinical trials (Paper II).

Aim 3. To identify all publications from the TADS study of antidepressant therapy in children and adolescents, and assess their reporting of adverse effects (Paper III).

Aim 4. To analyse adverse effects information and use of data from the TADS study in therapy guidelines on treatment of depression in children and adolescents (Paper IV).

3. Materials and methods

The research presented in the four papers in this thesis is based on literature searches and analyses of the included publications.

3.1 Methods of data extraction and evaluation

3.1.1 Study designs

Paper I is a quantitative study of 159 clinical trials with regard to risk of corticosteroid-induced gastrointestinal bleeding. Paper II is an in-depth analysis of the publications included in Paper I. Paper II contains quantitative and descriptive analyses, on fulfilment of quality criteria for adverse effects reporting, and descriptions of adverse effects, respectively. Paper III is a descriptive study on the reporting of adverse effects in the Treatment for Adolescents With Depression Study (TADS). Paper IV is a descriptive analysis on the information on adverse effects in clinical guidelines for treatment of depression in children and adolescents.

3.1.2 Literature search

For Papers I and II, clinical trials of corticosteroids were identified by a systematic literature search in Medline and EMBASE for the period 1 January 1983-30 June 2011, with an additional update per 30. March 2013. The search was limited to double-blinded, placebo-controlled randomised clinical trials of betamethasone, dexamethasone, methylprednisolone, prednisolone, prednisone, triamcinolone, cortisone or hydrocortisone. The Cochrane Database of Systematic Reviews was searched for additional references in therapy reviews of the most common diseases where corticosteroids are used. There were no language restrictions during the search, but articles in a Scandinavian language, English, or German were selected during the evaluation. The abstracts were reviewed by two researchers and any study that appeared relevant was assessed in full-text version. To be included, the article must mention gastrointestinal adverse effects, or be judged by the reviewers to have monitored patients for adverse effects in a way that could be expected to detect gastrointestinal bleeding. Because there is no uniform terminology to describe gastrointestinal bleeding or perforation as an endpoint in clinical trials, we included all cases

where an adverse effect was described as visible blood in stool, GI bleeding, haematemesis, melena and GI perforation. We did not include cases of dyspepsia, gastritis, duodenitis, epigastric pain, and necrotising enterocolitis.

In Paper III, publications arising from the TADS study were identified through literature searches in Medline, EMBASE, PsychInfo, and Google Scholar, using the phrases « TADS team» or «Treatment for adolescents with depression study». The searches included publications from the main authors of other TADS study articles, and searches for similar publications to already identified TADS references (the snowballing method). We also searched for TADS study references on the websites ClinicalTrials.gov, nimh.nih.gov, and the TADS website at Duke University, and by manual searching of reference lists in reviews and guidelines. The main search was ended on 5 September 2017, with an additional update search by January 2019.

In Paper IV, clinical practice guidelines on treatment of depression in children and adolescents were identified through literature searches in PubMed, EMBASE, and clinical guideline collections, in addition to manual searches in identified guidelines and references. PubMed and guideline collections were searched in the period 30.10.2018-17.02.2019. The EMBASE search was performed 5.12.2019. The age limits applied in PubMed and EMBASE were birth-18 years, and child or adolescent 13-17 years, respectively. We chose to include some well known international decision support tools, due to our belief that health care practitioners are using such tools in addition to clinical guidelines. Guidelines and clinical decision support tools are hereafter termed guidelines in this thesis.

3.1.3 Data extraction and assessment

In Paper I, at least two of the authors reviewed the clinical trial publications retrieved in the literature search. Data on study characteristics, publication year, inclusion and exclusion criteria, type and duration of corticosteroid use, indication, additional medications, patient monitoring procedures, descriptions of gastrointestinal adverse effects, and severity of disease (hospitalisation or ambulatory treatment) were extracted. Cases of gastrointestinal bleeding, gastroduodenal ulcers, melena, blood in stool, gastrointestinal perforation and haematemesis were included for analysis. Data from clinical trials where monitoring for adverse effects appeared comprehensive and other adverse effects were well described

were included, even if no gastrointestinal bleeding was mentioned, on the assumption that any cases would have been identified. Cases of dyspepsia, gastrointestinal pain, gastritis and necrotizing enterocolitis were not included in the analysis.

In Paper II, the 159 clinical trials reviewed in Paper I were assessed for monitoring procedures and descriptions of gastrointestinal bleeding. A data extraction form was developed based on the CONSORT Harms recommendations 10-point checklist, with two items modified to specify gastrointestinal adverse effects. Two reviewers assessed each study, extracted data on adverse effects descriptions, and assigned a score of 1 or 0 if the criterion was judged to be fulfilled or not. Criteria for scoring 1 or 0 were discussed by all authors. If the study fulfilled at least one of the criteria in a checklist item, the score would be 1. Fulfilment of all criteria gave a maximum score of 10.

In Paper III, identified TADS study publications were assessed by at least two reviewers, and were included in the analysis if they included some data on adverse effects. Data on type and frequency of adverse effects, patient group, and study stage were extracted. Adverse effects were included as described in the publications. We included worsening of depression as an adverse effect if described in the publications, even though some publications considered this a residual symptom rather than an adverse effect.

In Paper IV, clinical practice guidelines on depression in children and adolescents were analysed with regard to information about adverse effects. Statements on adverse effects were extracted and classified according to the MedDRA organ classification system (58) with adjustments (separate account of suicidal thought and behaviour, and libido changes classed with reproductive system disorders). Citations were classified and analysed by type and content for all guideline statements about adverse effects.

3.2 Statistical methods

Statistical analyses were performed using SPSS ver. 20 (Paper I) and ver. 23 (Paper II). RevMan ver. 5.2. was used for the meta-analysis (Paper I).

In Paper I, the proportion of patients with gastrointestinal bleeding or perforation in the corticosteroid and placebo groups was compared by calculating odds ratio (OR) and 95% confidence interval (CI), for the whole group and subgroups. Subgroup analyses were

performed comparing occurrence of gastrointestinal bleeding in hospitalized and ambulant patients, in eight disease groups, and in sensitivity analyses. In the sensitivity analyses, patient groups with known or suspected NSAID use, studies with peptic ulcer as exclusion criterion and studies with use of gastro protective drugs, were excluded. The Mantel-Hansel method with the random effects model was used for the meta-analysis.

In Paper II, the relationship between CONSORT Harms scores and the likelihood of reporting gastrointestinal bleeding was analysed by logistic regression. Correlations between scores on the CONSORT Harms checklist and year of study publication, setting (hospitalized or ambulant patients), funding (industry sponsored or not) and journal (major medical journal or not), were analysed using the Pearson chi-squared test. Differences in checklist scores were analysed using the independent samples t-test for equality of means. Reviewer agreement were analysed by use of Gwet's agreement coefficient with first-order chance correction AC1, referred to as Gwet's AC1 (173, 174).

Papers III and IV are descriptive studies and statistical analyses were not performed.

3.3 Ethical considerations and approvals

Studies I-IV are based on publicly available documents, including reports of clinical trials performed by other researchers. No patients were enrolled by the authors for this project, and ethical approvals were not necessary.

4. Summary of results

Paper I

We performed a systematic review and meta-analysis in order to establish whether corticosteroid treatment is associated with an increased risk of gastrointestinal bleeding and to reach a conclusion that could reduce previous confusion in this field. 159 publications of clinical trials published between 1983 and 22 May 2013 were identified through extensive literature searches. This comprised a total patient population of 33 253. In the group as a whole, 804 patients (2.4%) were reported to have had a gastrointestinal bleeding or perforation, with an incidence of 2.9% in the corticosteroid group and 2.0% in the placebo group. In patients treated with corticosteroids, the risk of gastrointestinal bleeding or perforation was increased by 40% (OR 1.43, 95% CI 1.22 to 1.66) compared to placebo. The risk differed between patient groups. In hospitalized patients, the risk increase was statistically significant, with bleeding or perforation occurring in 3.79% (472 of 12442 patients) and 2.64% (321 of 12160 patients) in the corticosteroid and placebo groups, respectively. In ambulatory patients, far fewer cases were reported. In this group, 11 of 8651 patients (0.13%) had a gastrointestinal bleeding or perforation, of which 8 belonged to the corticosteroid group. Overall, bleeding or perforation occurred in 3.22% (793 of 24 602) of patients (OR 1.42, 95% CI 1.22 to 1.66). We concluded that use of corticosteroids was associated with increased risk of gastrointestinal bleeding. The risk increase was statistically significant for hospitalized patients but not for ambulant patients.

Paper II

The 159 clinical trials analysed in Paper I varied considerably in their definitions and monitoring for gastrointestinal bleeding or perforation. Several diagnoses and laboratory values, of varying severity, were included. Most publications provided data on gastro-intestinal adverse effects including bleeding or perforation, but did not address adverse effects specifically. Other aspects of the adverse effects reporting are presented in Table 2. Most studies provided information on patient withdrawals (133/159 studies, 83.6%), absolute risk of gastrointestinal bleeding or perforation (130/159 studies, 81.8%), and method used to collect adverse effects data (118/159 studies, 74.2%). Relatively few studies

mentioned collection of adverse effect data in the introduction (48/159 studies, 30.2%), or described a plan for presenting and analysing information on harms (51/159 studies, 32.1%). Analysis of reviewer agreement showed large variations, but an overall fair to moderate agreement. The mean score for fulfilment of CONSORT Harms checklist criteria was 5.25 out of 10. We found the criteria to be ambiguous, and do not recommend use of the CONSORT Harms recommendation checklist to assess quality of harms reporting in clinical trial publications without individual judgment.

Paper III

We identified 48 publications as describing TADS study data. Eight publications gave some information about adverse effects observed in the trial, and they all mentioned risk of suicidal behaviour. Other psychiatric and somatic adverse effects were mentioned in some, but not all publications, and were reported in detail only for the initial study period of 12 weeks. Several well-known adverse effects of fluoxetine were not mentioned in any TADS publication. Other aspects of the adverse effects reporting are presented in Table 2. We found that publications from the TADS study do not present a full account of all adverse effects observed, and did not identify the full spectrum of known adverse effects.

Paper IV

We identified and analysed 19 clinical guidelines on treatment of depression in children and adolescents. We found considerable variation in the extent of adverse effects information provided. Risk of suicidality was discussed in all guidelines and reflected in the number of citations. Descriptions of other adverse effects varied widely and were not always referred to in guidelines, even when mentioned in the underlying citations. Risk of other psychiatric adverse effects was mentioned in most guidelines. Several guidelines mentioned none or a limited number of somatic adverse effects, though well-known and described in other sources. Eighteen of 19 guidelines referred to TADS data directly or indirectly, but some only referred to TADS with regard to suicidality, without mention of other adverse effects observed in the trial.

Table 2. Aspects of adverse effects reporting in assessed trials			
Trial phase	**Examples of variables**	**Corticosteroid trials (Papers I, II)**	**TADS (Paper III)**
Planning	Patient group Inclusion criteria Exclusion criteria Study duration	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 wk.
	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, possibly others
Assessments	Grouping of adverse effects Classification system Translation to medical codes	Different terminology	Ambiguous terminology for some symptoms
	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
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 study period
	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

5. Discussion

5.1 Discussion of the main findings

Our results show several risks of bias and weaknesses with regard to the adverse effects reporting, both in the clinical trials of corticosteroids and in the TADS study of anti-depressive therapy (Papers I-III). Study design, monitoring parameters, data analysis, publication thresholds, or other methodological limitations were elements that had a potentially large impact on the frequency and type of adverse effects reported in the publications. Clinical therapy guidelines on treatment of depression in children and adolescents varied considerably in their descriptions of adverse effects (Paper IV).

5.1.1 Aim 1. Risk of gastrointestinal bleeding or perforation due to systemic corticosteroid treatment (Paper I).

In the systematic review and meta-analysis presented in Paper I, we found that trial patients who had received corticosteroids had a 40% increased risk of a gastrointestinal bleeding compared to placebo (OR 1.43, 95% CI 1.22 to 1.66) (Paper I). The risk increase was statistically significant for hospitalized patients, and few cases were reported in ambulatory patients. The risk increase of 40%, translated to absolute numbers in the hospitalized patients in the meta-analysis, correspond to an incidence of 38 cases per 1000 patients in the corticosteroid group, and 26 patients per 1000 patients in the placebo group. The resulting risk difference is higher than 1 in 100 patients, which is the criterion for a common adverse effect.

Exclusion criteria

We found that a third of the trials had excluded patients with pre-existing peptic ulcer disease, however, several publications did not state whether patients with relevant underlying diseases were included in the study population (Paper I).

Additional medications and group analysis

We found that pre-existing or concomitant drug therapy, including NSAIDs or gastro-protective therapy, was described in some publications, but poorly described in others. This

present an element of bias, as patients may have had increased or reduced risk of gastrointestinal bleeding irrespective of corticosteroid treatment (Paper 1).

Missing information in publications

Many of the corticosteroid trials reported few or none adverse effects (Paper I); raising the question as to how incomplete and missing data should be interpreted. Non-mention of gastrointestinal bleeding could be due to non-occurrence, publication bias or non-monitoring. If monitoring of adverse effects appeared to be comprehensive and other adverse effects were well described, we assumed that any gastrointestinal bleeding would have been reported, and chose to set the number of cases to zero accordingly. This assumption may have been incorrect.

Existing reviews and meta-analyses

Previous reviews of corticosteroid-induced risk of gastrointestinal bleeding were published in 1976, 1982, and 1994 (143-145). They reached different conclusions, which is reflected in medical databases and textbooks, and is a major reason for our decision to perform the systematic review on more recent data. The previous systematic reviews have differed with regard to inclusion/exclusion criteria, time frame, and duration of treatment. Conn and Blitzer included 26 double-blinded studies published from 1950 to 1975 (143). The 1983 Messer review included 71 studies, of which 37 were double-blinded (145). Conn and Poynard included 93 double-blinded studies published from 1964 to 1982 (144), essentially the same time frame as Messer et al. The inclusion criteria were strongly debated because the reviews had included different studies to some extent, as exemplified by inclusion of both double-blinded and non-double-blinded studies (175, 176).

In our review, we excluded studies with single dose treatment (Paper I). Conn and Poynard included studies if any dose of steroids had been given (144). Conn and Blitzer excluded studies of short term treatment (few days) (143), while Messer et al. included studies with more than four days of treatment (145).

The reviews by Conn and Blitzer (143), and Conn and Poynard (144), did not identify a statistically significant increase in risk. In contrast, Messer et al. (145) found a statistically significant increase in risk, with an incidence of diagnosed peptic ulcers of 2.6% in the

corticosteroid group and 1.5% in the control group. Paper I included a more recent material of papers published between 1983 and 22 May 2013, and probably more robust data, due to the considerably larger group of patients than in previous reviews.

To our knowledge, Paper I is the first systematic review to perform a subgroup analysis by hospitalization or ambulant treatment. Previous reviews have had little focus on underlying disease severity, though Conn and Blitzer excluded studies in patients with highly lethal diseases (143), and Messer et al. found an increased risk of peptic ulcer in patients with predisposing illness (145). In Paper I, we found that a third of the studies had excluded patients with peptic ulcer disease. Both Messer et al. (145) and Conn and Blitzer (143) excluded studies where ulcerogenic medications such as NSAIDs had been administered to the control group, and Messer et al. excluded studies with antacid use (145). In the Conn and Poynard review, studies using medications with known risk of gastrointestinal bleeding were included and analysed as co-variables (144), as in Paper I.

Meta-analysis may be a valuable tool for identifying and quantifying adverse effects by increasing sample size and power beyond individual, limited trials (177). However, for analysis of adverse effects data across studies, several researchers have noted methodological challenges associated with inclusion of trials of highly variable quality, design, duration, and adverse effects assessment procedures (135, 136, 177-179). There are no clear guidelines as to what degree of heterogeneity is acceptable (15). Pooling of data, by adding events in each group and dividing by the number of patients, may be misleading due to differences in patient groups and comparisons in the underlying studies (180). In a meta-analysis, adverse effects data are compared using relative risks or odds ratios of the individual trials (131, 180), however, many meta-analyses include only published data on a summary level, and not individual patient data (15, 131, 135).

Since the publication of Paper I, no equivalent systematic reviews have been published on the risk of gastrointestinal bleeding in all patient groups. In 2019, Butler et al. published a systematic review and meta-analysis of the incidence of gastrointestinal bleeding in critically ill adults and found a statistically significant increased risk of clinically important bleeding, 2.3% in the corticosteroid group vs. 1.8% in the placebo group (RR 1.26) (181). This is lower than the overall risk found in Paper I, but the patient groups are not identical. Butler et al.

raise questions regarding bias and error risk analyses in Paper I (182), which are discussed in Section 5.2.1.

Systematic reviews rely heavily on published clinical trials, and are dependent upon the information in the included studies for comprehensiveness, relevance and trustworthiness (183). The different conclusions reached in the systematic reviews on corticosteroid-induced risk of gastrointestinal bleeding highlight weaknesses and pitfalls regarding performance and trust in systematic reviews.

5.1.2 Aim 2. Descriptions of gastrointestinal bleeding or perforation in corticosteroid trials, and assessment of the CONSORT Harms checklist as a quality tool (Paper II).

The 159 clinical trials that were included in the systematic review and meta-analysis (Paper I) were heterogeneous with regard to parameters such as indication, type of corticosteroid, treatment duration, study size, adverse effects monitoring, criteria for gastrointestinal bleeding, and extent of reporting (Paper II).

Monitoring and diagnostic criteria

We initially assumed gastrointestinal bleeding or perforation to be a defined and unambiguous outcome that could be clearly quantified. We found, however, that monitoring methods and diagnostic criteria for gastrointestinal bleeding varied considerably between trials (Paper II). Some researchers had examined all patients for occult faecal blood, while others reported only cases of haematemesis or melena requiring transfusion, monitored haematocrit, or performed general examinations without specifying methods or objectives. As pointed out by other researchers, few patients had been examined by endoscopy (176). The implications for performance of systematic reviews are discussed in Section 5.1.1.

Quality assessment using the CONSORT Harms criteria (Paper II)

The CONSORT Harms recommendations were developed to improve harms reporting in clinical trials, but have also been used as a tool for assessing the quality of reporting of adverse effects in clinical trials (2, 6, 71, 72, 96-99, 101-104). The checklist includes 10 items that should be part of the harms reporting. We found a mean score of 5.25 out of 10, meaning that several criteria were not fulfilled in many publications (Paper II). Other

researchers have found CONSORT Harms checklist scores of 3.0-6.7 (6, 96, 99, 103) or 50%-63% (72, 97, 102). We show that the items that were reported in most studies were patient withdrawals (83.6%), absolute risk of gastrointestinal bleeding or perforation (81.8%), and method used to collect adverse effects data (74.2%). Plans for collecting harms data, and presentation and analysis of harms data were mentioned to less extent (30.2% and 32.1%, respectively). Other authors have identified good reporting of patient monitoring, adverse effects definitions and statistical analysis, discussion of adverse effects in the result section, harms mentioned in title or abstract, or a good account of risks and withdrawals due to harm (2, 6, 72, 96, 99, 102, 103). Other researchers have identified shortcomings similar to our findings, with a large proportion of articles containing no information as to collection and measurements of adverse reactions, plans for analysis, or withdrawals due to adverse reactions (2, 6, 71, 96, 97, 99, 102-104). While the CONSORT Harms recommendation checklist may be well suited for planning and publication of trial data, routine application of the checklist to assess quality of harms reporting in a published trial is inadvisable due to ambiguities in the criteria, and qualified judgment of each study is necessary.

5.1.3 Aim 3. Identification of TADS trial publications, and assessment of adverse effects reporting (Paper III).

Publications

We identified 48 papers that presented data from the TADS trial (Paper III). Of those, eight mentioned adverse effects. Despite the number of papers, we found reporting of adverse effects from TADS to be incomplete beyond the first 12 weeks of treatment, despite the fact that TADS had a duration of up to 88 weeks including open phases after week 12. The TADS publications with information on adverse effects were judged to be associated with considerable publication bias. The only adverse effect described in all publications was risk of suicidal behaviour. The number of cases of suicidality varied between publications, partly due to later reassessment of cases. Other psychiatric and somatic adverse effects were reported in detail only for phase I at 12 weeks. Several well-known adverse effects of fluoxetine were not mentioned in any TADS publication. Our findings of incomplete publication of all adverse effects data and long-term data from the TADS reflect the findings of other researchers. In a study of 97 publications of second-generation antidepressants

trials in 2016, de Vries et al. found very poor reporting of serious adverse events, as only 36 publications mentioned serious adverse effects at all, and several reports were incomplete when compared to reviews from the U.S. Food and Drug Administration (FDA) (18). In a systematic review of SSRIs, Jakobsen et al. found that only 44 of 131 clinical trials provided data on the proportion of patients who had serious adverse events (184). In clinical trials of antipsychotics and antidepressants, more serious adverse effects were noted in clinical trial database registries or case report forms than appeared in the corresponding journal articles (4, 19). In many cases, published papers do not give information on whether withdrawals from a study were due to adverse effects (70, 71, 92, 185). Inclusion of non-published data in analyses of efficacy and safety has the potential to shift conclusions regarding the risk-benefit profile of SSRIs for paediatric depression (19, 186). There is an ongoing discussion about data quality with regard to antidepressants and suicidality risk, and Paper III has been cited in this context (187).

Publication of data from a single study in multiple publications is potentially problematic, even though it may be advantageous in addressing various aspects of the study. Multiple publications may lead to fragmentation of data (so-called salami publications), lack of a comprehensive overview, and the impression that safety has been studied more extensively than is actually the case (188, 189). The TADS trial is a case in point, as we identified 48 publications arising from this one trial, with several papers addressing adverse effects at various study stages (Paper III). The initial report on adverse effects was published 2004, and the comprehensive analysis of suicidal adverse effects in 2009.

Additional medications

We found extensive use of additional medications, and divergences from assigned treatment, in the TADS trial (Paper III). Many patients received an SSRI, other psychiatric medications, or cognitive therapy, in addition to the assigned treatment (190), and the drop-out rates were high (115, 191). Notably, the publication describing divergences from protocol treatments was published in 2009, five years after the original study publication.

Monitoring, diagnostic criteria, and coding of events

We found that the TADS trial applied severity thresholds for categorizing patient symptoms as adverse events, in that adverse events must cause clinically significant interference with functioning, need for medical attention or need for a medication. Limitation by thresholds is likely to have reduced the number of reported adverse effects due to insufficient severity (Paper III).

The TADS trial applied scoring tools for suicidality and mania, with subsequent dichotomisation of continuous parameters to binary outcomes (disease yes/no). Dichotomisation has been described as arbitrary, unnecessary, and simplistic (192, 193), as the results depend largely on selection of cut-off values, and information on degrees is lost. In the TADS trial (115, 191, 194), patients may have had worsening scores for suicidality or mania, but the overall effect of threshold values and dichotomization would be that not all adverse effects, appearing as worsening symptom scores, were registered and reported. Scoring of manic symptoms was described as inconsistent and varying between clinicians (195), with probable consequences for the number of cases.

For antidepressants in general, reanalyses of suicidality data has led to reclassifications and identification of additional events (19, 106, 196). Definitions of suicidality can be ambiguous, and judgements may vary. Misclassifications and coding of events into erroneous terms have been demonstrated, as for cases of suicidal behaviour originally coded as emotional lability, worsening depression or elevated liver enzymes (after a paracetamol overdose), and cases of akathisia as nervousness, agitation or irritability (106, 196). Posner et al. describe the process of reanalysing and reclassifying suicidality data from 25 antidepressant trials in children and adolescents, commissioned by the FDA (196). The reviewers assessed 427 events with potentially suicidal behaviour, of which 114 had been classified as possible suicidality by the investigating company. The reassessment identified an additional 26 cases of suicidal behaviour that had originally been classified differently. It was also found that 12 cases that had originally been classified as suicidal behaviour should be downgraded to a less serious diagnosis.

A network meta-analysis of 35 antidepressant clinical trials in children and adolescents (197) found an absence of reliable suicidality data and a generally low quality of evidence. In a

reanalysis (198), it was noted that the definition of tolerability in the meta-analysis, as dropouts only, excluded analysis of actual adverse effects reported in the trials. Risk-of-bias tools are available to assess the potential for bias in clinical trials (199), but specific aspects relevant to adverse effects monitoring are not well covered.

Statistical analysis

We found that cases of suicidal behaviour had been analysed as occurring in the placebo group even though the patients were receiving fluoxetine at the time. This is in agreement with Högberg et al. (17), who have criticized the TADS suicidality analysis (194) for underestimating suicidal risk by analysing cases of suicidality according to assigned treatment groups, even though several placebo patients were receiving active treatment at the time of the event. In a reanalysis based on treatments actually received, published in 2015, Högberg et al. found a statistically significant increased risk of suicidal events when placebo patients who had been receiving fluoxetine were assessed in the active treatment group (17).

According to the TADS protocol, group analyses would be performed as ITT (200). In Paper III, we pointed out that ITT analysis may introduce considerable bias under such circumstances, because it does not address actual treatment received. ITT analysis may minimise group differences (77, 201), and lower incidence estimates for observed adverse effects, through enlarging the denominator (number of exposed patients). The time periods applied in adverse effects analyses may have a large impact on findings, as many clinical trials have a short duration, and adverse effects may occur after the end of the trial. Patient follow up, after the trials has ended, may identify harms that were not evident during the trial (202, 203). In the TADS trial, a patient who terminated treatment due to suicidality was not included in the suicide attempts analysis because the attempt occurred after discontinuation (195). In a review of suicidality risk in paediatric antidepressant trials published in 2006, Hammad et al. included events occurring within one day of stopping treatment, with the result that several cases of suicidality were excluded (112). In a case of discontinuation due to suicidality and hospitalization, death occurred five days after end of treatment and was classified as a post-study event (106). It is conceivable that symptoms

occurring after end of treatment may be interpreted as worsening of the underlying disease, and not attributed to discontinuation symptoms or adverse effects of the medication.

5.1.4 Aim 4. Adverse effects information in therapy guidelines on treatment of depression in children and adolescents (Paper IV).

The clinical therapy guidelines analysed in Paper IV had a highly variable practice with regard to descriptions of adverse effects. Risk-benefit assessments appeared to be based on narrow definitions of risk, primarily suicidality. All 19 guidelines mentioned risk of suicidal behaviour, and most guidelines mentioned other psychiatric adverse effects, but several guidelines did not mention somatic adverse effects. Underlying studies, including TADS (115), and the SmPC for fluoxetine (204) describe more adverse effects than were included in guidelines. This finding was reflected in the number of citations for each type of adverse effects, with the highest number of citations referring to suicidality risk. The three systematic reviews among the five most frequent citations all focused on suicidality risk (114, 205, 206).

There is a close link between the concept of Evidence Based Medicine and clinical guidelines. Guidelines should be based on high-quality systematic reviews (23), but as discussed in Section 5.1.1, systematic reviews may reach different conclusions, probably due to methodological and publication issues. In addition, several systematic reviews may be available on a single topic, which may give readers the impression that the number of underlying primary studies is more extensive than is actually the case. For antidepressant therapy, the number of meta-analyses has been described as astonishing (140). In 2016, researchers identified 185 meta-analyses of antidepressants published from January 2007 to March 2014 (207). The majority of these had industry involvement in the form of employment or sponsorship and were less likely to include negative statements about treatment harms. The international, independent Cochrane network aims to produce high-quality systematic reviews as a basis for informed decision-making (208), but analysis of adverse event descriptions in Cochrane reviews have found considerable outcome bias by inadequate reporting of harms (134).

Several guidelines are authored by specialist groups that do not follow formal guidelines for guideline development, with the risk of guideline development by expert consensus rather than evidence (209). Conflicts of interest and publication bias are potential biases that may

affect the perception of risks and benefit by guideline developers and health professionals. For this reason, increased focus on transparency and declarations of conflicts of interest has led to more clearly identifiable sources of bias in guidelines and journal articles, though it has been argued that such declarations and statements do not cancel out inherent bias and may actually lead to bias acceptance (210, 211).

We have not identified any standards as to which adverse effects the guideline risk-benefit assessments should include; whether the most severe, the most common, or a composite endpoint risk of any adverse effect. In practice, guideline authors appear to be free to decide to what extent adverse effects will be included in the guideline, and the lengths and scope of the guidelines vary considerably. Selective citing, and our finding of guideline focus on suicidality to the exclusion of known and common adverse effects of antidepressants, is problematic due to the perception of guidelines as authoritative risk-benefit recommendations.

5.1.5 Impact of the findings

Our findings show that the clinical trials, systematic reviews, and guidelines that were analysed, showed considerable variations and shortcomings in descriptions of adverse effects associated with treatment, as illustrated in Figure 2. If this is representative for other therapeutic areas, the implications are that poor reporting of adverse effects in clinical trials, and the subsequent selective descriptions in systematic reviews and clinical guidelines, may give a biased presentation of treatment risk in documents aimed at health practitioners.

