BMJ Open

Influence of overstated abstract conclusions on clinicians: a web-based randomised controlled trial

Kiyomi Shinohara,[1] Takuya Aoki,[2] Ryuhei So,[1,3] Yasushi Tsujimoto,[2,4] Aya M Suganuma,[1] Morito Kise,[5] Toshi A Furukawa[1]

To cite: Shinohara K, Aoki T, So R, *et al*. Influence of overstated abstract conclusions on clinicians: a web-based randomised controlled trial. *BMJ Open* 2017;**7**:e018355. doi:10.1136/bmjopen-2017-018355

► 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-2017-018355).

Received 23 June 2017
Revised 12 October 2017
Accepted 13 October 2017

[1]Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan
[2]Department of Health care Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan
[3]Okayama Psychiatric Medical Center, Okayama, Okayama, Japan
[4]Department of Nephrology and Dialysis, Kyoritsu Hospital, Kawanishi, Hyogo, Japan
[5]Centre for Family Medicine Development, Japanese Health and Welfare Co-operative Federation, Tokyo, Japan

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

ABSTRACT

Objectives To investigate whether overstatements in abstract conclusions influence primary care physicians' evaluations when they read reports of randomised controlled trials (RCTs)

Design RCT setting: This study was a parallel-group randomised controlled survey, conducted online while masking the study hypothesis.

Participants Volunteers were recruited from members of the Japan Primary Care Association in January 2017. We sent email invitations to 7040 primary care physicians. Among the 787 individuals who accessed the website, 622 were eligible and automatically randomised into 'without overstatement' (n=307) and 'with overstatement' (n=315) groups.

Interventions We selected five abstracts from published RCTs with at least one non-significant primary outcome and overstatement in the abstract conclusion. To construct a version without overstatement, we rewrote the conclusion sections. The methods and results sections were standardised to provide the necessary information of primary outcome information when it was missing in the original abstract. Participants were randomly assigned to read an abstract either with or without overstatements and asked to evaluate the benefit of the intervention.

Outcome measures The primary outcome was the participants' evaluation of the benefit of the intervention discussed in the abstract, on a scale from 0 to 10. A secondary outcome was the validity of the conclusion.

Results There was no significant difference between the groups with respect to their evaluation of the benefit of the intervention (mean difference: 0.07, 95% CI −0.28 to 0.42, p=0.69). Participants in the 'without' group considered the study conclusion to be more valid than those in the 'with' group (mean difference: 0.97, 95% CI 0.59 to 1.36, P<0.001).

Conclusion The overstatements in abstract conclusions did not significantly influence the primary care physicians' evaluations of the intervention effect when necessary information about the primary outcomes was distinctly reported.

Trial registration number UMIN000025317; Pre-results.

Strengths and limitations of this study

- This is the first and only randomised controlled trial study that estimates the influence of overstatement in abstract conclusions.
- We evaluated the influence of overstatement among primary care physicians who were one of the major users of evidence.
- Although the number of participants was above our targeted sample size, a relatively low response rate limits the generalisability of our findings.
- As we focused on the influence of overstatement in abstract conclusions when necessary information about primary outcomes was reported in the methods and results sections, the effect of various other forms of inadequate reporting in abstracts should be further evaluated.

INTRODUCTION

Abstracts of reports of randomised controlled trials (RCTs) provide concise, educational and readily accessible information. They are particularly useful for primary care physicians because they deal with a wide range of patients and problems and need quick access to information regarding their practices. Sometimes abstracts are the only source of evidence they use.[1]

Abstract conclusions are the most crucial part of the whole abstract as they summarise the main results and provide interpretations.[2] A previous survey showed that primary care physicians paid the most attention to the conclusion.[3] The conclusion also guides primary physicians who are not confident in their skills in evidence-based medicine (EBM)[3 4] to interpret the results. Thus, a strong conclusion may alter the readers' interpretation of the whole study.

Unfortunately, the conclusion is the most frequently distorted section in abstracts.[5] Exaggerating the results of the trial, such as using spin[5] or overstatement,[6] is not uncommon. Examples of spin include omitting non-significant results of primary outcomes and focusing on significant secondary outcome or subgroup analysis.[5] Previous studies also found that 58% of RCTs with non-significant results,[5] and 70% of non-randomised

studies[7] had spin. Subsequent studies reported that spin, misleading information or overstatements were common in various subspecialties, such as rheumatology,[8] psychiatry,[9] wound care,[10] surgery[11 12] and oncology.[13–15]

This suggests that, as far as abstract conclusions are concerned, the quality of reporting is still poor despite the consolidated standards of reporting trials (CONSORT) guideline for abstracts.[2]

However, there has been limited evidence about the influence of such abstracts on the readers' interpretations in the real world. Only one RCT[16] investigated the extent of the impact of inappropriate reporting on readers' interpretations of the results. Boutron *et al*[16] randomised clinical researchers into two groups, and asked them to read an abstract with or without 'spin', which was defined by the authors as 'reporting the beneficial effect of the intervention as greater than shown by the results', to estimate how readers were influenced when they assessed the effectiveness of the intervention. The result showed that the participants who read abstract with spin were more likely to think that the intervention was beneficial for the patients than those who read the abstracts without spin.

