BMJ Open

Overstatements in abstract conclusions claiming effectiveness of interventions in psychiatry: a study protocol for a meta-epidemiological investigation

Aya M Suganuma, Kiyomi Shinohara, Hissei Imai, Nozomi Takeshima, Yu Hayasaka, Toshi A Furukawa

Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan

Correspondence to
Dr Kiyomi Shinohara;
shinohara.kiyomi.75r@st.kyoto-u.ac.jp

ABSTRACT

Introduction: Abstracts are the major and often the most important source of information for readers of the medical literature. However, there is mounting criticism that abstracts often exaggerate the positive findings and emphasise the beneficial effects of intervention beyond the actual findings mentioned in the corresponding full texts. In order to examine the magnitude of this problem, we will introduce a systematic approach to detect overstated abstracts and to quantify the extent of their prevalence in published randomised controlled trials (RCTs) in the field of psychiatry.

Methods and analysis: We will source RCTs published in 2014 from the Cochrane Register of Controlled Trials (CENTRAL) that claim effectiveness of any intervention for mental disorders. The abstract conclusions will be categorised into three types: superior (only stating significant superiority of intervention to control), limited (suggesting that intervention has limited superiority to control) and equal (claiming equal effectiveness of intervention as control). The full texts will also be classified as one of the following based on the primary outcome results: significant (all primary outcomes were statistically significant in favour of the intervention), mixed (primary outcomes included both significant and non-significant results) or all non-significant results. By comparing the abstract conclusion classification and that of the corresponding full text, we will assess whether each study exhibited overstatements in its abstract conclusion.

Ethics and dissemination: This trial requires no ethical approval. We will publish our findings in a peer-reviewed journal.

Trial registration number: UMIN000018668; Pre-results.

Strengths and limitations of this study

- Detects overstatements systematically without using subjective definition(s) and minimises investigators' arbitrary judgements.
- Provide practical suggestions for general readers on how to discern and detect overstated abstract conclusions when they first read abstracts.
- Our evaluation process follows the steps how articles are read generally, and thus our approach is easy to understand for general readers.
- Papers published in some journals are systematically excluded, for their reporting style does not have conclusions as a subheading under the abstract section.
- Owing to the nature of our framework, we will not be able to analyse other potential biases, such as selective reporting and publication bias.

time, clinicians often only read the abstracts.[1] Moreover, owing to the limited availability of free full texts,[2] abstracts are often the single source of information on the research for many readers. As a result, as stated in Consolidated Standards of Reporting Trials (CONSORT), 'readers often base their first assessment of a trial solely on the information in the abstract',[3] and thus abstracts potentially impact the interpretation of the results substantially.

As part of the recent effort to enhance the quality of reporting,[4] reformulation of the abstract structure[5] and the reporting guideline for abstracts have been proposed.[3] However, leading studies have reported that abstracts of research articles frequently contain bias. Boutron et al[6] studied 'spin' in the abstracts of 72 randomised controlled trials (RCTs) with non-significant primary outcomes. Their definition of spin included: (1) a focus on statistically significant results, (2) interpreting non-significant results as

INTRODUCTION

Abstracts of research articles are of significant importance for readers. They are often the primary source of information on the medical research. Partly owing to lack of

BMJ

showing treatment equivalence or comparable effectiveness, and (3) claiming or emphasising the beneficial effect of treatment despite non-significant results. They reported that 42 (58.3%) of the abstract conclusions exhibited some form of spin. Mathieu et al[7] analysed 105 trial reports in rheumatology about 'misleading' abstracts. They classified a trial as having misleading abstract if the study authors discussed only secondary outcomes or subgroup analyses in abstract, and stated conclusions contradictory to the actual study results. According to them, 24 (23%) of the studied reports had such misleading abstracts. For non-randomised trials, Lazarus et al[8] reported that 84% (107/128) of the studies contained at least one type of spin in their abstracts, with a high level of spin detected among 48% of the articles. Roest et al[9] reported that in the field of psychiatry, three trials had a spin out of 16 non-positive trials with second-generation antidepressants for anxiety disorders, and Amos[10] reviewed 38 published reports on early interventions for psychosis and found spin in 66% of them.