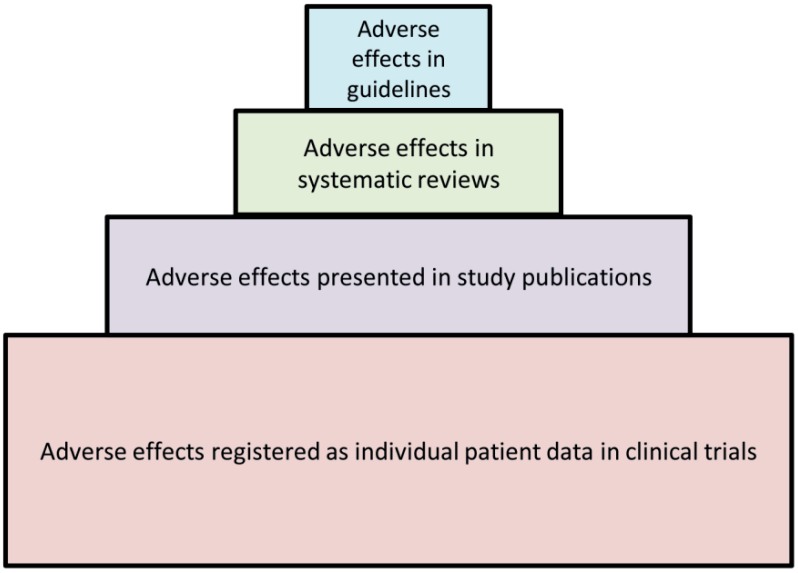

Figure 2. Differences in extent of adverse effects data provided at different documentation levels.

The systematic review and meta-analysis on corticosteroid bleeding risk (Paper I) identified a statistically significant risk increase of 40% compared to placebo. The majority of cases had occurred in hospitalized patients. The review has contributed to the knowledge on a long-debated topic where conclusions have differed, and has been cited 119 times by December 2020. On the basis of Paper I, The Stockholm Drug Committee has stated that use of corticosteroids in ambulatory patients does not warrant ulcer prophylaxis with proton pump inhibitors (212).

The analyses in Papers II and III contribute to the increasing awareness that clinical trial findings of adverse effects are strongly dependent on methodology and publication issues, which must be taken into account when assessing study results. Application of available tools for quality assessment of publications have a limited value and do not address differences in study designs, inclusion criteria, adverse effects definitions or extent of adverse effects descriptions.

The analysis of adverse effects descriptions in clinical depression guidelines (Paper IV) addresses a field that has not been researched previously, and where there are no established standards. Our findings of guideline differences and selective information on adverse effect risks are relevant to guideline societies, authors, health practitioners as guideline users, and patients as treatment recipients.

5.2 Methodological considerations

5.2.1 Analysis of gastrointestinal bleeding risk associated with corticosteroid use (Paper I)

In the literature search for clinical trials of corticosteroids (Paper I), we included randomized, double-blind, controlled trials. We excluded trials where only one dose had been administered; however, the exposure criterion can be debated. All corticosteroids were included, however, the type of corticosteroid, doses and treatment duration varied between studies. Indiscriminate inclusion of all corticosteroids and almost all treatment lengths has been practiced in earlier reviews (143-145), and we chose to apply essentially similar inclusion criteria. There was no uniform definition of gastrointestinal bleeding. We included cases with descriptions of blood in stool, gastrointestinal bleeding or perforation, haematemesis, and melena, while other gastrointestinal symptoms, such as epigastric pain, were excluded. The heterogeneity is an obvious source of bias in all reviews performed on this topic. It can be debated whether the variations are too great for a comprehensive analysis. In retrospect, additional analyses could have been performed on subgroups of studies with similar inclusion criteria, duration, and disease criteria. We cannot exclude the possibility that analysis by corticosteroid, treatment duration, or definition of gastrointestinal bleeding, would have resulted in a different conclusion. However, the incidence rates in placebo and treatment groups within each study are presumably based on identical criteria and reporting rates. We performed subgroup analyses according to treatment indication and whether patients were ambulatory or hospitalised, but some heterogeneity probably remained with regard to other variables.

Many clinical trials, including the studies included in Paper I, report few, or none, adverse events, which pose methodological challenges in meta-analyses (131). Some systematic reviews and meta-analyses have excluded trials that reported zero events of the adverse effects in question (15, 130), even though a finding of zero events may be highly relevant in assessments of risk. In previous systematic reviews on corticosteroid-induced gastrointestinal bleeding, both Messer et al., and Conn and Blitzer, excluded studies with none or insufficient information on adverse effects. Studies that described other adverse effects, but did not mention peptic ulcer or gastrointestinal bleeding, were included on the assumption that no cases had occurred (143, 145). In the Conn and Poynard review, studies

were included only if complications had been reported in both treatment and placebo groups, and statements that no complications had been observed were interpreted as zero cases of gastrointestinal bleeding (144).

In Paper I, the meta-analysis was performed using the Mantel-Haenszel random effect method, which make adjustments for heterogeneity. The Mantel-Haenszel method is recommended by Cochrane if there are few events in the study groups (213).

In a systematic review and meta-analysis of the risk of clinically important gastrointestinal bleeding associated with corticosteroid use in critically ill patients, Butler et al. raise some methodological issues with regard to Paper I (181, 182). They argue that our inclusion of infants with bronchopulmonary dysplasia had considerable impact on the results, and that exclusion of this patient group led to a reduced odds ratio. In our review, the subgroup analysis where patients with bronchopulmonary dysplasia had been excluded still resulted in a statistically increased risk of gastrointestinal bleeding, with OR 1.29 (95% CI 1.07 to 1.55). Like us, Butler et al. found an increased risk of clinically important gastrointestinal bleeding in critically ill patients, with RR 1.26 and 95% CI 1.01 to 1.57.

Butler et al. also comment on the omission of systematic assessments of risks of bias, random errors, and heterogeneity by funnel plot analysis, and GRADE assessments of included studies in Paper I (182). Funnel plot analysis is a scatter plot of the sample size and confidence intervals of the studies included in the meta-analysis, and is expected to result in a symmetrical funnel-like image, due to larger confidence intervals in the smaller trials, narrower confidence intervals in the larger trials, and an even distribution of results around the summary estimate. A skewed funnel plot can indicate publication bias or heterogeneity between studies (214). GRADE assessments are described in Section 1.5. These are formal issues that are increasingly expected of high-quality systematic reviews, however, at the time, our bias analysis by recording adverse effects monitoring methods, and definitions and selection criteria for adverse effects, were judged to be sufficient. It can be argued that this approach provide more precise descriptions of processes relevant to identification and descriptions when the review topic is adverse effects. Like Butler et al. (181), we noted considerable heterogeneity in the data, but publication bias associated with selective

reporting and different criteria for defining adverse effects will not be identified by funnel plot analysis.

5.2.2 Analysis of descriptions of gastrointestinal bleeding or perforation in corticosteroid trials, and assessment of the CONSORT Harms checklist as a quality tool (Paper II)

In Paper II, we noted considerable differences between publications with regard to descriptions of possible gastrointestinal bleeding. We applied identical inclusion criteria as in Paper I, described in Section 5.2.1. Any ambiguity in the case descriptions were discussed by at least two authors. The aim was to include all cases of diagnosed gastrointestinal bleeding; however, we cannot exclude the possibility that some cases of gastrointestinal bleeding were undiagnosed in the included trials.

Adverse effects data from the publications was extracted for further analysis by type and frequency. All identified publications were analysed by at least two authors independently, however, we cannot exclude the possibility that some clinical trials, or textual content may have been overlooked due to the search strategy, errors in judgment with regard to inclusion criteria, or failure to identify relevant text content.

To analyse whether presentation of adverse effects data followed the recommendations in the CONSORT Harms criteria, we developed a data extraction form to reduce variations in judgment, before text extraction by two authors independently. The CONSORT Harms recommendations were not developed as a scoring tool, and like several other researchers who have used the tool as a checklist (98, 99, 102), we found the recommendations ambiguous. Several criteria in the CONSORT Harms include more than one parameter, while the data extraction form was binary (criterion fulfilled 0/1). Interpretation of article texts with regard to the scoring system proved challenging and depended largely on reviewer judgments, and misunderstandings and errors in judgment cannot be ruled out.

Reviewer agreement was analysed by use of Gwet's agreement coefficient with first-order chance correction AC1 (Gwet's AC1), and showed large variations, though the overall agreement was fair to moderate. Analysis of reviewer agreement is often done by using the analysis known as Cohen's kappa, which has inherent and highly relevant methodological issues: With a high level of reviewer agreement, where both reviewers agree or disagree on

multiple scores, the resulting kappa statistic will be low due to imbalances in the 2x2 table, referred to as the kappa paradox (173, 174). The Gwet's AC1 has been found to be less affected by skewed distribution of agreement (174, 215).

5.2.3 Identification of TADS trial publications and assessment of adverse effects reporting (Paper III).

Identification of trial publications from the TADS trial (Paper III) proved to be challenging due to several potential authors, and difficulties in differentiating between publications of trial data and publications that cited the primary papers in the initial searches. We did not identify any comprehensive list of publications with data from the TADS trial on the ClinicalTrials.gov website or on the Duke University TADS trial website (https://tads.dcri.org/). In order not to miss any publications, we performed an extensive hand search, and selected papers where TADS was mentioned, for manual screening.

Adverse effects data from the publications was extracted for further analysis by type and frequency. All identified publications were analysed by at least two authors independently, however, we cannot exclude the possibility that some clinical trials, or textual content may have been overlooked due to the search strategy, errors in judgment with regard to inclusion criteria, or failure to identify relevant text content.

5.2.4 Analysis of clinical therapy guidelines on depression in children and adolescents (Paper IV)

For assessment of relevant guidelines (Paper IV), we performed extensive searches in literature databases and guideline collections. There are no comprehensive guideline registries or collections. Many guidelines are probably developed and distributed locally or nationally, without publication in international journals. Consequently, identification of relevant guidelines may prove difficult. The guidelines that were identified and included in Paper IV represent several countries and continents, and are probably representative with regard to adverse effects information in guidelines in the field of antidepressant therapy.

In the assessment of the guidelines, we evaluated the content and discussions of adverse effects related to medications. The information provided in the guidelines was classified by organ system, which provided a framework, but did not provide a detailed tool. In our

analysis, any mention of an adverse effect in an organ system would result in a score for fulfilment, and the analysis did not distinguish between mention of one, or many, adverse effects in that organ system. As an example, mention of mania would fulfil the criterion of mentioning psychiatric adverse effects, without including mention of lability, anxiety, or any of the many other possible adverse effects in that group. As a consequence, guidelines could receive scores even though the included information was scarce.

We evaluated the guidelines' presentation of risks versus benefits, by assessing whether the presentation was balanced and gave an extensive discussion of treatment risks. Presentation of risk may take several forms, including overall statements, detailed listings and frequencies of possible adverse effects, mention only of selected adverse effects, advice on handling any, or some, adverse effects, or specific monitoring procedures. To our knowledge, there are no criteria for how risk should be presented in guidelines, or what level or form of risk communication is expected. We exercised considerable judgment as to the sufficiency and balance of the information. For many guidelines, the evaluation was clearly negative as treatment risks were hardly discussed, however, the criteria are not clear, and readers may disagree with our classifications.

6. Conclusions

In our systematic review and meta-analysis (Paper I), we found that the risk of gastrointestinal bleeding or perforation was increased by 40% in patients treated with corticosteroids compared to placebo. The risk was increased in hospitalized patient, and not in ambulant patients due to very low occurrence of bleedings.

In our assessment of the quality of adverse effect reporting in clinical trials, we analysed reporting of gastrointestinal bleeding in corticosteroid trials (Paper II), and found that the clinical trials differed in their monitoring procedures and definitions of what should be classified as a gastrointestinal bleeding or perforation. There were considerable differences in severity thresholds for reporting in the trials. Quality criteria for reporting of harms in clinical trials (CONSORT Harms) have been established, but we, and other researchers, have found considerable shortcomings despite attempts to improve reporting of adverse effects (Paper II). The quality criteria are ambiguous, and do not ensure full reporting of all relevant data.

In the analysis of the reporting of adverse effects from a single study of antidepressant therapy in children and adolescents (TADS) (Paper III), we identified multiple publications and considerable risk of bias associated with reporting thresholds, inclusion criteria, adverse effects definitions, causality assessments, and selective publication of data. A full account of adverse effects was available only for the 12-week phase I of the study.

Analysis of adverse effects information in clinical therapy guidelines on depression in children and adolescents (Paper IV) showed considerable variation in the extent of adverse effects information provided. The guidelines focused on suicidality risk associated with antidepressants. Several acknowledged and common adverse effects were not mentioned in many of the guidelines.

7. Implications and future research

The need for improving the reporting of adverse effects in clinical trials is well documented, but the implications for systematic reviews, guidelines and overall perception of risk are still considerable.

The present clinical trial model, where individual case reports are assessed, interpreted and selected for publication by investigators, has inherent risks of bias. On accepting publication of a clinical trial, journal editors should ensure that descriptions of adverse effects are as complete as possible. The present quality criteria do not address this point sufficiently. In particular, sweeping statements to the effect that no relevant adverse effects were observed or that the treatment was well tolerated should not be accepted unless investigators can document comprehensive monitoring and zero findings. Such generic statements do not represent precise information to prescribers and patients (216). Patients may well have different experiences, and other views on severity and impact on daily life (217). Direct patient reporting by a mobile application is being tested for spontaneous reports (218), but has not been assessed in clinical trials.

Analysis of individual patient data from clinical trials by independent researchers has been shown to identify additional cases of adverse effects, and change risk assessments, compared to data summaries (19, 86). Post-trial follow-up of clinical trials has been proposed as a means to identify long-term risks that become apparent after the trials have ended. Many such studies have linked trial data with subsequent health records, with focus on major, predefined outcomes such as mortality or cancer (202, 203). Regulatory agencies foresee that identification of benefits and harms through traditional randomized clinical trials will need to change. The emergence of personalised medicine, and societal expectations of rapid approvals, will have the effect of reducing the extent and duration of clinical trials and increasing the need for scrutiny of marketed medications (219). Recent years have seen increasing focus on data sharing, where investigators agree to make trial data available to researchers (183, 220). Several data repositories have been established (221), and may eventually play an important role in identification and quantification of adverse effects. There are, however, legal restrictions concerning data sharing. In Europe,

the General Data Protection Regulation (GDPR) place considerable limitations on sharing and transfer of research data, an issue that has yet to be resolved (222).

Analysis of electronic health data, which may include data from clinical trials, is expected to play a major role in future analysis of adverse effects. This approach may potentially make better use of existing data, but pose considerable challenges in the areas of data structure and validation, analytic methodology, privacy issues, and risk of bias, among others (223). In the United States, FDA has established the Sentinel Initiative which monitors medication safety, primarily through analysis of health claims data (224). A similar project, called DARWIN EU, is presently being established in the European Union to set up a system for analysis of so-called “big data”, based on a real-world database network (225, 226). In Denmark, the Danish Medicines Agency has established the Data Analytics Centre (DAC), to analyse large-scale healthcare data with regard to adverse effects among other issues (227). However, such analyses may be more suitable for seeking answers to specific questions than to provide overall risk profiles of new medications.

The present guidelines for performing systematic reviews on adverse effects do not address the variations and shortcomings in underlying clinical trials to any extent. As reviewers have to work with what is, this poses a considerable problem, as we found in Paper II. Increased transparency with regard to the variations in underlying data, and use of more guarded language in statements of results, may at least alert readers as to uncertainties in the conclusions.

There are currently no standards or clear expectations of adverse effects descriptions in clinical therapy guidelines, and the question has not received much attention. The analysis presented in this thesis (Paper IV) may form a basis for a much needed discussion about expectations and content of clinical guidelines.

8. References

1. Ioannidis JP. Adverse events in randomized trials: neglected, restricted, distorted, and silenced. Arch Intern Med. 2009;169(19):1737-9.

2. Cornelius VR, Sauzet O, Williams JE, Ayis S, Farquhar-Smith P, Ross JR, et al. Adverse event reporting in randomised controlled trials of neuropathic pain: considerations for future practice. Pain. 2013;154(2):213-20.

3. Hartung DM, Zarin DA, Guise JM, McDonagh M, Paynter R, Helfand M. Reporting discrepancies between the ClinicalTrials.gov results database and peer-reviewed publications. Ann Intern Med. 2014;160(7):477-83.

4. 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.

5. Scharf O, Colevas AD. Adverse event reporting in publications compared with sponsor database for cancer clinical trials. J Clin Oncol. 2006;24(24):3933-8.

6. de Vries TW, van Roon EN. Low quality of reporting adverse drug reactions in paediatric randomised controlled trials. Arch Dis Child. 2010;95(12):1023-6.

7. Tang E, Ravaud P, Riveros C, Perrodeau E, Dechartres A. Comparison of serious adverse events posted at ClinicalTrials.gov and published in corresponding journal articles. BMC Med. 2015;13:189.

8. Hodkinson A, Gamble C, Smith CT. Reporting of harms outcomes: a comparison of journal publications with unpublished clinical study reports of orlistat trials. Trials. 2016;17(1):207.

9. Eldholm RS, Bergheim TS, Slordal L, Spigset O. What information is given on adverse drug reactions from new drugs? Tidsskr Nor Laegeforen. 2003;123(17):2414-7.

10. Tamura N, Ishiguro C, Matsuda T. Post-approval appending of CSARs to drug package inserts: an analysis of the types of adverse reactions and time to addition. Pharmacoepidemiol Drug Saf. 2015;24(2):166-75.

11. Pinnow E, Amr S, Bentzen SM, Brajovic S, Hungerford L, St George DM, et al. Postmarket Safety Outcomes for New Molecular Entity (NME) Drugs Approved by the Food and Drug Administration Between 2002 and 2014. Clin Pharmacol Ther. 2018;104(2):390-400.

12. Onakpoya IJ, Heneghan CJ, Aronson JK. Worldwide withdrawal of medicinal products because of adverse drug reactions: a systematic review and analysis. Crit Rev Toxicol. 2016;46(6):477-89.
13. Hennessy S, Strom BL. Improving postapproval drug safety surveillance: getting better information sooner. Annu Rev Pharmacol Toxicol. 2015;55:75-87.
14. Schick A, Miller KL, Lanthier M, Dal Pan G, Nardinelli C. Evaluation of Pre-marketing Factors to Predict Post-marketing Boxed Warnings and Safety Withdrawals. Drug Saf. 2017;40(6):497-503.
15. 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):559-70.
16. McGauran N, Wieseler B, Kreis J, Schuler YB, Kolsch H, Kaiser T. Reporting bias in medical research - a narrative review. Trials. 2010;11:37.
17. 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.
18. 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):1752-9.
19. Le Noury J, Nardo JM, Healy D, Jureidini J, Raven M, Tufanaru C, et al. Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. BMJ. 2015;351:h4320.
20. European Medicines Agency finalises review of antidepressants in children and adolescents [press release]. European Medicines Agency, 25.04.2005. Available from: https://www.ema.europa.eu/en/news/european-medicines-agency-finalises-review-antidepressants-childrenand-adolescents. Accessed: 30.12.2020.
21. FDA Launches a Multi-Pronged Strategy to Strengthen Safeguards for Children Treated With Antidepressant Medications [press release]. U.S. Food & Drug Administration, 15.10.2004. Available from: https://wayback.archive-it.org/7993/20170111185928/http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2004/ucm108363.htm. Accessed: 30.12.2020.
22. OCEBM Levels of Evidence Working Group. The Oxford Levels of Evidence v.2.1 Oxford, UK: Oxford Centre for Evidence-Based Medicine, University of Oxford; 2011.

Available from: https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence. Accessed: 30.12.2020.

23. Institute of Medicine (US) Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. In: Graham R, Mancher M, Wolman DM, Greenfield S, Steinberg E, editors. Clinical Practice Guidelines We Can Trust. 2011. Available from: ttps://www.ncbi.nlm.nih.gov/books/NBK209539/. Accessed: 29.12.2020.

24. Hartz I, Skurtveit S, Hjellvik V, Furu K, Nesvag R, Handal M. Antidepressant drug use among adolescents during 2004-2013: a population-based register linkage study. Acta Psychiatr Scand. 2016;134(5):420-9.

25. Avorn J. Two centuries of assessing drug risks. N Engl J Med. 2012;367(3):193-7.

26. Report of the Lancet commission appointed to investigate the subject of the administration of chloroform and other anesthetics from a clinical standpoint. The Lancet. 1893;141(3629):629-38.

27. Milestones in U.S. Food and Drug Law History: U.S. Food & Drug Administration; 2018. Available from: https://www.fda.gov/about-fda/fdas-evolving-regulatory-powers/milestones-us-food-and-drug-law-history. Accessed: 08.12.2020.

28. Doll R. Controlled trials: the 1948 watershed. BMJ. 1998;317(7167):1217-20.

29. Streptomycin treatment of pulmonary tuberculosis. Br Med J. 1948;2(4582):769-82.

30. Hill AB. Medical ethics and controlled trials. Br Med J. 1963;1(5337):1043-9.

31. Barr DP. Hazards of modern diagnosis and therapy: the price we pay. J Am Med Assoc. 1955;159(15):1452-6.

32. Lasagna L. The diseases drugs cause. Perspect Biol Med. 1964;7:457-70.

33. McBride WG. Thalidomide and congenital abnormalities. The Lancet. 1961;278(7216):1358.

34. Kelsey FO. Thalidomide update: regulatory aspects. Teratology. 1988;38(3):221-6.

35. Blythe J. UK government changes yellow card scheme. BMJ [Internet]. 2000; 321:658. Available from: https://www.bmj.com/content/321/7262/658.4. Accessed: 8.12.2020.

36. Putting patient safety first: Uppsala Monitoring Centre; [updated 17.09.2020]. Available from: https://www.who-umc.org/about-us/who-we-are/our-story/. Accessed: 08.12.2020.

37. Pedersen A, Lie AK. Da legemiddelkomiteene kom til Norge. Norsk Farmaceutisk Tidsskrift. 2013(9):24-8.

38. EMEA statement following withdrawal of Vioxx (rofecoxib) [press release]. European Medicines Agency, 06.10.2004. Available from: https://www.ema.europa.eu/en/news/european-medicines-agency-statement-following-withdrawal-vioxx-rofecoxib. Accessed: 07.12.2020.

39. Graham DJ, Campen D, Hui R, Spence M, Cheetham C, Levy G, et al. Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: nested case-control study. Lancet. 2005;365(9458):475-81.

40. Cohen D. European regulatory body recommends suspension of rosiglitazone. BMJ. 2010;341:c5291.

41. Blind E, Dunder K, de Graeff PA, Abadie E. Rosiglitazone: a European regulatory perspective. Diabetologia. 2011;54(2):213-8.

42. Aronson JK. Distinguishing hazards and harms, adverse drug effects and adverse drug reactions : implications for drug development, clinical trials, pharmacovigilance, biomarkers, and monitoring. Drug Saf. 2013;36(3):147-53.

43. Guideline on good pharmacovigilance practices (GVP): Annex I - Definitions. European Medicines Agency; 12.10.2017. Report No.: EMA/876333/2011 Rev 4*. Available from: https://www.ema.europa.eu/en/human-regulatory/post-authorisation/pharmacovigilance/good-pharmacovigilance-practices#final-gvp-annex-i---definitions-section. Accessed: 11.10.2019.

44. Glossary of pharmacovigilance terms: Uppsala Monitoring Centre; [updated 10.11.2020]. Available from: https://www.who-umc.org/global-pharmacovigilance/global-pharmacovigilance/glossary/. Accessed: 31.12.2020.

45. Guide on the interpretation of spontaneous case reports of suspected adverse reactions to medicines. European Medicines Agency; 22.06.2011. Report No.: EMA/CHMP/PhVWP/646186/2010. Available from: https://www.ema.europa.eu/en/documents/report/guide-interpretation-spontaneous-case-reports-suspected-adverse-reactions-medicines_en.pdf. Accessed: 06.02.2020.

46. The Importance of Pharmacovigilance: World Health Organization; 2002. Available from: https://apps.who.int/iris/handle/10665/42493. Accessed: 24.01.2020.

47. Glossary: Cochrane Network. Available from: https://community.cochrane.org/glossary. Accessed: 24.01.2020.

48. Aronson JK, Hauben M. Anecdotes that provide definitive evidence. BMJ. 2006;333(7581):1267-9.

49. Aronson JK. Adverse drug reactions - no farewell to harms. Br J Clin Pharmacol. 2007;63(2):131-5.

50. Singh S, Loke YK. Drug safety assessment in clinical trials: methodological challenges and opportunities. Trials. 2012;13:138.

51. Waller PC. Measuring the frequency of adverse drug reactions. Br J Clin Pharmacol. 1992;33(3):249-52.

52. Kaufman DW, Shapiro S. Epidemiological assessment of drug-induced disease. Lancet. 2000;356(9238):1339-43.

53. Strom BL. What Is Pharmacoepidemiology? In: Strom BL, Kimmel SE, Hennessy S, editors. Pharmacoepidemiology. 6 ed: John Wiley & Sons Ltd.; 2019. p. 1-26. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119413431.ch1. Accessed: 04.01.2021.

54. Strom BL. Overview of Electronic Databases in Pharmacoepidemiology. In: Strom BL, Kimmel SE, Hennessy S, editors. Pharmacoepidemiology. 6 ed: John Wiley & Sons Ltd.; 2019. p. 203-10. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119413431.ch11. Accessed: 04.01.2021.

55. Furu K, Wettermark B, Andersen M, Martikainen JE, Almarsdottir AB, Sørensen HT. The Nordic countries as a cohort for pharmacoepidemiological research. Basic Clin Pharmacol Toxicol. 2010;106(2):86-94.

56. Arlett P, Portier G, de Lisa R, Blake K, Wathion N, Dogne JM, et al. Proactively managing the risk of marketed drugs: experience with the EMA Pharmacovigilance Risk Assessment Committee. Nat Rev Drug Discov. 2014;13(5):395-7.

57. Periodic safety update reports (PSURs): European Medicines Agency. Available from: https://www.ema.europa.eu/en/human-regulatory/post-uthorisation/pharmacovigilance/periodic-safety-update-reports-psurs. Accessed: 14.02.2020.

58. A guideline on Summary of Product Characteristics (SmPC), European Commission. (September 2009). Available from:

https://ec.europa.eu/health//sites/health/files/files/eudralex/vol-2/c/smpc_guideline_rev2_en.pdf. Accessed: 13.02.2020.

59. Cornelius VR, Liu K, Peacock J, Sauzet O. Variation in adverse drug reactions listed in product information for antidepressants and anticonvulsants, between the USA and Europe: a comparison review of paired regulatory documents. BMJ Open. 2016;6(3):e010599.

60. Eriksson R, Aagaard L, Jensen LJ, Borisova L, Horluck D, Brunak S, et al. Discrepancies in listed adverse drug reactions in pharmaceutical product information supplied by the regulatory authorities in Denmark and the USA. Pharmacol Res Perspect. 2014;2(3):e00038.

61. Kesselheim AS, Franklin JM, Avorn J, Duke JD. Speaking the same language? International variations in the safety information accompanying top-selling prescription drugs. BMJ Qual Saf. 2013;22(9):727-34.

62. The Uppsala Monitoring Centre. The use of the WHO-UMC system for standardised case causality assessment: World Health Organization; [updated 05.06.2013]. Available from: https://www.who.int/publications/m/item/WHO-causality-assessment. Accessed: 16.03.2018.

63. Hill AB. The environment and disease: association or causation? Proc R Soc Med. 1965;58:295-300.

64. Miremont-Salamé G, Théophile H, Haramburu F, Bégaud B. Causality assessment in pharmacovigilance: The French method and its successive updates. Therapie. 2016;71(2):179-86.

65. Naranjo CA. A clinical pharmacologic perspective on the detection and assessment of adverse drug reactions. Drug Inf J. 1986;20(4):387-93.

66. Edwards RI. Causality Assessment in Pharmacovigilance: Still a Challenge. Drug Saf. 2017;40(5):365-72.

67. Horwitz RI, Singer BH. Why evidence-based medicine failed in patient care and medicine-based evidence will succeed. J Clin Epidemiol. 2017;84:14-7.

68. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. Lancet. 2017;390(10092):415-23.

69. Ioannidis JP, Mulrow CD, Goodman SN. Adverse events: the more you search, the more you find. Ann Intern Med. 2006;144(4):298-300.

70. 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.

71. Pitrou I, Boutron I, Ahmad N, Ravaud P. Reporting of safety results in published reports of randomized controlled trials. Arch Intern Med. 2009;169(19):1756-61.
72. Shukralla AA, Tudur-Smith C, Powell GA, Williamson PR, Marson AG. Reporting of adverse events in randomised controlled trials of antiepileptic drugs using the CONSORT criteria for reporting harms. Epilepsy Res. 2011;97(1-2):20-9.
73. Wahab IA, Pratt NL, Kalisch LM, Roughead EE. The detection of adverse events in randomized clinical trials: can we really say new medicines are safe? Curr Drug Saf. 2013;8(2):104-13.
74. Ennezat PV, Cosgrove S, Bouvaist H, Marechaux S, Guerbaai RA, Le Jemtel T, et al. From evidence-based medicine to personalized medicine, with particular emphasis on drug-safety monitoring. Arch Cardiovasc Dis. 2017;110(6-7):413-9.
75. Tsang R, Colley L, Lynd LD. Inadequate statistical power to detect clinically significant differences in adverse event rates in randomized controlled trials. J Clin Epidemiol. 2009;62(6):609-16.
76. Onakpoya IJ. Rare adverse events in clinical trials: understanding the rule of three. BMJ Evid Based Med. 2018;23(1):6.
77. CIOMS Working Group VI. Management of safety information from clinical trials. Geneva: Council for International Organizations of Medical Sciences (CIOMS); 2005. Available from: https://cioms.ch/wp-content/uploads/2017/01/Mgment_Safety_Info.pdf. Accessed: 14.01.2019.
78. Bent S, Padula A, Avins AL. Brief communication: Better ways to question patients about adverse medical events: a randomized, controlled trial. Ann Intern Med. 2006;144(4):257-61.
79. Allen EN, Chandler CI, Mandimika N, Leisegang C, Barnes K. Eliciting adverse effects data from participants in clinical trials [Internet]. Cochrane Database Syst Rev. 2018; 1(1):MR000039. Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.MR000039.pub2/full?highlightAbstract=datum%7Cdata%7Cin%7Cparticip%7Celiciting%7Cadvers%7Ctrial%7Celicit%7Ceffects%7Cclinical%7Ctrials%7Ceffect%7Cfrom%7Cclinic%7Cadverse%7Cparticipants. Accessed: 12.11.2019.

