BMJ Open

The influence of overstated abstract conclusions on clinicians: A randomized controlled trial: DOCTOR study (Do Overstated Conclusions Trick Our Readers?)

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-018355
Article Type:	Research
Date Submitted by the Author:	23-Jun-2017
Complete List of Authors:	Shinohara, Kiyomi; Kyoto University Graduate School of Medicine/Schoo of Public Health, Health Promotion and Human Behaviour Aoki, Takuya; Kyoto University Graduate School of Medicine / School of Public Health, Department of Healthcare Epidemiology So, Ryuhei ; Kyoto University Graduate School of Medicine / School of Public Health, Department of Health Promotion and Human Behavior Tsujimoto, Yasushi ; Kyoto University Graduate School of Medicine/Schoo of Public Health, Department of Healthcare Epidemiology Suganuma, Aya M; Kyoto University Graduate School of Medicine/Schoo of Public Health, Health Promotion and Human Behaviour Kise, Morito; Centre for Family Medicine Development, Japanese Health and Welfare Co-operative Federation Furukawa, Toshi; Kyoto University, Graduate School of Medicine and School of Public Health
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Evidence based practice
Keywords:	randomised controlled trials, general practice, overstatements, primary care physicians, reporting bias, clinical trial

Title:

The influence of overstated abstract conclusions on clinicians: A randomized controlled trial: DOCTOR study (Do Overstated Conclusions Trick Our Readers?)

Authors:

Kiyomi Shinohara[a],
Takuya Aoki [b],
Ryuhei So[a],
Yasushi Tsujimoto[b],
Aya M Suganuma [a],
Morito Kise[c],
Toshi A. Furukawa[a]

Affiliation of all authors:

a: Department of Health Promotion and Human Behavior Kyoto University Graduate School of Medicine / School of Public Health
Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501 Japan,
b: Department of Healthcare Epidemiology Kyoto University Graduate School of Medicine / School of Public Health
Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501 Japan
c: Centre for Family Medicine Development, Japanese Health and Welfare Co-operative Federation, Tokyo, Japan
Sanken Building 4th Floor
3-25-1, Hyakunin-cho, Shinjuku-ku
Tokyo Japan

Corresponding author:

Kiyomi Shinohara
Department of Health Promotion and Human Behavior
Kyoto University Graduate School of Medicine / School of Public Health
Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501 Japan
Phone: +81-75-753-9491
Fax: +81-75-753-4641
E-mail address: shinohara.kiyomi.75r@st.kyoto-u.ac.jp

Abstract

Objectives

To investigate whether the overstatements in abstract conclusions influence primary care physicians' evaluations when they read reports of a randomised controlled trial (RCT).

Design: Randomised controlled trial

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 from January to February in 2017. We sent e-mail invites to 7040 primary care physicians who had clinical experience of more than 2 years. Among the 787 individuals who accessed the website, 622 were eligible and automatically randomised to 'without overstatement' (n=307) and 'with overstatement' (n=315) group.

Interventions: Participants were randomly assigned to read an abstract of one of five RCT reports either with overstatements or without 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. Secondary outcomes were the validity of the conclusion and the interest in reading the full text.

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% confidence interval (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.

Trial registration number: the University hospital Medical Information Network-Clinical Trial Registry (UMIN000025317)

Strength and limitations of this study

- This is the first and only RCT study that estimates the influence of overstatement in abstract conclusions.
- We evaluated the influence of overstatement in primary care physicians who were one of the major users of evidence.
- Although the number of participants was above our targeted sample size, relatively low response rate limits the generalizability of our findings.
- Since we focused on the overstatement in abstract conclusions, the effect of other various 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. It is particularly useful for primary care physicians since 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 their 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 of evidence based medicine (EBM) [3 4] to interpret the results. Thus, a strong conclusion may alter the readers' interpretation of the whole study.

Unfortunately, abstract conclusions are the most frequently distorted section in abstracts [5]. Exaggerating the results of the trial, such as, using a 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]. According to previous studies, 58% of RCTs with non-significant results [5], 23% of RCTs in rheumatology [7], and 33 % of psychiatry trials [8] had spins, misleading information, or overstatements in their abstract conclusions. This suggests that as far as abstract conclusions are concerned, the quality of reporting is still poor, in spite of the 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. There was only one RCT [9] that investigated the extent of the impact of inappropriate reporting on readers' interpretations of the results. Boutron et al. [9] randomised clinical researchers into two groups, and asked them to read abstract with a 'spin', which was defined by the authors as 'reporting the beneficial effect of the intervention was greater than shown by the results', or without it 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 weight of spin in abstract conclusion in influencing the participants' interpretation was unclear because the investigators added changes to all sections of the abstracts. They used abstracts of

studies that had only non-significant primary outcomes. Then they either erased or added all the results of secondary outcomes while changing the wording. In other words, the study compared the fully spun abstracts and the 'paragon' analogues without any exaggeration. This approach may have overestimated the impact of spin. Moreover, the target population was clinical researchers with a publishing experience.

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

Methods

Setting and design

This online study was a double-blind RCT, and it was 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 Japan Primary Care Association (JPCA) by sending e-mail invites. The intervention was conducted on a website specifically designed for this study. Participants were randomised into two groups and asked to read and evaluate one of the ten abstracts (five pairs of two corresponding abstracts: one with and another without overstatement) of a report of an RCT. The trial was prospectively registered with the University hospital Medical Information Network-Clinical Trial Registry (UMIN000025317). We did not publish our protocol to avoid the risk of participants reading it.

Participants and recruiting

The target population was recruited from the members of the Japan Primary Care Association (JPCA). JPCA was established in 2010 to promote primary care specialty in Japan [10]. 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 [11] belong to the JPCA, and 5,836 members out of a total of 10851 are certificated as specialists in primary care.

We sent e-mail invites to JPCA members who had clinical experience of more than 2 years with registered e-mail address (the details of the recruiting process will be reported in a separate paper). Interested individuals were able to access the DOCTOR study website via the link in the e-mail. We added the code at the end of the link in

order to ensure that participants accessed the website via the given link. As an incentive, an Amazon gift card worth 3000 yen was given to 20 lucky draw winners.

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

Randomisation

When participants moved to the assessment page, they were randomly assigned to an abstract either with or without overstatements in the ratio of 1:1. The randomisation procedure was automatic (created by RS) through block randomisation (10 for each block).

Blinding

In the e-mail invites, 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 invite is included in the supplementary). They were therefore masked to the study hypothesis. The researchers (KS, TA, YT, and AS), except the website manager (RS), were blinded until the blind interpretations of the results were completed and signed off [12]. RS did not join the result analysis.

Selecting abstracts with overstatements

We selected five abstracts [13-17] (the text of the five abstracts is included in the supplementary material) from the pre-existing database of published reports in psychiatry RCTs dated between 2011 and 2014, which was collected from our previous study [6 18]. In order to avoid any bias arising from the participants' sub-specialty 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 intervention in abstract conclusion despite that some or all primary outcomes were not significant, 3) targeting a common mental illness that the primary care physicians are likely to encounter in clinical settings, and 4) impact factor of journal equal to or higher than two.

An overstatement was defined as the '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 that were selected, two abstracts only mentioned the superiority of 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 the treatment was partially effective even though all the primary outcomes were non-significant. Together they include all 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 without overstatements

In 'without overstatement' conclusion, we rewrote the conclusion (see an example in Table 1, and all the abstract conclusions are in Table 2) section of abstract following these rules: 1) when all primary outcomes were non-significant, we rewrote it 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 re-wrote it 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. The results of secondary outcomes and subgroup analysis were removed from conclusions.

In both groups, the names of intervention and control treatment were changed to anonymous 'intervention A' and 'control B' to minimise bias. To keep the conclusion consistent with other sections of abstract, we standardised the methods and results section. We clarified the primary outcomes and results (for example, odds ratio risk ratio, confidence interval, p-value) from the text if they were not stated in the original abstract. Except for the conclusion, 'without' and 'with' abstracts were identical. Also, we translated the texts into Japanese, and another researcher (SK), who was not involved in this study, checked the translation.

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 the intervention A is for the patients, on a scale of 0 to 10, 0 being not at all beneficial and 10 being conceivably most beneficial?' We also asked the following questions (score 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 of 0 to 10?.
- How much do you want to read the full text of this study on a scale of 0 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 [9]. They estimated the effect of spin by comparing the influence of the abstract '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 [19], we aimed for a sample size of 253 for a group, and 506 in total to detect between group effect size of 0.25 with a power of 90% and a two-sided alpha risk at 5%. At least 100 participants were allocated for each pair of abstracts (with and without).

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 intention-to-treat (ITT) population 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 only spent less than 30 seconds on the questionnaire. Our ITT population consisted of those who are willing to and do read the abstracts of scientific articles. TA and KS analysed the data using SPSS statistics 24 without knowing the allocation. To evaluate the influence of possible associated factors [3 20] on the interpretation, we conducted the following pre-specified subgroup analyses using the participants: 1) working clinics, 2) getting information only from pharmacological company, 3) with certification of 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 Guyatt et al. [12] 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.

Ethics

This study was approved by the Ethics Committee of Kyoto University Graduate School of Medicine and was conducted in accordance with the Declaration of Helsinki. We obtained an online consent of participation from each participant.

Results

We sent e-mail invites to 7040 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) group. A total of 281 doctors in the 'without' group and 286 in the 'with' group were included for the ITT analysis. Fifty-five individuals were excluded because they either spent less than 30 seconds on the webpage (n=14) or did not complete the survey (n=41). Most participants read and rated the abstract within four minutes (medium time: 162 seconds, 25 per trial; IQR, 114 – 236 seconds).

Table 3 shows the characteristics of the participants, and 76.5% were certified as primary care physicians. We classified their sub-specialty according to their certifications. The most common background was internal medicine. More than 60% of the participants had attended a course on EBM, and 40% of the physicians said that they read the conclusion section first when reading an abstract.

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) (Table 4).