While the above studies shed some light on the important characteristics of distorted abstract reporting, they have their own limitations. First of all, their strict judgement may leave some room for discussion. For example, Boutron et al considered a trial as having 'spin' even though the first sentence of its abstract conclusion stated 'this trial did not demonstrate significant improvement in the primary or secondary end points in the active treatment group vs the group receiving placebo'.[11] Second, their analyses covered a selected portion of the publication. Boutron et al excluded the majority of published trials due to their having statistically significant results. Also, both left out trials with ambiguous primary outcome(s) or of non-inferiority design. Lazarus et al only included non-randomised trials, and Roest et al and Amos only assessed trials on specific mental disorders. Lastly, they focused primarily on the investigators' perspectives and provided little practical suggestions for general readers on how to discern and detect overstated abstract conclusions when they first read abstracts.

We propose a systematic approach that minimises investigators' judgemental arbitrariness and potential bias in identifying and analysing overstated abstracts, and present results in a simple, reader-friendly table format. On the basis of our findings, we aim to provide insights into how to appraise the reported findings in abstract conclusions from the readers' viewpoint.

Objectives
We aim to evaluate the prevalence and patterns of overstated abstract conclusions in trials claiming effectiveness of interventions in psychiatry by comparing the abstract conclusion and the results of the corresponding full text. In addition, we will examine their predictors. The primary outcome of this research is the quantified prevalence of the overstated abstract conclusions, and our secondary outcome is the difference in the extent of

prevalence between those abstract conclusions that only mention good results of the intervention arm and those that include limited results.

METHODS AND ANALYSIS
Study selection
We will use the Cochrane Central Register of Controlled Trials (CENTRAL) to identify all RCTs claiming effectiveness of interventions for mental disorders published in the English language in 2014. We will use the Medical Subject Headings (MeSH) term 'mental disorders', MeSH-term subheadings 'drug therapy' and 'therapy', and publication type 'randomised controlled trial' (see table 1).

The selection will cover any kinds of interventions, from common pharmacological intervention to non-drug therapy such as aromatherapy and exercise. We will include reports whose abstract conclusions claim superior or equal effectiveness of intervention to control. We will focus on the primary outcome (if stated) or all outcomes (if none is declared primary) in the abstract conclusions. We will exclude reports in which the primary outcomes were declared and it was explicitly stated that they were not significant, because such description would be accurate and leave no room for overstatement. For instance, an abstract stating 'while the primary outcome was not significant, secondary measures showed relevant benefits of intervention over control' will be excluded. We will exclude studies without a conclusion or discussion section in the abstract, trials with more than two arms, and unpublished trials. Secondary analysis studies, feasibility studies and cost-effectiveness studies will not be included either.

Since this is a qualitative, descriptive research, it is not within our scope to assess its statistical significance. Therefore, while sample size calculation is not essential, we shall target an outcome with a margin of error of 10%± and a CI of 95%. Assuming that the proportion of overstated studies is 50%, the desired sample size is calculated to be around 100.

Date extraction
The selected studies will be divided into two sets and two pairs of assessors will analyse each set. To check for

---

Table 1 Searching strategy, Cochrane Central Register of Controlled Trials

Search terms
#1 MeSH descriptor: [Mental Disorders] explode all trees and with qualifier(s): [Therapy—TH]
#2 MeSH descriptor: [Mental Disorders] explode all trees and with qualifier(s): [Drug therapy—DT]
#3 'randomised controlled trial':pt Publication Year from 2014 to 2014 (Word variations have been searched)
#4 (#1 or #2) and #3

eligibility, each pair of assessors will screen the title and abstract of each candidate study independently in a given set, respectively. Afterwards, each pair will read the full text to decide whether the study meets the inclusion criteria. Then the pairs will swap the screened set. By doing so, we can ascertain that the assessors have no previous knowledge of the given trial when they categorise the abstract and the full text. The agreement within the pair will be reported at each step of assessments. This is to ensure that our analysis resembles the way that general readers read and evaluate a study, as well as to maintain the independence of appraisal from any influence of the full text on the abstract and vice versa.