80. Rief W, Nestoriuc Y, von Lilienfeld-Toal A, Dogan I, Schreiber F, Hofmann SG, et al. Differences in adverse effect reporting in placebo groups in SSRI and tricyclic antidepressant trials: a systematic review and meta-analysis. Drug Saf. 2009;32(11):1041-56.
81. Schroll JB, Maund E, Gotzsche PC. Challenges in coding adverse events in clinical trials: a systematic review. PLoS One. 2012;7(7):e41174.
82. Crepin S, Villeneuve C, Merle L. Quality of serious adverse events reporting to academic sponsors of clinical trials: far from optimal. Pharmacoepidemiol Drug Saf. 2016;25(6):719-24.
83. 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):72-6.
84. 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.
85. Tonéatti C, Saïdi Y, Meiffrédy V, Tangre P, Harel M, Eliette V, et al. Experience using MedDRA for global events coding in HIV clinical trials. Contemp Clin Trials. 2006;27(1):13-22.
86. Wallach JD, Wang K, Zhang AD, Cheng D, Grossetta Nardini HK, Lin H, et al. Updating insights into rosiglitazone and cardiovascular risk through shared data: individual patient and summary level meta-analyses. BMJ. 2020;368:I7078.
87. Zhou Y, Ke C, Jiang Q, Shahin S, Snapinn S. Choosing Appropriate Metrics to Evaluate Adverse Events in Safety Evaluation. Ther Innov Regul Sci. 2015;49(3):398-404.
88. Hernan MA, Hernandez-Diaz S. Beyond the intention-to-treat in comparative effectiveness research. Clin Trials. 2012;9(1):48-55.
89. Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med. 2004;141(10):781-8.
90. Altman DG. Better reporting of randomised controlled trials: the CONSORT statement. BMJ. 1996;313(7057):570-1.
91. Ioannidis JP, Lau J. Improving safety reporting from randomised trials. Drug Saf. 2002;25(2):77-84.
92. Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. JAMA. 2001;285(4):437-43.

93. Loke YK, Derry S. Reporting of adverse drug reactions in randomised controlled trials - a systematic survey. BMC Clin Pharmacol. 2001;1:3.

94. Wang SM, Han C, Lee SJ, Jun TY, Patkar AA, Masand PS, et al. Efficacy of antidepressants: bias in randomized clinical trials and related issues. Expert Rev Clin Pharmacol. 2018;11(1):15-25.

95. Roest AM, de Jonge P, Williams CD, de Vries YA, Schoevers RA, Turner EH. Reporting Bias in Clinical Trials Investigating the Efficacy of Second-Generation Antidepressants in the Treatment of Anxiety Disorders: A Report of 2 Meta-analyses. JAMA Psychiatry. 2015;72(5):500-10.

96. Maggi CB, Griebeler IH, Dal Pizzol TS. Information on adverse events in randomised clinical trials assessing drug interventions published in four medical journals with high impact factors. Int J Risk Saf Med. 2014;26(1):9-22.

97. Peron J, Maillet D, Gan HK, Chen EX, You B. Adherence to CONSORT adverse event reporting guidelines in randomized clinical trials evaluating systemic cancer therapy: a systematic review. J Clin Oncol. 2013;31(31):3957-63.

98. Smith SM, Wang AT, Katz NP, McDermott MP, Burke LB, Coplan P, et al. Adverse event assessment, analysis, and reporting in recent published analgesic clinical trials: ACTTION systematic review and recommendations. Pain. 2013;154(7):997-1008.

99. Smith SM, Chang RD, Pereira A, Shah N, Gilron I, Katz NP, et al. Adherence to CONSORT harms-reporting recommendations in publications of recent analgesic clinical trials: an ACTTION systematic review. Pain. 2012;153(12):2415-21.

100. Mhaskar RS, Reljic T, Wao H, Kumar A, Miladinovic B, Djulbegovic B. Treatment-related harms: what was planned and what was reported? National Cancer Institute's Co-operative group phase III randomized controlled trials: a systematic review. J Clin Epidemiol. 2014;67(3):354-6.

101. 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.

102. Bagul NB, Kirkham JJ. The reporting of harms in randomized controlled trials of hypertension using the CONSORT criteria for harm reporting. Clin Exp Hypertens. 2012;34(8):548-54.

103. Haidich A-B, Birtsou C, Dardavessis T, Tirodimos I, Arvanitidou M. The quality of safety reporting in trials is still suboptimal: Survey of major general medical journals. J Clin Epidemiol. 2011;64(2):124-35.
104. Lee PE, Fischer HD, Rochon PA, Gill SS, Herrmann N, Bell CM, et al. Published randomized controlled trials of drug therapy for dementia often lack complete data on harm. J Clin Epidemiol. 2008;61(11):1152-60.
105. 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.
106. Sharma T, Guski LS, Freund N, Gotzsche PC. Suicidality and aggression during antidepressant treatment: systematic review and meta-analyses based on clinical study reports. BMJ. 2016;352:i65.
107. Fergusson D, Doucette S, Glass KC, Shapiro S, Healy D, Hebert P, et al. Association between suicide attempts and selective serotonin reuptake inhibitors: systematic review of randomised controlled trials. BMJ. 2005;330(7488):396.
108. Jonville-Bera AP, Giraudeau B, Autret-Leca E. Reporting of drug tolerance in randomized clinical trials: when data conflict with authors' conclusions. Ann Intern Med. 2006;144(4):306-7.
109. Jack RH, Hollis C, Coupland C, Morriss R, Knaggs RD, Butler D, et al. Incidence and prevalence of primary care antidepressant prescribing in children and young people in England, 1998-2017: A population-based cohort study. PLoS Med. 2020;17(7):e1003215.
110. The Norwegian Prescription Database [Internet]. Norwegian Institute of Public Health. Available from: http://www.reseptregisteret.no/. Accessed: 11.12.2020.
111. CHMP meeting on Paroxetine and other SSRIs [press release]. European Medicines Agency, 09.12.2004. Available from: https://www.ema.europa.eu/en/news/chmp-meeting-paroxetine-other-ssris. Accessed: 14.12.2020.
112. Hammad TA, Laughren T, Racoosin J. Suicidality in pediatric patients treated with antidepressant drugs. Arch Gen Psychiatry. 2006;63(3):332-9.
113. Dubicka B, Hadley S, Roberts C. Suicidal behaviour in youths with depression treated with new-generation antidepressants: meta-analysis. Br J Psychiatry. 2006;189:393-8.
114. Bridge JA, Iyengar S, Salary CB, Barbe RP, Birmaher B, Pincus HA, 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):1683-96.

115. March JS, Silva S, Petrycki S, Curry J, Wells K, Fairbank J, 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):807-20.

116. Walkup JT. Antidepressant efficacy for depression in children and adolescents: Industry- and NIMH-funded studies. Am J Psychiatry. 2017;174(5):430-7.

117. NIH Guide: Effectiveness of treatment for adolescents with major depression. National Institute of Mental Health, 28.04.1998. https://grants.nih.gov/grants/guide/notice-files/not98-059.html. Accessed: 20.03.2017.

118. Hengartner MP, Ploderl M. Reply to the Letter to the Editor: "Newer-Generation Antidepressants and Suicide Risk: Thoughts on Hengartner and Ploderl's Re-Analysis". Psychother Psychosom. 2019;88(6):373-4.

119. 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):247-8.

120. Hengartner MP. Methodological Flaws, Conflicts of Interest, and Scientific Fallacies: Implications for the Evaluation of Antidepressants' Efficacy and Harm. Front Psychiatry. 2017;8:275.

121. Healy D, Le Noury J, Jureidini J. Paediatric antidepressants: Benefits and risks. Int J Risk Saf Med. 2019;30(1):1-7.

122. Gunnell D, Saperia J, Ashby D. Selective serotonin reuptake inhibitors (SSRIs) and suicide in adults: meta-analysis of drug company data from placebo controlled, randomised controlled trials submitted to the MHRA's safety review. BMJ. 2005;330(7488):385.

123. Chandler J, Cumpston M, Thomas J, Higgins JPT, Deeks JJ, Clarke MJ. Chapter I: Introduction. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions. 6 ed. 2019. Available from: https://training.cochrane.org/handbook/current/chapter-i. Accessed: 15.07.2020.

124. Courtney DB, Bennett K, Szatmari P. The Forest and the Trees: Evidence-Based Medicine in the Age of Information. J Am Acad Child Adolesc Psychiatry. 2019;58(1):8-15.

125. Clarke M, Brice A, Chalmers I. Accumulating research: a systematic account of how cumulative meta-analyses would have provided knowledge, improved health, reduced harm and saved resources. PLoS One. 2014;9(7):e102670.
126. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA. 2000;283(15):2008-12.
127. Mercuri M, Baigrie B, Upshur REG. Going from evidence to recommendations: Can GRADE get us there? J Eval Clin Pract. 2018;24(5):1232-9.
128. Chou R, Helfand M. Challenges in systematic reviews that assess treatment harms. Ann Intern Med. 2005;142(12 Pt 2):1090-9.
129. Cornelius VR, Perrio MJ, Shakir SA, Smith LA. Systematic reviews of adverse effects of drug interventions: a survey of their conduct and reporting quality. Pharmacoepidemiol Drug Saf. 2009;18(12):1223-31.
130. Zorzela L, Golder S, Liu Y, Pilkington K, Hartling L, Joffe A, et al. Quality of reporting in systematic reviews of adverse events: systematic review. BMJ. 2014;348:f7668.
131. Warren FC, Abrams KR, Golder S, Sutton AJ. Systematic review of methods used in meta-analyses where a primary outcome is an adverse or unintended event. BMC Med Res Methodol. 2012;12:64.
132. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). J Clin Epidemiol. 2011;64(4):407-15.
133. Aronson JK. Metameta-analysis. Br J Clin Pharmacol. 2005;60(2):117-9.
134. 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:g6501.
135. 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):389-97.
136. Loke YK, Price D, Herxheimer A. Systematic reviews of adverse effects: framework for a structured approach. BMC Med Res Methodol. 2007;7:32.

137. Mahady SE, Schlub T, Bero L, Moher D, Tovey D, George J, et al. Side effects are incompletely reported among systematic reviews in gastroenterology. J Clin Epidemiol. 2015;68(2):144-53.
138. Yavchitz A, Ravaud P, Altman DG, Moher D, Hrobjartsson A, Lasserson T, 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.
139. Naci H, Ioannidis JP. How good is "evidence" from clinical studies of drug effects and why might such evidence fail in the prediction of the clinical utility of drugs? Annu Rev Pharmacol Toxicol. 2015;55:169-89.
140. Ioannidis JP. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. Milbank Q. 2016;94(3):485-514.
141. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339:b2535.
142. Zorzela L, Loke YK, Ioannidis JP, Golder S, Santaguida P, Altman DG, et al. PRISMA harms checklist: improving harms reporting in systematic reviews. BMJ. 2016;352:i157.
143. Conn HO, Blitzer BL. Nonassociation of adrenocorticosteroid therapy and peptic ulcer. N Engl J Med. 1976;294(9):473-9.
144. Conn HO, Poynard T. Corticosteroids and peptic ulcer: meta-analysis of adverse events during steroid therapy. J Intern Med. 1994;236(6):619-32.
145. 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):21-4.
146. Saag KG, Furst DE. Major side effects of systemic glucocorticoids [Internet]. UpToDate. Available from: https://www.uptodate.com/contents/major-side-effects-of-systemic-glucocorticoids/print?search=glucocorticoids&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1. Accessed: 20.05.2020.
147. Prednisolon Alternova. Summary of Product Characteristics [Internet]. Alternova AS. 20.08.2019. Available from: https://www.legemiddelsok.no/_layouts/15/Preparatomtaler/Spc/12-8942.pdf. Accessed: 31.12.2020.
148. Dexametason Abcur. Summary of Product Characteristics [Internet]. Abcur AB. 28.04.2020. Available from:

https://www.legemiddelsok.no/_layouts/15/Preparatomtaler/Spc/11-8815.pdf. Accessed: 10.06.2020.

149. Prednisolone [Internet]. IBM Micromedex. Available from: https://www.micromedexsolutions.com/micromedex2/librarian/. Accessed: 10.06.2020.

150. Corticosteroids, general use [Internet]. British National Formulary, Medicines Complete. Royal Pharmaceutical Society. Available from: https://www.medicinescomplete.com/#/content/bnf/_111415959?hspl=corticosteroids#content%2Fbnf%2F_111415959%23section_111415959-1. Accessed: 10.06.2020.

151. Prednisolone Adverse Reactions [Internet]. Clinical Pharmacology. Elsevier. Available from: https://www.clinicalkey.com/pharmacology/monograph/504?sec=monadve#1504. Accessed: 10.06.2020.

152. Shekelle P. Overview of clinical practice guidelines [Internet]. UpToDate. Available from: https://www.uptodate.com/contents/overview-of-clinical-practice-guidelines?search=guidelines&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1#H7. Accessed: 24.06.2019.

153. Qaseem A, Forland F, Macbeth F, Ollenschläger G, Phillips S, van der Wees P. Guidelines International Network: toward international standards for clinical practice guidelines. Ann Intern Med. 2012;156(7):525-31.

154. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011;64(4):383-94.

155. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence - publication bias. J Clin Epidemiol. 2011;64(12):1277-82.

156. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011;64(4):401-6.

157. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. CMAJ. 2010;182(18):E839-42.

158. The AGREE II Instrument: The AGREE Next Steps Consortium; [updated 12.2017]. Available from: http://www.agreetrust.org. Accessed: 15.08.2019.

159. Zafra-Tanaka JH, Goicochea-Lugo S, Villarreal-Zegarra D, Taype-Rondan A. Characteristics and quality of clinical practice guidelines for depression in adults: a scoping review. BMC Psychiatry. 2019;19(1):76.
160. Lin I, Wiles LK, Waller R, Goucke R, Nagree Y, Gibberd M, et al. Poor overall quality of clinical practice guidelines for musculoskeletal pain: a systematic review. Br J Sports Med. 2018;52(5):337-43.
161. Bennett K, Courtney D, Duda S, Henderson J, Szatmari P. An appraisal of the trustworthiness of practice guidelines for depression and anxiety in children and youth. Depress Anxiety. 2018;35(6):530-40.
162. Deng Y, Luo L, Hu Y, Fang K, Liu J. Clinical practice guidelines for the management of neuropathic pain: a systematic review. BMC Anesthesiol. 2016;16:12.
163. Castellani A, Girlanda F, Barbui C. Rigour of development of clinical practice guidelines for the pharmacological treatment of bipolar disorder: systematic review. J Affect Disord. 2015;174:45-50.
164. Acuna-Izcaray A, Sanchez-Angarita E, Plaza V, Rodrigo G, de Oca MM, Gich I, et al. Quality assessment of asthma clinical practice guidelines: a systematic appraisal. Chest. 2013;144(2):390-7.
165. Keating D, McWilliams S, Schneider I, Hynes C, Cousins G, Strawbridge J, et al. Pharmacological guidelines for schizophrenia: a systematic review and comparison of recommendations for the first episode. BMJ Open. 2017;7(1):e013881.
166. Nuckols TK, Anderson L, Popescu I, Diamant AL, Doyle B, Di Capua P, et al. Opioid prescribing: a systematic review and critical appraisal of guidelines for chronic pain. Ann Intern Med. 2014;160(1):38-47.
167. Lenzer J, Hoffman JR, Furberg CD, Ioannidis JP. Ensuring the integrity of clinical practice guidelines: a tool for protecting patients. BMJ. 2013;347:f5535.
168. Kung J, Miller RR, Mackowiak PA. Failure of clinical practice guidelines to meet institute of medicine standards: Two more decades of little, if any, progress. Arch Intern Med. 2012;172(21):1628-33.
169. Guyatt G, Akl EA, Hirsh J, Kearon C, Crowther M, Gutterman D, et al. The vexing problem of guidelines and conflict of interest: a potential solution. Ann Intern Med. 2010;152(11):738-41.

170. Neuman J, Korenstein D, Ross JS, Keyhani S. Prevalence of financial conflicts of interest among panel members producing clinical practice guidelines in Canada and United States: cross sectional study. BMJ. 2011;343:d5621.
171. 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):1712-5.
172. Ploderl M, Hengartner MP. Guidelines for the pharmacological acute treatment of major depression: conflicts with current evidence as demonstrated with the German S3-guidelines. BMC Psychiatry. 2019;19(1):265.
173. Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. Br J Math Stat Psychol. 2008;61(Pt 1):29-48.
174. Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC Med Res Methodol. 2013;13:61.
175. Conn HO, Poynard T. Adrenocorticosteroid therapy and peptic-ulcer disease. N Engl J Med. 1984;310(3):201-2.
176. Conn HO, Poynard T. Adrenocorticosteroid administration and peptic ulcer: a critical analysis. J Chronic Dis. 1985;38(6):457-68.
177. Moller MH, Ioannidis JPA, Darmon M. Are systematic reviews and meta-analyses still useful research? We are not sure. Intensive Care Med. 2018;44(4):518-20.
178. Huang HY, Andrews E, Jones J, Skovron ML, Tilson H. Pitfalls in meta-analyses on adverse events reported from clinical trials. Pharmacoepidemiol Drug Saf. 2011;20(10):1014-20.
179. Berlin JA, Crowe BJ, Whalen E, Xia HA, Koro CE, Kuebler J. Meta-analysis of clinical trial safety data in a drug development program: answers to frequently asked questions. Clin Trials. 2013;10(1):20-31.
180. Lièvre M, Cucherat M, Leizorovicz A. Pooling, meta-analysis, and the evaluation of drug safety. Curr Control Trials Cardiovasc Med. 2002;3(1):6.
181. Butler E, Møller MH, Cook O, Granholm A, Penketh J, Rygård SL, et al. The effect of systemic corticosteroids on the incidence of gastrointestinal bleeding in critically ill adults: a systematic review with meta-analysis. Intensive Care Med. 2019;45(11):1540-9.

182. Butler E, Møller MH, Cook O, Granholm A, Penketh J, Rygård SL, et al. Corticosteroids and risk of gastrointestinal bleeding in critically ill adults: Protocol for a systematic review. Acta Anaesthesiol Scand. 2018;62(9):1321-6.
183. Lehman R, Loder E. Missing clinical trial data. BMJ. 2012;344:d8158.
184. Jakobsen JC, Katakam KK, Schou A, Hellmuth SG, Stallknecht SE, Leth-Moller K, et al. Selective serotonin reuptake inhibitors versus placebo in patients with major depressive disorder. A systematic review with meta-analysis and Trial Sequential Analysis. BMC Psychiatry. 2017;17(1):58.
185. Favier R, Crepin S. The reporting of harms in publications on randomized controlled trials funded by the "Programme Hospitalier de Recherche Clinique," a French academic funding scheme. Clin Trials. 2018;15(3):257-67.
186. Whittington CJ, Kendall T, Fonagy P, Cottrell D, Cotgrove A, Boddington E. Selective serotonin reuptake inhibitors in childhood depression: systematic review of published versus unpublished data. Lancet. 2004;363(9418):1341-5.
187. Spielmans GI, Spence-Sing T, Parry P. Duty to Warn: Antidepressant Black Box Suicidality Warning Is Empirically Justified. Front Psychiatry. 2020;11:18.
188. Ebrahim S, Montoya L, Kamal El Din M, Sohani ZN, Agarwal A, Bance S, et al. Randomized trials are frequently fragmented in multiple secondary publications. J Clin Epidemiol. 2016;79:130-9.
189. Spielmans GI, Biehn TL, Sawrey DL. A case study of salami slicing: pooled analyses of duloxetine for depression. Psychother Psychosom. 2010;79(2):97-106.
190. Kennard BD, Silva SG, Tonev S, Rohde P, Hughes JL, Vitiello B, et al. Remission and recovery in the Treatment for Adolescents with Depression Study (TADS): acute and long-term outcomes. J Am Acad Child Adolesc Psychiatry. 2009;48(2):186-95.
191. March JS, Silva S, Petrycki S, Curry J, Wells K, Fairbank J, et al. The Treatment for Adolescents With Depression Study (TADS): long-term effectiveness and safety outcomes. Arch Gen Psychiatry. 2007;64(10):1132-43.
192. Senn S. Statistical pitfalls of personalized medicine. Nature. 2018;563(7733):619-21.
193. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ. 2006;332(7549):1080.

194. Vitiello B, Silva SG, Rohde P, Kratochvil CJ, Kennard BD, Reinecke MA, et al. Suicidal events in the Treatment for Adolescents With Depression Study (TADS). J Clin Psychiatry. 2009;70(5):741-7.
195. Emslie G, Kratochvil C, Vitiello B, Silva S, Mayes T, McNulty S, et al. Treatment for Adolescents with Depression Study (TADS): safety results. J Am Acad Child Adolesc Psychiatry. 2006;45(12):1440-55.
196. 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):1035-43.
197. Cipriani A, Zhou X, Del Giovane C, Hetrick SE, Qin B, Whittington C, et al. Comparative efficacy and tolerability of antidepressants for major depressive disorder in children and adolescents: a network meta-analysis. Lancet. 2016;388(10047):881-90.
198. Munkholm K, Paludan-Muller AS, Boesen K. Considering the methodological limitations in the evidence base of antidepressants for depression: a reanalysis of a network meta-analysis. BMJ Open. 2019;9(6):e024886.
199. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:l4898.
200. The TADS Team. Protocol RFP-NIH-NIMH 98-DS-0008 Treatment for Adolescents with Depression Study (TADS): Duke Clinical Research Institute; 2000. Available from: tads.dcri.org. Accessed: 20.03.2017.
201. Zink RC, Marchenko O, Sanchez-Kam M, Ma H, Jiang Q. Sources of Safety Data and Statistical Strategies for Design and Analysis: Clinical Trials. Ther Innov Regul Sci. 2018;52(2):141-58.
202. Fitzpatrick T, Perrier L, Shakik S, Cairncross Z, Tricco AC, Lix L, et al. Assessment of Long-term Follow-up of Randomized Trial Participants by Linkage to Routinely Collected Data: A Scoping Review and Analysis. JAMA Netw Open. 2018;1(8):e186019.
203. Llewellyn-Bennett R, Edwards D, Roberts N, Hainsworth AH, Bulbulia R, Bowman L. Post-trial follow-up methodology in large randomised controlled trials: a systematic review. Trials. 2018;19(1):298.
204. Prozac. Summary of Product Characteristics [Internet]. Eli Lilly and Company Limited. 18.03.2018. Available from:

https://www.medicines.org.uk/emc/product/3768/smpc#PRODUCTINFO. Accessed: 31.05.2018.

205. Hetrick SE, McKenzie JE, Cox GR, Simmons MB, Merry SN. Newer generation antidepressants for depressive disorders in children and adolescents [Internet]. Cochrane Database Syst Rev. 2012;11:CD004851. Available from: https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.CD004851.pub3/full?highlightAbstract=generation%7Cadolescents%7Cin%7Cfor%7Cgenerat%7Cantidepressants%7Cdepress%7Cdepressive%7Cnewer%7Cdisorders%7Cdisord%7Cchildren%7Cfour%7Cdepressiv%7Cadolesc%7Cantidepress%7Cchild. Accessed: 27.03.2017.

206. Barbui C, Esposito E, Cipriani A. Selective serotonin reuptake inhibitors and risk of suicide: a systematic review of observational studies. CMAJ. 2009;180(3):291-7.

207. Ebrahim S, Bance S, Athale A, Malachowski C, Ioannidis JP. Meta-analyses with industry involvement are massively published and report no caveats for antidepressants. J Clin Epidemiol. 2016;70:155-63.

208. Our vision: Cochrane. Available from: https://www.cochrane.org/about-us. Accessed: 19.06.20.

209. LeFevre M. From Authority- to Evidence-Based Medicine: Are Clinical Practice Guidelines Moving us Forward or Backward? Ann Fam Med. 2017;15(5):410-2.

210. Cosgrove L, Shaughnessy AF, Wheeler EE, Krimsky S, Peters SM, Freeman-Coppadge DJ, et al. From caveat emptor to caveat venditor: time to stop the influence of money on practice guideline development. J Eval Clin Pract. 2014;20(6):809-12.

211. Ahn R, Woodbridge A, Abraham A, Saba S, Korenstein D, Madden E, et al. Financial ties of principal investigators and randomized controlled trial outcomes: cross sectional study. BMJ. 2017;356.

212. Region Stockholms läkemedelskommittés expertgrupp för mag- och tarmsjukdomar. Steroidbehandling motiverar inte ulkusprofylax med protonpumpshämmare. Janusinfo Region Stockholm; [updated 15.09.2020]. Available from: https://www.janusinfo.se/behandling/expertgruppsutlatanden/magochtarmsjukdomar/magochtarmsjukdomar/steroidbehandlingmotiverarinteulkusprofylaxmedprotonpumpshammare.5.78ae827d1605526e94b61df2.html. Accessed: 21.12.2020.

213. Deeks JJ, Higgins JPT, Altman DG. Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane

Handbook for Systematic Reviews of Interventions. 6.1 ed: Cochrane; 2020. Available from: www.training.cochrane.org/handbook. Accessed: 28.10.2020.

214. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ. 2011;343:d4002.

215. Govatsmark RE, Sneeggen S, Karlsaune H, Slørdahl SA, Bønaa KH. Interrater reliability of a national acute myocardial infarction register. Clin Epidemiol. 2016;8:305-12.

216. Sacks CA, Miller PW, Longo DL. Talking about Toxicity - "What We've Got Here Is a Failure to Communicate". N Engl J Med. 2019;381(15):1406-8.

217. Rolfes L, Haaksman M, van Hunsel F, van Puijenbroek E. Insight into the Severity of Adverse Drug Reactions as Experienced by Patients. Drug Saf. 2019.

218. Defer G, Le Caignec F, Fedrizzi S, Montastruc F, Chevanne D, Parienti JJ, et al. Dedicated mobile application for drug adverse reaction reporting by patients with relapsing remitting multiple sclerosis (Vigip-SEP study): study protocol for a randomized controlled trial. Trials. 2018;19(1):174.

219. Lumpkin MM, Eichler HG, Breckenridge A, Hamburg MA, Lönngren T, Woods K. Advancing the science of medicines regulation: the role of the 21st-century medicines regulator. Clin Pharmacol Ther. 2012;92(4):486-93.

220. Open data: BMJ. Available from: https://www.bmj.com/open-data. Accessed: 29.10.2020.

221. Banzi R, Canham S, Kuchinke W, Krleza-Jeric K, Demotes-Mainard J, Ohmann C. Evaluation of repositories for sharing individual-participant data from clinical studies. Trials. 2019;20(1):169.

222. Bovenberg J, Peloquin D, Bierer B, Barnes M, Knoppers BM. How to fix the GDPR's frustration of global biomedical research. Science. 2020;370(6512):40-2.

223. Moore TJ, Furberg CD. Electronic Health Data for Postmarket Surveillance: A Vision Not Realized. Drug Saf. 2015;38(7):601-10.

224. CDER Conversation: The FDA's Sentinel Initiative. U.S. Food & Drug Administration; [updated 21.11.2017]. Available from: https://www.fda.gov/drugs/news-events-human-drugs/cder-conversation-fdas-sentinel-initiative. Accessed: 05.01.2021.

225. Ten recommendations to unlock the potential of big data for public health in the EU [press release]. European Medicines Agency, 20.01.2020. Available from:

https://www.ema.europa.eu/en/news/ten-recommendations-unlock-potential-big-data-public-health-eu. Accessed: 05.01.2021.

226. Big data: European Medicines Agency. Available from: https://www.ema.europa.eu/en/about-us/how-we-work/big-data. Accessed: 05.01.2021.

227. DAC Data Analytics Centre: Lægemiddelstyrelsen. Available from: https://laegemiddelstyrelsen.dk/en/about/DAC/. Accessed: 05.01.2021.