Secondary outcomes and subgroup analyses

However, there was a significant difference between the groups on their perception of

the validity of the conclusion (mean difference: 0.97; 95% CI, 0.59 to 1.36; P<0.001) (Fig. 2). Those in the 'without overstatement' group considered the abstract to be more valid than the 'with overstatement' group (effect size calculated by Cohen's d was 0.41). No significant difference was found when asked if they wanted to read the full text. In both groups, majority of the doctors referred to the results section to make an assessment

We conducted sub-group 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 resource of information (only pharmacological company, n=43, mean difference: 0.06; 95% CI, -1.36 to 1.48; P=0.93), certificated primary care physician (n=434, mean difference: -0.01; 95% CI, -0.41 to 0.39; P=0.96), or having no experience of being 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% confidence interval of the estimated effect (effect size: 0.031; 95% CI, -0.13 to 0.20) rules out the existence of even a small effect. The participants read the conclusion section first, but referred to the results section mostly for their interpretation. They tend to judge the overstated conclusion as less valid than those without the overstatement. These results suggested that primary care physicians who belong to 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 sub-group 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 report, which helped them to recognise the inconsistency between the result and the conclusion.

Compared to the previous study, our results differed in some aspects. In Boutron's study [9], they showed the interpretation of abstracts was affected by spin. The 'abstracts with spin' group considered the intervention as more beneficial than the 'without spin' group, and '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 that the abstract was less methodologically rigorous than the 'without spin' group. This was consistent with our results.

There are several possible reasons for these differences. First, the level of spin was much higher in Boutron's study than this study. While Boutron et al. rewrote the whole abstract, we only made such changes in the conclusion section because our aim was to measure the influence of the conclusion section. In addition, if abstracts did not report, we added information on the primary outcomes in the methods and results section for both groups. Furthermore, the baseline characteristics of the participants differed. While all the participants in Boutron's study were experienced clinical researchers, we chose primary care physicians as our target. This study showed that primary care physicians were not affected by overstatement in conclusion even though they have little experience in clinical research.

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 [21]; 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 were one of the major users of evidence. They encounter clinical queries in daily clinical practice and use evidence to make the best decision for their patients [22]. 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 represented the JPCA members completely. The relatively low response rate of 11.1% (787/7040) limits the generalizability of our findings. Two things should be noted. Firstly, we chose the JPCA as our recruiting pool because the members were considered to be active users of scientific evidence in their primary care practice. Moreover, the responders to our invitation were potentially avid readers of scientific reports, which is the reason they volunteered for this assessment, and therefore they have better critical appraisal skills of abstracts than other JPCA members. 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 were not measured. In our study, we added necessary information of primary outcomes in the methods and results section as CONSORT statement[2] recommends. 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 judgment on the validity of misleading conclusions. However, it does not mean that overstatements can be overlooked. The conclusion section of abstracts should be written solely on the basis of the primary outcome results. The impact of inappropriate writing style in clinical settings should be further researched.

Acknowledgement

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

Contributors

All authors (KS, TA, RS, YT, AMS, MK and TAF) of the paper have contributed 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 web-site. 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.

Conflicts of interest

KS has received grant from Japan Primary Care Association. 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. AMS, TA, YT, MK, and RS reported no competing interests.

Data sharing statement

Additional unpublished data is still being analyzed for another research and only available to the members of the study team.

1. Burke DT, Judelson AL, Schneider JC, DeVito MC, Latta D. Reading habits of practicing physiatrists. Am J Phys Med Rehabil 2002;**81**(10):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**(1):e20.
3. Hisham R, Ng CJ, Liew SM, Hamzah N, Ho GJ. Why is there variation in the practice of evidence-based medicine in primary care? A qualitative study. BMJ Open 2016;**6**(3):e010565.
4. Zwolsman S, te Pas E, Hooft L, Wieringa-de Waard M, van Dijk N. Barriers to GPs' use of evidence-based medicine: a systematic review. Br J Gen Pract 2012;**62**(600):e511-21.
5. Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. JAMA 2010;**303**(20):2058-64.
6. Suganuma AM, Shinohara K, Imai H, Takeshima N, Hayasaka Y, Furukawa TA. Overstatements in abstract conclusions claiming effectiveness of interventions in psychiatry: a study protocol for a meta-epidemiological investigation. BMJ Open 2016;**6**(4):e009832.
7. Mathieu S, Giraudeau B, Soubrier M, Ravaud P. Misleading abstract conclusions in randomized controlled trials in rheumatology: comparison of the abstract conclusions and the results section. Joint Bone Spine 2012;**79**(3):262-7.
8. Kiyomi Shinohara , Aya M. Suganuma, Hissei Imai, Nozomi Takeshima, Yu Hayasaka, Toshi A. Furukawa. Overstatements in abstract conclusions claiming effectiveness of interventions in psychiatry: a meta-epidemiological investigation.
9. Boutron I, Altman DG, Hopewell S, Vera-Badillo F, Tannock I, Ravaud P. 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**(36):4120-6.
10. Japan Primary Care Association. https://www.primary-care.or.jp/jpca_eng/index.html.
11. Aoki T, Inoue M, Nakayama T. Development and validation of the Japanese version of Primary Care Assessment Tool. Fam Pract 2016;**33**(1):112-17.
12. Jarvinen TL, Sihvonen R, Bhandari M, et al. Blinded interpretation of study results can feasibly and effectively diminish interpretation bias. J Clin Epidemiol 2014;**67**(7):769-72.
13. Sternfeld B, Guthrie KA, Ensrud KE, et al. Efficacy of exercise for menopausal symptoms: a randomized controlled trial. Menopause 2014;**21**(4):330-8.
14. 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**(4):398-414.

15. Freund-Levi Y, Jedenius E, Tysen-Bäckström AC, Lärksäter M, Wahlund LO, Eriksdotter M. Galantamine versus risperidone treatment of neuropsychiatric symptoms in patients with probable dementia: an open randomized trial. The American journal of geriatric psychiatry 2014;**22**(4):341-8.
16. 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**(2):132-9.
17. Lam RW, Parikh SV, Ramasubbu R, et al. Effects of combined pharmacotherapy and psychotherapy for improving work functioning in major depressive disorder. The British Journal of Psychiatry 2013;**203**(5):358-65.
18. Shinohara K, Tajika A, Imai H, Takeshima N, Hayasaka Y, Furukawa TA. Protocol registration and selective outcome reporting in recent psychiatry trials: new antidepressants and cognitive behavioural therapies. Acta Psychiatr Scand 2015;**132**(6):489-98.
19. Cohen J. *Statistical Power Analysis in the Behavioral Sciences.* Erlbaum, 1988.
20. DeJong C, Aguilar T, Tseng C, Lin GA, Boscardin W, Dudley R. PHarmaceutical industry–sponsored meals and physician prescribing patterns for medicare beneficiaries. JAMA Internal Medicine 2016;**176**(8):1114-10.
21. 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.
22. de Vries YA, Roest AM, Franzen M, Munafo MR, Bastiaansen JA. 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**(14):2971-79.

Table1. An example of the abstracts (underlined where extra text added)

TITLE

Intervention A for menopausal symptoms: a randomized 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 flash, sweating, and poor circulation participated in a randomized controlled trial conducted in three sites: 106 women randomized to exercise and 142 women randomized 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 Generalized 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

CONCLUCIONS

Intervention A was not more effective than control B in terms of frequent vasomotor symptoms (VMS) such as hot flash, sweating in postmenopausal women.

Control B is the standard treatment for menopausal symptoms.

Table2. Five sample abstracts and their two versions of conclusions

Study	symptoms or illness	Conclusion in the original abstract	Conclusion without overstatement (rewritten by investigators)
Samus 2014	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.
Oosterbaan 2013	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.
Sternfeld 2014	Menopausal Symptoms	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.	Intervention A was not more effective than control B in terms of frequent vasomotor symptoms (VMS) such as hot flash, sweating in postmenopausal women.
Levi 2014	Neuropsychiatric symptoms in patients with probable dementia	These results support that intervention A, with its benign safety profile, can be used as first-line treatment of NPSD symptoms, 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 2013	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.