Although their trial demonstrated that spin in the abstract had a small impact (effect size=0.24), it left several questions unanswered. First, the level of influence of spin in the abstract conclusion on the participants' interpretation remained unclear because the investigators added changes to all sections of the abstracts. In their study, they either erased or added all the results of secondary outcomes while changing the wording. In other words, they investigated the general influence of spin in an abstract by comparing it with its 'paragon' counterpart. Moreover, the target population was clinical researchers with publishing experience. Therefore, the influence of spin in the abstract conclusion on other types of evidence users remains unknown.

This study aims to determine the influence of the overstatements in abstract conclusions on general clinical practice by focusing on the primary care physicians who read reports of RCTs.

METHODS

Setting and design

This online study was a double-blind RCT conducted from January to February in 2017. The participants were masked to the study hypothesis, and the investigators (except RS who constructed the random sequence) were masked from the allocation. We recruited volunteers from members of the Japan Primary Care Association (JPCA) by sending email invitations. The intervention was conducted on a website specifically designed for this study. Participants were randomised into two groups and asked to read and evaluate 1 of the 10 abstracts (five pairs of two corresponding abstracts: one with and another without overstatement) of an RCT report. The trial was prospectively registered with the University Hospital Medical Information Network—Clinical Trial Registry (UMIN000025317) and now it is at "sutdy completed" stage. We had submitted the protocol including a statistical analysis plan to the JPCA before commencement but did not publish it to avoid the risk of participants reading it.

Participants and recruiting

The target population was recruited from the members of the JPCA. The JPCA was established in 2010 to the promote primary care specialty in Japan.[17] It is the largest organisation for primary care physicians in the country, and has been promoting evidence-based practice among its members. Currently, over 10 000 doctors working in various types of medical institutions[18] belong to the JPCA, and 5836 out of a total of 10 851 members are certified as specialists in primary care.

We sent email invitations to JPCA members who had more than 2 years of clinical experience with registered email addresses. (The details of the recruiting process will be reported in a separate paper.) We excluded clinicians with less than 2 years' experience because our target population was primary care physicians, and doctors usually choose their specialty after 2 years of clinical training in Japan. Interested individuals could access the DOCTOR study website via the link in the email. We added a code at the end of the link to ensure that participants accessed the website via the given link. As an incentive, an Amazon gift card worth 3000 yen (US$26.6) was given to 20 drawing winners.

The inclusion criteria for participants were as follows: JPCA member, medical doctor currently in clinical practice, more than 2 years of clinical practice experience and access to up-to-date clinical research knowledge. We asked how respondents learnt about the recent clinical trials, and individuals who did not respond with any information source were excluded. Screening questions were on the leading page on the website. We excluded those who work at research laboratories or educational institutions.

Randomisation and allocation concealment

When participants moved to the assessment page, they were randomly assigned an abstract either with or without overstatements with a 1:1 ratio. The block randomisation (10 for each block) was automatically performed using a computer-generated random sequence (created by RS). The allocation concealment was maintained through the automatic random allocation process.

Blinding

In the email invitations, participants were notified that this study aimed to investigate the impression of the abstracts and that they would be asked to score one randomly selected abstract numerically. (The English version of the invitation is included in the online supplementary appendix 1). Thus, they were masked to the study hypothesis. The researchers (KS, TA, YT and AS), excluding the website manager (RS), were blinded until the blind interpretations of the results

were completed and signed off.[19] RS did not join the result analysis.

Selecting abstracts with overstatements

We selected five abstracts[20–24] (the text of the five abstracts is included in the online supplementary appendix 2) from the pre-existing database of published reports in psychiatry RCTs dated between 2011 and 2014, which was collected from our previous study.[6 25] To avoid any bias arising from the participants' subspecialty expertise (such as internal medicine or surgery), we chose reports from psychiatry.

The abstracts were selected based on the following criteria: (1) superiority RCT with two arms, (2) claiming effectiveness of an intervention in the abstract conclusion despite some or all primary outcomes not being significant, (3) targeting a common mental illness primary care physicians are likely to encounter in clinical settings and (4) having a journal impact equal to or higher than two.

An overstatement was defined as 'inconsistency between the results of primary outcomes in full-text and those deduced from the abstract conclusion'.[6] While spin is any technique embellishing the results across whole reports, an overstatement specifically refers to exaggerations in the abstract conclusion.