9. Errata

Navn kandidat: Tone Westergren

Avhandlingstittel: Reporting of adverse effects in clinical trials, systematic reviews, and guidelines. How events are lost along the evidence chain

Forkortelser for type rettelser:

Cor – korrektur

Celtf – endring av sidelayout eller tekstformat

Side	Linje	Referanse	Originaltekst	Type rettelse	Korrigert tekst
9	7		… TADS- studien …	Cor	… TADS-studien …|
9	17		… kvalitets vurdere …	Cor	… kvalitetsvurdere …
12	15		… clinical trials has …	Cor	… clinical trials have …
13	8		… a medication's …	Cor	… a medication's …
20	25		… individual patients data …	Cor	… individual patients' data …
23	28		… observational studies, …	Cor	… observational studies; …
29	10		… Dukes University, …	Cor	… Duke University, …
54	23	9	… Hvilken informasjon gis om bivirkninger av nye legemidler? …	Cor	… What information is given on adverse drug reactions from new drugs? …
56	32	37	…Tidsskrift 2013…	Cor	…Tidsskrift. 2013…
57	9	40	… of rosiglitazone. 2010;341:c5291. …	Cor	… of rosiglitazone. BMJ. 2010;341:c5291. …

10. Papers I-IV

Paper I

BMJ Open

Corticosteroids and risk of gastrointestinal bleeding: a systematic review and meta-analysis

Sigrid Narum,[1,2] Tone Westergren,[3] Marianne Klemp[4,5]

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

► Prepublication history and additional material is available. To view please visit the journal (http://dx.doi.org/10.1136/bmjopen-2013-004587).

Received 30 November 2013
Revised 23 April 2014
Accepted 25 April 2014

For numbered affiliations see end of article.

Correspondence to
Sigrid Narum;
sigrid.narum@diakonsyk.no

ABSTRACT

Objective: To assess whether corticosteroids are associated with increased risk of gastrointestinal bleeding or perforation.

Design: Systematic review and meta-analysis of randomised, double-blind, controlled trials comparing a corticosteroid to placebo for any medical condition or in healthy participants. Studies with steroids given either locally, as a single dose, or in crossover studies were excluded.

Data sources: Literature search using MEDLINE, EMBASE and Cochrane Database of Systematic Reviews between 1983 and 22 May 2013.

Outcome measure: Outcome measures were the occurrence of gastrointestinal bleeding or perforation. Predefined subgroup analyses were carried out for disease severity, use of non-steroidal anti-inflammatory drugs (NSAIDs) or gastroprotective drugs, and history of peptic ulcer.

Results: 159 studies (N=33 253) were included. In total, 804 (2.4%) patients had a gastrointestinal bleeding or perforation (2.9% and 2.0% for corticosteroids and placebo). Corticosteroids increased the risk of gastrointestinal bleeding or perforation by 40% (OR 1.43, 95% CI 1.22 to 1.66). The risk was increased for hospitalised patients (OR 1.42, 95% CI 1.22 to 1.66). For patients in ambulatory care, the increased risk was not statistically significant (OR 1.63, 95% CI 0.42 to 6.34). Only 11 gastrointestinal bleeds or perforations occurred among 8651 patients in ambulatory care (0.13%). Increased risk was still present in subgroup analyses (studies with NSAID use excluded; OR 1.44, 95% CI 1.20 to 1.71, peptic ulcer as an exclusion criterion excluded; OR 1.47, 95% CI 1.21 to 1.78, and use of gastroprotective drugs excluded; OR 1.42, 95% CI 1.21 to 1.67).

Conclusions: Corticosteroid use was associated with increased risk of gastrointestinal bleeding and perforation. The increased risk was statistically significant for hospitalised patients only. For patients in ambulatory care, the total occurrence of bleeding or perforation was very low, and the increased risk was not statistically significant.

Strengths and limitations of this study

- This systematic review and meta-analysis includes published results from 159 trials with a total of 33 253 participants.
- The strength of this systematic review is the size due to the inclusion of a large number of randomised controlled trials that allowed for subgroup analyses.
- Limitations are the possible loss of relevant studies due to the selected search strategy, the quality of adverse event reporting in the primary studies and the heterogeneity in the patient populations.

INTRODUCTION

The association between corticosteroid use and gastrointestinal (GI) adverse effects, including bleeding or perforation, has been a source of debate since the 1950s.[1–3] Since GI bleeding and perforation are rare events, no single randomised controlled trial has been large enough to show any increased risk for GI bleeding with the use of corticosteroids. Adverse effects and studies of rare events can often be effectively investigated in observational studies. Thus controlled, observational studies may be the method of choice to detect rare adverse effects. For corticosteroid use, several observational studies have been performed to clarify whether corticosteroids do induce GI bleeding or not, but there is still uncertainty whether this adverse effect is a result of corticosteroid use, use of other medications, underlying disease or other causes.[4–7]

This lack of evidence is reflected in the literature. In databases and in product monographs for corticosteroids, peptic ulcer disease and GI bleeding may or may not be described as possible adverse effects.[8–13] Similarly, in clinical recommendations, an association between corticosteroid use and peptic ulcer has been described as unlikely, and the value of antiulcer prophylaxis has been questioned due to a low bleeding risk.[8–13] Although many gastroenterologists consider corticosteroids as not having ulcerogenic properties, a recent survey has shown that corticosteroids are still considered

ulcerogenic by a majority of physicians and that a majority of practitioners would treat corticosteroid users with ulcer prophylaxis.[14] This uncertainty may have consequences for clinical recommendations and treatment guidelines, and is the main reason why we performed this systematic review.[15–18]

GI bleeding, bleeding peptic ulcer and perforation are feared complications of peptic ulcer disease, associated with considerable morbidity and mortality.[19 20] Non-steroidal anti-inflammatory drugs (NSAID) use and *Helicobacter pylori* infection are the most important risk factors for peptic ulcer disease. Bleeding or perforation is also seen as complications to stress ulcers among patients with critical illness in intensive care units. GI bleeding and perforation are assumed to occur when ulcers erode into underlying vessels. The mechanism by which corticosteroids might induce GI bleeding or perforation has not been fully established, but corticosteroids may impair tissue repair, thus leading to delayed wound healing.[8] In addition, the anti-inflammatory and analgesic properties of corticosteroids may mask symptoms of gastroduodenal ulcers and ulcer complications and thus possibly delay diagnosis.

The aim of this systematic review was to examine whether use of systemic corticosteroids was associated with an increased risk of peptic ulcer complications such as GI bleeding or perforation. Since observational studies have not been conclusive, we have chosen to include published studies with a randomised, controlled design.

METHODS

Search strategy and selection criteria

A systematic literature search was performed to identify randomised, double-blind, placebo controlled trials in which any systemic corticosteroid (defined as oral, intravenous or intramuscular) or a placebo had been administrated to randomly selected groups of patients in the treatment of a medical disorder or to healthy participants.

We searched the databases MEDLINE and EMBASE with no language restrictions between 1983 (since date of the latest review by Conn and Poynard)[1] and 30 June 2011 using the following text words: (β methasone/ or dexamethasone/ or methylprednisolone/ or prednisolone/ or prednisone/ or triamcinolone/ or cortisone/ or hydrocortisone/). The search was limited to randomised controlled trials, humans, double blind.mp and placebo.mp. An updated search was performed on 22 May 2013. For the full search strategy, see online supplementary file 1. An additional search was performed in the Cochrane Database of Systematic Reviews for corticosteroids and the following text words: traumatic injury, sepsis/septic shock, meningitis, bronchopulmonary dysplasia, liver diseases, lung diseases and rheumatoid arthritis. Only results fully reported in journal articles in English, German or any Scandinavian language were considered for inclusion. Whenever a title or abstract suggested that a randomised, double-blind, placebo controlled trial comparing a corticosteroid to placebo had been performed, the full text version was reviewed for documentation of GI adverse events. Articles with documentation of GI adverse effects or with assessment of adverse event monitoring described in the methods section were included. Titles, abstracts and full-text articles were evaluated and reviewed for inclusion by at least two of the authors. Disagreements were resolved by consensus among the authors.

Methodological quality assessment of eligible trials was carried out by including only randomised, double-blind studies.[21] In most studies, there was no specific description of randomisation and allocation concealment, blinding methods or handling of withdrawals. Authors' description of randomisation and double blinding was assumed to be valid. We used intention-to-treat data when available. All types of comedications were allowed if administered systematically to both groups or as a part of standard care. No medical disorder or age groups were excluded. When medications known to induce GI symptoms, such as NSAIDs or acetylsalicylic acid (ASA), had been used, they were analysed as covariables. We excluded trials with a crossover design because of potential difficulties in assessment between the treatment groups. Trials in which the steroid was given as a single dose were also excluded due to the generally short follow-up.

Data extraction and outcomes reporting

For the diagnosis of complications of gastroduodenal ulcers, such as occult or visible blood in stool, GI bleeding, haematemesis, melena and GI perforation, the investigators' diagnoses were accepted as valid without requiring specific criteria or methods. Outcomes like dyspepsia, gastritis, duodenitis and epigastric pain were not included, and nor was necrotising enterocolitis. For assessment of GI bleeding or perforation as an adverse effect, the number of events should be reported in the results section as text or in a table. Events reported as percentages only were calculated into numbers by us. In some trials, other adverse effects were reported in the results section but no GI bleeding was listed. These studies were included only if adverse event monitoring was described in the methods section or if it was judged reasonable to expect from the adverse event monitoring system that any GI adverse effects would have been recorded.

We recorded information on study characteristics and demographics such as publication year, corticosteroid use, indication for treatment, use of concomitant medications, description of adverse effects, study size, duration of treatment and follow-up. Severity of disease was assessed by assuming that patients needing hospitalisation were sicker than patients in ambulatory care. Information regarding exclusion from study by ongoing, recent or a history of peptic ulcer disease was also recorded. Risk of bias was assessed by recording which methods were used for monitoring, definition and description of adverse effects, randomisation and selection criteria.

Statistical analysis

The relative frequencies of the adverse effects were compared in the placebo and the corticosteroid group(s) using conventional statistics and meta-analysis. Subgroup analyses were performed for different predefined variables, such as for concomitant NSAID use, for use of gastroprotective drugs (proton pump inhibitors, H2 blockers or antacids) and for disease severity.

All meta-analytic calculations were made with RevMan (V.5.2) using the Mantel-Haenszel method with the random effects model. For other statistics, SPSS (V.20) was used. For binary outcomes, we calculated ORs and 95% CIs. All analyses were two tailed, with an α of 0.05.

RESULTS

Literature search and study selection

The search process identified 3483 records from database searches and 15 studies were retrieved by hand searching. A total of 159 articles fitted our inclusion criteria and were included in the review. Further details regarding study inclusion and exclusion are shown in figure 1. We performed an updated search on 22 May 2013 and retrieved three additional studies reporting confirmed GI bleeding events. The new studies did not change the results.

Characteristics of included studies

In this systematic review, 159 studies were included. The main medical conditions were severe infections, lung diseases, traumatic injuries and prevention of bronchopulmonary dysplasia in premature infants. Further details regarding the disease groups are shown in table 1.

The corticosteroids used were dexamethasone (55), prednisolone (30), methylprednisolone (29), prednisone (22), hydrocortisone (16) and other steroids or combinations (7). The sample size ranged from 15 to 10 008 people, with a median sample size of 86. The median duration of treatment was 8.5 days (range 1–1095 days), and the median follow-up period was 56 days (range 1–1155 days). There was a trend towards shorter duration of treatment and follow-up during hospital treatment (6 and 33 days) compared with ambulant treatment (14 and 58 days; p=0.061 and 0.057, respectively). The adverse effects were described as any form of bleeding in 59 studies (upper/lower, minor, haematemesis, melena, visible/occult blood in stool), perforation in seven studies (perforated gastric ulcer, ileum perforation) and bleeding and perforation in six studies. The definition of GI bleeding varied between the studies, from bleeding requiring transfusion to occult blood in stool.

Altogether, 72 (45.3%) studies reported GI bleeding or perforation as an adverse effect (67 hospitalised, 5 ambulant). In the 87 studies without reporting of any GI bleeding or perforation, peptic ulcer was described in only four studies.

Use of concomitant medication was described in 135 studies (84.9%). In addition, use of concomitant medication was likely in many of the remaining 24 studies, as a consequence of diagnoses such as acute respiratory distress syndrome, bronchopulmonary dysplasia and traumatic injury to the head or spine. Use of medication

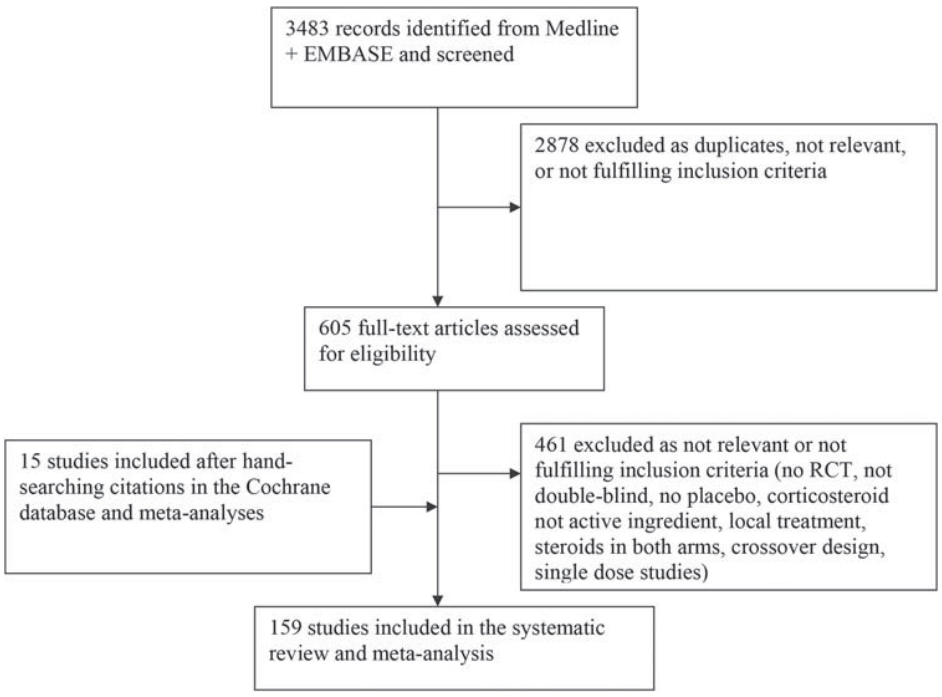

Figure 1 Flowchart for the selection of eligible studies.

Table 1 Medical conditions in which corticosteroids were tested, with number of studies, number of participants and number of adverse effects

	Hospitalised					Ambulant					Total
		Number of participants		Number of adverse effects			Number of participants		Number of adverse effects		Number of participants
Disease	Number of studies	Ster	Plac	Ster	Plac	Number of studies	Ster	Plac	Ster	Plac	Sum
Traumatic injury (brain, spinal cord, multiple)	9	5821	5790	95	75	0	–	–	–	–	11 611
Meningitis	18	1589	1549	110	91	0	–	–	–	–	3138
Sepsis/septic shock	7	482	449	32	28	0	–	–	–	–	931
Bronchopulmonary dysplasia	21	1508	1487	155	85	0	–	–	–	–	2995
Liver diseases*	4	150	114	26	15	3	705	709	5	1	1678
Lung diseases %	20	1149	1105	8	3	7	537	544	0	0	3335
Rheumatoid arthritis	0	–	–	–	–	5	283	279	1	2	562
Miscellaneous†	24	1743	1666	46	24	41	2806	2788	2	0	9003
Sum	103	12 442	12 160	472	321	56	4331	4320	8	3	33 253

Grouping by treatment level was based on statements in the reports and, if there was no indication of treatment level, on clinical judgement. Patients with traumatic injury, meningitis, sepsis/septic shock and bronchopulmonary dysplasia were defined as hospitalised.
*Hepatitis, liver cirrhosis, acute hepatic failure. % Asthma, ARDS, bronchiolitis, chronic obstructive pulmonary disease, pneumonia, tuberculosis, ventilator weaning.
†Miscellaneous diseases as stated in the original reports (number of studies in brackets): acute otitis media, adhesive capsulitis, allergic rhinitis, Alzheimer's disease, Behçet's syndrome, Bell's palsy (2), carpal tunnel syndrome, cerebral infarction, chronic fatigue syndrome, coronary artery bypass grafting (2), cysticercus granuloma with seizures, depression, Duchenne's muscular dystrophy, emesis (9), erysipelas, facial nerve paralysis (2), glaucoma, Grave's orbitopathy, Guillain-Barré syndrome (2), healthy postmenopausal women, Henoch Schonlein purpura (2), herpes zoster (3), IgA nephropathy, intracerebral haemorrhage (2), leprosy, lumbar disc surgery, migraine headaches, multiple sclerosis (3), myocardial infarction (2), postinfectious irritable bowel syndrome, preeclampsia, (pre)terminal cancer (2), aphthous stomatitis, sinonasal polyposis, sinusitis, Sjøgren's syndrome, Sydenham's Chorea in children, tetanus, tonsillectomy (2), tuberculous pericarditis in HIV, typhoid fever, urticaria, vestibular neuritis, withdrawal headache.
ARDS, acute respiratory distress syndrome; Plac, placebo; Ster, corticosteroids.

for any pre-existing diseases was sparsely described. Concomitant use of NSAIDs/ASA was described in 19 studies (bronchopulmonary dysplasia, rheumatoid arthritis, miscellaneous and sepsis in 9 studies, 5 studies, 4 studies and 1 study, respectively), and use of gastroprotective drugs was described in 14 studies. In addition, use of concomitant drugs 'according to standard clinical practice', etc, which may potentially include use of gastroprotective drugs, was described in 12 studies.

Peptic ulcer, ongoing, recent or previous, was an exclusion criterion in 53 (33.3%) of the studies. In the majority of studies (85, 53.5%), the authors reported no effect of corticosteroids on the primary efficacy endpoint. Study-specific characteristics are shown in table 2 and in online supplementary file 2.

Risk of GI bleeding or perforation

The analysis included 33 253 participants (16 773 received corticosteroids and 16 480 received placebo). Of those, 804 patients (480 receiving a corticosteroid and 324 receiving a placebo) were reported to have a GI bleeding or perforation, which comprises 2.4% of the study participants (2.9% and 2% for corticosteroids and placebo, respectively). Overall, meta-analysis of all the included studies showed a 40% increased OR of experiencing GI bleeding or perforation among corticosteroid users compared with placebo users (OR 1.43, 95% CI 1.22 to 1.66; figure 2, and see online supplementary file 3). Subgroup analysis for each disease group showed a trend towards an increased risk of GI bleeding or perforation in seven out of eight subgroups, but the result was statistically significant only for premature infants in prevention of bronchopulmonary dysplasia (1.83, 1.37 to 2.43).

Sensitivity analyses

Data from subgroup analyses are shown in table 3.

Subgroup analysis of studies with hospitalised patients showed an increased risk of developing GI bleeding or perforation (OR 1.42, 95% CI 1.22 to 1.66). There was also a trend towards increased risk for patients in ambulatory care (1.63, 0.42 to 6.34), but this result was not significant. When the studies with documentation of concomitant NSAID use were excluded, a significant difference between corticosteroid and placebo with respect to GI bleeding or perforation was still present (1.44, 1.20 to 1.71). When all studies of premature infants in prevention of bronchopulmonary dysplasia were excluded from the analysis (assuming NSAIDs were given in all studies), the results were lower but still significant (1.29, 1.07 to 1.55). When studies with peptic ulcer as an exclusion criterion and studies with concomitant use of gastroprotective drugs were subsequently excluded from the analyses, there was little change in the risk of bleeding or perforation in the remaining studies (table 3). The majority of the adverse effects occurred in hospitalised patients. Only 11 GI bleedings or perforations occurred among 8651 patients in

Table 2 Study-specific characteristics

Summary of study characteristics	Studies total	Studies with bleeding	Studies without bleeding	p Values
Studies included (%)	159	72 (45.3)	87 (54.7)	
Year of publication, median		1998	1999	0.109
Description of adverse effect (%)				
Bleeding		59 (81.9)	0	
Perforation		7 (9.7)	0	
Bleeding and perforation		6 (8.3)	0	
Peptic ulcer only			4	
Level of care (%)				
Hospitalised	103	67 (93.1)	36 (41.4)	<0.001
Ambulant	56	5 (6.9)	51 (58.6)	
Use of concomitant medication (%)				
No concomitant medication described	24	11 (15.3)	13 (14.9)	
Concomitant medication described	135	61 (84.7)	74 (85.1)	
NSAIDs/ASA	19	11 (15.3)	8 (9.2)	0.326
Gastroprotective drugs	14	12	2	0.002
Exclusion criteria (%)				
Recent/ongoing peptic ulcer	36	14 (19.4)	22 (25.3)	0.237
Previous/history of peptic ulcer	17	6 (8.3)	11 (12.6)	
Study size, number of participants				
Median (IQR)	86 (49.0–181.0)	100 (60.3–246.5)	70 (40.0–128.0)	0.104
Duration of treatment, days				
Median (IQR)	8.5 (3.3–28.0)	6.0 (3.0–12.0)	14 (4.0–45.0)	0.061
Duration of follow-up, days				
Median (IQR)	56 (21.0–243.8)	33 (21.0–180.0)	58 (19.5–286.5)	0.057

ASA, acetylsalicylic acid; NSAIDs, non-steroidal anti-inflammatory drugs; PPIs, proton pump inhibitors.

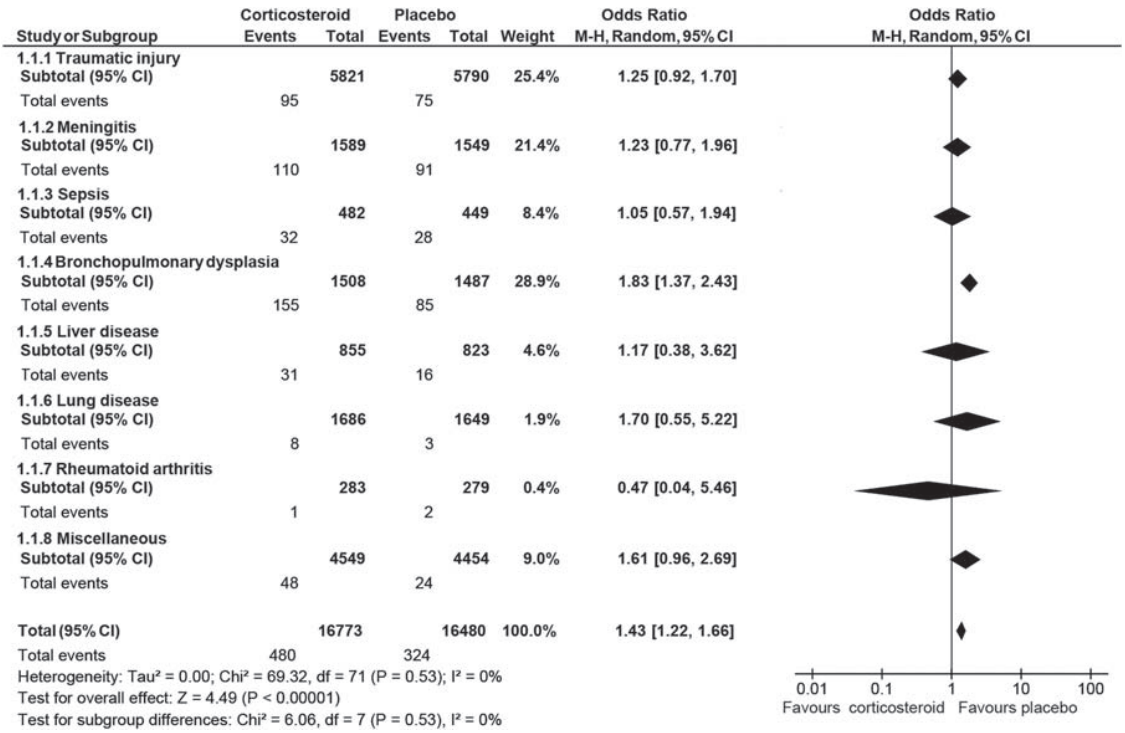

Figure 2 Summary of pooled results. Gastrointestinal bleeding in corticosteroid users versus placebo users. The Mantel-Haenszel (M-H) method with a random effects model was used.

ambulatory care (0.13%), compared with 793 GI bleeds or perforations among 24 602 hospitalised patients (3.22%; p<0.001; table 1). The absolute risk of experiencing GI bleeding, events per 1000 patients, was 1.8 for ambulant patients given steroids, compared with 0.7 for ambulant patients given placebo (table 3). In contrast, hospitalised patients had a much higher risk, 37.9/1000 for steroids and 26.4/1000 for placebo.

DISCUSSION

The overall findings of this systematic review show that the use of corticosteroids may increase the OR by 40% for GI bleeding or perforation. The increased risk, however, was limited to hospitalised patients. For patients in ambulatory care, who had a very low absolute occurrence of GI bleeding or perforation, the increased risk was not statistically significant. The results persisted when high-risk/low-risk patients (concomitant NSAID use, previous peptic ulcer as an exclusion criterion and use of gastroprotective drugs) were excluded, indicating the robustness of the results.

Comparison with other studies

Previously published meta-analyses addressing whether corticosteroid use predisposes people to GI bleeding or perforation have shown conflicting results.[1–3] In two

Table 3 Summary of subgroup analyses

	Number of studies	Number of patients	OR (95% CI)	Events steroids/ placebo	Events per 1000 patients steroids/ placebo
Hospitalised	103	24 602	1.42 (1.22 to 1.66)	472/321	37.9/26.4
Ambulant	56	8651	1.63 (0.42 to 6.34)	8 / 3	1.8/0.7
NSAID use not documented	140	30 874	1.44 (1.20 to 1.71)	372/248	23.9/16.2
NSAID use documented	19	2379	1.30 (0.81 to 2.07)	108/76	90.2/64.4
Peptic ulcer as an exclusion criterion not documented	106	25 760	1.47 (1.21 to 1.78)	421/284	32.5/22.1
Peptic ulcer as an exclusion criterion documented	53	7493	1.26 (0.81 to 1.96)	59/40	15.4/10.9
Gastroprotective drugs not documented	145	31 759	1.42 (1.21 to 1.67)	442/299	27.6/19.0
Gastroprotective drugs documented	14	1494	1.29 (0.62 to 2.69)	38/25	50.6/33.6
Bronchopulmonary dysplasia excluded	138	30 258	1.29 (1.07 to 1.55)	325/239	21.3/15.9

NSAID, non-steroidal anti-inflammatory drug.

meta-analyses, Conn and colleagues[1 2] concluded that there was no increased risk of peptic ulcer, GI bleeding or perforation by corticosteroid use. In contrast, Messer *et al*[3] found an increased incidence of peptic ulcer and GI bleeding. In a subgroup analysis by Conn and Blitzer,[2] however, there was a significantly higher rate of GI bleeding from an unknown site among corticosteroid users compared with controls. In his second paper, steroid users had more GI adverse effects (ulcers, symptoms of ulcers, bleeding, erosions and perforation) than controls, but because of subgroup analyses only and no pooling of results, no differences emerged as statistically significant.[1] These meta-analyses of randomised controlled trials, which included published literature up to 1983, show how different inclusion criteria, selection criteria, data handling and interpretation of results may give totally different results and conclusions. Newer Cochrane meta-analyses have addressed the question in selective patient populations (meningitis, traumatic brain injury and preterm infants). These analyses show a trend[22–24] or a statistically significant increase[25] in the risk ratio of experiencing GI bleeding, with the included studies and results being similar to the subgroups in our study.

In our study, we included the literature published from 1983 until now. With 33 253 participants from double-blind, randomised, controlled trials, this is the largest meta-analysis analysing whether corticosteroids increase the risk of GI bleeding. Owing to the large size of our study, findings that were seen as trends in other reviews or went unnoticed because of many subgroup analyses have emerged as a significant increase in risk, despite the non-significant increase in occurrence in all subgroups except prevention of bronchopulmonary dysplasia in premature infants. Surprisingly, peptic ulcers were hardly listed as an adverse effect in the included studies, in contrast to the studies in the previous reviews by Conn and Messer. One explanation may be the differences in disease panorama and the discovery and treatment of *H. pylori*. The true occurrence of peptic ulcer may also have been underestimated in the studies because of the heavy medication and intensive care treatment.

Strengths and limitations of this review

In many reviews, the use of narrow inclusion criteria and wide exclusion criteria makes the population homogeneous, but with rare events there is a high risk of insignificant results. In our analysis, inclusion of all studies with a relevant design, including those with concomitant medications and studies with zero events, may reflect more realistic treatment conditions and may contribute to the validity of the findings. Owing to the large size of included studies in our review, we were able to perform predefined subgroup analyses assessing the severity of disease (ie, assessed as hospitalised or as ambulant treatment), use of NSAIDs or gastroprotective drugs and documentation of peptic ulcer as exclusion criteria. To the best of our knowledge, this is the first systematic review to indicate that disease severity might influence the risk of GI bleeding or perforation in corticosteroid users.

The main limitations of this review are the possible loss of relevant studies due to the selected search strategy, the quality of the included trials and the heterogeneity of the included patient populations. However, we believe the findings to be robust, despite this, due to the large number of included studies and participants. Randomised controlled trials are designed to show the effect of treatment, not to detect adverse effects, which in many studies were sparsely reported or not reported at all. However, since we included only double-blind studies with placebo control, we suspect similar under-reporting in both study groups. To minimise the risk of bias according to adverse effect detection and reporting, we recorded the methods used for monitoring adverse effects and how the adverse effect was defined in the primary studies. We found diversity in the definitions of GI bleeding (ie, from occult blood in stool to GI bleeding requiring transfusion or hospital stay). In addition, differences in the methods used for monitoring adverse effects may explain the risk differences found in the sensitivity analyses. A more rigorous follow-up of patients in intensive care units may thus explain some of the risk differences found between hospitalised patients and patients in ambulatory care. This makes comparisons of absolute risk differences between different disease groups difficult.