Fig1. Flow diagram of participants

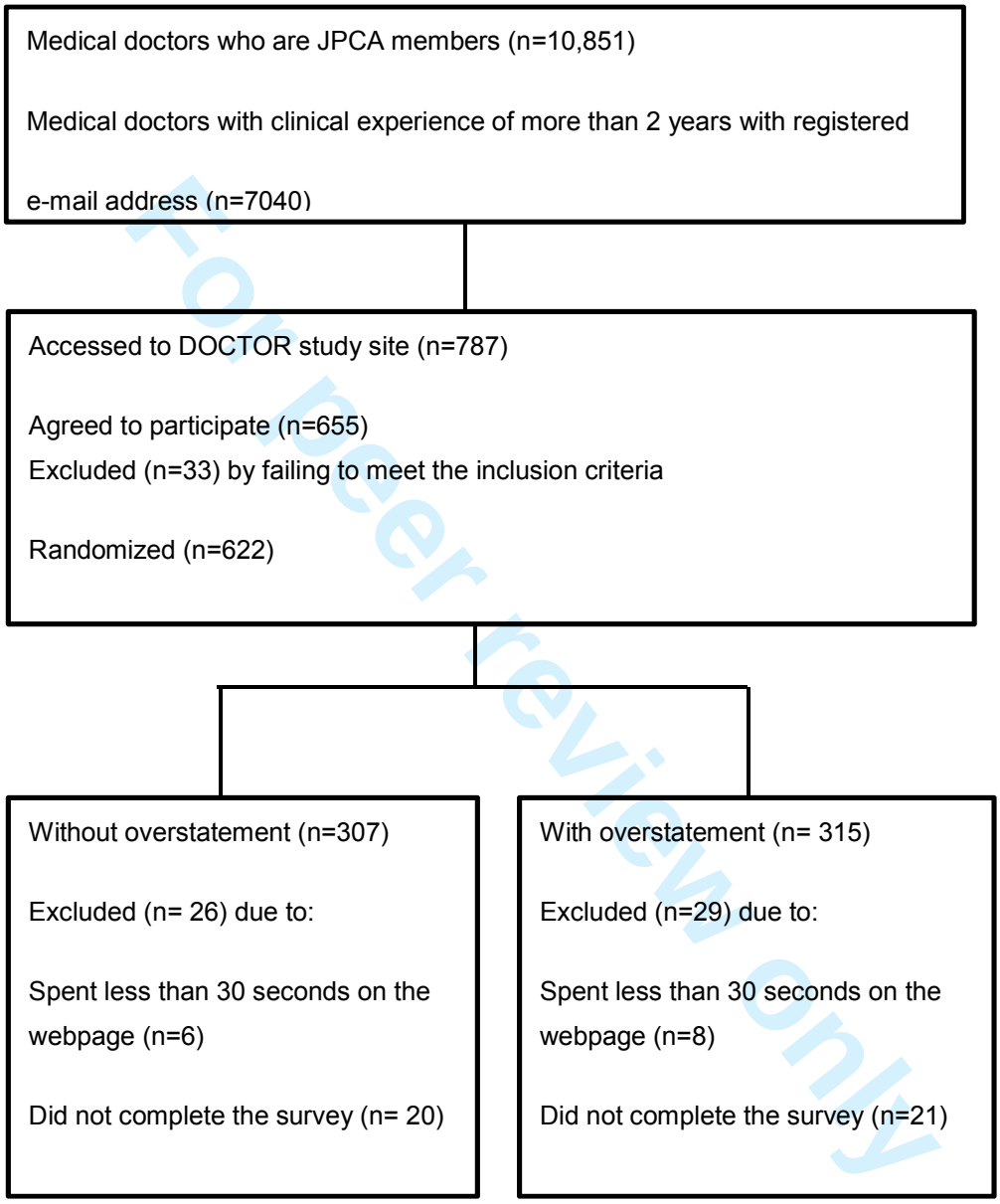


Table 3. Characteristics of participants

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 to 24	median 16.0 IQR 11 to 24	median 16.0 IQR 11 to 24
Work place			
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	115(40.9)	130 (45.5)	245 (43.2)
Surgery	26(9.3)	26 (9.1)	52 (9.2)
Emergency medicine	15 (5.3)	14 (4.9)	29 (5.1)
Pediatrics	6 (2.1)	5 (1.7)	11 (1.9)
Others	28 (10.0)	23 (8.0)	51 (9.0)
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	108 (38.4)	105 (36.7)	213 (37.6)

Background			
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)
5 or more	129 (45.9)	112 (39.2)	241 (42.5)

Clinical background data was available with participants who have sub-specialty certifications.

¶ multiple answers allowed

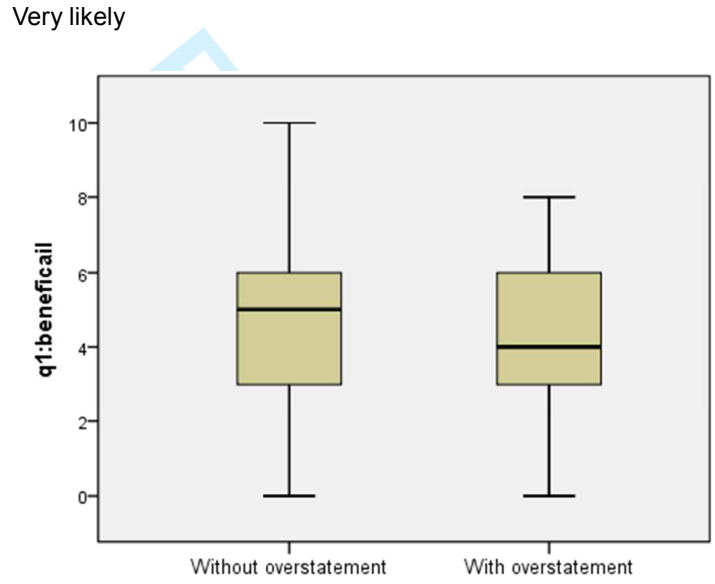
Abbreviation: OS: overstatement; IQR: interquartile range percentiles; PI: principle investigator

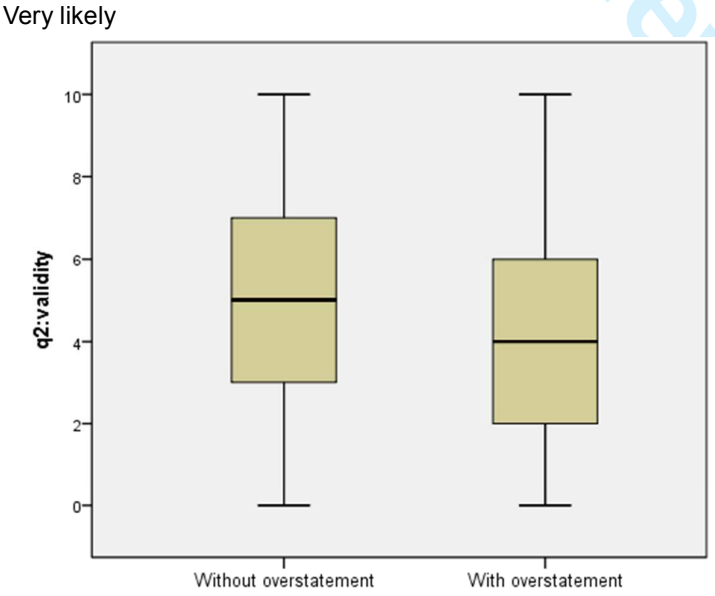
Table 4. Impression of the abstract

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 n=567(95%CI)
How beneficial do you think the 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

Fig 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 percentiles.





CONTENTS OF SUPPLEMENTAL MATERIAL

Appendix1: Invitation e-mail

Subject: Win an Amazon gift card by participating in a 5-minute survey on EBM

This email is important.
My name is Morito Kise from Centre for Family Medicine Development, Japanese Health and Welfare Co-operative Federation, Tokyo, Japan

I am sending this email to invite you to participate in a clinical trial targeting clinicians. This research is a collaborative effort between Japan Primary Care Association (JPCA) and Kyoto University and aims to investigate the application of published articles among clinical practitioners. It is funded by Japan Primary Care Association, and has been approved by the board of committees.
For those JPCA members with more than three years of clinical experience, we would kindly ask you to read ONE abstract of a medical article and evaluate it on a scale of 0 to 10. The estimated time to complete the whole process is 5 minutes.

As a token of appreciation, we give away Amazon gift cards worth 3000 yen to 20 of the participants. The prize winners will be notified at the end of the survey.

▼▼▼Please click the link below to participate.▼▼▼
http://doctor-study.net/abstud y/public/base/index/0124B
It can be also accessed via your smartphone. The deadline is on the 31st of January, 2017.

This project investigates how clinical practitioners assess abstracts of scientific reports. It is funded by JPCA, and has been approved the Ethics Committee of Kyoto University. No personal particulars may be used to identify any individuals nor any results may be associated with particular individuals. The data obtained may be used, after blinding, for secondary research purposes. No information will be given to other organisations or individuals. Results of this investigation will be reported and published publicly but only after a blinding. Prize winners will be asked to provide their email and work addresses. The information will not be used for any other purposes. It is possible to drop out after you start.

Again, we would appreciate it greatly if you could give us your time for five minutes. Thank you for your cooperation

---

Morito Kise, MD
Centre for Family Medicine Development, Japanese Health and Welfare Co-operative Federation, Tokyo, Japan

Takuya Aoki, MD MMA
Department of Healthcare Epidemiology
Kyoto University Graduate School of Medicine / School of Public Health Morito Kise, MD

Kiyomi Shinohara, MD PhD
Department of Health Promotion and Human Behavior
Kyoto University Graduate School of Medicine / School of Public Health

Please contact:
doctor.study.pc@gmail.com

Kyoto University Graduate School of Medicine Administration Office Department of Conflict of Interest
(tel) 075-753-4305
(E-mail) 060rieki-sohan@mail2. adm.kyoto-u.ac.jp

Appendix 2 – Full text of five abstracts

We added the shaded part to the original abstract.

134

TITLE: Intervention A for menopausal symptoms: a randomized 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 flash, sweating, and poor circulation participated in a randomized controlled trial conducted in three sites: 106 women randomized to exercise and 142 women randomized 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 Generalized 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.

(Without OS)

Intervention A was not more effective than control B in terms of frequent vasomotor symptoms (VMS) such as hot flash, sweating in postmenopausal women.

Control B is the standard treatment for menopausal symptoms.

253

TITLE: Intervention A versus control B treatment of neuropsychiatric symptoms in patients with probable dementia: an open randomized trial

OBJECTIVES: to examine the effect of intervention A and control B on neuropsychiatric symptoms in dimentia (NPSD) and global function

METHODS: Using a randomised controlled and open-blind, once centre trial at an in-and outpatient clinic at a university hospital, we studied 100 adults with probablu dementia and NPSD. Participants received treatment A (N=50) or control B (N=50) for 12 weeks. The primary outcome was effects on NPSD, the difference between baseline and 12 weeks, assessed by the Neuropsychiatric Inventory (NPI). Secondary measures included the Mini-MEntal State Examination (MMSE), clinical demential rating, clinical global impressiona nd Simpson Angus scales. All tests were performed before and after treatment.

RESULTS: Outcome measures were analyzed using analysis of covariance. 91 patients (67% women, mean age 79+/- 7.5 years) with initial NPI score of 51 (+/- 25.8) and MMSE of 20.1 (+/- 4.6) completed the trial. Both intervention A and control B resulted in improved NPSD symptoms and were equally effective in treating several NPI domains (the differences at 12 weeks intervention A: 16.7±15.6, control B: 17.9±16.3, p=0.06). However, control B showed a significant treatment advantage in the NPI domains irritation and agitation, F(1, 97) = 5.2, p=0.02. Intervention A also ameliorated cognitive functions where MMSE scores increased 2.8 points compared with baseline (95% CI: 1.96-3.52). No treament-related severe side effects occurred.

CONCLUSION: These results support that intervention A, with its benign safety profile, can be used as first-line treatment of NPSD symptoms, unless symptoms of irritation and agitation are prominent, where control B is more efficient.

(Without OS)

Intervention A was not more effective than control B in terms of neuropsychiatric symptoms in patients with dementia.

Control B is a generally used antipsychotics.