In the five sample abstracts selected, two only mentioned the superiority of the intervention to the control in the conclusions. In fact, one had non-significant results and the other had mixed results (significant and non-significant) in their primary outcomes. The remaining three had conclusions that emphasised the partial superiority of the intervention with respect to the control. They stated that the treatment was partially effective even though all the primary outcomes were non-significant. Together, they include different levels of overstatement from completely misleading to less informative (not mentioning non-significant primary outcome) conclusions. They were checked independently by two or more investigators (KS, AS and RS)

Constructing abstracts with and without overstatements

We constructed abstracts in line with the following prespecified guidelines. First, we rewrote the conclusion to make a conclusion without overstatement following these rules. (1) When all primary outcomes were non-significant, we rewrote it the conclusion as 'Intervention A was not more effective than control B in terms of …'. (2) When one primary outcome (PO1) was significant but the other (PO2) was non-significant, we rewrote it the conclusion as 'Intervention A was more effective than control B in terms of PO1, but not more effective in PO2' according to the order in the original abstract. We also removed the results of secondary outcomes and subgroup analysis from the conclusions. (See an example in box, and all the abstract conclusions are in table 1.)

Second, we standardised the methods and results sections. We explicitly stated the primary outcomes and results (for example, OR, risk ratio, CI, P value) from the text if they were not stated in the original abstract. Therefore, all abstracts had the information necessary for participants to understand the results of the primary outcomes from the method and results sections. This modification was necessary to keep the conclusion consistent with the other sections of the abstract. Without this step, the conclusion of an abstract without overstatement would be inconsistent with other sections of the same abstract because the conclusion of an abstract without overstatement would now be reconstructed based on the actual primary outcomes that were not mentioned in the

Box An example of the abstracts (italics where extra text added, bold where changed in the 'without overstatement' group)

Title
Intervention A for menopausal symptoms: a randomised controlled trial

Objective
This study aims to determine the efficacy of intervention A for alleviating vasomotor and other menopausal symptoms.

Methods
Late perimenopausal and postmenopausal sedentary women with frequent vasomotor symptoms (VMS) *such as hot flush, sweating and poor circulation* participated in a randomised controlled trial conducted in three sites: 106 women randomised to exercise and 142 women randomised to usual activity. VMS frequency and bother were recorded on daily diaries at baseline and on weeks 6 and 12. Intent-to-treat analyses compared between-group differences in changes in VMS frequency and bother, sleep symptoms (Insomnia Severity Index and Pittsburgh Sleep Quality Index), and mood (Patient Health Questionnaire-8 and Generalised Anxiety Disorder-7 Questionnaire). *Primary outcomes were VMS frequency and bother mean frequency or bother of VMS at 6 and 12 weeks.*

Results
At the end of week 12, changes in VMS frequency in intervention A group (mean change −2.4 VMS/d, 95% CI −3.0 to −1.7) and VMS bother (mean change on a four-point scale −0.5, 95% CI −0.6 to −0.4) were not significantly different from those in control B group (−2.6 VMS/d, 95% CI −3.2 to −2.0, P=0.43, −0.5 points, 95% CI −0.6 to −0.4, P=0.75). The exercise group reported greater improvement in insomnia symptoms (P=0.03), subjective sleep quality (P=0.01) and depressive symptoms (P=0.04), but differences were small and not statistically significant when p values were adjusted for multiple comparisons. Results were similar when considering treatment-adherent women only.

Conclusions
These findings provide strong evidence that 12 weeks of intervention A do not alleviate VMS but may result in small improvements in sleep quality, insomnia and depression in midlife sedentary women.
Control B is the standard treatment for menopausal symptoms.

'Without' overstatement version conclusions
Intervention A was not more effective than control B in terms of frequent VMS such as hot flush, sweating in postmenopausal women.
Control B is the standard treatment for menopausal symptoms.

Table 1 Five sample abstracts and their two versions of conclusion

Study year	Symptoms or illness	Conclusion in the original abstract	Conclusion without overstatement (rewritten by investigators)
Sternfeld *et al*, 2014[20]	Menopausal symptoms	These findings provide strong evidence that 12 weeks of intervention A do not alleviate vasomotor symptoms (VMS) but may result in small improvements in sleep quality, insomnia and depression in midlife sedentary women.	Intervention A was not more effective than control B in terms of frequent VMS such as hot flush, sweating in postmeopausal women.
Freund-Levi *et al*, 2014[22]	Patients with neuropsychiatric symptoms of probable dementia (NPSD)	These results support that intervention A, with its benign safety profile, can be used as first-line treatment of NPSD, unless symptoms of irritation and agitation are prominent, where control B is more efficient.	Intervention A was not more effective than control B in terms of neuropsychiatric symptoms in patients with dementia.
Lam *et al*, 2013[24]	Major depressive disorder	Intervention A with escitalopram significantly improved some self-reported work functioning outcomes, but not symptom-based outcomes, compared with escitalopram and control B.	Intervention A with escitalopram was not more effective than control B with escitalopram in terms of depressive symptoms in patients with major depression.
Oosterbaan *et al*, 2013[23]	Common mental disorders	Intervention A resulted in an earlier treatment response compared with control B	Intervention A was not more effective than control B in terms of treatment response or remission in patients with common mental illness.
Samus *et al*, 2014[21]	Elders with memory disorders	Intervention A delivered by non-clinical community workers trained and overseen by geriatric clinicians led to delays in transition from home, reduced unmet needs and improved self-reported QOL.	Intervention A was more effective than control B in terms of delay in transition from home, but not more effective in terms of reducing unmet needs in elders with memory disorders.