We aimed to include all disease groups, but still some groups may be under-represented (ie, rheumatoid arthritis, organ transplanted patients) since corticosteroid use is standard treatment and is no longer compared with placebo in randomised controlled trials. Patients included in randomised controlled trials may differ from patients excluded from trial participation, and may be healthier, without previous peptic ulcer. This may underestimate the true effect of corticosteroids on GI bleeding and perforation within the population. In the majority of the included studies, the use of concomitant medications was described. Concomitant medication was related to the study indication (eg, treatment of trauma, meningitis, sepsis, bronchopulmonary dysplasia, etc), in contrast to medications for coexisting diseases, which were hardly mentioned. Concomitant use of gastroprotective drugs and descriptions of supportive care according to standard clinical practice, which may include the use of gastroprotective drugs, was declared only in a minority of the studies. In addition, the potential under-reporting and undisclosed use of gastroprotective drugs may have underestimated the true risk of having GI bleeding with steroid use. Undisclosed use of gastroprotective drugs may especially apply to ambulant treated patients with dyspepsia. Owing to the short-term treatment and inclusion of only double-blind studies, we assume that the effect of the possible under-reporting and undisclosed use of gastroprotective drugs was not

substantial. Despite the heterogeneity of the included studies and a potential of under-reporting of adverse effects, there is a consistency across the analyses of an increased frequency of GI bleeding and perforation among patients given steroids compared with patients given placebo. This indicates the robustness of the analysis.

Clinical implications of this review

Our analysis shows that the increased risk of GI bleeding or perforation applied to hospitalised patients only, indicating that additional factors to corticosteroid therapy, such as disease severity or advanced medical treatment, may make some patients more vulnerable to adverse events to corticosteroid use. One possible explanation is that the bleedings and perforations seen among hospitalised patients may be complications to the stress ulcers seen in critically ill patients.

Owing to diagnoses or illnesses like traumatic injury, meningitis and sepsis, we suspected a substantial portion of the hospitalised patients to have been critically ill. To scrutinise this further, we aimed to do separate analyses of critically ill patients or treatment in intensive care units, but lack of descriptions of critical illness or treatment in intensive care units in the included studies made us use hospitalisation and ambulant treatment as surrogate markers for disease severity.

Stress ulcers occur in response to severe physiological stress in critically ill patients. Although the mechanism is not completely understood, it involves decreased mucosal blood flow and subsequent tissue ischaemia, resulting in breakdown of mucosal defences, allowing physiological factors to produce injury and ulceration.[26] Many risk factors for stress ulcer bleeding have been proposed,[26 27] but only mechanical ventilation and coagulopathy have been documented as independent risk factors. Despite this evidence, several studies have shown that acid-suppressive therapy is used as stress ulcer prophylaxis in hospital wards and outpatient settings.[15–17] This has been described as an inappropriate use of acid-suppressive therapy. An explanation to this overuse may be the discrepancy between product monographs and databases/clinical recommendations in assessment of peptic ulcer disease and GI bleeding as possible adverse effects to corticosteroids.[8 11–13]

Our analysis also showed increased risk of GI bleeding or perforation among patients in ambulatory care, but the result was not significant due to a very low occurrence of GI bleeding and perforation. According to our results, the data are insufficient to conclude whether corticosteroids are associated with GI bleeding or perforation among patients in ambulatory care. It seems reasonable to conclude that the absolute risk of GI bleeding is very low in the ambulatory setting.

Author affiliations
[1]Center for Psychopharmacology, Diakonhjemmet Hospital, Oslo, Norway
[2]Department of Pharmacology, Oslo University Hospital, Oslo, Norway
[3]Department of Pharmacology, Regional Medicines Information & Pharmacovigilance Centre (RELIS), Oslo University Hospital, Oslo, Norway
[4]Norwegian Knowledge Centre for the Health Services, Oslo, Norway
[5]Department of Pharmacology, University of Oslo, Oslo, Norway

Contributors SN, TW and MK conceived the study, performed the systematic review and data extraction, analysed the data and drafted the manuscript. All authors had full access to the data and take responsibility for the integrity of the data and accuracy of the analysis. SN is the guarantor.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The dataset is available from the corresponding author.

REFERENCES

1. Conn HO, Poynard T. Corticosteroids and peptic ulcer: meta-analysis of adverse events during steroid therapy. *J Intern Med* 1994;236:619–32.
2. Conn HO, Blitzer BL. Nonassociation of adrenocorticosteroid therapy and peptic ulcer. *N Engl J Med* 1976;294:473–9.
3. Messer J, Reitman D, Sacks HS, *et al*. Association of adrenocorticosteroid therapy and peptic-ulcer disease. *N Engl J Med* 1983;309:21–4.
4. Nielsen GL, Sorensen HT, Mellemkjoer L, *et al*. Risk of hospitalization resulting from upper gastrointestinal bleeding among patients taking corticosteroids: a register-based cohort study. *Am J Med* 2001;111:541–5.
5. Hernandez-Diaz S, Rodriguez LA. Steroids and risk of upper gastrointestinal complications. *Am J Epidemiol* 2001;153:1089–93.
6. Piper JM, Ray WA, Daugherty JR, *et al*. Corticosteroid use and peptic ulcer disease: role of nonsteroidal anti-inflammatory drugs. *Ann Intern Med* 1991;114:735–40.
7. de Abajo FJ, Gil MJ, Bryant V, *et al*. Upper gastrointestinal bleeding associated with NSAIDs, other drugs and interactions: a nested case-control study in a new general practice database. *Eur J Clin Pharmacol* 2013;69:691–701.
8. Corticosteroids: Introduction (Martindale: The Complete Drug Reference). 2010. [cited 2013 Feb 18]. http://online.lexi.com/lco/action/doc/retrieve/docid/martindale_f/1353587
9. Saag KG, Furst DE. Major side effects of systemic glucocorticoids. 2012 October 11. [cited 5 Mar 2013]. http://www.uptodate.com
10. Joint Formulary Committee. British National Formulary. Side-effects of corticosteroids. 2013. [cited 14 Aug 2013]. http://www.medicinescomplete.com
11. Summary of product characteristics; Dexamethasone Tablets BP 2.0 mg. 2012 November 13. [cited 14 Aug 2013]. http://www.mhra.gov.uk/home/groups/spcpil/documents/spcpil/con1369200053458.pdf
12. Summary of product characteristics; Prednisolone 5 mg Tablets. 2011 July 5. [cited 14 Aug 2013]. http://www.mhra.gov.uk/home/groups/spcpil/documents/spcpil/con1364186684018.pdf
13. Gold Standard, Inc. Prednisone. *Clinical pharmacology* 2012 September 8. [cited 11 Oct 2013]. http://www.clinicalpharmacology.com
14. Martinek J, Hlavova K, Zavada F, *et al*. "A surviving myth"—corticosteroids are still considered ulcerogenic by a majority of physicians. *Scand J Gastroenterol* 2010;45:1156–61.
15. Heidelbaugh JJ, Inadomi JM. Magnitude and economic impact of inappropriate use of stress ulcer prophylaxis in non-ICU hospitalized patients. *Am J Gastroenterol* 2006;101:2200–5.
16. Zink DA, Pohlman M, Barnes M, *et al*. Long-term use of acid suppression started inappropriately during hospitalization. *Aliment Pharmacol Ther* 2005;21:1203–9.
17. Munson JC, Wahl PM, Daniel G, *et al*. Factors associated with the initiation of proton pump inhibitors in corticosteroid users. *Pharmacoepidemiol Drug Saf* 2012;21:366–74.

18. van der Goes MC, Jacobs JW, Boers M, *et al*. Monitoring adverse events of low-dose glucocorticoid therapy: EULAR recommendations for clinical trials and daily practice. *Ann Rheum Dis* 2010;69: 1913–19.
19. Ahsberg K, Ye W, Lu Y, *et al*. Hospitalisation of and mortality from bleeding peptic ulcer in Sweden: a nationwide time-trend analysis. *Aliment Pharmacol Ther* 2011;33:578–84.
20. Lanas A, Garcia-Rodriguez LA, Polo-Tomas M, *et al*. The changing face of hospitalisation due to gastrointestinal bleeding and perforation. *Aliment Pharmacol Ther* 2011;33:585–91.
21. Juni P, Witschi A, Bloch R, *et al*. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282:1054–60.
22. Alderson P, Roberts I. Corticosteroids for acute traumatic brain injury. *Cochrane Database Syst Rev* 2005;(1):CD000196.
23. Brouwer MC, McIntyre P, de Grans J, *et al*. Corticosteroids for acute bacterial meningitis. *Cochrane Database Syst Rev* 2010;(9): CD004405.
24. Halliday HL, Ehrenkranz RA, Doyle LW. Late (>7 days) postnatal corticosteroids for chronic lung disease in preterm infants. *Cochrane Database Syst Rev* 2009;(1):CD001145.
25. Halliday HL, Ehrenkranz RA, Doyle LW. Early (<8 days) postnatal corticosteroids for preventing chronic lung disease in preterm infants. *Cochrane Database Syst Rev* 2010;(1):CD001146.
26. Stollman N, Metz DC. Pathophysiology and prophylaxis of stress ulcer in intensive care unit patients. *J Crit Care* 2005;20:35–45.
27. Cook DJ, Fuller HD, Guyatt GH, *et al*. Risk factors for gastrointestinal bleeding in critically ill patients. Canadian Critical Care Trials Group. *N Engl J Med* 1994;330:377–81.

Paper II

Journal of Clinical Epidemiology 94 (2018) 19–26

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Characterization of gastrointestinal adverse effects reported in clinical studies of corticosteroid therapy

Tone Westergren[a,*], Sigrid Narum[b], Marianne Klemp[c]

[a]*Deptartment of Pharmacology, Regional Medicines Information & Pharmacovigilance Centre (RELIS), Oslo University Hospital Rikshospitalet, PO Box 4950 Nydalen, 0424 Oslo, Norway*

[b]*Centre for Psychopharmacology, Diakonhjemmet Hospital, PO Box 23 Vinderen, 0319 Oslo, Norway*

[c]*Department of Pharmacology, University of Oslo, PO Box 1057 Blindern, Oslo 0316, Norway*

Accepted 30 October 2017; Published online 4 November 2017

Abstract

Objectives: To examine whether 159 studies included in a previous meta-analysis reported on gastrointestinal bleeding or perforation in accordance with the CONSORT extension for reporting harms outcomes (CONSORT Harms recommendations checklist); whether differences were associated with funding source, journal, or publication year; and whether the CONSORT Harms checklist is a suitable tool for evaluation of adverse effects reporting.

Study Design and Setting: Articles were assessed for fulfillment of the CONSORT Harms recommendations, funding source, publication type, and year. Agreement between reviewers was assessed by comparing scores for each study.

Results: The mean CONSORT Harms score was 5.25 out of 10 (standard deviation ± 2.09). Most studies included information on participant withdrawals (133 studies, 83.6%), absolute risk of gastrointestinal bleeding or perforation (130 studies, 81.8%), and how harms-related information was collected (118 studies, 74.2%). Reporting of gastrointestinal bleeding or perforation increased with higher scores (odds ratio 1.173, $P = 0.042$). There was no significant association between CONSORT Harms score achieved and publication year or funding source, but there was a trend toward higher scores in studies published in the major medical journals (score difference 0.78, $P = 0.052$). Definitions of gastrointestinal bleeding differed between studies. Reviewer agreement was fair to moderate with large variations.

Conclusion: Few studies in the systematic review received high scores using the CONSORT Harms criteria. Most studies reported on the most important criteria regarding risk of gastrointestinal bleeding or perforation. Reviewer agreement showed large variations due to imprecise texts and ambiguous criteria. Routine scoring according to fulfillment of the CONSORT Harms recommendations would be inadvisable without qualified judgment.

Keywords: Gastrointestinal hemorrhage; Glucocorticoids; Pharmacovigilance; Adverse drug reaction reporting systems/standards; Guideline adherence; Systematic review

1. Introduction

Most randomized clinical trials are designed to evaluate efficacy of drug treatment and therefore provide better assessments of benefits than risks. However, comprehensive and reliable data on both benefits and risks are necessary to make a balanced risk/benefit assessment. Safety and risk of adverse effects cannot be thoroughly explored in short-term studies that include only a limited patient group. Shortcomings in adverse effects monitoring and reporting may lead to inadequate assessments and lower estimates of serious harm [1,2]. The Declaration of Helsinki [3], developed by the World Medical Association, states that medical research may only be conducted if the importance of the objective outweighs the risk to the research subjects. Failure to identify relevant risks may lead to research projects with an unacceptable risk/benefit balance. If problems of unsystematic monitoring or reporting of adverse effects are added to inconsistent or heterogeneous data, it may be

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of interest: None.

* Corresponding author. Tel.: +47 23075439; fax: + 47 23075510.

E-mail address: twesterg@ous-hf.no (T. Westergren).

https://doi.org/10.1016/j.jclinepi.2017.10.018

What is new?

Key findings

- Studies included in a previous review on the risk of gastrointestinal bleeding or perforation during corticosteroid treatment were analyzed with regard to quality of adverse effects monitoring and reporting. The studies were assessed and scored using the CONSORT Harms criteria with 10 recommendations.
- The mean score was 5.3/10, which means that several CONSORT Harms criteria were not met for many of the studies.
- Only 59/159 studies were identified as having addressed and monitored gastrointestinal adverse effects judging from the study descriptions. However, the absolute risk of gastrointestinal bleeding or perforation was found in 130/159 studies. Gastrointestinal adverse effects were reported in studies that did not specify the intention to address them.
- Reporting of gastrointestinal bleeding or perforation was higher in studies with higher CONSORT Harms criteria scores, compared to studies with lower scores. Exclusion of studies with low scores would have led to exclusion of relevant findings of cases of gastrointestinal bleeding or perforation.

What this adds to what was known?

- This is the first in-depth analysis of adverse effects monitoring and reporting in studies that were included in a systematic review of risk of adverse effects. Data on adverse effects could be found in most studies, although several aspects of adverse effects reporting were heterogeneous and unsystematic with regard to definitions, method of monitoring, and data analysis. The study provides an insight into the realities of summarizing literature on adverse effects.

What is the implication and what should change now?

- Use of checklists is advocated for quality assessment of included clinical trials in reviews and meta-analyses. Routine scoring of clinical studies using CONSORT Harms criteria for harms assessment would be inadvisable without adding qualified judgment on the study in question. Published clinical studies generally do not fulfill all criteria in the CONSORT Harms checklist. Too narrow inclusion criteria may eliminate studies that are suboptimal with regard to adverse effects reporting but still give relevant information on adverse effects. Conclusions about adverse effects made in systematic reviews should take the variability and heterogeneous reporting of adverse effects in the underlying data into account.

impossible to draw conclusions regarding risk from single or pooled clinical studies and to perform systematic reviews for risk/benefit assessment [4]. Weaknesses in the original reporting of adverse effects will be magnified when those reports form the basis of meta-analyses and systematic reviews.

The Cochrane Handbook for Systematic Reviews of Interventions emphasizes the need for careful scrutiny of the studies' intensity of monitoring adverse effects and clarity of reporting [5]. The PRISMA statement addresses improvements in quality and transparency of systematic reviews by way of minimum standards for reporting [6], and a PRISMA Harms extension for systematic reviews has been developed [7]. Use of the GRADE approach for grading quality of evidence [8] is recommended by the *British Medical Journal* and the Cochrane Collaboration among others but does not provide the tools for a detailed examination of the adverse effects reporting. Other methods have been proposed to address the quality of adverse effects or harms reporting. Both the CONSORT group and Cochrane Adverse Effects Methods group advocate the use of checklists when including clinical studies for methodology review or meta-analysis [5,9]. These are lists of recommendations describing what information should be included in various parts of the article. A commonly cited example is the CONSORT checklist [10], an initiative to improve the reporting of clinical trials, with an added 10 recommendations for reporting harms published in 2004 [11], often referred to as the CONSORT Harms criteria. Others have developed extended, more detailed versions [12]. The McMaster tool for assessing quality of harms assessment and reporting in study reports (McHarm) covers many of the same recommendations as the CONSORT checklist [13]. As yet, there are no universally endorsed instruments for assessing risk of bias with regard to adverse effects or harms in clinical trials or systematic reviews.

We have previously published a systematic review and meta-analysis of corticosteroid use and risk of gastrointestinal bleeding or perforation [14], including only randomized, double-blinded studies. During the review process, it became clear that the included 159 studies varied widely in their descriptions and methods of adverse effects reporting and definitions of gastrointestinal bleeding, although they all fulfilled our inclusion criteria. We have analyzed the studies to examine whether they reported on adverse effects in accordance with the CONSORT extension for reporting harms outcomes (referred to as CONSORT Harms criteria) [11]; to examine whether any differences could be linked to variables such as funding source, journal quality, or publication year; and to evaluate whether the CONSORT Harms criteria are a suitable tool for evaluation of the quality of adverse effects reporting in clinical trials.

2. Methods

2.1. Study data and criteria assessments

One hundred fifty-nine articles included in a previous systematic review and meta-analysis of corticosteroids and risk of gastrointestinal bleeding or perforation (referred to as gastrointestinal bleeding in the rest of the article) were included in the analysis [14]. A standardized checklist and data extraction form was prepared based on the CONSORT Harms recommendations. The criteria were discussed by all authors to arrive at a common understanding.

We collected data on 10 different outcomes using the CONSORT Harms recommendations (Table 1, recommendations 1−10). Two recommendations (3 and 8) were modified to include gastrointestinal adverse events only, to reflect whether the study specified assessments of adverse gastrointestinal effects associated with study treatment, as this was the adverse effect addressed in the meta-analysis [14]. The relevant text from the articles was extracted and scored as 0 or 1 by two of the authors independently (S.N. and T.W.) by interpreting the checklist criteria in relation to the article text. Several CONSORT Harms criteria included two or more parameters. If the article met any one of the criteria that CONSORT included for a topic, it was counted as fulfilled for that topic, as has been practiced elsewhere [15]. The scores were discussed and a final score was decided. The reviewers were not blinded to the name of the journal or the authors.

All articles were assessed for reporting of gastrointestinal adverse effects, funding source, publication type, and year. Studies scoring 8, 9, or 10 were classified as high-score studies. Studies scoring 3 or less were classified as low-score studies. To see if publication of the CONSORT Harms extension in 2004 had led to improved adverse effects reporting, studies were grouped according to publication year ($\leq$2004, $\geq$2005). Studies with industry coauthorship or donations of product or money were classified as industry sponsored. Studies published by one of the five major medical journals (*Lancet*, *British Medical Journal*, *New England Journal of Medicine*, *Journal of the American Medical Association*, and *Annals of Internal Medicine*) were analyzed separately.

2.2. Reviewer agreement

Agreement between reviewers was used as an indicator of the ease of use and suitability of the CONSORT Harms recommendations. Interrater agreement for each study was analyzed using Gwet's agreement coefficient with first-order chance correction, AC1 (value 0-1) [16]. Interrater agreement for each CONSORT Harms criterion across the 159 studies was analyzed using Gwet's AC1 [17].

2.3. Statistical analysis

We calculated correlations using the Pearson chi-square test and differences in scores using the *t*-test for equality of means. Logistic regression analysis was used to examine the relationship between CONSORT Harms criteria scores and the likelihood of reporting gastrointestinal bleeding. Correlations, score comparisons, and logistic regressions were analyzed using IBM SPSS Statistics (version 23). All analyses were two tailed, with an α of 0.05.

3. Results

3.1. Study scores using CONSORT Harms criteria

The 159 clinical studies each received a total score for 10 different criteria, giving a total of 3,180 criteria assessments in the two separate reviewer evaluations and 1,590 criteria assessments evaluated for the final score. All discrepancies were resolved during the final discussion, and no cases were referred to the third author.

In the final assessment, the studies received a mean score of 5.25 out of a maximum of 10 (standard deviation [SD] $\pm$ 2.09). For studies without a subgroup analysis (excluding recommendation 9), the mean score was 5.15 (SD $\pm$ 1.97) out of a maximum of 9. Most studies did not include a subgroup analysis.

The distribution of criteria scores among the studies is shown in Fig. 1. Logistic regression analysis showed a higher reporting of gastrointestinal bleeding with increasing CONSORT Harms criteria scores (odds ratio [OR] 1.17, 95% confidence interval 1.01−1.37, $P = 0.042$). The odds of reporting cases of gastrointestinal bleeding were three times higher for high-score studies compared to low-score studies (OR 3.43, 95% confidence interval 1.17−10.04).

The recommendations with the highest scores were recommendation 6—participant withdrawals (133 studies, 83.6%), 8—absolute risk of gastrointestinal adverse events (130 studies, 81.8%), and 4—clarify how harms-related information was collected (118 studies, 74.2%). The recommendations with the lowest scores were recommendation 9—subgroup analysis (16 studies, 10.1%), 2—collection of harms data mentioned in introduction (48 studies, 30.2%), and 5—plan for presenting and analyzing information on harms (51 studies, 32.1%). The scores according to the CONSORT Harms recommendations are presented in Table 2.

Fifty-nine studies (37.1%) did address and monitor for gastrointestinal adverse events, either specifically or as part of a comprehensive clinical examination (recommendation 3). The remaining 100 studies (62.9%) did not address gastrointestinal adverse effects or did not describe a clinical examination of sufficient extent. Despite this, the absolute risk of gastrointestinal adverse events (recommendation 8) could be found in 130/159 studies (81.8%). This number includes studies with zero observed gastrointestinal adverse

Table 1. Scoring criteria

Recommendation 1. If the study collected data on harms and benefits, the title or abstract should so state
Definition: Score 1 if any mention of harms, adverse events, side effects, toxicity, or complications, excluding those clearly due to lack of treatment effects or underlying disease. If not, score 0.
Recommendation 2. If the study collected data on harms and benefits, the introduction should so state
Definition: Score 1 if any mention of harms, adverse events, side effects, toxicity, or complications, excluding those clearly due to lack of treatment effects or underlying disease. If not, score 0.
Recommendation 3. List addressed gastrointestinal adverse events with definitions for each
Definition: Score 1 if any gastrointestinal adverse event was specified as an outcome to be addressed or if the clinical examination described is perceived as comprehensive enough to discover overt gastrointestinal adverse effects and any other major events. If not, score 0.
Recommendation 4. Clarify how harms-related information was collected
Definition: Score 1 if method of collection or system of monitoring for harms is specified. If not, score 0.
Recommendation 5. Describe plan for presenting and analyzing information on harms
Definition: Score 1 if harms analysis is specified, or if the general method of result analysis appeared to have been applied to harms data. If not, score 0.
Recommendation 6. Describe for each arm the participant withdrawals that are due to harm and the experience with the allocated treatments
Definition: Score 1 if withdrawals due to adverse events were specified. If not, score 0.
Recommendation 7. Provide the denominators for analyses on harm
Definition: Score 1 if denominators are described. If not, score 0.
Recommendation 8. Present the absolute risk of each gastrointestinal adverse event and present appropriate metrics for recurrent events, continuous variables, and scale variable
Definition: Score 1 if absolute risk can be found for any gastrointestinal adverse effect. If not, score 0.
Recommendation 9. Describe any subgroup analyses and explanatory analyses for harms
Definition: Score 1 if any subgroup analysis for adverse drug reactions was done. If not, score 0.
Recommendation 10. Provide a balanced discussion of benefits and harms with emphasis on study limitations, generalizability and other sources of information on harms
Definition: Score 1 if the discussion is perceived as balanced and study limitations are discussed. If not, score 0.

Adapted from CONSORT Harms recommendations [11].

effects, which in several cases had to be interpreted from lists of observed adverse effects or statements of no detected adverse effects. In 29 studies (18.2%), the number of patients included in the risk analysis was not described. However, cases of gastrointestinal bleeding were reported in five of those publications. In studies where gastrointestinal bleeding was addressed or observed, the definitions and descriptions varied widely. A detailed description is provided in Supplementary Materials.

Twenty-four studies (15.1%) received a score of 8, 9, or 10 and were classified as high-score studies. Those studies included 4,510 patients (2,277 receiving steroid, 2,233 receiving placebo), of which 16 studies (66.7%) reported cases of gastrointestinal bleeding (155 cases in the steroid group, 92 cases in the placebo group). Twelve of the 24 high-score studies (50%) concerned prevention of bronchopulmonary dysplasia in pediatric patients and contributed a major proportion of cases of gastrointestinal bleeding (120 cases in 1,066 corticosteroid-treated patients, 69 cases in 1,047 placebo-treated patients).

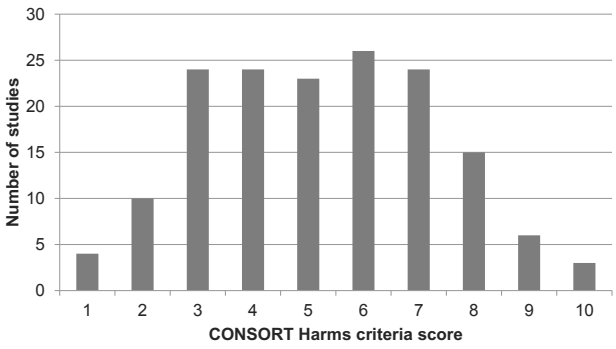

Fig. 1. Studies grouped by CONSORT Harms criteria scores (N = 159).

Thirty-eight studies (23.9%) received a score of 3 or less, indicating that few of the CONSORT Harms criteria were met in the publications (low-score studies). Those studies included 6,605 patients (3,312 receiving steroid, 3,293 receiving placebo), of which 14 studies (36.8%) reported cases of gastrointestinal bleeding (41 in the steroid group, 22 in the placebo group). Twenty-four studies (63.2%) did not report any cases of bleeding. The main reasons for achieving low scores were that adverse effects were not mentioned in title, abstract, or introduction; plans for presenting and analyzing harms were not described; or discussions were not perceived as balanced. None of the low-score studies received a score for addressing gastrointestinal adverse effects. Still, in most of the low-score studies, it was possible to present an absolute risk of gastrointestinal bleeding (25/38, 65.8%) and participant withdrawals due to harm (23/38, 60.5%).

3.2. CONSORT Harms criteria scores relating to key variables

We found no significant correlation between the CONSORT Harms score and publication year, ambulant or hospitalized patients, or funding source (industry sponsored or not) (Table 3). There was a trend toward higher scores for

Table 2. Studies which fulfilled CONSORT Harms criteria, $N = 159$

CONSORT Harms criterion	No. (%)
1. If the study collected data on harms and benefits, the title or abstract should so state	93 (58.5)
2. If the study collected data on harms and benefits, the introduction should so state	48 (30.2)
3. List addressed adverse gastrointestinal events with definitions for each	59 (37.1)
4. Clarify how harms-related information was collected	118 (74.2)
5. Describe plan for presenting and analyzing information on harms	51 (32.1)
6. Describe for each arm the participant withdrawals that are due to harm and the experience with the allocated treatments	133 (83.6)
7. Provide the denominators for analyses on harms	102 (64.2)
8. Present the absolute risk of each gastrointestinal adverse event and present appropriate metrics for recurrent events, continuous variables, and scale variable	130 (81.8)
9. Describe any subgroup analyses and explanatory analyses for harms	16 (10.1)
10. Provide a balanced discussion of benefits and harms with emphasis on study limitations, generalizability, and other sources of information on harms	85 (53.5)

studies published in the major medical journals, with mean score 5.86 vs. 5.08 in other journals ($P = 0.052$). The studies with the highest scores (score ≥ 8) had 33.3% (8/24) industry sponsoring, compared to 54.8% (74/135) industry sponsoring for the rest of the studies ($P = 0.052$).

To see if reporting had improved in the most recent years, the reporting after 2007 was analyzed separately. Studies published in 2007−30.6.2011 ($N = 26$) had a mean score of 4.88. Studies published in the major medical journals in 2007−30.6.2011 ($N = 7$) had a mean score of 5.29.

3.3. Qualitative assessment

Several studies collected data on adverse effects, including gastrointestinal, without mentioning the fact in title, abstract, or introduction. In many studies, adverse effect monitoring had obviously been performed without mention of intention or method. Risk of gastrointestinal adverse effects had often been considered beforehand, as evidenced by exclusion criteria such as previous peptic ulceration, but not mentioned in methods, results, or discussion sections. Plans for presenting and analyzing information on harms were often not specified. Information on adverse effects was in many cases presented less systematically than efficacy outcomes and could be found in various sections of the publications. In some studies, efficacy and harm were analyzed in the same way; in other studies, statistical methods were applied to efficacy outcome only. Some studies limited adverse effects reporting to the most common or most serious cases. Denominators were sometimes specified for efficacy only, not for adverse effects, and could only be found by inference by comparing adverse effects tables with text. Several studies presented adverse effect data as percentages. If the denominator for adverse effect analysis was not clearly stated, the absolute risk could not be found. In several studies that quantified withdrawals, the reason was not always stated but could be inferred by interpreting the text in relation to the withdrawal data. Conclusions of safety, such as ''no safety problems,'' were sometimes drawn despite underpowered study design and unsystematic addressing of adverse effects.