1000382

TITLE: Effects of intervention A for improving work functioning in major depressive disorder

BACKGROUND: Major depressive disorder is associated with significant impairment in occupational functioning and reduced productivity, which represents a large part of the overall burden of depression.

AIMS: To examine symptom-based and work functioning outcomes with intervention A treatment of major depressive disorder.

METHOD: Employed patients with a DSM-IV diagnosis of major depressive disorder were treated with escitalopram 10-20 mg/day and randomized to intervention A (n = 48) or control B (n = 51). Primary outcome was the Montgomery-Asberg Depression Rating Scale (MADRS), administered by masked evaluators via telephone. Secondary outcome was self-rated work functioning scales completed online.

RESULTS: After 12 weeks, there were no significant between-group differences in change in MADRS score [effect size (Cohen's d) 0.16, P=0.60] or in response /remission (response: ≥ 50% improvement in MADRS scores, remission: MADRS ≤ 12). However, participants in intervention A had significantly greater improvement on some measures of work functioning than the control B.

CONCLUSIONS: Intervention A with escitalopram significantly improved some self-reported work functioning outcomes, but not symptom-based outcomes, compared with escitalopram and control B.

(Without OS)
Intervention A with escitalopram was not more effective than control B with escitalopram in terms of depressive symptoms in patients with major depression.

Control B is the standard treatment for depression.

1000385

TITLE: Intervention A v. control as usual for common mental disorders: 8-month, cluster randomized controlled trial

AIMS: To evaluate the effectiveness of intervention A in the treatment of common mental disorders.

METHOD: An 8-month cluster randomized controlled trial comparing intervention A to control B. Primary outcomes were the percentage of patients responding to and remitting on Clinical Global Impression of Improvement Scale (CGI-I) after treatment.

RESULTS: Twenty general practitioners (GPs) and 8 psychiatric nurses were randomised to provide intervention A or control B. The GPs recruited 163 patients [intervention A (n=94)、treatment B (n=64)] of whom 85% completed the post-test measurements. At 4-month mid-test intervention A was superior to control B: 74.7% (n = 68) v. 50.8% (n = 31) responders (P = 0.003). At 8-month post-test and 12-month follow-up no significant differences were found as the patients in control B group improved as well [response at 8-month: 80.2% (n = 73) vs. 67.2% (n = 41), *P*=0.072; remission at 8 month: 58.9% (n = 53) vs. 51.7% (n = 31), *P*=0.383].

CONCLUSIONS: Intervention A resulted in an earlier treatment response compared with control B.

(Without OS)
Intervention A was not more effective than control B in terms of treatment response or remission in patients with common mental illness.

Control B is the standard treatment for common mental illness.

135

TITLE: Intervention A for elders with memory disorders: the pilot randomized trial

OBJECTIES: To assess whether intervention A delays time to transition from home (to a hospital or nursing home) and reduces unmet needs in elders with memory disorders.

DESIGN: 18-month randomized controlled trial of 303 community-living elders.
SETTING: 28 postal code areas of Baltimore, MD.
PARTICIPANTS: Age 70+, with a cognitive disorder, community-living, English-speaking, and having a study partner available.

INTERVENTION: 18-month intervention A. Care monitoring by an interdisciplinary team.

MEASUREMENTS: Primary outcomes were time to transfer from home and total percent of unmet care needs at 18 months (measured on Johns Hopkins Dementia Care Needs Assessment).

RESULTS: Intervention participants had a significant delay in time to all-cause transition from home and the adjusted hazard of leaving the home was decreased by 37% (HR = 0.63, 95% CI 0.42 to 0.94) compared to control participants. While there was no significant group difference in reduction of total percent of unmet needs from baseline to 18 months (p=0.054), the intervention group had significant reductions in the proportion of unmet needs in safety and legal/advance care domains relative to controls. Participants in intervention A group had a significant improvement in self-reported quality of life (QOL) relative to control participants. No group differences were found in proxy- rated QOL, neuropsychiatric symptoms, or depression.

Conclusions—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.

(Without OS)
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.

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	P1
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	P2
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	P4
	2b	Specific objectives or hypotheses	P5
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	P6 Randomisation P5 Setting and design
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	NA
Participants	4a	Eligibility criteria for participants	P5 Participants and recruiting
	4b	Settings and locations where the data were collected	P5 Setting and design
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	P6 Selecting abstracts with overstatements P7 Constructing abstracts without overstatements
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	P7 Outcomes
	6b	Any changes to trial outcomes after the trial commenced, with reasons	P8 Statistical analysis
Sample size	7a	How sample size was determined	P8 Sample size

	7b	When applicable, explanation of any interim analyses and stopping guidelines	NA
Randomisation:			
Sequence	8a	Method used to generate the random allocation sequence	P6 Randomisation
generation	8b	Type of randomisation; details of any restriction (such as blocking and block size)	P6 Randomisation
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	P6 Randomisation
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	P6 Randomisation
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	P6 Blinding P8 Blinded data interpretation
	11b	If relevant, description of the similarity of interventions	P7 Constructing abstracts without overstatements
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	P8 Statistical analysis
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	P8 Statistical analysis
Results			
Participant flow (a diagram is strongly	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	P9 Results, Fig1
recommended)	13b	For each group, losses and exclusions after randomisation, together with reasons	P9 Results, Fig1
Recruitment	14a	Dates defining the periods of recruitment and follow-up	P5 Setting and design
	14b	Why the trial ended or was stopped	P9 Results
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	Table 3
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	Fig1
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	P9 Primary outcome, Secondary outcomes and subgroup analyses, Table 4

	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	NA
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	P9 Secondary outcomes and subgroup analyses
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	NA
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	P11 Limitations and strengths
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	P11 Limitations and strengths
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	P10 Discussion
Other information			
Registration	23	Registration number and name of trial registry	P1
Protocol	24	Where the full trial protocol can be accessed, if available	NA
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	P12 Funding

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

BMJ Open

The influence of overstated abstract conclusions on clinicians: A randomized controlled trial: DOCTOR study (Do Overstated Conclusions Trick Our Readers?)

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-018355.R1
Article Type:	Research
Date Submitted by the Author:	01-Sep-2017
Complete List of Authors:	Shinohara, Kiyomi; Kyoto University Graduate School of Medicine/Schoo of Public Health, Health Promotion and Human Behaviour Aoki, Takuya; Kyoto University Graduate School of Medicine / School of Public Health, Department of Healthcare Epidemiology So, Ryuhei ; Kyoto University Graduate School of Medicine / School of Public Health, Department of Health Promotion and Human Behavior Tsujimoto, Yasushi ; Kyoto University Graduate School of Medicine/Schoo of Public Health, Department of Healthcare Epidemiology Suganuma, Aya M; Kyoto University Graduate School of Medicine/Schoo of Public Health, Health Promotion and Human Behaviour Kise, Morito; Centre for Family Medicine Development, Japanese Health and Welfare Co-operative Federation Furukawa, Toshi; Kyoto University, Graduate School of Medicine and School of Public Health
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Evidence based practice
Keywords:	randomised controlled trials, general practice, overstatements, primary care physicians, reporting bias, clinical trial

Title:

The influence of overstated abstract conclusions on clinicians: A randomized controlled trial: DOCTOR study (Do Overstated Conclusions Trick Our Readers?)

Authors:

Kiyomi Shinohara[a],
Takuya Aoki [b],
Ryuhei So[a],
Yasushi Tsujimoto[b],
Aya M Suganuma [a],
Morito Kise[c],
Toshi A. Furukawa[a]

Affiliation of all authors:

a: Department of Health Promotion and Human Behavior Kyoto University Graduate School of Medicine / School of Public Health
Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501 Japan,
b: Department of Healthcare Epidemiology Kyoto University Graduate School of Medicine / School of Public Health
Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501 Japan
c: Centre for Family Medicine Development, Japanese Health and Welfare Co-operative Federation, Tokyo, Japan
Sanken Building 4th Floor
3-25-1, Hyakunin-cho, Shinjuku-ku
Tokyo Japan

Corresponding author:

Kiyomi Shinohara
Department of Health Promotion and Human Behavior
Kyoto University Graduate School of Medicine / School of Public Health
Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501 Japan
Phone: +81-75-753-9491
Fax: +81-75-753-4641
E-mail address: 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: Randomised controlled trial

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 e-mail 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 sample 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 section were standardized by adding the information of primary outcome information if 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% confidence interval [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

Strength and limitations of this study

- This is the first and only RCT 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 generalizability of our findings.
- Since 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 since 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-randomized 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].

According to previous studies, 58% of RCTs with non-significant results [5], 23% of RCTs in rheumatology [8], and 33 % of psychiatry trials [9] had spin, misleading information, or overstatements in their abstract conclusions. This suggests that, as far as abstract conclusions are concerned, the quality of reporting is still poor despite the 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 e-mail 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). 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 Japan Primary Care Association (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 5,836 out of a total of 10851 members are certified as specialists in primary care.

We sent e-mail invitations to JPCA members who had more than two years of clinical experience with registered e-mail addresses. (The details of the recruiting process will be reported in a separate paper). We excluded clinicians with less than two years' experience because our target population was primary care physicians, and doctors usually choose their specialty after two years of clinical training in Japan. Interested individuals could access the DOCTOR study website via the link in the e-mail. 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 was given to 20 drawing winners.

The inclusion criteria for participants were as follows: JPCA member, medical doctor currently in clinical practice, more than two years of clinical practice experience, and access to up-to-date clinical research knowledge. We asked how respondents learned 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 the a computer-generated random sequence (created by RS). The allocation concealment was maintained through the automatic random allocation process.

Blinding

In the e-mail 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 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 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' sub-specialty 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 differentlevels 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 pre-specified guidelines. First, we rewrote the conclusion to make a conclusion 'without overstatement' following these rules.: 1) Wwhen all primary outcomes were non-significant, we rewrote it the conclusion as 'Iintervention 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 re-wrote ithe conclusion t as 'iIntervention 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. (Ssee an example in Table 1;, and all the abstract conclusions are in Table 2.)