Following the screening and check, we will collect information from each study in the following three steps: categorisation of abstracts, classification of primary outcomes, and assessing inconsistency between the two. In each step, we will extract the relevant data from each study using the Excel spreadsheet specially made for this trial. The data include: the type of intervention, targeted mental illness, the region where the study was conducted, the number of randomised patients, study design, primary outcomes supposed in the abstract conclusion, the results of primary and secondary outcomes in the full text. Data extraction, categorisation of abstracts and the evaluation of the primary outcomes will be done independently by the two teams consisting of two or three assessors. Any disagreements in the team may be resolved by discussions with a third party. The citation will be recorded in each study PDF document and kept as a reference.

Categorisation of abstracts

First, we will categorise each abstract conclusion into one of the three types according to the level of effectiveness it claims without reading their full text. If a trial only stated significant effectiveness of the intervention in its abstract conclusion, it will be classified as 'superior'. On the other hand, a trial reporting significant and non-significant findings with respect to the intervention's efficacy will be considered as 'limited' (eg, 'treatment significantly improved quality of life than the control, but not depression', or 'treatment improved quality of life and anxiety, and had limited effect on depression'). Trials claiming equal effectiveness of intervention to control (eg, 'intervention was equally effective as control') will be categorised as 'equal'. Note that our judgement will be based solely on the abstract conclusion, regardless of the primary outcome results discussed in the results section of the abstract or full text.

Classification of results of primary outcome in full text

Second, we will assess the level of statistical evidence for their findings in primary outcome(s) in full text, and identify them into one of the three: significant (all primary outcomes were statistically significant), mixed (primary outcome included statistically significant and

non-significant results), and all non-significant. Note that no results of any secondary analysis or subgroup analysis results will be taken into consideration when determining the category.

We will define those trials as having an 'ambiguous primary outcome' if they did not explicitly state any outcome(s) as 'primary' or 'main analysis', except when they only measured a single outcome. In such cases, the single outcome will be considered the primary outcome. When a trial did not specify the primary time point, the following time points will be regarded as primary: end of treatment in trials studying the effectiveness of the acute treatment, the last follow-up in trials studying maintenance of the effectiveness, and the end of treatment and last follow-up in trials studying both, or in case the objective was unclear. We will also check if a trial was designed as a superiority trial or non-inferiority trial.

Furthermore, we will call a trial lacking vital information a 'sub-quality trial'. For example, a trial without statistical analysis of the main comparison, or having no assessment at the end of treatment will be classified as 'sub-quality trial'.

Assessing inconsistency between abstract and results in full text

By comparing the classification of the abstract conclusion and that of the full text for each study, we can determine whether its abstract conclusion was overstated and find the patterns of overstatements as shown in table 2. Naturally, a trial with an abstract conclusion categorised as 'superior' should have statistically significant results in all of their declared primary outcomes.

Table 2 Overstatement(s) in abstract conclusion	
Primary outcome (supposed) in abstract conclusion	**Primary outcome in full text**
Not overstated	
1. Superior	Significant: All significant primary outcomes
2. Limited	Mixed: significant and non-significant primary outcomes
3. Equal	Non-significant in non-inferiority design
Overstated	
4. Superior	4.1 Mixed or Non-significant: Including non-significant primary outcomes
	4.2 Ambiguous primary outcomes
5. Limited	5.1 Non-significant: all non-significant primary outcomes
	5.2 Ambiguous primary outcomes
6. Equal	6.1 Non-significant in superiority design
	6.2 Ambiguous primary outcomes

A study without declared primary outcomes in its full text will be included in the overstated group.

Similarly, a 'limited' abstract conclusion should correspond to mixed results. An 'Equal' abstract conclusion category must only be populated by non-inferiority trials showing effectiveness of treatment at least as much as the control or worse only by a prespecified amount. Note that non-inferiority trials are designed and conducted using a specific methodology under a specific design, such as the sample size calculation and equivalence margin prespecification. The full-text results should show that the upper limit of 95% CIs for the difference between intervention and control lies below that equivalence margin.[12]

For a study that does not fall into any of the above patterns, it is regarded as having an overstated abstract conclusion.