3.4. Reviewer agreement

In the analysis by two separate reviewers, the mean CONSORT Harms criteria scores were 5.19 (SD ± 2.13) and 6.06 (SD ± 2.11), respectively, for the 159 studies. Interrater agreement for each study, calculated as Gwet's AC1, had a mean value of 0.56 (SD ± 0.29) and a median value of 0.62 (range −0.28 to 1.00). The 15 studies with slight or poor reviewer agreement coefficients (Gwet's AC1 < 0.2) received significantly lower CONSORT Harms scores than studies with higher degrees of agreement (3.87 vs. 5.40) ($P = 0.007$). Interrater agreement for each CONSORT Harms criterion through all 159 studies, using Gwet's AC1 agreement coefficient, showed a mean value of 0.58 (SD ± 0.15) and a median value of 0.57 (range 0.32-0.82). Agreements between reviewers differed with regard to individual criteria. The criteria with the three lowest Gwet's AC1 scores were recommendations 7, 9, and 10 (0.42, 0.32, and 0.46, respectively).

Details can be found in Supplementary Materials.

4. Discussion

4.1. CONSORT harms criteria score, main findings

We examined the reporting of gastrointestinal adverse effects in 159 published randomized controlled trials which were included in a published meta-analysis addressing risk of gastrointestinal bleeding associated with corticosteroid use [14]. The studies had undergone quality assessment and fulfilled the criteria for inclusion in a systematic review. However, analysis of the publications, using criteria proposed in CONSORT Harms adjusted for gastrointestinal adverse effects, showed that few studies received high scores. Adverse effects monitoring and reporting varied greatly, and most of the studies did not fulfill several criteria. Only 24 studies (15.1%) received a score of 8 or more. The mean CONSORT Harms criteria score of 5.3 for all 10 criteria corresponds generally to that found by Maggi et al. [18]. Exclusion of criterion 9 had

Table 3. Correlation between CONSORT Harms scores and key variables

Key variable	No (%)	Mean CONSORT Harms score	Score difference (95% confidence interval)	*P*
Industry sponsored	82/159 (51.6)	5.00	0.52 (−0.13 to 1.17)	0.118
Not industry sponsored	77/159 (48.4)	5.52		
Score ≥ 8, industry sponsored	8/24 (33.3)	8.25	0.38 (−0.27 to 1.02)	0.239
Score ≥ 8, not industry sponsored	16/24 (66.7)	8.63		
Published in major medical journal	35/159 (22.0)	5.86	0.78 (−0.01 to 1.56)	0.052
Published in other journals	124/159 (78.0)	5.08		
Published 1983–2004	114/159 (71.7)	5.26	0.04 (−0.69 to 0.77)	0.912
Published 2005–30.6.2011	45/159 (28.3)	5.22		
Ambulant patients	56/159 (35.2)	5.32	0.11 (−0.58 to 0.80)	0.757
Hospitalized patients	103/159 (64.8)	5.21		

only a limited effect on the overall mean score, reflecting that relatively few studies had received a score on this criterion.

Some of the criteria were fulfilled for most of the studies but, in many cases, to a limited degree where information had to be inferred by the reviewers. The present study gives no indications as to why criteria were not fulfilled. Most of the 159 studies focused on treatment efficacy. Adverse effects were generally given little space and were, for most studies, not a prespecified end point. Another possible explanation may be journal text limitations, although space limitations should not be an excuse to exclude information on this highly important issue when reporting on results of a clinical study. It remains to be seen whether adverse effects reporting will improve with increasing use of electronic publications.

4.2. Weaknesses in monitoring and reporting gastrointestinal bleeding

Occurrence of gastrointestinal bleeding was assumed to be an objective and unambiguous adverse effect that would have been described if observed in the studies. Most studies did not address the risk of gastrointestinal bleeding specifically, although several studies did record gastrointestinal bleeding and discussed the risk in the introduction section. Definitions of gastrointestinal bleeding varied widely and cases could possibly be hidden within broader diagnostic groups such as ''gastrointestinal reactions.'' This may be an even greater problem with more subjective adverse effects.

Some of the CONSORT Harms criteria may be less critical than others when it comes to the facts of whether the study did address gastrointestinal adverse effects and whether any adverse effects were reported. Many studies did monitor adverse effects, including gastrointestinal, with little mention of intention or method. This was a major reason for interrater differences on recommendation 4. It can be argued that the most important recommendations regarding actual findings of gastrointestinal bleeding risk are recommendations 6–8 (withdrawals, denominators, and absolute risks), although it is reasonable to expect any intention to look for adverse effects to be mentioned in the abstract or introduction. Information on absolute risks was given in 130 studies (81.8%), although not always clearly stated. In addition, some studies reported cases of gastrointestinal bleeding without describing absolute risk. One hundred thirty-three studies (83.6%) described withdrawals and experience with the allocated treatments to some extent. Recommendations 1–5 (stating of intention and plans for analyzing harms data) were not always fulfilled, even when adverse effects were described in the results sections. Subgroup analysis (recommendation 9) is obviously not a quality criterion for reporting harms if not part of the study. Several of the studies that received low scores using the CONSORT Harms criteria nevertheless gave an impression of thoroughness and awareness of the risk of adverse effects, despite the fact that little space was allotted to adverse effect descriptions in the publication.

Superficial descriptions of adverse effects and use of cutoff valuations such as ''serious'' or ''frequency >5%'' make it possible to avoid describing all adverse effects that occurred. It has been argued that it is safer to assume that adverse effects were not ascertained or not recorded than to assume that the prevalence or incidence was zero if the adverse effect is not mentioned specifically [19]. However, in a clinical trial, there are risks of several adverse effects and it would be unreasonable to expect authors to mention all those that did not occur, unless they were addressed specifically.

Studies with low quality of reporting of harms, as assessed using the CONSORT Harms criteria, might have a correspondingly lower chance of finding adverse effects, from either poor study design or poor monitoring.

Inclusion of only those studies that described active or comprehensive adverse effects monitoring would have eliminated 100 of 159 studies and 63 cases of gastrointestinal bleeding from our systematic review [14]. If mentioning of adverse effects in title, abstract, or introduction sections had been a selection criterion, 57 studies would have been lost for analysis. Exclusion of studies with low scores would have led to exclusion of relevant findings of cases of gastrointestinal bleeding.

4.3. CONSORT Harms scores in relation to key variables

We found no clear correlation between publication year (before or after publication of the CONSORT Harms criteria) and the reporting of adverse effects. This reflects most previous findings [20–23], whereas Haidich et al. [24] found a somewhat increased reporting of harms from 2003 to 2006. In our analysis, studies published in 2007–30.6.2011 had a lower mean score than studies published in the period preceding publication of the CONSORT Harms criteria, indicating that the reporting of adverse effects did not improve over time.

Contrary to expectations, there was a relatively small score difference between studies published in major medical journals and other journals. There was a trend toward a higher mean score for these studies. Haidich et al. [24] analyzed randomized clinical trials published in the five major medical journals and found mean scores of 5.8 and 6.7 for studies published in 2003 and 2006, respectively. This corresponds generally with the mean score of 5.86 found in our study, but in our study, the scores appeared to decline over time. In an analysis of studies published in four major medical journals in 2009, Maggi et al. [18] found that most studies did not incorporate the CONSORT Harms recommendations sufficiently.

In contrast to previous studies, where industry-funded studies have shown better safety reporting than nonindustry studies [18,20,24,25], we found a trend toward worse safety reporting in studies that were supported or funded by the pharmaceutical industry. This may be due to our broad definition of sponsoring or the fact that most studies were published before 2004 and were probably not performed for regulatory purposes, as the corticosteroid used had been on the market for several years.

4.4. Reviewer agreement

Analysis of initial reviewer agreement for each study and for each CONSORT Harms criterion across studies showed fair-to-moderate agreement with large variations. Low agreement was mainly caused by differences in interpretation of information in the article texts and difficulties in determining whether a criterion was sufficiently fulfilled or not. In addition, many of the CONSORT Harms criteria include several questions within one recommendation. Some authors have addressed the ambiguity by splitting some of the original recommendations into several, more precise subcategories [20,22,26,27], in some cases with option of half credits [24,27]. Because of the heterogeneity of the studies regarding the methods descriptions and the presentation of data, a more detailed approach using a more specific checklist would probably not have reduced the necessity for judgment or resulted in greater agreement between reviewers. The use of half credits if a criterion was partly fulfilled might have resulted in more specific scores, but there would still be an element of judgment regarding the degree of fulfillment of each criterion.

Because subgroup analysis of adverse effects is rarely done, other authors have excluded CONSORT Harms recommendation 9 from assessment [20,26,27]. Subgroup analysis of harm was done in several of the studies included in our review but with focus on harm as a result of disease or treatment failure. This was a major reason for score discrepancies between reviewers.

4.5. Limitations

Several studies reported adverse effects without mentioning gastrointestinal bleeding. As all the studies did address or report adverse effects to some extent, we concluded that no gastrointestinal bleeding occurred in those studies. This assumption may be mistaken, as a lack of reports does not necessarily mean that the adverse effects did not occur [9]. In studies where gastrointestinal bleeding was not observed, the nonoccurrence cannot necessarily be expected to be commented on unless the adverse effect was expected or looked for. There is, however, an uncertainty if the risk profile is not described in detail. Our finding of higher reporting of gastrointestinal bleeding with increasing CONSORT Harms criteria scores might indicate underreporting of adverse events in the low-score studies.

We scored the studies through assessment by at least two authors. However, application of the CONSORT Harms criteria to clinical studies involves considerable judgment. Other reviewers may differ in their opinion as to what should constitute a score of 0 or 1.

The recommendations of CONSORT [10] and CONSORT Harms [11] were developed to improve the quality of clinical study reporting and were not intended as a validated tool for assessing the methodological quality of studies. A validated tool is not available at the present time.

5. Conclusion

Analysis of clinical studies included in a previous review and meta-analysis, using criteria proposed in CONSORT Harms adjusted for gastrointestinal adverse effects, showed that few studies received high scores. Reporting of gastrointestinal bleedingincreased with increasing CONSORT Harms score. Application of the CONSORT Harms criteria to the clinical studies involved considerable judgment, because of the multiple items within several of the criteria and the highly variable adverse effects reporting in the studies. So far, no clear assessment method has been proposed to describe studies adequately without risking eliminating studies with relevant findings. In our opinion, routine scoring by CONSORT Harms criteria for harms assessment would be inadvisable without adding qualified judgment on the study in question.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2017.10.018.

References

[1] Ross SD. Drug-related adverse events: a readers' guide to assessing literature reviews and meta-analyses. Arch Intern Med 2001;161:1041−6.

[2] Yazici Y. Some concerns about adverse event reporting in randomized clinical trials. Bull NYU Hosp Jt Dis 2008;66(2):143−5.

[3] The World Medical Association (WMA). WMA Declaration of Helsinki−Ethical principles for medical research involving human subjects. Available at https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/. Accessed November 21, 2017.

[4] Saltman D, Jackson D, Newton PJ, Davidson PM. In pursuit of certainty: can the systematic review process deliver? BMC Med Inform Decis Mak 2013;13:25.

[5] Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (Updated March 2011). The Cochrane Collaboration; 2011. Available at. http://handbook.cochrane.org/Chapter 14: Adverse effects. Accessed November 12, 2014.

[6] PRISMA transparent reporting of systematic reviews and meta-analyses. PRISMA Checklist. 2009. Available at http://www.prisma-statement.org/. Accessed November 12, 2014.

[7] Zorzela L, Loke YK, Ioannidis JP, Golder S, Santaguida L, Altman D, et al. PRISMA Harms: improving harms reporting in systematic reviews. BMJ 2016;352:i157.

[8] Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. BMJ 2004;328:1490.

[9] Loke YK, Price D, Herxheimer A. Systematic reviews of adverse effects: framework for a structured approach. BMC Med Res Methodol 2007;7:32.

[10] CONSORT Transparent reporting of trials. CONSORT checklist. Available at http://www.consort-statement.org. Accessed November 19, 2013.

[11] Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med 2004;141:781−8.

[12] Pitrou I, Boutron I, Ahmad N, Ravaud P. Reporting of safety results in published reports of randomized controlled trials. Arch Intern Med 2009;169:1756−61.

[13] Agency for Healthcare Research and Quality (AHRQ). Methods guide for effectiveness and comparative effectiveness reviews. AHRQ Publication No 10(14)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2014. Available at https://www.ncbi.nlm.nih.gov/books/NBK47095/. Accessed November 12, 2014.

[14] Narum S, Westergren T, Klemp M. Corticosteroids and risk of gastrointestinal bleeding: a systematic review and meta-analysis. BMJ Open 2014;4(5):e004587.

[15] Lee PE, Fischer HD, Rochon PA, Gill SS, Herrmann N, Bell CM, et al. Published randomized controlled trials of drug therapy for dementia often lack complete data on harm. J Clin Epidemiol 2008;61:1152−60.

[16] Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. Br J Math Stat Psychol 2008;61(Pt 1):29−48.

[17] Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC Med Res Methodol 2013;13:61.

[18] Maggi CB, Griebeler IH, Dal Pizzol Tda S. Information on adverse events in randomised clinical trials assessing drug interventions published in four medical journals with high impact factors. Int J Risk Saf Med 2014;26(1):9−22.

[19] Chou R, Aronson N, Atkins D, Ismaila AS, Santaguida P, Smith DH, et al. AHRQ series paper 4: assessing harms when comparing medical interventions: AHRQ and the effective health-care program. J Clin Epidemiol 2010;63:502−12.

[20] Shukralla AA, Tudur-Smith C, Powell GA, Williamson PR, Marson AG. Reporting of adverse events in randomised controlled trials of antiepileptic drugs using the CONSORT criteria for reporting harms. Epilepsy Res 2011;97(1−2):20−9.

[21] Nuovo J, Sather C. Reporting adverse events in randomized controlled trials. Pharmacoepidemiol Drug Saf 2007;16(3):349−51.

[22] Carmichael K, Nolan SJ, Weston J, Tudur Smith C, Marson AG. Assessment of the quality of harms reporting in non-randomised studies and randomised controlled studies of topiramate for the treatment of epilepsy using CONSORT criteria. Epilepsy Res 2015;114:106−13.

[23] Meister R, von Wolff A, Mohr H, Nestoriuc Y, Härter M, Hölzel L, et al. Adverse event methods were heterogeneous and insufficiently reported in randomized trials on persistent depressive disorder. J Clin Epidemiol 2016;71:97−108.

[24] Haidich AB, Birtsou C, Dardavessis T, Tirodimos I, Arvanitidou M. The quality of safety reporting in trials is still suboptimal: survey of major general medical journals. J Clin Epidemiol 2011;64:124−35.

[25] Smith SM, Chang RD, Pereira A, Shah N, Gilron I, Katz NP, et al. Adherence to CONSORT harms-reporting recommendations in publications of recent analgesic clinical trials: an ACTTION systematic review. Pain 2012;153(12):2415−21.

[26] Bagul NB, Kirkham JJ. The reporting of harms in randomized controlled trials of hypertension using the CONSORT criteria for harm reporting. Clin Exp Hypertens 2012;34:548−54.

[27] Breau RH, Gaboury I, Scales CD Jr, Fesperman SF, Watterson JD, Dahm P. Reporting of harm in randomized controlled trials published in the urological literature. J Urol 2010;183:1693−7.

Paper III

BMJ Open

Critical appraisal of adverse effects reporting in the 'Treatment for Adolescents With Depression Study (TADS)'

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

To cite: 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**:e026089. doi:10.1136/bmjopen-2018-026089

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2018-026089).

Received 16 August 2018
Revised 17 January 2019
Accepted 5 February 2019

[1]Regional Medicines Information and Pharmacovigilance Centre (RELIS Sør-Øst), Department of Pharmacology, Oslo University Hospital, Oslo, Norway
[2]Center for Psychopharmacology, Diakonhjemmet Hospital, Oslo, Norway
[3]Department of Pharmacology, University of Oslo, Oslo, Norway

Correspondence to
Tone Westergren;
twesterg@ous-hf.no

ABSTRACT

Objective To identify all publications from the 'Treatment for Adolescents With Depression Study (TADS)' and assess the findings regarding occurrence of any adverse effects in the treatment groups both for the short-term and long-term study stages.

Design Descriptive analysis of TADS publications with any information on adverse effects.

Results We identified 48 publications describing various aspects of the TADS, in which 439 adolescent patients received treatment with fluoxetine, cognitive–behavioural therapy, cognitive–behavioural therapy plus fluoxetine or placebo. Eight publications were assessed as providing some data on adverse effects. Risk of suicidal behaviour was the only adverse effect that was addressed in all publications. Several psychiatric and physical adverse effects were reported during the first 12 weeks, but not mentioned in reports from later study stages. Common adverse effects of fluoxetine, such as weight changes or sexual problems, were not identified or mentioned in the publications.

Conclusions The TADS publications do not present a comprehensive assessment of treatment risk with fluoxetine in adolescents, especially for more than 12 weeks of treatment. Risk of suicidality was the only adverse effect that was reported over time. Reporting of adverse effects was incomplete with regard to the long-term safety profile of fluoxetine.

Strengths and limitations of this study

- ► This is the first systematic assessment of adverse effects reporting in publications from the Treatment for Adolescents With Depression Study (TADS).
- ► The analysis encompasses all adverse events mentioned in publications from the TADS.
- ► An extensive literature search was conducted and we believe that all relevant studies have been identified.
- ► We cannot exclude the possibility that some publications may have been overlooked.

INTRODUCTION

The safety profile of selective serotonin reuptake inhibitors (SSRIs) in adolescents has been extensively debated. Several systematic reviews have analysed what is known about the risk of suicidal behaviour[1–3] as well as other psychiatric and somatic adverse risks and the perceived benefit/risk balance. The reviews have highlighted considerable variations in assessment, definitions and reporting of adverse effects in the clinical trials.

The Norwegian Regional Medicines Information and Pharmacovigilance Centres and the Center for Psychopharmacology at Diakonhjemmet Hospital regularly receive queries from hospital doctors and general practitioners regarding the safety of fluoxetine (FLX) and other SSRIs in adolescent patients.

One of the major clinical studies of efficacy and safety of FLX in adolescents is the 'Treatment for Adolescents With Depression Study (TADS)', which is often referred to in textbooks and reviews.

In 1998, the US National Institute of Mental Health (NIMH) issued a request for proposals (RFP-NIH-NIMH 98-DS-0008) with the objective of launching a clinical trial to address the effectiveness of treatment for adolescents with major depression.[4] The subsequent study, 'TADS' was coordinated by the Department of Psychiatry and Behavioral Sciences and the Duke Clinical Research Institute, both at Duke University Medical Center, collaborating with and funded by NIMH,[5] and carried out in the period 2000–2003.[6] The study included 439 youths who were randomised to one of four treatment groups; (1) FLX, (2) cognitive–behavioural therapy (CBT), (3) cognitive–behavioural therapy plus fluoxetine (COMB) or (4) placebo (PBO) for 12 weeks (stage I).[6] Double-blind treatment was performed among patients treated with FLX and PBO only, while patients treated with CBT with or without FLX received open treatment. Stage II and III were maintenance phases for the

active treatment groups, with the option of intensifying treatment for partial responders. Patients in the PBO group were offered open active treatment of FLX, CBT or both. Stage IV consisted of an additional year of open follow-up.[5]

The two primary outcome measures in the TADS were Children's Depression Rating Scale-Revised (CDRS-R) total scores, and responder status on the Clinical Global Impressions-Improvement scale. According to protocol, all analyses would be performed by intention to treat (ITT), regardless of later events.

Adverse events during the acute and maintenance phases were defined as secondary outcomes.[7] Patients were monitored for safety regarding affective disorders, need for mental health treatment, need for concomitant medications, occurrence of adverse events and serious adverse events and use of adjunctive services and attrition prevention. Most assessments were based on both patient and parent information.[8]

The TADS has been described as the largest and arguably the highest quality acute-phase randomised PBO controlled trial of an antidepressant drug for adolescent depression.[9] We understand from the protocol and monitoring procedures that the TADS team intended to evaluate the tolerability of treatment, and that the study was expected to provide improved insight into the potential adverse effects of antidepressant treatment in this age group, due to its study size and duration. Several publications from the TADS have addressed risks of adverse effects. Despite this, concerns have been raised regarding under-reporting of suicidal risk,[10] study size and an increased risk of psychiatric adverse effects.[11]

In the TADS, adverse events were defined as an unfavourable medical change that occurred after beginning or during the study that might or might not be related to or caused by the study drug or CBT treatment. This was further specified as any medical event that caused clinically significant interference with functioning (eg, headache that caused school absence or otherwise caused clinically significant activity restriction), any event that required medical attention, and any medical event associated with impairment in functioning and induced the patient to take a concomitant medication. Conditions that did not lead to clinically significant interference with functioning or did not require medical attention were not defined as adverse events.[7 8] The protocol specified that new-onset psychiatric symptoms, such as emerging mania or panic attacks, would be recorded if they caused clinically significant interference with functioning.[8] It follows that such conditions would not be recorded unless a certain severity threshold was reached.

Harm-related adverse events were defined as involving harm to self, which could include a non-suicidal event. Examples given are cutting, worsening of suicidal ideation, suicide attempt or harm to others. Suicide-related adverse events were defined as worsening suicidal ideation and/or suicide attempt. Adverse event forms were to be used throughout the study and it must be assumed that such data were collected, as well as clinical scoring data for possible psychiatric adverse events.

Our objective in the present study was to identify all publications from the TADS and assess the findings regarding occurrence of any adverse effects in the treatment groups both for the short-term and long-term study stages. The TADS was chosen because of the non-industrial funding and because it is considered as a high-quality study.[9]

METHODS

Literature search

Publications from the TADS were identified through searches in PubMed, EMBASE, PsycINFO, Google Scholar, ClinicalTrials.gov, NIMH website nimh.nih.gov, the Duke Clinical Research Institute TADS website (http://tads.dcri.org), by hand searching of references in identified publications, and by searching other publications by the main authors (snowballing). Search terms in Google Scholar were either «TADS team» or «Treatment for adolescents with depression study». Search term in PsycINFO was «Treatment for adolescents with depression study». Search term in PubMed was the phrase Treatment for adolescents with depression study. The initial publications with data from the TADS study were identified and used to search for similar publications, limited to 2004 to 1 September 2017, Clinical Trial or Randomized Controlled Trial and age group Child 0–18. Search term in Embase was «Treatment for adolescents with depression study». The final main search in all databases was conducted on 5 September 2017. An additional literature search in PubMed for any recent TADS publications was conducted in February 2018 and updated in January 2019.

Inclusion and exclusion criteria

Identified TADS publications were assessed and classified according to publication topic and reported outcomes. Inclusion criteria: All publications that reported on results from the TADS and provided some information on adverse effects. Publications on efficacy or non-primary or non-secondary outcomes were excluded if they gave no information on adverse events.

Data assessment

Adverse effects were defined as psychiatric or somatic diagnoses or complaints arising during treatment, as described in the publications. In addition, we have included worsening of depression as an adverse effect if described in the publications. Publications describing any adverse events during treatment were analysed in detail regarding the types and frequency estimates of adverse events. Two researchers (TW and SN) evaluated each publication independently. All researchers (TW, SN and MK) discussed any ambiguity and the data extraction tables.

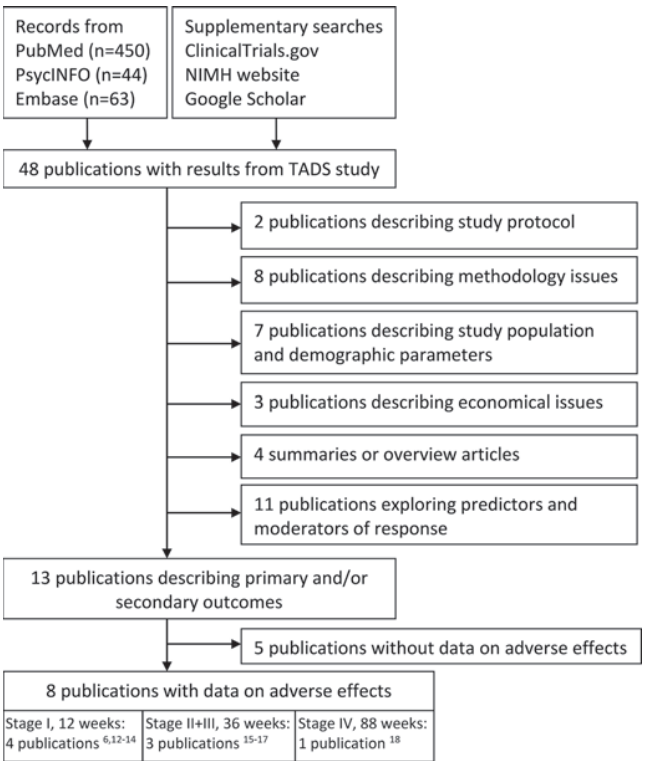

Figure 1 Selection and characteristics for publications from the TADS. NIMH, National Institute of Mental Health; TADS, Treatment for Adolescents With Depression Study.

Patient and public involvement

Patients or the public were not involved in this literature review.

RESULTS

We identified 48 publications that reported on the study protocol and/or various outcomes in the TADS population. The selection process and publication characteristics are described in figure 1.

Eight publications were assessed as providing at least some data on adverse effects,[6 12–18] of which four publications reported possible adverse effects for subgroups of patients only; patients who responded to treatment,[13] patients originally assigned to PBO treatment,[16] patients who had at least one suicidal event[17] and patients using attrition prevention services,[14] respectively. Reporting of adverse effects was most detailed in the two initial results publications from stage I (0–12 weeks),[6 12] and included a wide range of adverse effects, including several psychiatric and gastrointestinal reactions. One stage I publication did not address adverse effects explicitly; however, symptoms that may be associated with adverse effects were described as residual symptoms of depression.[13]

The publications that reported on adverse effects in the later study stages II–IV listed few adverse effects except suicidal behaviour (table 1). The publication that purported to report on long-term effectiveness and safety outcomes only included reporting of suicide-related adverse events.[15]

Patient population and treatment modifications during the study

In the TADS, 439 patients were randomised to one of the four treatment groups. By the end of stage I (12 weeks), 351 patients remained for assessment, of them 270 patients in active treatment groups. The rest of the patients had either withdrawn their consent, or been classified as premature terminators due to need for additional treatment.[6 15] It is not specified to what extent drop-outs or premature terminations were due to adverse events in the initial study population and if those adverse events were included in the reports. By week 36 (end of stage III), 178 patients remained in the group to which they had been randomised, specifically 68 for COMB, 55 for FLX and 55 for CBT.[15] Patients who terminated their assigned treatment prematurely did in many cases continue their assessments and were included in the ITT analyses for their original group, although they received an active treatment other than that specified for the group they were assigned to.[12 15 19] Between 34% and 46% of patients in the monotherapy groups did not remain in their assigned treatment arm by the end of stage II, and 43 of the 111 patients (38%) in the CBT group were receiving another SSRI or antidepressant by the end of stage III (36 weeks).[19]

Reporting of suicidality in TADS publications

Suicidality symptoms were monitored using an affective disorders screening procedure (ADS), Reynolds Adolescent Depression Scale, a revised CDRS-R, a Suicide Ideation Questionnaire-Junior (SIQ-Jr) as well as adverse event/serious adverse event forms. All the TADS publications classified as reporting adverse effects[6 12–18] describe the risk of suicidal events, defined as discrete episodes of suicidal ideation, suicidal attempts or preparatory acts towards an imminent attempt. Injury to self was not included if there was no suicidal intent. Reporting of suicidal events and risk is described in the online supplementary file. Data on suicidality are presented as either counts of discrete episodes, mean scores, score changes or proportion of patients reaching threshold values on scoring tools.

By week 12, CDRS-R item 13 scores are reported as per cent of patients with score ≥2 for the total study population,[6] per cent of patients with score worsening ≥1 point and per cent of patients with score increase from 1-2 to ≥5 for each treatment group.[12] SIQ-Jr scores are reported as per cent of patients with scores ≥31 for the total study population[6] and each treatment group,[15] per cent of patients with score increase to ≥31[12] and mean score for each treatment group.[6 12]

By week 36, CDRS-R scores are not described in any of the publications. For SIQ-Jr scores, results are described for patients who had completed the SIQ-Jr assessment at week 36 and for a smaller number of patients who

Table 1 Reporting of adverse events in publications from the TADS

Reported event	Stage 1 (12 weeks)				Stage 2+3 (36 weeks)			Stage 4 (88 weeks)
	TADS team[6]	Emslie *et al*[12]	Kennard *et al*[13]*	May *et al*[14]†	TADS team[15]	Kennard *et al*[16]‡	Vitiello *et al*[17]§	TADS team[18]
Harm-related adverse event	x	x		x				
Suicide-related adverse event	x	x	x¶	x	x	x	x	x
Attempted suicide	x	x		x	x	x	x	
Homicidality		x		x				
Mania	x	x		x			x	
Hypomania	x	x		x				
Elevated mood	x	x						
Trouble attention/concentration		x	x¶					
Racing thoughts		x						
Excessive talking/talking very fast		x						
Increase in activities		x						
Impulsivity		x						
Hypersensitivity**	x	x						
Irritability	x	x	x¶				x	
Anger	x	x						
Worsening of depression	x	x		x		x	x	x
Psychomotor			x¶					
Guilt			x¶					
Mood			x¶					
Interest			x¶					
Crying	x	x						
Agitation	x	x						
Akathisia	x	x						
Nervousness	x	x						
Restlessness	x	x						
Hyperactivity	x	x						
Panic attacks	x	x						
Anxiety	x	x						
Excessive sweating		x						
Difficulty breathing		x						
Hearing problems		x						
Somnolence/feeling drowsy	x	x						
Insomnia/sleeplessness	x	x					x	

Continued

Table 1 Continued

Reported event	Stage 1 (12 weeks)				Stage 2+3 (36 weeks)			Stage 4 (88 weeks)
	TADS team[6]	Emslie *et al*[12]	Kennard *et al*[13]*	May *et al*[14]†	TADS team[15]	Kennard *et al*[16]‡	Vitiello *et al*[17]§	TADS team[18]
Sleep	x	x	x¶					
Nightmare	x	x						
Night sweats	x							
Sedation	x	x						
Fatigue	x		x¶					
Tremor	x	x						
Behaviour/feeling abnormal	x	x						
Social problems				x			x	
Headache	x	x						
Upper abdominal pain	x	x						
Stomach pain		x						
Diarrhoea	x	x						
Influenza/sinusitis	x							
Cold, sore throat, cough/wheeze		x						
Allergies		x						
Dry mouth		x						
Nausea/vomiting	x	x						
Fever		x						
Muscle aches or cramps		x						
Joint pain		x						
Numbness or tingling arms or legs		x						
Weight			x¶					
Chest pain		x						
Racing/pounding heart, skip beats		x						
Urination frequency or pain		x						
Constipation, feeling bloated		x						
Skin rash/hives		x						

*Reporting limited to responders subgroup, regardless of treatment arm.
†Reporting limited to subgroup of patients seeking attrition prevention.
‡Reporting limited to ITT placebo group.
§Reporting limited to patients with a suicidal event.
¶Reported as residual symptoms of depression.
**Understood as mood hypersensitivity.
ITT, intention to treat; TADS, Treatment for Adolescents With Depression Study.

both completed the assessment and were still in their assigned treatment group.[15] Results are presented as the percentage of patients with score ≥31 for each treatment group. Patients with score increases and mean scores are not reported.