Second, we standardized the methods and results sections. We explicitly stated the primary outcomes and results (for example, odds ratio, risk ratio, confidence interval, p-value) from the text if they were not stated in the original abstract. Therefore, all abstracts had necessary 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 original abstract. Also Additionally, this standardization 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 minimize bias. 2) We added a few words for explanation when there was a medical term that seemed unfamiliar to primary care physicians (e.g. vasomotor symptoms [(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 tThe 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 enroll 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 seconds 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 pre-specified 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 Guyatt 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.

Ethics

This study was approved by the Ethics Committee of Kyoto University Graduate School of Medicine and was conducted in accordance with the Declaration of Helsinki. We obtained an online consent for participation from each participant.

Results

We sent e-mail invitations to 7,040 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). Supplementary appendix 3 provides further breakdown per abstract.

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

Table 3 shows the participant characteristics; 76.5% were certified as primary care physicians. We classified their sub-specialties 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 4).

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) (Fig. 2). Those in the 'without overstatement' group considered the abstract to be more valid

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]; 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 [29]. 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 generalizability of our findings. Two things should be noted. Firstly, 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 organization 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. 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 were not measured. In our study, we added necessary information of primary outcomes in the methods and results section as CONSORT statement[2] recommends. Actually, most of participants answered that they read abstracts regularly. This suggests 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 of on primary outcomes in the methods and results sections as recommended by a CONSORT statement[2] recommends. 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 overgeneralize the association between the type or level of overstatement and their 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 enough sufficient sample size to explore such relationships. There are various types of inappropriate, misleading reporting. 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 judgment 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.

Acknowledgement

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

Contributors

All authors (KS, TA, RS, YT, AMS, MK and TAF) of the paper have contributed 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 web-site. 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.

Conflicts of interest

KS has received grant from Japan Primary Care Association. 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. AMS, TA, YT, MK, and RS reported no competing interests.

Data sharing statement

No additional data available.

1. Burke DT, Judelson AL, Schneider JC, DeVito MC, Latta D. Reading habits of practicing physiatrists. Am J Phys Med Rehabil 2002;**81**(10):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**(1):e20.
3. Hisham R, Ng CJ, Liew SM, Hamzah N, Ho GJ. Why is there variation in the practice of evidence-based medicine in primary care? A qualitative study. BMJ Open 2016;**6**(3):e010565.
4. Zwolsman S, te Pas E, Hooft L, Wieringa-de Waard M, van Dijk N. Barriers to GPs' use of evidence-based medicine: a systematic review. Br J Gen Pract 2012;**62**(600):e511-21.
5. Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. JAMA 2010;**303**(20):2058-64.
6. Suganuma AM, Shinohara K, Imai H, Takeshima N, Hayasaka Y, Furukawa TA. Overstatements in abstract conclusions claiming effectiveness of interventions in psychiatry: a study protocol for a meta-epidemiological investigation. BMJ Open 2016;**6**(4):e009832.
7. Lazarus C, Haneef R, Ravaud P, Boutron I. 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, Ravaud P. Misleading abstract conclusions in randomized controlled trials in rheumatology: comparison of the abstract conclusions and the results section. Joint Bone Spine 2012;**79**(3):262-7.
9. Kiyomi Shinohara , Aya M. Suganuma, Hissei Imai, Nozomi Takeshima, Yu Hayasaka, Toshi A. Furukawa. Overstatements in abstract conclusions claiming effectiveness of interventions in psychiatry: a meta-epidemiological investigation.
10. Lockyer S, Hodgson R, Dumville JC, Cullum N. "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 2016.
12. Fleming PS. Evidence of spin in clinical trials in the surgical literature. Ann Transl Med 2016;**4**(19):385.

13. Vera-Badillo FE, Shapiro R, Ocana A, Amir E, Tannock IF. Bias in reporting of end points of efficacy and toxicity in randomized, clinical trials for women with breast cancer. Ann Oncol 2013;**24**(5):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**(7):1949-50.
16. Boutron I, Altman DG, Hopewell S, Vera-Badillo F, Tannock I, Ravaud P. 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**(36):4120-6.
17. Japan Primary Care Association. https://www.primary-care.or.jp/jpca_eng/index.html.
18. Aoki T, Inoue M, Nakayama T. Development and validation of the Japanese version of Primary Care Assessment Tool. Fam Pract 2016;**33**(1):112-17.
19. Jarvinen TL, Sihvonen R, Bhandari M, et al. Blinded interpretation of study results can feasibly and effectively diminish interpretation bias. J Clin Epidemiol 2014;**67**(7):769-72.
20. Sternfeld B, Guthrie KA, Ensrud KE, et al. Efficacy of exercise for menopausal symptoms: a randomized controlled trial. Menopause 2014;**21**(4):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**(4):398-414.
22. Freund-Levi Y, Jedenius E, Tysen-Bäckström AC, Lärksäter M, Wahlund LO, Eriksdotter M. Galantamine versus risperidone treatment of neuropsychiatric symptoms in patients with probable dementia: an open randomized trial. The American journal of geriatric psychiatry 2014;**22**(4):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**(2):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. The British Journal of Psychiatry 2013;**203**(5):358-65.
25. Shinohara K, Tajika A, Imai H, Takeshima N, Hayasaka Y, Furukawa TA. Protocol registration and selective outcome reporting in recent psychiatry trials: new antidepressants and cognitive behavioural therapies. Acta Psychiatr Scand

2015;**132**(6):489-98.

26. Cohen J. *Statistical Power Analysis in the Behavioral Sciences* . Erlbaum, 1988.

27. DeJong C, Aguilar T, Tseng C, Lin GA, Boscardin W, Dudley R. PHarmaceutical industry–sponsored meals and physician prescribing patterns for medicare beneficiaries. JAMA Internal Medicine 2016;**176**(8):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, Munafo MR, Bastiaansen JA. 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**(14):2971-79.

Table1. An example of the abstracts (underlined where extra text added, double underlined where changed in the 'without overstatement' group)

TITLE Intervention A for menopausal symptoms: a randomized 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 randomized controlled trial conducted in three sites: 106 women randomized to exercise and 142 women randomized 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 Generalized 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 vasomotor symptoms (VMS) such as hot flush, sweating in postmenopausal women. Control B is the standard treatment for menopausal symptoms.

Table2. 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)
1.Sternfeld 2014	Menopausal Symptoms	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.	Intervention A was not more effective than control B in terms of frequent vasomotor symptoms (VMS) such as hot flush, sweating in postmeopausal women.
2.Levi 2014	Neuropsychiatric symptoms in patients with probable dementia	These results support that intervention A, with its benign safety profile, can be used as first-line treatment of NPSD symptoms, 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.
3.Lam 2013	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.
4.Oosterbaan 2013	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.

5.Samus 2014	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.

Table 3. 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 to 24	median 16.0 IQR 11 to 24	median 16.0 IQR 11 to 24
Work place			
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)
Pediatrics	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	108 (38.4)	105 (36.7)	213 (37.6)

Background			
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)
5 or more	129 (45.9)	112 (39.2)	241 (42.5)

Clinical background data was available with participants who have sub-specialty certifications.

¶ multiple answers allowed

Abbreviation: OS: overstatement; IQR: interquartile range percentiles; PI: principle investigator

Table 4. 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

Fig 1. Flow diagram of participants

Fig 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

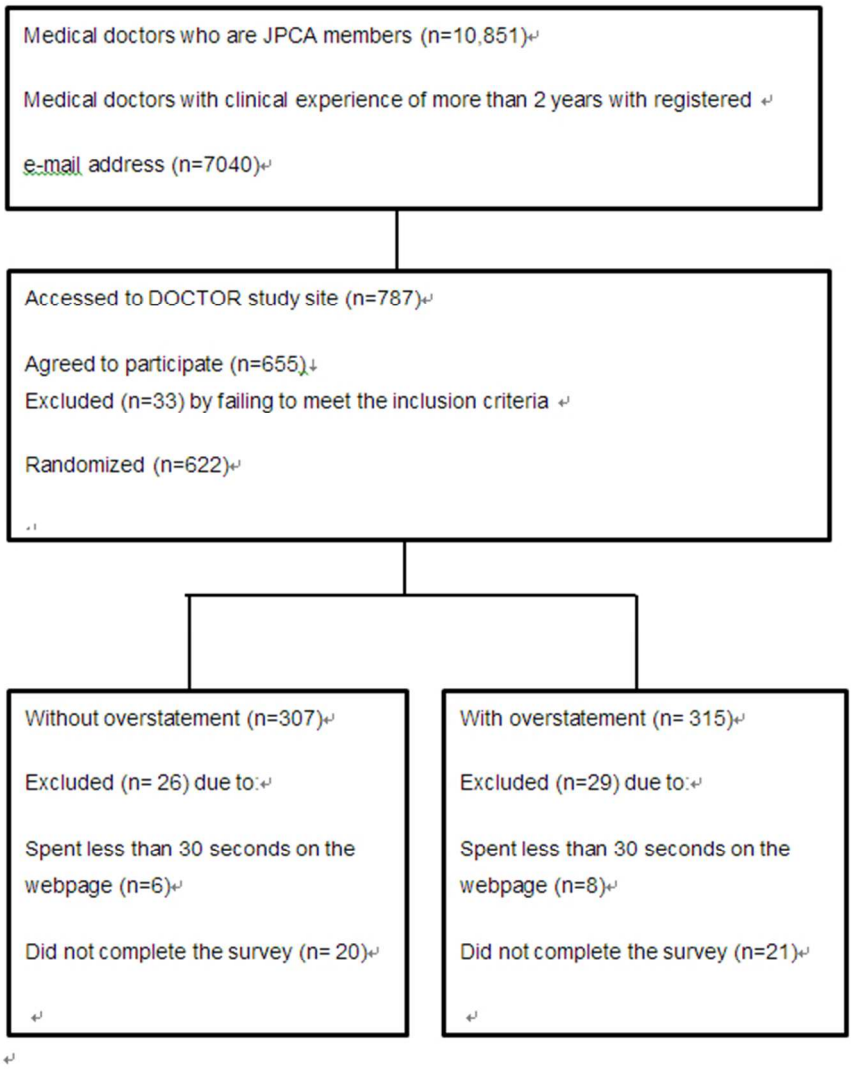


Fig 1. Flow diagram of participants

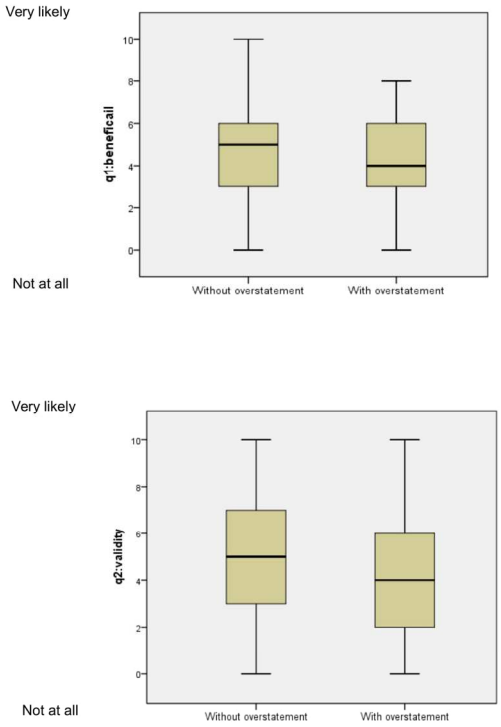


Fig 2. Evaluation of the beneficial effect and validity of the intervention discussed in the abstract. The answers to q1 "How beneficial do you think the intervention A is to patients?", and q2"How valid is this conclusion in your opinion?" given in a scale of 0 (not at all) to 10 (very likely). Boxes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