Outcomes

Primary outcome

The outcome will be summarised in table 2 as a proportion of the total studies used. Our primary outcome will be the proportion of overstated abstract conclusions, that is, the sum of 4.1–6.2.over all studies.

Secondary outcome

We will compare between 'superior' and 'limited', namely 4.1 vs 5.1 expressed as a percentage of total number of studies, respectively. It allows us to predict that given a type of abstract conclusion, which is more likely to be overemphasised.

Subgroup analyses

We will investigate the existence of any factors that potentially influence these inconsistencies, such as impact factors (IF) of the journal in which the trials are published, source(s) of funding and the sample size. For our purposes, the top 10 impact factor journals as ranked in *Journal Citation Report* (2014) in general medicine, psychiatry and psychology will be regarded as high IF journals.

Amendments

In the case of any amendments added to this protocol, each amendment will be listed with the date of change accompanied by a description of the change and the rationale.

ETHICS AND DISSEMINATION

This trial requires no ethical approval because we use secondary data from published trial reports. This protocol has been registered in the University hospital Medical Information Network (UMIN) Clinical Trials Registry. We will publish our findings in a peer-reviewed journal and also may present them at conferences.

DISCUSSION

A RCT is methodologically the most robust strategy to investigate the effectiveness of an intervention and can be a primary source for reliable evidence; results derived from RCTs have a strong influence on clinical judgement. Moreover, proper reporting in the abstract of articles is crucial to disseminating the results to the public. However, previous studies have demonstrated that the outcomes are occasionally positively misrepresented, by overstating the favourable outcomes and understating the unfavourable findings. Our study evaluates overstatements observed in abstracts of published articles in the field of psychiatry. Using a new approach specifically designed to minimise the influence of investigators' judgements and to maximise the generalisability to the general setting when articles are read, we aim to evaluate and quantify the magnitude of this problem. Since this approach is comparable to a systematic review, we attach the PRISMA-P checklist for reference.[13 14]

REFERENCES

1. Burke DT, Judelson AL, Schneider JC, et al. Reading habits of practicing physiatrists. *Am J Phys Med Rehabil* 2002;81:779–87.
2. Kurata K, Morioka T, Yokoi K, et al. Remarkable growth of open access in the biomedical field: analysis of PubMed articles from 2006 to 2010. *PLoS ONE* 2013;8:e60925.
3. Hopewell S, Clarke M, Moher D, et al. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. *PLoS Med* 2008;5:e20.
4. Schulz KF, Altman DG, Moher D, et al. CONSORT 2010 changes and testing blindness in RCTs. *Lancet* 2010;375:1144–6.
5. Bauchner H, Henry R, Golub RM. The restructuring of structured abstracts: Adding a table in the results section. *JAMA* 2013;309:491–2.
6. Boutron I, Dutton S, Ravaud P, et al. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. *JAMA* 2010;303:2058–64.
7. Mathieu S, Giraudeau B, Soubrier M, et al. Misleading abstract conclusions in randomized controlled trials in rheumatology: comparison of the abstract conclusions and the results section. *Joint Bone Spine* 2012;79:262–7.
8. Lazarus C, Haneef R, Ravaud P, et al. Classification and prevalence of spin in abstracts of non-randomized studies evaluating an intervention. *BMC Med Res Methodol* 2015;15:85.
9. Roest AM, de Jonge P, Williams CD, et al. 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;72:500–10.

10. Amos AJ. A review of spin and bias use in the early intervention in psychosis literature. *Prim Care Companion CNS Disord* 2014;16: PCC.13r01586.

11. Hoyles RK, Ellis RW, Wellsbury J, *et al.* A multicenter, prospective, randomized, double-blind, placebo-controlled trial of corticosteroids and intravenous cyclophosphamide followed by oral azathioprine for the treatment of pulmonary fibrosis in scleroderma. *Arthritis Rheum* 2006;54:3962–70.

12. Piaggio G, Elbourne DR, Altman DG, *et al.* Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA* 2006;295:1152–60.

13. Moher D, Shamseer L, Clarke M, *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;4:1.

14. Shamseer L, Moher D, Clarke M, *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;349: g7647.