QOL, quality of life.

original abstract. Additionally, this standardisation made it possible to estimate the influence of overstatement in the conclusion when the methods and results reported essential information.

Third, we changed the names of the intervention and control treatments to anonymous 'intervention A' and 'control B' to minimise bias. We added a few words for explanation when there was a medical term that seemed unfamiliar to primary care physicians (eg, VMS): hot flush, sweating and poor circulation). Finally, we translated the texts into Japanese. Except for the conclusion, abstracts 'with' or 'without' overstatement were identical.

We made established two pairs of investigators, and each pair did modification and translation of a half of the abstracts ('with' and 'without' overstatement). Then, the other pair then checked whether they were following the guidelines. Another researcher (SK), who was not involved in this study, checked the translation. Any disagreement was resolved by discussion among investigators.

Outcomes

Our primary outcome was the numerical evaluation, which was scored by participants, of the effectiveness of the intervention discussed in the given abstract: 'How beneficial do you think intervention A is for the patients, on a scale from 0 to 10, 0 being not at all beneficial and 10 being conceivably most beneficial?' We also asked the following questions (scored 0 to 10 with 0 being not at all and 10 being very likely).

- How valid is this conclusion in your opinion on a scale from 0 to 10?
- How much do you want to read the full text of this study on a scale from to 10?
- When you answered the above questions, which part of the abstract did you refer to the most? (background/ methods/ results/conclusion)

Sample size

We referred to the effect size of 0.25 obtained in the previous study.[16] They estimated the effect of spin by comparing the influence of the abstracts 'with' and 'without' spin on clinical researchers. Although our target population differed from the previous study, considering that the effect of 0.2 represented a small effect,[26] we aimed for a sample size of 253 per group, and 506 in total to detect a between group effect size of 0.25 with a power of 90% and a two-sided alpha risk at 5%. Given that we had prepared five pairs of abstracts with or without overstatement, we intended to enrol 100 or more participants for each pair.

Statistical analysis

For the main analysis, we used a linear mixed effects model with a fixed factor (for the intervention) and a

random intercept for the abstract to account for the clustering effects of the abstracts (each abstract had two versions: with or without overstatements). The model accounted for the correlation within abstracts by using an unstructured covariance matrix. We excluded the following subjects from our analysis before proceeding to the study analyses and therefore without knowledge of any outcomes: (1) those who were erroneously allocated by the web system although they did not satisfy the eligibility criteria and (2) those who were eligible and were randomised but did not complete the questionnaire or spent less than 30 s on the questionnaire. TA and KS analysed the data using SPSS statistics 24 without knowing the allocation. To evaluate the influence of possible associated factors[3 27] on the interpretation, we conducted the following prespecified subgroup analyses using the participants: (1) working clinics, (2) getting information only from a pharmacological company, (3) with certification of a primary care physician and (4) having an experience of being the principal researcher (this is post hoc).

Blinded data interpretation

Blinded interpretation of study results was the approach recommended by Järvinen *et al*[19] to reduce interpretation bias. Following their suggestion, we interpreted the results blindly before breaking the randomisation code. Thus, we prepared two interpretations of the results based on two scenarios: (1) assuming group A was with overstatement and group B was without overstatement and (2) assuming group A was without overstatement and group B was with overstatement. After agreeing that there would be no further change, we broke the randomisation code and chose the correct interpretations.

Ethical considerations

This study was conducted in accordance with the Declaration of Helsinki. We obtained an online consent for participation from each participant.

RESULTS

We sent email invitations to 7040 JPCA members (figure 1). After sending one reminder, we reached the targeted sample size of 510. Among the 787 individuals who accessed the website, 622 were eligible and randomly assigned to without overstatement (n=307) and with overstatement (n=315) groups. A total of 281 doctors in the 'without' group and 286 in the 'with' group were included for the analysis. The number of participants allocated to each pair 'with' or 'without' overstatement was as follows: abstract pattern 1 (n=116), 2 (n=109), 3 (n=115), 4 (n=113) and 5 (n=114). Online supplementary appendix 3 provides further breakdown per abstract.

Fifty-five individuals were excluded because they either spent less than 30 s on the webpage (n=14) or did not complete the survey (n=41). Most participants read and rated the abstract within 4 min (medium time: 162 s, IQR: 114–236 s).