583x825mm (72 x 72 DPI)

CONTENTS OF SUPPLEMENTAL MATERIAL

Appendix1: Invitation e-mail

Subject: Win an Amazon gift card by participating in a 5-minute survey on EBM

This email is important.
My name is Morito Kise from Centre for Family Medicine Development, Japanese Health and Welfare Co-operative Federation, Tokyo, Japan

I am sending this email to invite you to participate in a clinical trial targeting clinicians. This research is a collaborative effort between Japan Primary Care Association (JPCA) and Kyoto University and aims to investigate the application of published articles among clinical practitioners. It is funded by Japan Primary Care Association, and has been approved by the board of committees.
For those JPCA members with more than three years of clinical experience, we would kindly ask you to read ONE abstract of a medical article and evaluate it on a scale of 0 to 10. The estimated time to complete the whole process is 5 minutes.

As a token of appreciation, we give away Amazon gift cards worth 3000 yen to 20 of the participants. The prize winners will be notified at the end of the survey.

▼▼▼Please click the link below to participate.▼▼▼
http://doctor-study.net/abstud y/public/base/index/0124B
It can be also accessed via your smartphone. The deadline is on the 31st of January, 2017.

This project investigates how clinical practitioners assess abstracts of scientific reports. It is funded by JPCA, and has been approved the Ethics Committee of Kyoto University. No personal particulars may be used to identify any individuals nor any results may be associated with particular individuals. The data obtained may be used, after blinding, for secondary research purposes. No information will be given to other organisations or individuals. Results of this investigation will be reported and published publicly but only after a blinding. Prize winners will be asked to provide their email and work addresses. The information will not be used for any other purposes. It is possible to drop out after you start.

Again, we would appreciate it greatly if you could give us your time for five minutes. Thank you for your cooperation

Morito Kise, MD
Centre for Family Medicine Development, Japanese Health and Welfare Co-operative Federation, Tokyo, Japan

Takuya Aoki, MD MMA
Department of Healthcare Epidemiology
Kyoto University Graduate School of Medicine / School of Public Health Morito Kise, MD

Kiyomi Shinohara, MD PhD
Department of Health Promotion and Human Behavior
Kyoto University Graduate School of Medicine / School of Public Health

Please contact:
doctor.study.pc@gmail.com

Kyoto University Graduate School of Medicine Administration Office Department of Conflict of Interest
(tel) 075-753-4305
(E-mail) 060rieki-sohan@mail2. adm.kyoto-u.ac.jp

Appendix 2 – Full text of five abstracts

We added the shaded part to the original abstract.

134

TITLE: Intervention A for menopausal symptoms: a randomized 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 flash, sweating, and poor circulation participated in a randomized controlled trial conducted in three sites: 106 women randomized to exercise and 142 women randomized 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 Generalized 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.

(Without OS)

Intervention A was not more effective than control B in terms of frequent vasomotor symptoms (VMS) such as hot flash, sweating in postmenopausal women.

Control B is the standard treatment for menopausal symptoms.

253

TITLE: Intervention A versus control B treatment of neuropsychiatric symptoms in patients with probable dementia: an open randomized trial

OBJECTIVES: to examine the effect of intervention A and control B on neuropsychiatric symptoms in dimentia (NPSD) and global function

METHODS: Using a randomised controlled and open-blind, once centre trial at an in-and outpatient clinic at a university hospital, we studied 100 adults with probablu dementia and NPSD. Participants received treatment A (N=50) or control B (N=50) for 12 weeks. The primary outcome was effects on NPSD, the difference between baseline and 12 weeks, assessed by the Neuropsychiatric Inventory (NPI). Secondary measures included the Mini-MEntal State Examination (MMSE), clinical demential rating, clinical global impressiona nd Simpson Angus scales. All tests were performed before and after treatment.

RESULTS: Outcome measures were analyzed using analysis of covariance. 91 patients (67% women, mean age 79+/- 7.5 years) with initial NPI score of 51 (+/- 25.8) and MMSE of 20.1 (+/- 4.6) completed the trial. Both intervention A and control B resulted in improved NPSD symptoms and were equally effective in treating several NPI domains (the differences at 12 weeks intervention A: 16.7±15.6, control B: 17.9±16.3, p=0.06). However, control B showed a significant treatment advantage in the NPI domains irritation and agitation, F(1, 97) = 5.2, p=0.02. Intervention A also ameliorated cognitive functions where MMSE scores increased 2.8 points compared with baseline (95% CI: 1.96-3.52). No treament-related severe side effects occurred.

CONCLUSION: These results support that intervention A, with its benign safety profile, can be used as first-line treatment of NPSD symptoms, unless symptoms of irritation and agitation are prominent, where control B is more efficient.

(Without OS)

Intervention A was not more effective than control B in terms of neuropsychiatric symptoms in patients with dementia.

Control B is a generally used antipsychotics.

1000382

TITLE: Effects of intervention A for improving work functioning in major depressive disorder

BACKGROUND: Major depressive disorder is associated with significant impairment in occupational functioning and reduced productivity, which represents a large part of the overall burden of depression.

AIMS: To examine symptom-based and work functioning outcomes with intervention A treatment of major depressive disorder.

METHOD: Employed patients with a DSM-IV diagnosis of major depressive disorder were treated with escitalopram 10-20 mg/day and randomized to intervention A (n = 48) or control B (n = 51). Primary outcome was the Montgomery-Asberg Depression Rating Scale (MADRS), administered by masked evaluators via telephone. Secondary outcome was self-rated work functioning scales completed online.

RESULTS: After 12 weeks, there were no significant between-group differences in change in MADRS score [effect size (Cohen's d) 0.16, P=0.60] or in response /remission (response: ≥ 50% improvement in MADRS scores, remission: MADRS ≤ 12). However, participants in intervention A had significantly greater improvement on some measures of work functioning than the control B.

CONCLUSIONS: Intervention A with escitalopram significantly improved some self-reported work functioning outcomes, but not symptom-based outcomes, compared with escitalopram and control B.

(Without OS)
Intervention A with escitalopram was not more effective than control B with escitalopram in terms of depressive symptoms in patients with major depression.

Control B is the standard treatment for depression.

1000385

TITLE: Intervention A v. control as usual for common mental disorders: 8-month, cluster randomized controlled trial

AIMS: To evaluate the effectiveness of intervention A in the treatment of common mental disorders.

METHOD: An 8-month cluster randomized controlled trial comparing intervention A to control B. Primary outcomes were the percentage of patients responding to and remitting on Clinical Global Impression of Improvement Scale (CGI-I) after treatment.

RESULTS: Twenty general practitioners (GPs) and 8 psychiatric nurses were randomised to provide intervention A or control B. The GPs recruited 163 patients [intervention A (n=94)、treatment B (n=64)] of whom 85% completed the post-test measurements. At 4-month mid-test intervention A was superior to control B: 74.7% (n = 68) v. 50.8% (n = 31) responders (P = 0.003). At 8-month post-test and 12-month follow-up no significant differences were found as the patients in control B group improved as well [response at 8-month: 80.2% (n = 73) vs. 67.2% (n = 41), *P*=0.072; remission at 8 month: 58.9% (n = 53) vs. 51.7% (n = 31), *P*=0.383].

CONCLUSIONS: Intervention A resulted in an earlier treatment response compared with control B.

(Without OS)
Intervention A was not more effective than control B in terms of treatment response or remission in patients with common mental illness.

Control B is the standard treatment for common mental illness.

135

TITLE: Intervention A for elders with memory disorders: the pilot randomized trial

OBJECTIES: To assess whether intervention A delays time to transition from home (to a hospital or nursing home) and reduces unmet needs in elders with memory disorders.

DESIGN: 18-month randomized controlled trial of 303 community-living elders.
SETTING: 28 postal code areas of Baltimore, MD.
PARTICIPANTS: Age 70+, with a cognitive disorder, community-living, English-speaking, and having a study partner available.

INTERVENTION: 18-month intervention A. Care monitoring by an interdisciplinary team.

MEASUREMENTS: Primary outcomes were time to transfer from home and total percent of unmet care needs at 18 months (measured on Johns Hopkins Dementia Care Needs Assessment).

RESULTS: Intervention participants had a significant delay in time to all-cause transition from home and the adjusted hazard of leaving the home was decreased by 37% (HR = 0.63, 95% CI 0.42 to 0.94) compared to control participants. While there was no significant group difference in reduction of total percent of unmet needs from baseline to 18 months (p=0.054), the intervention group had significant reductions in the proportion of unmet needs in safety and legal/advance care domains relative to controls. Participants in intervention A group had a significant improvement in self-reported quality of life (QOL) relative to control participants. No group differences were found in proxy- rated QOL, neuropsychiatric symptoms, or depression.