Suicidal events are presented for all three treatment groups, and reported for ITT and observed cases groups. The frequency of suicidal events was calculated using the group size according to the original randomisation, with no reference to the reduction in study group sizes.[15]

The publication by Vitiello *et al*[17] analyses suicidal events in more detail. Patients with high or increased scores, but not classified as having an event, were not included in the analysis. Nine cases of suicidal behaviour were presented as occurring in the PBO group, even though the patients were using FLX at the time and the PBO period had ended. The paper reports on the number of cases, but does not include results from the suicidality scoring tools CDRS-R Item 13 and SIQ-Jr. The number of suicidal episodes was greater than it appears, as seven patients had more than one episode,[17] and only the most severe episode was included in the analysis.

The long term phase IV publication[18] present SIQ-Jr scores for a total of 66 patients who had at least one stage IV assessment. The paper refers to the baseline ITT groups of 327 patients (excluding PBO), but due to withdrawals any changes in scores may be biased, and reflect a selected study population rather than a treatment effect.

Reporting of psychiatric adverse effects/mania across TADS publications

The TADS group found higher rates for psychiatric adverse events in patients receiving FLX than in patients receiving CBT or PBO.[6 12] The psychiatric adverse events included symptoms classified as mania spectrum, irritability/depression spectrum, agitation spectrum, anxiety or other. Of these, mania spectrum symptoms were described in greater detail in the 2006 safety publication.[12] We have therefore assessed and summarised the reporting of mania spectrum symptoms across the TADS publications (table 2).

Mania spectrum symptoms (mania, hypomania and elevated mood) were monitored using an ADS procedure, as well as adverse event or serious adverse event forms. Due to the adverse event definition threshold, new cases of emerging mania were not recorded unless the symptoms caused clinically significant interference with functioning.[7]

Mania spectrum symptoms were mentioned in three of the four publications that reported on adverse effects in TADS during 0–12 weeks of treatment (stage I). The initial 2004 publication by the TADS group reported a total of seven patients with mania spectrum symptoms as an adverse effect; four in the FLX group, one in the COMB group, none in the CBT group and two in the PBO group.[6] In the 2006 safety results publication,[12] occurrences of mania spectrum symptoms were reported based on both spontaneous reports and assessment by physician using a formal symptom checklist (ADS mania items). According to this publication, six patients spontaneously reported a mania spectrum disorder; four in the FLX group, one in the COMB group and one in the PBO group. On the ADS mania scoring scale, however, 65 of 424 patients across all treatment groups reportedly had an increase of 3 points or more. The absolute score increase for each patient or treatment group is not provided. The analysis of patients with at least one suicidal event (n=44) describes mean ADS mania score prior to the suicidal event for 31 of the 44 patients during 36 weeks of treatment.[17]

We did not identify any publication describing mania spectrum symptoms in the entire study population that received treatment for more than 12 weeks (stages II–IV) (table 2).

The publications from stage II–IV failed to mention psychiatric adverse effects that were identified during stage I, such as restlessness, nervousness and sleep difficulties (table 1).

Other adverse effects

Adverse effects other than suicidality were summed up by the TADS team in 2004,[6] reported in further detail in 2006[12] and mentioned in the two other publications from study stage I to a varying extent.[13 14 19] According to the most extensive publication with regard to safety data at 12 weeks,[12] sedation, insomnia, vomiting and upper abdominal pain occurred at least twice as often in patients receiving FLX with or without CBT than with PBO. We did not identify any publication describing non-psychiatric adverse effects in the study population that received treatment for more than 12 weeks (stages II–IV) (table 1).

Adverse effects of FLX, as acknowledged at present, are listed in table 3. The adverse effects are classified according to whether they were reported in any of the eight TADS publications or not. Several well-known adverse effects of FLX were not reported in the TADS publications, among them weight and appetite changes. Effects on sexual functioning are not mentioned in this group of young patients.

DISCUSSION

The TADS protocol included a threshold limit on what would be considered an adverse event, specifying that the event must cause clinically significant interference with functioning, require medical attention or cause a need to take medication.[6] As an example, emerging mania was not recorded unless symptoms exceeded this threshold.[7] It must be assumed that this reduced the number of reported adverse effects, which may not have been severe enough to reduce daily functioning or cause a need for additional treatment. We have not been able to find a published version of the questionnaires that were used and consequently do not have information as to which adverse effects were specifically asked for. The protocol does not define how the scoring parameters for adverse events should be analysed. The number of suicidal events

Table 2 Reporting of mania spectrum symptoms in publications from the TADS

	Stage 1 (12 weeks)				Stage 2+3 (36 weeks)			Stage 4 (88 weeks)
Reporting parameter	**TADS team[6]**	**Emslie *et al*[12]**	**Kennard *et al*[13]**	**May *et al*[14]**	**TADS team[15]**	**Kennard *et al*[16]**	**Vitiello *et al*[17]**	**TADS team[18]**
ADS Mania subscale score		Baseline: All: 2.4±2.3 COMB: 2.6±2.4 FLX: 2.2±2.2 CBT: 2.5±2.4 PBO: 2.2±2.3 12 weeks: All: 0.9±1.4 COMB: 0.5±0.8 FLX: 1.1±1.0 CBT: 1.0±1.2 PBO: 1.1±0.1					Baseline: 2.5±2.2 Prior to suicidal event: 1.6±2.2 Mean change before event: −0.6±2.3	
ADS Mania subscale score increase (≥3 points)		All: 65/424 (15.3%) COMB: 20% (n=21) FLX: 14.2% (n=15) CBT: 12.3% (n=13) PBO: 15.0% (n=16)						
Patients with attrition prevention due to mania/hypomania				1.28% (1/78)				
Mania	COMB: n=0 FLX: n=1 CBT: n=0 PBO: n=1	FLX: n=1						
Hypomania	COMB: n=1 FLX: n=2 CBT: n=0 PBO: n=1	COMB: n=1 FLX: n=2 PBO: n=1						
Elevated mood	COMB: n=0 FLX: n=1 CBT: n=0 PBO: n=0	FLX: n=1						

ADS, affective disorders screening; CBT, cognitive behavioural therapy; COMB, cognitive behavioural therapy plus fluoxetine; FLX, fluoxetine; PBO, placebo; TADS, Treatment for Adolescents With Depression Study.

Table 3 TADS reporting of presently acknowledged common adverse effects of fluoxetine[30]

Mentioned in publications from the TADS *	Not mentioned in publications from the TADS
Insomnia	Decreased appetite, incl. anorexia
Sleep disorder	Weight decreased
Abnormal dreams, incl. nightmares	Tension
Anxiety	Libido decreased, incl. loss of libido
Somnolence, incl. hypersomnia, sedation	Gynaecological bleeding, incl. menstrual bleeding disorders
Nervousness	Erectile dysfunction
Restlessness	Ejaculation disorder
Headache	Dizziness
Disturbance in attention	Dysgeusia
Tremor	Lethargy
Palpitations	Vision blurred
Diarrhoea	ECG QT prolonged
Nausea	Flushing, incl. hot flushes
Vomiting	Yawning
Dry mouth	Dyspepsia
Rash	Chills
Urticaria (hives)	Feeling jittery
Pruritus	
Hyperhidrosis	
Arthralgia	
Frequent urination	
Fatigue	

*Not necessarily identified as an adverse effect of fluoxetine treatment.
TADS, Treatment for Adolescents With Depression Study.

is described, but other parameters, such as absolute or worsening scores on risk assessment scales, are not consistently reported. An example is the SIQ-Jr scores, where week 12 publications report mean scores and number of patients with score increase to ≥31,[6 12] while the follow-up publication by week 36 reported per cent of patients with SIQ-Jr score ≥31.[15] Scoring of mania symptoms is described as inconsistent and varying between clinicians.[12] It is conceivable that some patients may have had worsening scores without passing the threshold score for suicidality or mania, respectively. Conversion into dichotomous scales, as was done for SIQ-Jr scores ≥31 and ADS Mania subscale score change increase ≥3 points, does not give insight into the magnitude in case of increased scores.

All analyses were planned as ITT, regardless of later events.[7] Nine cases of suicidal behaviour were presented as occurring in the PBO group[17] although the patients were using FLX at the time and the PBO period had ended. As pointed out by Högberg *et al*,[10] the risk of suicidal behaviour will not appear to be increased for FLX compared with PBO if patients using FLX are assessed in the PBO group. ITT analyses of adverse events may be biased towards finding no differences between groups.[20] This is especially relevant in studies with large drop-out rates and in study groups where patients received add-on treatment that differed from the assigned medication, as was the case in the TADS.[19] Other authors have questioned whether the TADS may have under-reported adverse effects due to small numbers and patients leaving the study early.[11] Use of ITT analyses will have led to underestimation of the frequency of psychiatric and other adverse events, a fact which has been little discussed.

Risk of suicidal behaviour was the only adverse effect that was addressed during all four treatment stages. Several psychiatric and physical adverse effects were reported during the first 12 weeks, but not mentioned in publications from the further treatment stages. Examples are sedation, insomnia, vomiting and upper abdominal pain, which occurred in more than 2% of patients in the first 12 weeks.[12] The 2% occurrence is described as infrequent (≤5%), but should more correctly be classified as common.[21] The risk of psychiatric adverse events such as mania, irritability, agitation and anxiety is given as 11% in the FLX group and 5.6% in the COMB group.[12] In the review by Jane Garland *et al*, the occurrence of emotional/behavioural adverse effects is given as 10%–25%,[3] but the numbers may not be comparable due to different inclusion criteria. Other adverse effects of SSRI treatment, such as appetite changes, weight changes and sexual problems, are not mentioned in any publication. Growth issues were not addressed. Changes in weight or appetite may have occurred without reaching the severity threshold. Sexual adverse effects may not have been relevant to many patients at the time due to their age, or may not have been forthcoming in interviews, especially as many patients were interviewed in the company of caregivers.[12] Risk of sexual adverse effects was discussed in the adverse event monitoring protocol[22] and procedures in case of pregnancies were established,[23] so it is reasonable to assume a that certain proportion of patients were sexually active. Prolonged treatment into adulthood may well increase the relevance of such concerns.

To our knowledge, this is the first systematic assessment of adverse effects reporting in publications from the TADS. We conducted an extensive literature search and believe that all relevant studies have been identified, however, we cannot exclude the possibility that some publications may have been overlooked. Our findings regarding adverse effect reporting and potential for bias are based on analysis of only one study and do not give information on adverse effects reporting or bias in other studies of SSRIs in adolescents. However, discrepancies and weaknesses in the reporting of

adverse events in such studies have also been noted by other authors.[24 25] We have not had access to primary data.

A previous assessment of the adverse effects reporting in TADS focused on the occurrence of suicidal events and increased risk of suicidal behaviour[10] and this is reflected in the most recent Cochrane review.[1] Like Högberg *et al*,[10] we have noted the misleading PBO group classification of patients with a suicidal event who were using FLX at the time. Our analysis encompasses all adverse events mentioned in publications from the TADS. Gaps and discrepancies in coding, transcription and reporting of harms in clinical trials have been reported, and the number of adverse events may differ between study reports and published papers.[24 26] Several barriers to accurate harms reporting[24] are relevant to the TADS, notably the severity threshold, conversions from continuous to dichotomous outcomes, individual judgements of association between event and medication, handling of adverse events in patients who discontinued treatment and the extensive use of concomitant medications. In future studies, the potential for bias may be substantially reduced by avoiding severity thresholds and defining a consistent method of describing adverse effects such as suicidal risk and mania score worsening. Occurrence or worsening of mania and other psychiatric adverse effects for individual patients should be reported in more detail. We would also suggest that if risk is presented as percentages, it should be calculated based on the number of patients who were receiving treatment at the time the adverse event occurred. This will be of particular importance in studies with large drop-out rates and treatment changes. The full spectrum of adverse effects should be reported for all study stages. A plan for data sharing should be in place to facilitate reanalysis and evaluation by other researchers, as practised by the *BMJ*.[27]

Due to its long duration (36 weeks) and follow-up (1 year), the TADS could have provided valuable information on the long-term occurrence of adverse effects both in frequency and severity. The adverse effects profile of FLX in the TADS has only been reported in detail for stage 1, where approximately 200 patients received FLX for 12 weeks. The raw data from the trial have been requested[28] and planned for release into the public domain,[29] but we have not been able to ascertain that these have been made publicly available. The incomplete reporting of adverse effects in a major study like TADS may lead to bias and erroneous conclusions regarding the safety profile of FLX when given to minors. The risk of suicidal behaviour has been the subject of many discussions and regulatory actions, but there has been considerably less focus on other clinically important adverse effects. This may have clinically important implications, since the benefit/risk estimations regarding FLX use in adolescents will be biased. If adverse effects are not acknowledged as such, there is a risk that symptoms may be misinterpreted and treated as more serious illnesses.

Contributors SN suggested the research question. All authors discussed and defined the project. TW and SN researched the literature and made the initial assessments. All authors discussed the publications included in the study, including interpretation and presentation of results. TW drafted and finalised the manuscript as lead author. SN and MK commented on the draft and revised the manuscript at all stages. All authors agreed to the final version of the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES

1. Hetrick SE, McKenzie JE, Cox GR, *et al*. Newer generation antidepressants for depressive disorders in children and adolescents. *Cochrane Database Syst Rev* 2012;11:Cd004851.
2. Locher C, Koechlin H, Zion SR, *et al*. Efficacy and safety of selective serotonin reuptake inhibitors, serotonin-norepinephrine reuptake inhibitors, and placebo for common psychiatric disorders among children and adolescents: a systematic review and meta-analysis. *JAMA Psychiatry* 2017;74:1011–20.
3. Jane Garland E, Kutcher S, Virani A, *et al*. Update on the use of SSRIs and SNRIs with children and adolescents in clinical practice. *J Can Acad Child Adolesc Psychiatry* 2016;25:4–10.
4. National Institute of Mental Health. NIH Guide: Effectiveness of treatment for adolescents with major depression. 1998 https://grants.nih.gov/grants/guide/notice-files/not98-059.html (accessed 20 Mar 2017).
5. Treatment for Adolescents With Depression Study Team. Treatment for Adolescents With Depression Study (TADS): rationale, design, and methods. *J Am Acad Child Adolesc Psychiatry* 2003;42:531–42.
6. March J, 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;292:807–20.
7. The TADS Team Duke University Medical Center. Protocol RFP-NIH-NIMH 98-DS-0008 Treatment for Adolescents with Depression Study (TADS): Duke Clinical Research Institute. 2000 tads.dcri.org (accessed 20 Mar 2017).
8. The TADS Team Duke University Medical Center. Adverse event monitoring and adjunctive services and attrition prevention (ASAP) in the NIMH Treatment of Adolescent Depression Study (TADS) vs 4.1: Duke Clinical Research Institute. 2005 tads.dcri.org (accessed 20 Mar 2017).
9. Walkup JT. Antidepressant efficacy for depression in children and adolescents: industry- and NIMH-funded studies. *Am J Psychiatry* 2017;174:430–7.
10. Högberg G, Antonuccio DO, Healy D. Suicidal risk from TADS study was higher than it first appeared. *Int J Risk Saf Med* 2015;27:85–91.
11. Jureidini J, Tonkin A, Mansfield PR. TADS study raises concerns. *BMJ* 2004;329:1343–4.
12. Emslie G, Kratochvil C, Vitiello B, *et al*. Treatment for Adolescents with Depression Study (TADS): safety results. *J Am Acad Child Adolesc Psychiatry* 2006;45:1440–55.
13. Kennard B, Silva S, Vitiello B, *et al*. Remission and residual symptoms after short-term treatment in the Treatment of Adolescents with Depression Study (TADS). *J Am Acad Child Adolesc Psychiatry* 2006;45:1404–11.
14. May DE, Kratochvil CJ, Puumala SE, *et al*. A manual-based intervention to address clinical crises and retain patients in the Treatment of Adolescents With Depression Study (TADS). *J Am Acad Child Adolesc Psychiatry* 2007;46:573–81.
15. March JS, Silva S, Petrycki S, *et al*. The Treatment for Adolescents With Depression Study (TADS): long-term effectiveness and safety outcomes. *Arch Gen Psychiatry* 2007;64:1132–43.
16. Kennard BD, Silva SG, Mayes TL, *et al*. Assessment of safety and long-term outcomes of initial treatment with placebo in TADS. *Am J Psychiatry* 2009;166:337–44.

17. Vitiello B, Silva SG, Rohde P, *et al*. Suicidal events in the Treatment for Adolescents With Depression Study (TADS). *J Clin Psychiatry* 2009;70:741–7.
18. March J, Silva S, Curry J, *et al*. The Treatment for Adolescents With Depression Study (TADS): outcomes over 1 year of naturalistic follow-up. *Am J Psychiatry* 2009;166:1141–9.
19. Kennard BD, Silva SG, Tonev S, *et al*. Remission and recovery in the Treatment for Adolescents with Depression Study (TADS): acute and long-term outcomes. *J Am Acad Child Adolesc Psychiatry* 2009;48:186–95.
20. CIOMS Working Group VI. Management of safety information from clinical trials. Geneva: Council for International Organizations of Medical Sciences (CIOMS). 2005 https://cioms.ch/wp-content/uploads/2017/01/Mgment_Safety_Info.pdf (accessed 14 Jan 2019).
21. Uppsala Monitoring Centre. Glossary of pharmacovigilance terms. 2018 https://www.who-umc.org/global-pharmacovigilance/global-pharmacovigilance/glossary/ (accessed 14 Jan 2019).
22. The TADS Team Duke University Medical Center. Treatment for adolescents with depression study (TADS) Pharmacotherapy Treatment Manual Final Version (5.1). 2000 http://tads.dcri.org/tads-manuals (accessed 10 Jan 2019).
23. The TADS Team Duke University Medical Center. Procedures for managing pregnancy risk in the NIMH treatment of adolescent depression study (TADS). 2001 http://tads.dcri.org/tads-manuals (accessed 11 Jan 2019).
24. 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.
25. 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.
26. Maund E, Tendal B, Hróbjartsson A, *et al*. Benefits and harms in clinical trials of duloxetine for treatment of major depressive disorder: comparison of clinical study reports, trial registries, and publications. *BMJ* 2014;348:g3510.
27. Loder E, Groves T. The BMJ requires data sharing on request for all trials. *BMJ* 2015;350:h2373.
28. Mann H. Public release of raw data from the TADS trial. *BMJ* 2005;330:730.
29. March JS. TADS Group. Authors of TADS study reply to letter raising concerns. *BMJ* 2005;330:730.3–1.
30. Prozac. Summary of Product Characteristics: Eli Lilly and Company Limited. 2018 https://www.medicines.org.uk/emc/product/3768/smpc#PRODUCTINFO (accessed 31 May 2018).

Paper IV

BMJ Open

Adverse effects information in clinical guidelines on pharmacological treatment of depression in children and adolescents: a systematic review

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

To cite: 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**:e036412. doi:10.1136/bmjopen-2019-036412

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2019-036412).

Received 13 December 2019
Revised 15 April 2020
Accepted 20 May 2020

[1]Regional Medicines Information and Pharmacovigilance Centre (RELIS Sør-Øst), Dept. of Pharmacology, Oslo University Hospital, Oslo, Norway
[2]Dept. of Pharmacology, Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[3]Center for Psychopharmacology, Diakonhjemmet Hospital, Oslo, Norway

Correspondence to
Tone Westergren;
twesterg@ous-hf.no

ABSTRACT

Objectives To analyse to what extent clinical practice guidelines on drug treatment of depression in children and adolescents mention the risk of adverse effects, to characterise the citations in the guidelines and to assess to what extent data from a major study (Treatment for Adolescents With Depression Study, TADS) was used as basis for information about adverse effects.

Design Systematic review of clinical guidelines and clinical decision support tools.

Data sources PubMed, EMBASE, guideline collections, Health libraries.

Eligibility criteria We included national guidelines on depression in children and adolescents from European and/or English-speaking countries, published in English, German, French or any Scandinavian language since 2008. We also included well-known, international clinical decision support tools.

Data extraction and synthesis Guidelines were examined by all authors to identify and classify information on adverse effects. Citations for statements on adverse effects were extracted and classified by category. The extent of citations about suicidality risk versus other adverse effects was assessed.

Results 19 guidelines were assessed. All guidelines discussed risk of suicidal behaviour connected with use of antidepressants. Most guidelines mentioned some other psychiatric adverse effects. Several guidelines did not include information on well-known and common somatic adverse effects. Most references concerned risk of suicidality. Adverse effects identified in underlying studies were not always presented. The TADS study was referred to, directly or indirectly, by 18/19 guidelines, but some only referred to TADS with regard to suicidality without citing the study's findings of somatic adverse effects. No guideline commented on the lack of long-term adverse effects data from TADS.

Conclusions Guidelines for treatment of depression in children and adolescents vary widely regarding information on adverse effects. Many guidelines do not provide information on common somatic adverse effects. There is no consensus as to what extent risks of adverse effects connected with use of antidepressants should be described in guidelines.

Strengths and limitations of this study

- ► The search for guidelines was extensive.
- ► Inclusion of guidelines was not limited to English texts.
- ► The main limitation is that some guidelines may have been overlooked and that the cut-off for guideline inclusion can be debated on a geographical and local level.
- ► The guideline committee mandate and scope may have varied between guidelines.
- ► Guidelines were searched for all relevant statements about adverse effects and corresponding citations, but some statements or text excerpts may have been overlooked.

INTRODUCTION

The use of antidepressants in children and adolescents with depression is the topic of several national and international guidelines which appear to differ considerably in extent, quality and information on potential harms.

As defined by the US Institute of Medicine in 2011, clinical practice guidelines are statements that include recommendations intended to optimise patient care that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options. Guidelines should provide a clear description of potential benefits and harms for each recommendation.[1 2] However, compliance with standards has been shown to be low for several parameters, including more focus on benefits than on potential harms.[3] Other researchers have assessed the quality of clinical guidelines regarding benefits of antidepressants[4 5] using the Appraisal of Guidelines for Research& Evaluation II (AGREE II) tool.[6] We have not identified any specific appraisals of the adverse effects information provided in guidelines on antidepressant therapy.

For fluoxetine, which is the suggested medication of first choice, few high-quality

clinical trials are available to assess risk–benefit.[7] The largest randomised controlled clinical trial on fluoxetine in children and adolescents with depression is a National Institute of Mental Health (NIMH)-funded study (Treatment for Adolescents With Depression Study (TADS)),[8] which is considered a high-quality trial.[9] We have previously analysed the adverse effects reporting in the TADS study[10] and found that little data had been published concerning adverse effects profile during treatment for more than 12 weeks, with the exception of suicidality.

Based on our previous study,[10] we had reason to believe that adverse effects descriptions in many guidelines on treatment of depression in children and adolescents would focus primarily on suicidality, with less attention given to the risk of other adverse effects and the risk-benefit balance. The adverse effects profile of the most commonly used medications, the selective serotonin reuptake inhibitors (SSRIs), is generally assumed to be similar in adults and children/adolescents and similar between the different SSRIs. This includes psychiatric effects such as suicidality, mania, anxiety, agitation and sleep disorders, as well as gastrointestinal effects such as nausea, diarrhoea and anorexia/weight loss. In this review, we aimed to analyse to what extent adverse effects data on SSRIs were mentioned in clinical practice guidelines on treatment of depression in children and adolescents. We also aimed to characterise the documentation provided as references in the clinical guidelines, and to assess to what extent data from the TADS study, with the identified data gaps, was used as basis for information about adverse effects.

METHODS

To identify guidelines and clinical evidence summarisations on treatment of depression in children and adolescents, a search was performed according to the Norwegian Health Library guidelines on literature searches for development of clinical procedures.[11]

Literature database search

Searches in PubMed and guideline collections were carried out in the period 30 October 2018 to 17 February 2019. An EMBASE search with time limits 2008–2019 was performed 5 December 2019.

Several PubMed searches were performed, using the search terms:

- ("depressive disorder"[MeSH Terms] OR ("depressive"[All Fields] AND "disorder"[All Fields]) OR "depressive disorder"[All Fields] OR "depression"[All Fields] OR "depression"[MeSH Terms]) AND (("guideline"[Publication Type] OR "guidelines as topic"[MeSH Terms] OR "guideline"[All Fields]) OR ("practice guideline"[Publication Type] OR "practice guidelines as topic"[MeSH Terms] OR "clinical practice guideline"[All Fields])) AND ("humans"[MeSH Terms] AND ("infant"[MeSH Terms] OR "child"[MeSH Terms] OR "adolescent"[MeSH Terms]))
- "Depressive Disorder/drug therapy"[MAJR] AND "Antidepressive Agents/therapeutic use"[MAJR] AND (Guideline[ptyp] OR Practice Guideline[ptyp])
- ((practice guidelines as topic) AND adolescent psychiatry) AND depressive disorder

Filters: Humans, Child: birth-18 years

EMBASE search terms were: (exp adolescent depression/ OR exp major depression/ OR exp depression/) AND practice guideline/. Limits: (child <unspecified age>or adolescent <13 to 17 years>) and yr="2008–2019"

Searches in clinical guidelines collections

Clinical treatment guidelines were identified through searches in Guidelines International Network, McMaster Plus, Epistemonikos, UpToDate, BMJ Best Practice, DynaMed Plus, International Network of Agencies for Health Technology Assessment, National Institute for Health and Care Excellence (UK), Cochrane library, the Norwegian Health Library, Sundhetsstyrelsen (Denmark), Socialstyrelsen (Denmark), Center for Clinical Guidelines (Denmark), Socialstyrelsen (Sweden), AHRQ website (USA), IQWIQ website (Germany), Psychenet website (Germany), guidelines.gov (USA), SIGN website (Scotland), Health Canada website, CADTH website (Canada). In addition, a manual reference search was performed on identified studies and guidelines. The guideline search and selection process is described in a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow chart.[12]

Inclusion and exclusion criteria

We included national guidelines from European and/or English-speaking countries, published in English, German, French and any Scandinavian language since 2008, covering the last 10 years. We also included well-known and widely used international decision support tools. Local hospital or county guidelines were not included.

The guidelines and decision support tools (hereafter referred to as guidelines) should specifically address depression in children and adolescents. In cases where we retrieved multiple guidelines from the same organisation, the most recent version was analysed.

Data extraction

The guidelines were examined to identify information on safety or adverse effects. All authors made an overall assessment of the full guidelines. Text excerpts and corresponding references concerning safety or adverse effects were initially extracted by TW. Data extraction from all guidelines was checked by SN and MK. Text excerpts were assessed in detail by all authors. Information on adverse effects provided in each guideline was analysed and classified by organ system, in accordance

with the classification used in the product monograph for Prozac (fluoxetine) by Eli Lilly.[13] In order to distinguish suicidality or self-harm from other psychiatric adverse effects, these were registered as a separate category. Analysis was limited to adverse effects occurring during treatment and did not include mention of withdrawal reactions.

We assessed whether the guidelines provided information on the quality of the underlying evidence (high, moderate, low or very low) and strengths of recommendations (strong or weak) by use of the Grading of Recommendations Assessments, Development, and Evaluation (GRADE) tool[14 15] or similar criteria for quality ratings.

We also evaluated the guidelines regarding presentation and risk–benefit assessment. To our knowledge, there are no explicit criteria for such evaluations. The guidelines were assessed by all authors independently, according to whether the guideline provided an overall discussion on risks, made an attempt to assess risks and benefits together, gave information on handling adverse effects, provided frequency estimates on risks or gave recommendations on monitoring patients. Any discrepancies in judgement were discussed by all authors.