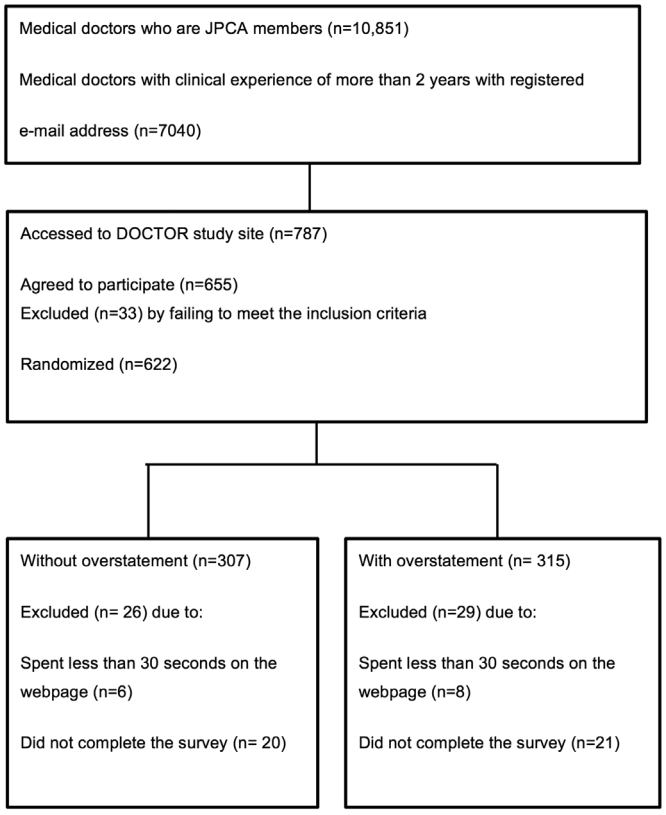


Figure 1 Flow diagram of participants. JPCA, Japan Primary Care Association.

Table 2 shows the participant characteristics; 76.5% were certified as primary care physicians. We classified their subspecialties according to their certifications. The most common background was internal medicine. More than 60% of the participants had attended a course on EBM. About 40% of the physicians said the first section they read was the conclusion; only 11% of them read the results section first. There was no substantial difference between the two groups.

Primary outcomes

There was no statistically significant difference between the groups with regard to the interpretation of the benefits of the intervention discussed in the given abstracts (mean difference: 0.07, 95% CI −0.28 to 0.42, P=0.69, effect size calculated by Cohen's d: 0.031) (table 3).

Secondary outcomes and subgroup analyses

However, there was a significant difference between the groups in their perception of the validity of the conclusion (mean difference: 0.97, 95% CI 0.59 to 1.36, P<0.001) (figure 2). Those in the without overstatement group considered the abstract to be more valid than those in the with overstatement group (effect size calculated by Cohen's d was 0.41). No significant difference was found when asked if they wanted

Table 2 Characteristics of participants

Characteristics of doctors	Without OS , n=281 (%)	With OS, n=286 (%)	Total, n=567(%)
Male	241 (85.8)	243 (85.0)	484 (85.4)
Years of practice	Median 15.0 IQR 11–24	Median 16.0 IQR 11–24	Median 16.0 IQR 11–24
Workplace			
Hospitals (public and private)	131 (46.6)	165 (57.7)	296 (52.2)
Clinics	97 (34.5)	80 (28.0)	177 (31.2)
University hospitals	46 (16.4)	40 (14.0)	86 (15.2)
Nursing homes	2 (0.7)	0 (0)	2 (0.4)
Others	5 (1.8)	1 (0.3)	6 (1.1)
Certification/degree*			
Primary care physician	216 (76.9)	218 (76.2)	434 (76.5)
PhD	88 (31.3)	93 (32.5)	181 (31.9)
Other certification	167 (59.4)	180 (62.9)	347 (61.2)
Clinical background*			
Internal medicine	123 (43.8)	135 (47.2)	258 (45.5)
Surgery	26 (9.3)	26 (9.1)	52 (9.2)
Emergency medicine	15 (5.3)	14 (4.9)	29 (5.1)
Paediatrics	6 (2.1)	5 (1.7)	11 (1.9)
Others	21 (7.5)	22 (7.7)	43 (7.6)
Source of information*			
Brochures/lectures sponsored by pharmaceutical companies	153 (54.4)	165 (57.7)	318 (56.1)
Journal club	81 (28.8)	83 (29.0)	164 (28.9)
Searching evidence/medical journals	187 (66.5)	193 (67.5)	380 (67.0)
Secondary information	191 (68.0)	199 (69.6)	390 (68.8)
Others	21 (7.5)	9 (3.1)	30 (5.3)
Ever attended an EBM workshop	181 (64.4)	186 (65.0)	367 (64.7)
Experience of PI	94 (33.5)	106 (37.1)	200 (35.3)
The first section to read when studying abstracts			
Background	108 (38.4)	105 (36.7)	213 (37.6)
Methods	24 (8.5)	25 (8.7)	49 (8.6)
Results	35 (12.5)	30 (10.5)	65 (11.5)
Conclusion	114 (40.6)	126 (44.1)	240 (42.3)
The number of abstract read in the last month			
0	22 (7.8)	26 (9.1)	48 (8.5)
1	23 (8.2)	31 (10.8)	54 (9.5)
2–4	107 (38.1)	117 (40.9)	224 (39.5)
Five or more	129 (45.9)	112 (39.2)	241 (42.5)

Clinical background data were available with participants who have subspecialty certifications.
*Multiple answers allowed.
EBM, evidence-based medicine; OS, overstatement; PI, principle investigator.