Conclusions—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.

(Without OS)
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.

Appendix 3-The results of each abstract

abstract	overstatement		q1	q2	q3
Sternfeld 2014	without OS (N=56)	average	2.98	5.64	3.54
		SD	2.385	2.393	2.703
	with OS (N=60)	average	3.4	3.12	2.78
		SD	1.942	2.415	2.300
Levi 2014	without OS (N=58)	average	4	4.09	3.47
		SD	2.362	2.430	2.773
	with OS (N=51)	average	3.94	3.29	3.18
		SD	2.240	2.452	2.613
Lam 2013	without OS (N=57)	average	4.37	4.54	3.3
		SD	2.143	2.646	2.464
	with OS (N=58)	average	3.97	4.36	3.19
		SD	2.255	2.375	2.806
Oosterbaan 2013	without OS (N=53)	average	3.96	4.58	3.02
		SD	2.038	2.365	2.162
	with OS (N=60)	average	3.92	3.53	3.4
		SD	2.149	2.174	2.402
Samus 2014	without OS (N=57)	average	5.56	5.37	4.26
		SD	1.711	1.789	2.489
	with OS (N=57)	average	5.3	5.07	4.53
		SD	1.861	1.850	2.726

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	P1
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	P2
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	P4
	2b	Specific objectives or hypotheses	P5
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	P6 Randomisation P5 Setting and design
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	NA
Participants	4a	Eligibility criteria for participants	P5 Participants and recruiting
	4b	Settings and locations where the data were collected	P5 Setting and design
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	P6 Selecting abstracts with overstatements P7 Constructing abstracts without overstatements
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	P7 Outcomes
	6b	Any changes to trial outcomes after the trial commenced, with reasons	P8 Statistical analysis
Sample size	7a	How sample size was determined	P8 Sample size

	7b	When applicable, explanation of any interim analyses and stopping guidelines	NA
Randomisation:			
Sequence	8a	Method used to generate the random allocation sequence	P6 Randomisation
generation	8b	Type of randomisation; details of any restriction (such as blocking and block size)	P6 Randomisation
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	P6 Randomisation
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	P6 Randomisation
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	P6 Blinding P8 Blinded data interpretation
	11b	If relevant, description of the similarity of interventions	P7 Constructing abstracts without overstatements
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	P8 Statistical analysis
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	P8 Statistical analysis
Results			
Participant flow (a diagram is strongly	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	P9 Results, Fig1
recommended)	13b	For each group, losses and exclusions after randomisation, together with reasons	P9 Results, Fig1
Recruitment	14a	Dates defining the periods of recruitment and follow-up	P5 Setting and design
	14b	Why the trial ended or was stopped	P9 Results
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	Table 3
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	Fig1
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	P9 Primary outcome, Secondary outcomes and subgroup analyses, Table 4

	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	NA
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	P9 Secondary outcomes and subgroup analyses
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	NA
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	P11 Limitations and strengths
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	P11 Limitations and strengths
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	P10 Discussion
Other information			
Registration	23	Registration number and name of trial registry	P1
Protocol	24	Where the full trial protocol can be accessed, if available	NA
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	P12 Funding

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.

BMJ Open

The influence of overstated abstract conclusions on clinicians: A web-based randomized controlled trial

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-018355.R2
Article Type:	Research
Date Submitted by the Author:	12-Oct-2017
Complete List of Authors:	Shinohara, Kiyomi; Kyoto University Graduate School of Medicine/Schoo of Public Health, Health Promotion and Human Behaviour Aoki, Takuya; Kyoto University Graduate School of Medicine / School of Public Health, Department of Healthcare Epidemiology So, Ryuhei ; Kyoto University Graduate School of Medicine / School of Public Health, Department of Health Promotion and Human Behavior Tsujimoto, Yasushi ; Kyoto University Graduate School of Medicine/Schoo of Public Health, Department of Healthcare Epidemiology Suganuma, Aya M; Kyoto University Graduate School of Medicine/Schoo of Public Health, Health Promotion and Human Behaviour Kise, Morito; Centre for Family Medicine Development, Japanese Health and Welfare Co-operative Federation Furukawa, Toshi; Kyoto University, Graduate School of Medicine and School of Public Health
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Evidence based practice
Keywords:	randomised controlled trials, general practice, overstatements, primary care physicians, reporting bias, clinical trial

Title:

The influence of overstated abstract conclusions on clinicians: A web-based randomized controlled trial

Authors:

Kiyomi Shinohara[a],
Takuya Aoki [b],
Ryuhei So[a],
Yasushi Tsujimoto[b],
Aya M Suganuma [a],
Morito Kise[c],
Toshi A. Furukawa[a]

Affiliation of all authors:

a: Department of Health Promotion and Human Behavior Kyoto University Graduate School of Medicine / School of Public Health
Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501 Japan,
b: Department of Healthcare Epidemiology Kyoto University Graduate School of Medicine / School of Public Health
Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501 Japan
c: Centre for Family Medicine Development, Japanese Health and Welfare Co-operative Federation, Tokyo, Japan
Sanken Building 4th Floor
3-25-1, Hyakunin-cho, Shinjuku-ku
Tokyo Japan

Corresponding author:

Kiyomi Shinohara
Department of Health Promotion and Human Behavior
Kyoto University Graduate School of Medicine / School of Public Health
Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501 Japan
Phone: +81-75-753-9491
Fax: +81-75-753-4641
E-mail address: 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: Randomised controlled trial

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 e-mail 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 section were standardized 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% confidence interval [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

Strength and limitations of this study

- This is the first and only RCT 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 generalizability 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 since 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-randomized 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 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 e-mail 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). 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 Japan Primary Care Association (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 5,836 out of a total of 10851 members are certified as specialists in primary care.

We sent e-mail invitations to JPCA members who had more than two years of clinical experience with registered e-mail addresses. (The details of the recruiting process will be reported in a separate paper). We excluded clinicians with less than two years' experience because our target population was primary care physicians, and doctors usually choose their specialty after two years of clinical training in Japan. Interested individuals could access the DOCTOR study website via the link in the e-mail. 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 (26.6 US dollars) was given to 20 drawing winners.

The inclusion criteria for participants were as follows: JPCA member, medical doctor currently in clinical practice, more than two years of clinical practice experience, and access to up-to-date clinical research knowledge. We asked how respondents learned 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 e-mail 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 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 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' sub-specialty 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 pre-specified 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 re-wrote 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 Table 1; and all the abstract conclusions are in Table 2.)

Second, we standardized the methods and results sections. We explicitly stated the primary outcomes and results (for example, odds ratio, risk ratio, confidence interval, p-value) from the text if they were not stated in the original abstract. Therefore, all

abstracts had necessary 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 original abstract. Additionally, this standardization 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 minimize bias. We added a few words for explanation when there was a medical term that seemed unfamiliar to primary care physicians (e.g. vasomotor symptoms [(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 enroll 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 seconds 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 pre-specified 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 Jarvinen 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.

Ethics

This study was approved by the Ethics Committee of Kyoto University Graduate School of Medicine and was conducted in accordance with the Declaration of Helsinki.

We obtained an online consent for participation from each participant.

Results

We sent e-mail invitations to 7,040 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). Supplementary appendix 3 provides further breakdown per abstract.

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

Table 3 shows the participant characteristics; 76.5% were certified as primary care physicians. We classified their sub-specialties 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 4).

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) (Fig. 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 to read the full text. In both groups, the majority of the doctors referred to the results section to make

an assessment

We conducted sub-group 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 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% confidence interval 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. On the other hand, 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 sub-group 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 for 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]; 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 [29]. 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 generalizability of our findings. Two things should be noted. Firstly, 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 organization 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 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 overgeneralize 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 judgment 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.

Acknowledgement

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

Contributors

All authors (KS, TA, RS, YT, AMS, MK and TAF) of the paper have contributed 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 web-site. 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.

Conflicts of interest

KS has received grant from Japan Primary Care Association. 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. AMS, TA, YT, MK, and RS reported no competing interests.

Data sharing statement

No additional data available.

1. Burke DT, Judelson AL, Schneider JC, DeVito MC, Latta D. Reading habits of practicing physiatrists. Am J Phys Med Rehabil 2002;**81**(10):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**(1):e20.
3. Hisham R, Ng CJ, Liew SM, Hamzah N, Ho GJ. Why is there variation in the practice of evidence-based medicine in primary care? A qualitative study. BMJ Open 2016;**6**(3):e010565.
4. Zwolsman S, te Pas E, Hooft L, Wieringa-de Waard M, van Dijk N. Barriers to GPs' use of evidence-based medicine: a systematic review. Br J Gen Pract 2012;**62**(600):e511-21.
5. Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. JAMA 2010;**303**(20):2058-64.
6. Suganuma AM, Shinohara K, Imai H, Takeshima N, Hayasaka Y, Furukawa TA. Overstatements in abstract conclusions claiming effectiveness of interventions in psychiatry: a study protocol for a meta-epidemiological investigation. BMJ Open 2016;**6**(4):e009832.
7. Lazarus C, Haneef R, Ravaud P, Boutron I. 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, Ravaud P. Misleading abstract conclusions in randomized controlled trials in rheumatology: comparison of the abstract conclusions and the results section. Joint Bone Spine 2012;**79**(3):262-7.
9. Shinohara K, Suganuma AM, Imai H, Takeshima N, Hayasaka Y, Furukawa TA. Overstatements in abstract conclusions claiming effectiveness of interventions in psychiatry: A meta-epidemiological investigation. PLoS One 2017;**12**(9):e0184786.
10. Lockyer S, Hodgson R, Dumville JC, Cullum N. "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 2016.
12. Fleming PS. Evidence of spin in clinical trials in the surgical literature. Ann Transl Med 2016;**4**(19):385.