Classification of adverse effects

References that were provided as basis for statements on adverse effects were extracted from all guidelines, and the most cited references were identified. Adverse effects were mainly classified by System Organ Class according to the MedDRA classification, as used in European Summaries of Product Characteristics.[16] In this classification, appetite disorders are classified under Metabolism and nutrition disorders; sleep disorders, restless, anxiety, mania and mood disorders are classified under psychiatric disorders and headache, dizziness, and somnolence are classified under nervous system disorders. Suicidal thoughts and behaviour are classified under psychiatric disorders in MedDRA, but has been noted separately in this review. Reproductive system and breast disorders include sexual dysfunction and erectile and ejaculation disorders. Libido changes are classified under Psychiatric disorders in MedDRA, we have, however, included any such information in the Reproductive system and breast disorders category, in order to show a comprehensive view on all sexual adverse effects. An overview of the System Organ Classes and the included adverse effects can be found in the online supplementary material table S1.

The references were assessed and classified by category. Analysis by category and type of adverse effects was performed to assess to what extent the citations were used to discuss risk of suicidality versus other adverse effects.

Patient and public involvement

No patients were involved in this review.

RESULTS

Inclusion of guidelines

Nineteen guidelines were included in the final analysis (table 1). A PRISMA flow diagram describing the search and screening process is shown in figure 1.

Adverse effects mentioned in guidelines

We identified 20 adverse effects categories in the summary of product characteristics for Prozac.[13] The guidelines' mention of these adverse effects categories were highly variable (table 2).

The 19 guidelines mentioned an average of 5.3 adverse effects categories (median 4.0, range 1–15). All 19 guidelines included information on suicidal risk. Fifteen of 19 mentioned other types of psychiatric adverse reactions, including mention of manic switching or manic/hypomanic episodes in seven guidelines.

Seven of 19 guidelines limited their information on adverse effects to suicidality and/or psychiatric reactions. Three guidelines mentioned risk of suicidal behaviour only (figure 2). Nervous system adverse effects including headache was mentioned by 10/19 guidelines, as was also the case for gastrointestinal reactions. Several possible adverse effects were only mentioned in a few guidelines, as an example, the risk of sexual adverse effects was mentioned in 5/19 guidelines. Four guidelines mentioned adverse effects in 10 organ categories or more.[17–20]

The overall assessments of somatic and nervous adverse effects varied widely, as illustrated by different descriptions of the SSRI risk profile. While one guideline stated that studies have shown somatic adverse effects to have small significance,[21] another guideline stated that SSRI treatment causes significantly more nausea, diarrhoea, anorexia and stimulatory side effects (agitation, insomnia and anxiety) than tricyclic antidepressants.[22] In our evaluation of the guidelines regarding presentation and risk–benefit assessment, we judged that 9/19 guidelines could be classified as having a sufficiently extensive and balanced consideration of adverse effects in their overall recommendations, while 10 guidelines did not (see online supplementary material table S2). Nine of 19 guidelines presented grading of evidence and strength of recommendations by the GRADE tool or similar, though the evidence grading mostly reflected efficacy documentation (see online supplementary material table S3).

References in guidelines as basis for adverse effects information

We identified 124 specific references as basis for statements in the 19 guidelines (see online supplementary material table S4). Some guidelines mentioned general sources of information, such as 'FDA warnings', 'Cochrane reviews', or 'Product monographs'. The guidelines had a mean number of adverse effects references of 9.3 (range 0–32).

Table 1 Clinical practice guidelines on treatment of depression in children and adolescents

No	Title	Publisher	Country	Year	Category
1	Guidelines for Adolescent Depression in Primary Care: Part II. Treatment, Ongoing Management.[30]	American Academy of Pediatrics	USA	2018	Guideline
2	Depression in children and adolescents.[31]	DynaMed Plus, EBSCO Health	USA	2018	Decision support tool
3	Pediatric unipolar depression and pharmacotherapy: Choosing a medication.[17]	UpToDate	USA	2017	Decision support tool
4	Clinical Guidelines for the Management of Adults with Major Depressive Disorder: Section 6.[32]	Canadian Network for Mood and Anxiety Treatments	Canada	2016	Guideline
5	Anxiety and Depression in Children and Youth - Diagnosis and Treatment.[18]	British Columbia Guidelines Protocols Advisory Committee	Canada	2010	Guideline
6	Depression in children.[24]	BMJ Best Practice	UK	2018	Decision support tool
7	Depression in children and young people: identification and management (CG28).[33]	NICE (UK) National Institute for Health and Care Excellence	UK	2017	Guideline
8	Evidence-based guidelines for treating depressive disorders with antidepressants.[22]	British Association for Psychopharmacology	UK	2015	Guideline
9	Treating depression in young people.[34]	Orygen National Centre of Excellence in Youth Mental Health	Australia	2017	Guideline
10	Identification of Common Mental Disorders and Management of Depression in Primary Care.[35]	New Zealand Guidelines Group/Ministry of Health	New Zealand	2008	Guideline
11	Manifestations dépressives à l'adolescence : repérage, diagnostic, prise en charge en soins de premier recours.[19]	Haute Authorité de Santé	France	2014	Guideline
12	S3-Leitlinie. Behandlung von depressiven Störungen bei Kindern und Jugendlichen.[20]	Deutsche Gesellschaft Kinder- und Jugendpsychiatrie, Psychosomatik, Psychotherapie	Germany	2013	Guideline
13	Håndbok for barn og unges psykiske helse.[36]	Center for Child and Adolescent Mental Health. Eastern, Southern Norway	Norway	2019	Guideline
14	Pediatriveilederen. Psykiske lidelser og psykososiale tilstander. 12.3 Angst og depresjon.[37]	Norwegian Pediatric Association	Norway	2018	Guideline
15	Vejledning om medikamentel behandling af børn og unge med psykiske lidelser.[38]	Danish Health Authority	Denmark	2013	Guideline
16	Landsdækkende klinisk retningslinje vedrørende udredning og behandling af depression hos børn og unge.[39]	Børne-og Ungdomspsykiatrisk Selskab	Denmark	2011	Guideline
17	Nationella riktlinjer för vård vid depression och ångestsyndrom.[21]	The National Board of Health and Welfare	Sweden	2017	Guideline
18	Childhood depression.[40]	EBM Guidelines Duodecim	Finland	2018	Guideline
19	Depression of adolescents.[25]	EBM Guidelines Duodecim	Finland	2018	Guideline

EBM, evidence based medicine; NICE, National Institute for Health and Care Excellence.

The 124 references were cited 177 times, as some references had multiple citations (table 3). The largest group of references was single studies, followed by review articles, statements or warnings from medicinal authorities and systematic reviews (see online supplementary material table S4). Most references to review articles, authorities, systematic reviews and guidelines concerned risk of suicidality. For the single studies, however, 22/52 citations (42,3%) referred to adverse effects in other organ systems (see online supplementary material table S4). These included 14 clinical trials. Somatic adverse effects were referred to in 3.2 organ categories for each citation (mean value), minimum value 1, maximum value 10.

Forty-one of 177 citations concerned adverse effects other than suicidality and/or other psychiatric adverse effects (see online supplementary material table S4). Most of those referred to somatic adverse effects in few organ systems (mean value 2.8, median 2.0, minimum value 1, maximum value 10).

Overall, 116 citations concerned risk of suicidal behaviour and 54 citations concerned risk of other psychiatric adverse effects, while fewer citations

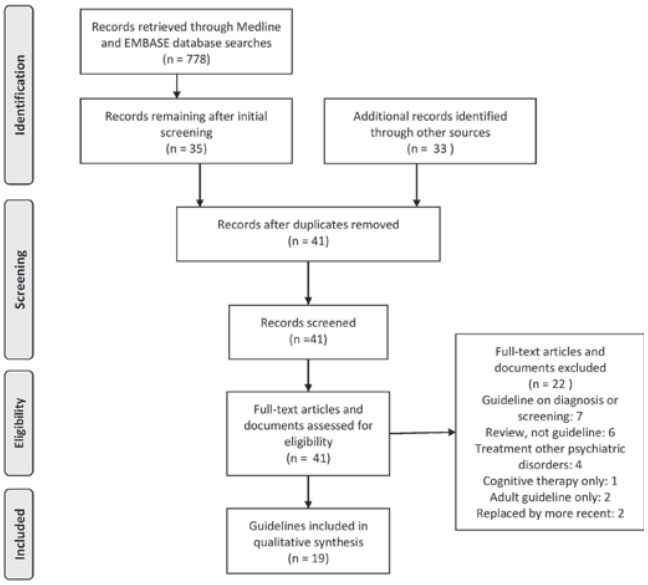

Figure 1 Identification of guidelines for treatment of depression in children and adolescents. PRISMA flowchart. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

concerned somatic adverse effects. For some known adverse effects according to the product monograph,[13] no guidelines provided any citations (figure 3).

Data from the TADS study as basis for information about adverse effects

Of the 19 guidelines, seven referred directly to publications from the TADS study, while 11 referred to sources that cited TADS as part of their assessment. One guideline did not provide any references and consequently did not refer to the TADS study.

Adverse effects data from the TADS study were published by March *et al*[8] and in more detail by Emslie *et al* in 2006.[23] Both publications describe a broad range of adverse effects, including risk of suicidal behaviour, psychiatric adverse effects such as mania, sedation or sleeping problems, abdominal pain, diarrhoea and vomiting during the first 12 weeks of treatment. These articles were cited by seven and two guidelines, respectively. Most guidelines referring to TADS, directly or indirectly, described psychiatric, nervous, gastrointestinal and respiratory adverse effects. However, three guidelines[20 24 25] that cited the TADS publication from March 2004[8] did not cite the study's findings of somatic adverse effects. No guideline commented on the lack of long-term adverse effects data from TADS beyond 12 weeks.

DISCUSSION

Our analysis of 19 treatment guidelines of depression in children and adolescents shows that the guidelines vary widely regarding information on adverse effects. The most commonly used medications, the SSRIs, may have multiple adverse effects involving several organ systems, but few guidelines described the full adverse effect profile. Several guidelines mentioned only risk of suicidal behaviour with or without mention of other psychiatric adverse effects. In many guidelines where adverse effects are mentioned, the frequency or the benefit-risk assessment is missing. Overall, somatic adverse effects were mentioned to little extent in the guidelines, and mention was often limited to few organ systems. The reporting of adverse effects in the guidelines appears to be selective and arbitrary. Several guidelines give the impression that suicidality is the only safety issue of consequence, and that there is a minor and limited risk of other adverse effects. We have not been able to identify a current, accepted standard for inclusion of adverse effects information in guidelines. It is possible that the varying descriptions of adverse effects in the guidelines are due to a lack of consensus as to what should be included. It can be argued that a full spectrum of adverse effects should be described. However, limitations may be necessary due to readability and format, in which case selection criteria should be stated.

All guidelines indicated suicidality as a major risk factor, reflecting that this has been a major topic of safety discussions. This was also the case for the underlying literature references. Most citations concerned suicidality with or without other psychiatric adverse effects. Far fewer citations concerned adverse effects in other organ systems. The fact that most references to review articles, authorities, systematic reviews and guidelines concerned risk of suicidality indicate that suicidality risk has been the main focus when addressing adverse effects of SSRIs. There were indications of selective citing, where the underlying articles provided more details of adverse effects than was referred to in the guidelines. The most frequently cited reference, the Cochrane database review by Hetrick *et al*,[7] provides details on several adverse effects observed in the included studies, but most guidelines referred to this review only in the context of suicidality risk. Likewise, the TADS publication by March *et al*[8] describes several somatic adverse effects which were not mentioned in many guidelines that cited the study on suicidality risk. A later publication on the range of adverse effects observed in the TADS study[23] was cited by two guidelines only. We found that most of the known somatic adverse effects were not mentioned in the majority of the guidelines, thereby giving users a biased and skewed impression of the risk of harms. This lack of adverse effect information may affect treatment of children and adolescents with depression.

Guidelines should include descriptions of benefits and harms for the recommendations, however, this is a general statement and does not specify to what extent possible harms and adverse drug reactions should be described. The recommended tool for assessment of guideline quality, the AGREE II tool[6] includes the criterion 'The health benefits, side effects and risks

Table 2 Adverse effects categories mentioned in guidelines

Guideline	At least one adverse effect in category mentioned in guideline ■Yes □No																			
	Suicidality	Other psychiatr.	Blood, lymphat	Immune	Endocrine	Metabol, nutrit	Nervous system	Eye	Ear	Cardiac	Vascular	Respiratory	Gastrointestin	Hepatobiliary	Skin	Muskuloskeletal	Renal, urinary	Reprod, breasts	Gen eral	Investigations
Manifestations dépressives à l'adolescence. Haute Authorité Santé, France[19]	■	■	□	□	■	■	■	■	□	■	■	□	■	■	■	■	■	■	■	□
Behandlung von depressiven Störungen bei Kindern und Jugendlichen. Deutsche Gesellschaft Kinder Jugendpsychiatrie[20]	■	■	■	□	□	■	■	■	■	■	□	■	■	□	■	■	□	□	■	□
Pediatric unipolar depression and pharmacotherapy: UpToDate, USA[17]	■	■	□	□	□	■	■	□	□	■	■	■	■	□	□	■	□	■	□	□
Anxiety and Depression in Children and Youth. British Columbia, Canada[18]	■	■	□	□	□	■	■	□	□	■	□	□	■	□	■	■	■	□	□	■
Evidence-based guidelines for treating depressive disorders with antidepressants. British Ass. Psychopharmacol.[22]	■	■	□	□	■	■	□	□	□	■	□	□	■	□	■	□	□	■	□	□
Depression in children. BMJ Best Practice, UK[24]	■	■	□	□	□	□	■	□	□	□	■	□	■	□	□	■	□	■	□	□
Guidelines for Adolescent Depression in Primary Care: Part II. American Acad Pediatrics[30]	■	■	□	□	□	□	■	□	□	□	□	□	■	□	■	□	□	■	□	□
Depression in children and young people: identification and management (CG28). NICE, UK[33]	■	■	□	□	□	□	■	□	□	□	□	□	■	□	■	□	□	□	■	□
Vejledning om medikamentel behandling af børn og unge med psykiske lidelser. Danish Health Authority[38]	■	■	□	□	□	□	■	□	□	□	□	□	■	□	□	□	□	□	□	□
Landsdækkende klinisk retningslinje vedrørende udredning og behandling af depression hos børn og unge. BUP-DK[39]	■	■	□	□	□	□	■	□	□	□	□	□	□	□	□	□	□	□	■	□
Nationella riktlinjer för vård vid depression och ångestsyndrom. National Board of Health and Welfare, Sweden[21]	■	■	□	□	□	□	□	□	□	■	□	□	□	■	□	□	□	□	□	□
Håndbok for barn og unges psykiske helse. Center Child Adolescent Mental Health, Norway[36]	■	□	□	□	□	□	■	□	□	□	□	□	■	□	□	□	□	□	□	□
Treating depression in young people. Orygen, Australia[34]	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□
Identification of Common Mental Disorders and Management of Depression in Primary Care. New Zealand Ministry of Health[35]	■	■	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□	□

Continued

Table 2 Continued

Guideline	At least one adverse effect in category mentioned in guideline ■Yes □No																		
	Suicidality	Other psychiatr.	Blood, lymphat	Immune	Endocrine	Metabol, nutrit	Nervous system	Eye	Ear	Cardiac	Vascular	Respiratory	Gastrointestin	Hepatobiliary	Skin	Muskuloskeletal	Renal, urinary	Reprod, breasts	General	Investigations
Pediatriveilederen. Angst og depresjon Norwegian Pediatric Association[37]	■	■																		
Depression of adolescents. Duodecim, Finland[25]	■	■																		
Childhood depression. Duodecim, Finland[40]	■																			
Clinical Guidelines for the Management of Adults with Major Depressive Disorder. CANMAT, Canada[32]	■	■																		
Depression in children and adolescents. DynaMed Plus, USA[41]	■	■																		

BUP-DK, Børne-og Ungdomspsykiatrisk Selskab; CANMAT, Canadian network for mood and anxiety treatments; NICE,National Institute for Health and Care Excellence.

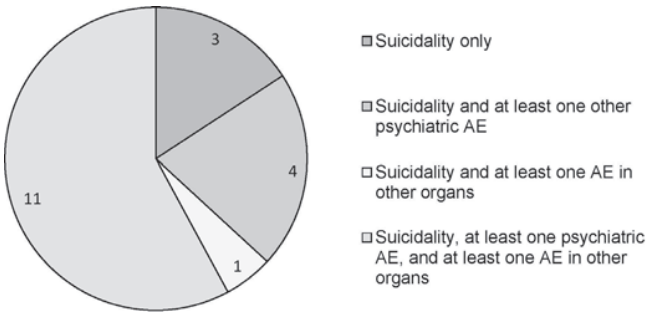

Figure 2 Information on adverse effects (AE) in clinical treatment guidelines (n=19).

have been considered in formulating the recommendations'. This should include descriptions of supporting data and reports of side effects, reports of the balance between benefits and side effects, and recommendations reflecting considerations of both benefits and side effects.[26] The tool does not, however, specify what should be considered a sufficient level of risk information. We have not identified any quality assessment studies of adverse effect information in clinical guidelines.

For systematic reviews, which sum up the literature and form a basis for guidelines, a framework has been proposed to include relevant harms data in a more comprehensive way.[27] However, research still show failings in the reporting of harms in systematic reviews.[28 29] Despite having procedures and checklists for developing guidelines, there is no guarantee that adverse effects identified in underlying clinical trials will be reflected in the finished guideline.

In conclusion, we found that many guidelines on treatment of depression in children and adolescents did not provide a thorough risk assessment with information on well-known and common adverse effects. There is currently no international standard regarding the extent of adverse effects information that should be included in guidelines. Development of such standards would give clinicians better accounts of risks and benefits as basis for therapy decisions.

Strengths and limitations

We conducted an extensive search for guidelines, however, some older guidelines were not found in electronic full text due to closure of the US National Guideline Clearinghouse. In many cases, we were able to identify updates that were available elsewhere. Inclusion of guidelines was not limited to English texts. Due to our geographical location and language issues, many included guidelines are of European or American origin. The cut-off for guideline inclusion can be debated on a geographical and local level. We have included guidelines on a national level, but did not aim to include guidelines from all countries. Local hospital or county guidelines were not included. There is, however, a possibility that we have not identified all relevant guideline collections, and that some guidelines may have been overlooked. We included

Table 3 Most cited references

Reference	No guidelines citing (n=19)
Hetrick SE, McKenzie JE, *et al*. Newer generation antidepressants for depressive disorders in children and adolescents. Cochrane Database 2012[7]	9
March J, Silva 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[8]	7
Bridge JA, Iyengar S, *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[41]	6
NICE 2015:CG28 Depression in children and young people: identification and management in primary, community and secondary care[33]	5
Barbui C, Esposito E, Cipriani A. Selective serotonin reuptake inhibitors and risk of suicide: a systematic review of observational studies. CMAJ 2009[42]	5

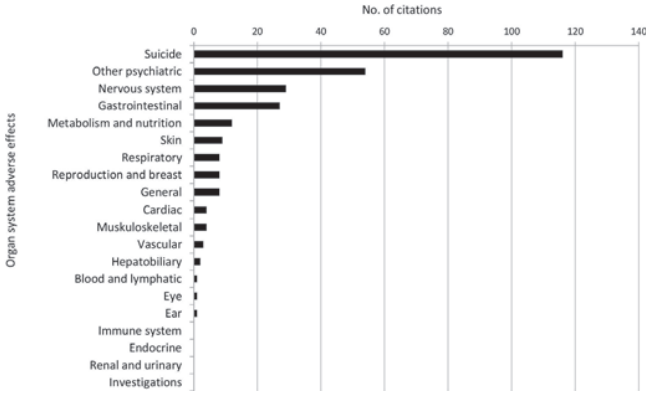

Figure 3 Number of citations in guidelines as basis for statements on adverse effects in organ systems.

three well-known, international decision support tools as they are probably being used by professionals to an increasing extent, however, other similar tools may be commercially available. Guidelines were examined for text extracts concerning adverse effects, and the corresponding reference was noted. We cannot exclude the possibility that some statements or references may have been overlooked. Finally, we did not aim to perform a formal assessment of the overall quality of the guidelines and cannot assess whether this is correlated with the extent of adverse effects information provided in the guideline. It is possible that guideline committee mandates and local expectations varied between guidelines.

Contributors The research question was developed by all authors. All authors discussed and defined the project. TW researched the literature and made the initial assessments. All authors discussed the guidelines included in the study, including interpretation and classification of information in the guidelines. TW drafted and finalised the manuscript as lead author. SN and MK commented on the draft and revised the manuscript at all stages.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Tables of references used in guidelines and their corresponding text statements are available on request.

ORCID iD

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

REFERENCES

1 Institute of Medicine (US). *Committee on standards for developing trustworthy clinical practice guidelines. clinical practice guidelines we can trust*. Washington DC: National Academies Press (US), 2011.
2 Shekelle P. Overview of clinical practice guidelines, 2019. Available: https://www.uptodate.com/contents/overview-of-clinical-practice-guidelines?search=guidelines&source=search_result&selectedTitle=1~150&usage_type=default&display_rank=1#H7 [Accessed 24 Jun 2019].
3 Kung J, Miller RR, Mackowiak PA. Failure of clinical practice guidelines to meet institute of medicine standards: two more decades of little, if any, progress. *Arch Intern Med* 2012;172:1628–33.
4 Bennett K, Courtney D, Duda S, *et al*. An appraisal of the trustworthiness of practice guidelines for depression and anxiety in children and youth. *Depress Anxiety* 2018;35:530–40.
5 Zafra-Tanaka JH, Goicochea-Lugo S, Villarreal-Zegarra D, *et al*. Characteristics and quality of clinical practice guidelines for depression in adults: a scoping review. *BMC Psychiatry* 2019;19:76.
6 Brouwers MC, Kho ME, Browman GP, *et al*. AGREE II: advancing guideline development, reporting and evaluation in health care. *CMAJ* 2010;182:E839–42.
7 Hetrick SE, McKenzie JE, Cox GR, *et al*. Newer generation antidepressants for depressive disorders in children and adolescents. *Cochrane Database Syst Rev* 2012;11:Cd004851.
8 March J, 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;292:807–20.
9 Walkup JT. Antidepressant efficacy for depression in children and adolescents: Industry- and NIMH-funded studies. *Am J Psychiatry* 2017;174:430–7.
10 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:e026089.
11 The Norwegian Electronic Health Library. Litteratursøk. Available: http://www.helsebiblioteket.no/fagprosedyrer/lage-og-oppdatere-fagprosedyrer/litteratursok [Accessed Jan 2019].
12 Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;339:b2535.

13 Datapharm. Summary of product characteristics; Prozac: Eli Lilly and Company Limited, 2018. Available: https://www.medicines.org.uk/emc/product/3768/smpc#PRODUCTINFO [Accessed 31 May 2018].
14 Guyatt G, Oxman AD, Akl EA, *et al*. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;64:383–94.
15 Andrews J, Guyatt G, Oxman AD, *et al*. GRADE guidelines: 14. going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol* 2013;66:719–25.
16 European Commission. A guideline on summary of product characteristics (SmPC), 2009. Available: https://ec.europa.eu/health/documents/eudralex/vol-2_en
17 Moreland CS, Bonin L. Pediatric unipolar depression and pharmacotherapy: choosing a medication, 2019. Available: https://www.uptodate.com/contents/pediatric-unipolar-depression-and-pharmacotherapy-choosing-a-medication [Accessed 14 Feb 2019].
18 Guidelines & Protocols Advisory Committee, Medical Services Commission, British Columbia. *Anxiety and depression in children and youth—diagnosis and treatment*. British Columbia: British Columbia Medical Association, Ministry of Health Services, 2010. https://www2.gov.bc.ca/gov/content/health/practitioner-professional-resources/bc-guidelines/anxiety-and-depression-in-youth
19 HAS - Haute Authorité de Santé. Manifestations dépressives l'adolescence: repérage, diagnostic et prise en charge en soins de premier recours, 2014. Available: https://www.has-sante.fr/portail/jcms/c_1782013/fr/manifestations-depressives-a-l-adolescence-reperage-diagnostic-et-prise-en-charge-en-soins-de-premier-recours [Accessed 17 Feb 2019].
20 Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften. Behandlung von Depressiven Störungen bei Kindern und Jugendlichen, 2013. Available: https://www.awmf.org/leitlinien/detail/ll/028-043.html [Accessed 31 Jan 2019].
21 Socialstyrelsen. Nationella riktlinjer för vård vid depression och ångestsyndrom. Egentlig depression, lindrig till medelsvår, långvarig, barn och ungdomar, 2017. Available: https://roi.socialstyrelsen.se/riktlinjer/nationella-riktlinjer-for-vard-vid-depression-och-angestsyndrom/2; https://www.socialstyrelsen.se/regler-och-riktlinjer/nationella-riktlinjer/om-nationella-riktlinjer/sa-tar-vi-fram-riktlinjerna/ [Accessed 31 Jan 2019].
22 Cleare A, Pariante CM, Young AH, *et al*. Evidence-based guidelines for treating depressive disorders with antidepressants: a revision of the 2008 British association for psychopharmacology guidelines. *J Psychopharmacol* 2015;29:459–525.
23 Emslie G, Kratochvil C, Vitiello B, *et al*. Treatment for adolescents with depression study (TADS): safety results. *J Am Acad Child Adolesc Psychiatry* 2006;45:1440–55.
24 BMJ Best Practice. Depression in children. Available: https://bestpractice.bmj.com/topics/en-gb/785?q=Depression [Accessed 30 Oct 2018].
25 Laukkanen E. Depression of adolescents: Duodecim medical publications Ltd, 2018. Available: https://www.ebm-guidelines.com/dtk/ebmg/home?id=ebm00703 [Accessed 30 Oct 2018].
26 The AGREE Next Steps Consortium. Appraisal of guidelines for research & evaluation II (AGREE II), 2017. Available: https://www.agreetrust.org/agree-ii/ [Accessed 15 Aug 2019].
27 Loke YK, Price D, Herxheimer A, *et al*. Systematic reviews of adverse effects: framework for a structured approach. *BMC Med Res Methodol* 2007;7:32.
28 Zorzela L, Golder S, Liu Y, *et al*. Quality of reporting in systematic reviews of adverse events: systematic review. *BMJ* 2014;348:f7668.
29 Saini P, Loke YK, Gamble C, *et al*. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ* 2014;349:g6501.
30 Cheung AH, Zuckerbrot RA, Jensen PS, *et al*. Guidelines for adolescent depression in primary care (GLAD-PC): part II. treatment and ongoing management. *Pediatrics* 2018;141:e20174082.
31 Kramer S, van Zuuren EJ, Ehrlich A. Depression in children and adolescents: DynaMed plus, EBSCO. Available: http://www.dynamed.com/topics/dmp~AN~T906140/Depression-in-children-and-adolescents#Overview-and-Recommendations [Accessed 3 Feb 2019].
32 MacQueen GM, Frey BN, Ismail Z, *et al*. Canadian network for mood and anxiety treatments (CANMAT) 2016 clinical guidelines for the management of adults with major depressive disorder: section 6. special populations: youth, women, and the elderly. *Can J Psychiatry* 2016;61:588–603.
33 NICE National Institute for Health and Care Excellence. Depression in children and young people: identification and management (CG28), 2017. Available: https://www.nice.org.uk/guidance/cg28 [Accessed 3 Feb 2019].
34 Orygen The National Centre of Excellence in Youth Mental Health. Treating depression in young people 2017. Available: https://www.orygen.org.au/Education-Training/Resources-Training/Resources/Free/Clinical-Practice/Treating-depression-in-yp [Accessed 5 May 2019].
35 New Zealand Guidelines Group. Identification of common mental disorders and management of depression in primary care. an evidence-based best practice guideline, 2008. Available: https://www.health.govt.nz/publication/identification-common-mental-disorders-and-management-depression-primary-care [Accessed 31 Jan 2019].
36 Regionsenter for barn og unges psykiske helse, Helseregion Øst og Sør. Mer om medikamentell behandling av barn og unge diagnostisert med depresjon. In: *Håndbok for barn og unges psykiske helse*, 2019.
37 Norsk barnelegeforening. Pediatriveilederen. Psykiske lidelser og psykososiale tilstander. 12.3 Angst og depresjon, 2018. Available: https://www.helsebiblioteket.no/pediatriveiledere?menuitemkeylev1=5962&menuitemkeylev2=5975&key=144656 [Accessed 1 Feb 2019].
38 Retsinformation. Vejledning om medikamentel behandling af børn og unge med psykiske lidelser: sundhedsstyrelsen, 2013. Available: https://www.retsinformation.dk/Forms/R0710.aspx?id=146409 [Accessed 31 Jan 2019].
39 Sørensen MJ, Kalvåg MB, Nøhr-Jensen P, *et al*. Landsdækkende klinisk retningslinje vedrørende udredning og behandling af depression hos børn og unge. Børne- og ungdomspsykiatrisk Selskab I Danmark, 2011. Available: http://www.bupnet.dk/kliniske-retningslinjer/depression-hos-b%C3%B8rn-og-unge [Accessed 31 Jan 2019].
40 Puura K. Childhood depression: Duodecim medical publications Ltd, 2016. Available: https://www.ebm-guidelines.com/dtk/ebmg/home?id=ebm00682 [Accessed 30 Oct 2018].
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;297:1683–96.
42 Barbui C, Esposito E, Cipriani A. Selective serotonin reuptake inhibitors and risk of suicide: a systematic review of observational studies. *CMAJ* 2009;180:291–7.