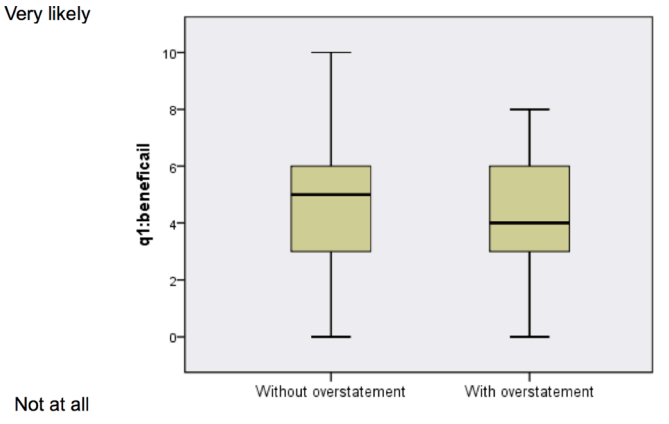
to read the full text. In both groups, the majority of the doctors referred to the results section to make an assessment.

We conducted subgroup analyses, but no significant differences were found with regard to the interpretation of the benefits of the intervention based on the workplace (clinic, n=177, mean difference: 0.04, 95% CI −0.67 to 0.74, P=0.91), general source of information (only pharmacological company, n=43, mean difference: 0.06, 95% CI −1.36 to 1.48, P=0.93), being a certified primary care physician (n=434, mean difference: −0.01, 95% CI −0.41 to 0.39, P=0.96) or having

Table 3 Impression of the abstract

Questions (answers given in a scale of 0–10 with 0 least)	Without OS, n=281 (SD)	With OS, n=286 (SD)	Mean difference, n=567 (95% CI)	Effect size (Cohen's d) n=567 (95% CI)
How beneficial do you think intervention A is for the patients?	4.18 (2.29)	4.10 (2.17)	0.07 (–0.28 to 0.42)	0.031 (–0.13 to 0.20)
How valid is this conclusion in your opinion?	4.84 (2.40)	3.88 (2.36)	0.97*** (0.59 to 1.36)	0.41 (0.24 to 0.57)
How much do you want to read the full text of this study?	3.52 (2.55)	3.41 (2.62)	0.10 (–0.32 to 0.53)	0.039 (–0.13 to 0.20)
When you answered the above questions, which part of the abstract did you refer to the most?				
Background	2 (0.7)	5 (1.7)		
Methods	58 (20.6)	59 (20.6)		
Results	181 (64.4)	174 (60.8)		
Conclusion	40 (14.2)	48 (16.8)		

***P<0.001.



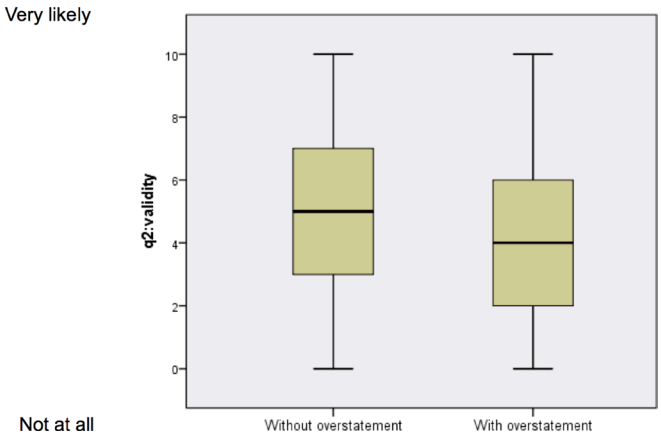


Figure 2 Evaluation of the beneficial effect and validity of the intervention discussed in the abstract. The answers to q1 "How beneficial do you think the intervention A is to patients?", and q2 "How valid is this conclusion in your opinion?" given in a scale of 0 (not at all) to 10 (very likely). Boxes showed the median score (horizontal rule) with 25th and 75th percentile.

no experience as a principal researcher (n=367, mean difference: −0.10, 95% CI −0.53 to 0.34, P=0.66).

DISCUSSION

We showed that primary care physicians were not influenced by overstatement in the conclusion section if the abstract contained necessary information on the primary outcomes. The 95% CI of the estimated effect (effect size by Cohen's d: 0.031, 95% CI −0.13 to 0.20) rules out the existence of even a small effect. In the baseline questionnaire, 42% of participants answered that they read the conclusion section first when reading abstracts. However, more than 60% of them referred to the results section for their interpretation of the given abstract. They tended to judge the overstated conclusion as less valid than those without overstatement. These results suggested that primary care physicians who belonged to the JPCA with up-to-date knowledge of clinical trials were not misled by overstatements in abstract conclusions if the method and results section reported sufficient information. Our subgroup analysis showed that factors such as the workplace, types of information resources or experience of being a principal investigator would make little difference. These results suggest that the participants had good critical appraisal skills of research reports, which helped them to recognise the inconsistency between the result and the conclusion.

Our results differed in some respects from the previous study. Boutron and colleagues' study[16] showed that the interpretation of abstracts was affected by spin. The 'abstracts with spin' group considered the intervention more beneficial than the without spin group, and the 'with spin' group was more interested in reading the full text. This was contrary to our main findings. On the other hand, the abstracts with spin group interpreted the abstract as less methodologically rigorous than the without spin group. This was consistent with our results.

However, we must consider some differences in design between Bourton *et al*'s study and this study. First, the level of spin was much higher in their study than in this study. Boutron *et al* aimed to investigate the impact of spin in the abstract generally, so they removed all spin from the abstract and compared this 'perfect' abstract with the original one. On the other hand, in our study, the difference between 'with' and 'without' groups was limited in the conclusion section because our aim was to estimate the influence of overstatement in the conclusion section. Thus, we added the information on the primary outcomes in the methods and results sections of both groups. Second, the baseline characteristics of the participants differed. While all the participants in the study of Boutron *et al* were experienced clinical researchers, we chose primary care physicians as our target. Although the participants in this study had little experience in clinical research, they were regular users of medical literature (90% of participants had read more than one abstract in the previous month). Most participants were eager to learn EBM and had some knowledge of critical appraisal. In addition, 60% referred to the results section when making clinical interpretation. Therefore, their study and ours are more complementary than contradictory.

Limitations and strengths

Our strength is that this is the first and only RCT study that estimates the influence of overstatement in abstract conclusions. Authors of scientific articles like to use promising, positive words[28 29]; nonetheless, we demonstrated that overstated conclusions did not affect the readers' interpretations of the results if sufficient information was provided in other sections. Second, we evaluated the influence of overstatement in primary care physicians, who are among the major users of evidence. They encounter clinical queries in daily clinical practice and use evidence to make the best decisions for their patients. Therefore, it is important to clarify whether primary care physicians are susceptible to overstatement in abstract conclusions. The results showed that primary care physicians with up-to-date knowledge of trial/research information were not misled by an overstated conclusion.

There are some limitations. While the number of participants was above our targeted sample size, it may not have completely represented the JPCA members. The relatively low response rate of 11.1% (787/7040) limits the generalisability of our findings. Two things should be noted. First, we chose the JPCA as our recruiting pool because that the members were considered representative of active users of scientific evidence in their primary care practice. The JPCA is the only organisation that certifies clinicians as primary care physicians, and they regularly conduct workshop on EBM. However, those who responded to our invitation were potentially avid readers of scientific reports, which is the reason they volunteered for this assessment, and, therefore, they may have better critical appraisal skills for abstracts than other JPCA members. Actually, most of participants answered that they read abstracts regularly. This suggests that they were not representative of all primary care physicians in Japan. Furthermore, the effect of overstatements in the abstracts that did not report the necessary information of primary outcomes or other various forms of inadequate reporting was not measured. In our study, we added essential information on primary outcomes in the methods and results sections as recommended by a CONSORT statement.[2] More than 60% of the participants stated that they mainly refer to the results to evaluate the abstract. In contrast, only around 15% based their assessment on the conclusion. This means that adequate reporting of the results is necessary for interpretation of the abstract. Finally, we should not overgeneralise the association between the type or level of overstatement and its impact on interpretation. We chose five abstracts at different levels of overstatement as a sample, but the selection did not cover all levels of spin or all types of spin. Neither did we have sufficient sample size to explore such relationships. The influence of biased reporting on clinical decisions should be further researched.

In conclusion, our findings suggested that sensible and well-read clinicians are capable of discerning the inconsistency between results and conclusion and of making a sound judgement on the validity of misleading conclusions when primary outcomes are appropriately reported in the methods and results sections. However, this does not mean that overstatements can be overlooked. The conclusion sections of abstracts should be written solely based on the primary outcome results. The impact of inappropriate writing style in clinical settings should be further researched.

Acknowledgements We thank those who participated in this study, A Igaki for organising and sending invitation e-mails, and S Kishimoto, for double checking translation of abstracts. We would like to thank Editage (www.editage.jp) for English language editing.

Contributors All authors of the paper have contributed to the conception or design of the work, development of the intervention and the acquisition or interpretation of data. KS, TA, RS, YT and AMS were involved in drafting the work. MK and TAF revised it critically for important intellectual content. RS designed and developed the study website. TA and KS analysed the data. All authors gave the final approval of the manuscript before submission.

Funding This work was supported by Japan Primary Care Association (grant number 28-01-001) to KS.

Competing interests TAF has received lecture fees from Eli Lilly, Janssen, Meiji, Mitsubishi-Tanabe, MSD and Pfizer and consultancy fees from Takeda Science Foundation. He has received research support from Mochida and Mitsubishi-Tanabe.

Ethics approval The Ethics Committee of Kyoto University Graduate School of Medicine.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data available.

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. 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.
3. Hisham R, Ng CJ, Liew SM, *et al*. Why is there variation in the practice of evidence-based medicine in primary care? A qualitative study. *BMJ Open* 2016;6:e010565.
4. Zwolsman S, te Pas E, Hooft L, *et al*. Barriers to GPs' use of evidence-based medicine: a systematic review. *Br J Gen Pract* 2012;62:511–21.
5. 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.
6. Suganuma AM, Shinohara K, Imai H, *et al*. Overstatements in abstract conclusions claiming effectiveness of interventions in psychiatry: a study protocol for a meta-epidemiological investigation. *BMJ Open* 2016;6:e009832.
7. 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.
8. 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.
9. Shinohara K, Suganuma AM, Imai H, *et al*. Overstatements in abstract conclusions claiming effectiveness of interventions in psychiatry: a meta-epidemiological investigation. *PLoS One* 2017;12:e0184786.
10. Lockyer S, Hodgson R, Dumville JC, *et al*. "Spin" in wound care research: the reporting and interpretation of randomized controlled trials with statistically non-significant primary outcome results or unspecified primary outcomes. *Trials* 2013;14:371.
11. Arunachalam L, Hunter IA, Killeen S. Reporting of randomized controlled trials with statistically nonsignificant primary outcomes published in high-impact surgical journals. *Ann Surg* 2017;265:1141–5.
12. Fleming PS. Evidence of spin in clinical trials in the surgical literature. *Ann Transl Med* 2016;4:385.
13. Vera-Badillo FE, Shapiro R, Ocana A, *et al*. Bias in reporting of end points of efficacy and toxicity in randomized, clinical trials for women with breast cancer. *Ann Oncol* 2013;24:1238–44.
14. Vera-Badillo FE, Napoleone M, Krzyzanowska MK, *et al*. Bias in reporting of randomised clinical trials in oncology. *Eur J Cancer* 2016;61:29–35.
15. Vera-Badillo FE, Tannock IF. Spin and bias: the tip of the iceberg. *Ann Oncol* 2013;24:1949–50.
16. Boutron I, Altman DG, Hopewell S, *et al*. Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: the SPIIN randomized controlled trial. *J Clin Oncol* 2014;32:4120–6.
17. Japan Primary Care Association.https://www.primary-care.or.jp/jpca_eng/index.html (accessed 6 Nov 2017).
18. Aoki T, Inoue M, Nakayama T. Development and validation of the Japanese version of primary care assessment tool. *Fam Pract* 2016;33:112–7.
19. Järvinen TL, Sihvonen R, Bhandari M, *et al*. Blinded interpretation of study results can feasibly and effectively diminish interpretation bias. *J Clin Epidemiol* 2014;67:769–72.
20. Sternfeld B, Guthrie KA, Ensrud KE, *et al*. Efficacy of exercise for menopausal symptoms: a randomized controlled trial. *Menopause* 2014;21:330–8.
21. Samus QM, Johnston D, Black BS, *et al*. A multidimensional home-based care coordination intervention for elders with memory disorders: the maximizing independence at home (MIND) pilot randomized trial. *Am J Geriatr Psychiatry* 2014;22:398–414.
22. Freund-Levi Y, Jedenius E, Tysen-Bäckström AC, *et al*. Galantamine versus risperidone treatment of neuropsychiatric symptoms in patients with probable dementia: an open randomized trial. *Am J Geriatr Psychiatry* 2014;22:341–8.
23. Oosterbaan DB, Verbraak MJ, Terluin B, *et al*. Collaborative stepped care v. care as usual for common mental disorders: 8-month, cluster randomised controlled trial. *Br J Psychiatry* 2013;203:132–9.
24. Lam RW, Parikh SV, Ramasubbu R, *et al*. Effects of combined pharmacotherapy and psychotherapy for improving work functioning in major depressive disorder. *Br J Psychiatry* 2013;203:358–65.
25. Shinohara K, Tajika A, Imai H, *et al*. Protocol registration and selective outcome reporting in recent psychiatry trials: new antidepressants and cognitive behavioural therapies. *Acta Psychiatr Scand* 2015;132:489–98.
26. Cohen J. *Statistical power analysis in the behavioral sciences*. Hillsdale, NJ: Erlbaum, 1988.
27. DeJong C, Aguilar T, Tseng CW, *et al*. Pharmaceutical industry-sponsored meals and physician prescribing patterns for medicare beneficiaries. *JAMA Intern Med* 2016;176:1114–10.
28. Vinkers CH, Tijdink JK, Otte WM. Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: retrospective analysis. *BMJ* 2015;351:h6467.
29. de Vries YA, Roest AM, Franzen M, *et al*. Citation bias and selective focus on positive findings in the literature on the serotonin transporter gene (5-HTTLPR), life stress and depression. *Psychol Med* 2016;46:2971–9.