13. Vera-Badillo FE, Shapiro R, Ocana A, Amir E, Tannock IF. Bias in reporting of end points of efficacy and toxicity in randomized, clinical trials for women with breast cancer. Ann Oncol 2013;**24**(5):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**(7):1949-50.
16. Boutron I, Altman DG, Hopewell S, Vera-Badillo F, Tannock I, Ravaud P. 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**(36):4120-6.
17. Japan Primary Care Association. https://www.primary-care.or.jp/jpca_eng/index.html.
18. Aoki T, Inoue M, Nakayama T. Development and validation of the Japanese version of Primary Care Assessment Tool. Fam Pract 2016;**33**(1):112-17.
19. Jarvinen TL, Sihvonen R, Bhandari M, et al. Blinded interpretation of study results can feasibly and effectively diminish interpretation bias. J Clin Epidemiol 2014;**67**(7):769-72.
20. Sternfeld B, Guthrie KA, Ensrud KE, et al. Efficacy of exercise for menopausal symptoms: a randomized controlled trial. Menopause 2014;**21**(4):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**(4):398-414.
22. Freund-Levi Y, Jedenius E, Tysen-Bäckström AC, Lärksäter M, Wahlund LO, Eriksdotter M. Galantamine versus risperidone treatment of neuropsychiatric symptoms in patients with probable dementia: an open randomized trial. The American journal of geriatric psychiatry 2014;**22**(4):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**(2):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. The British Journal of Psychiatry 2013;**203**(5):358-65.
25. Shinohara K, Tajika A, Imai H, Takeshima N, Hayasaka Y, Furukawa TA. Protocol registration and selective outcome reporting in recent psychiatry trials: new antidepressants and cognitive behavioural therapies. Acta Psychiatr Scand

2015;**132**(6):489-98.

26. Cohen J. *Statistical Power Analysis in the Behavioral Sciences* . Erlbaum, 1988.

27. DeJong C, Aguilar T, Tseng C, Lin GA, Boscardin W, Dudley R. PHarmaceutical industry–sponsored meals and physician prescribing patterns for medicare beneficiaries. JAMA Internal Medicine 2016;**176**(8):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, Munafo MR, Bastiaansen JA. 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**(14):2971-79.

Table1. An example of the abstracts (underlined where extra text added, double underlined where changed in the 'without overstatement' group)

TITLE

Intervention A for menopausal symptoms: a randomized 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 randomized controlled trial conducted in three sites: 106 women randomized to exercise and 142 women randomized 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 Generalized 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 vasomotor symptoms (VMS) such as hot flush, sweating in postmenopausal women.

Control B is the standard treatment for menopausal symptoms.

Table2. 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)
1.Sternfeld 2014	Menopausal Symptoms	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.	Intervention A was not more effective than control B in terms of frequent vasomotor symptoms (VMS) such as hot flush, sweating in postmeopausal women.
2.Levi 2014	Neuropsychiatric symptoms in patients with probable dementia	These results support that intervention A, with its benign safety profile, can be used as first-line treatment of NPSD symptoms, 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.
3.Lam 2013	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.
4.Oosterbaan 2013	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.

5.Samus 2014	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.

Table 3. 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 to 24	median 16.0 IQR 11 to 24	median 16.0 IQR 11 to 24
Work place			
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)
Pediatrics	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	108 (38.4)	105 (36.7)	213 (37.6)

Background			
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)
5 or more	129 (45.9)	112 (39.2)	241 (42.5)

Clinical background data was available with participants who have sub-specialty certifications.

¶ multiple answers allowed

Abbreviation: OS: overstatement; IQR: interquartile range percentiles; PI: principle investigator

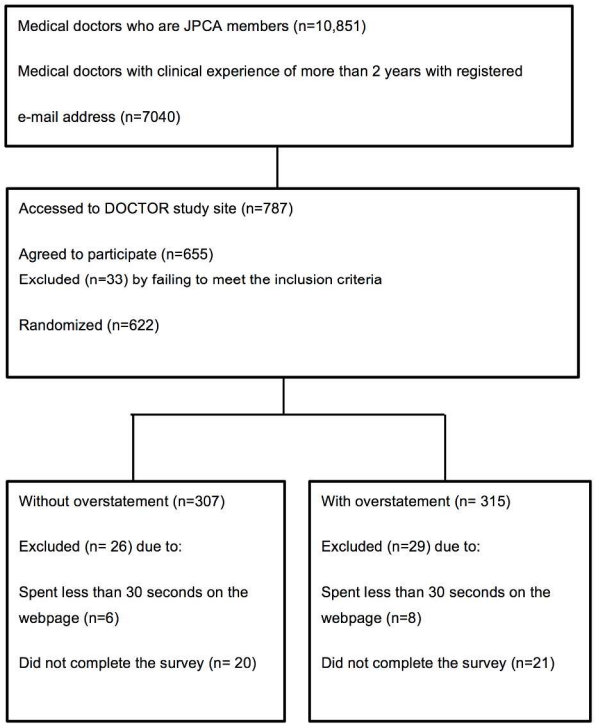
Table 4. Impression of the abstract

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

Fig 1. Flow diagram of participants

Fig 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



1

Fig1. Flow diagram of participants

209x296mm (300 x 300 DPI)

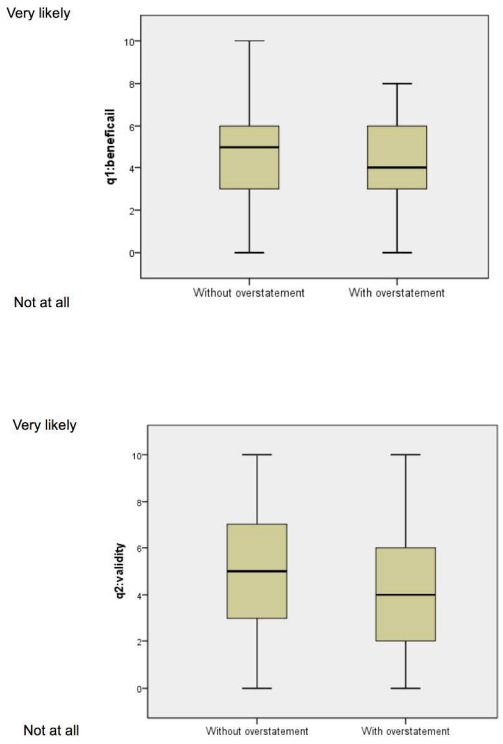


Fig 2. Evaluation of the beneficial effect and validity of the intervention discussed in the abstract. The answers to q1 "How beneficial do you think the intervention A is to patients?", and q2"How valid is this conclusion in your opinion?" given in a scale of 0 (not at all) to 10 (very likely). Boxes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

209x296mm (300 x 300 DPI)

CONTENTS OF SUPPLEMENTAL MATERIAL

Appendix1: Invitation e-mail

Subject: Win an Amazon gift card by participating in a 5-minute survey on EBM

This email is important.
My name is Morito Kise from Centre for Family Medicine Development, Japanese Health and Welfare Co-operative Federation, Tokyo, Japan

I am sending this email to invite you to participate in a clinical trial targeting clinicians. This research is a collaborative effort between Japan Primary Care Association (JPCA) and Kyoto University and aims to investigate the application of published articles among clinical practitioners. It is funded by Japan Primary Care Association, and has been approved by the board of committees.
For those JPCA members with more than three years of clinical experience, we would kindly ask you to read ONE abstract of a medical article and evaluate it on a scale of 0 to 10. The estimated time to complete the whole process is 5 minutes.

As a token of appreciation, we give away Amazon gift cards worth 3000 yen to 20 of the participants. The prize winners will be notified at the end of the survey.

▼▼▼Please click the link below to participate.▼▼▼
http://doctor-study.net/abstud y/public/base/index/0124B
It can be also accessed via your smartphone. The deadline is on the 31st of January, 2017.

This project investigates how clinical practitioners assess abstracts of scientific reports. It is funded by JPCA, and has been approved the Ethics Committee of Kyoto University. No personal particulars may be used to identify any individuals nor any results may be associated with particular individuals. The data obtained may be used, after blinding, for secondary research purposes. No information will be given to other organisations or individuals. Results of this investigation will be reported and published publicly but only after a blinding. Prize winners will be asked to provide their email and work addresses. The information will not be used for any other purposes. It is possible to drop out after you start.

Again, we would appreciate it greatly if you could give us your time for five minutes. Thank you for your cooperation

Morito Kise, MD
Centre for Family Medicine Development, Japanese Health and Welfare Co-operative Federation, Tokyo, Japan

Takuya Aoki, MD MMA
Department of Healthcare Epidemiology
Kyoto University Graduate School of Medicine / School of Public Health Morito Kise, MD

Kiyomi Shinohara, MD PhD
Department of Health Promotion and Human Behavior
Kyoto University Graduate School of Medicine / School of Public Health

Please contact:
doctor.study.pc@gmail.com

Kyoto University Graduate School of Medicine Administration Office Department of Conflict of Interest
(tel) 075-753-4305
(E-mail) 060rieki-sohan@mail2. adm.kyoto-u.ac.jp

Appendix 2 – Full text of five abstracts

We added the shaded part to the original abstract.

134

TITLE: Intervention A for menopausal symptoms: a randomized 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 flash, sweating, and poor circulation participated in a randomized controlled trial conducted in three sites: 106 women randomized to exercise and 142 women randomized 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 Generalized 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.

(Without OS)

Intervention A was not more effective than control B in terms of frequent vasomotor symptoms (VMS) such as hot flash, sweating in postmenopausal women.

Control B is the standard treatment for menopausal symptoms.

253

TITLE: Intervention A versus control B treatment of neuropsychiatric symptoms in patients with probable dementia: an open randomized trial

OBJECTIVES: to examine the effect of intervention A and control B on neuropsychiatric symptoms in dimentia (NPSD) and global function

METHODS: Using a randomised controlled and open-blind, once centre trial at an in-and outpatient clinic at a university hospital, we studied 100 adults with probablu dementia and NPSD. Participants received treatment A (N=50) or control B (N=50) for 12 weeks. The primary outcome was effects on NPSD, the difference between baseline and 12 weeks, assessed by the Neuropsychiatric Inventory (NPI). Secondary measures included the Mini-MEntal State Examination (MMSE), clinical demential rating, clinical global impressiona nd Simpson Angus scales. All tests were performed before and after treatment.

RESULTS: Outcome measures were analyzed using analysis of covariance. 91 patients (67% women, mean age 79+/- 7.5 years) with initial NPI score of 51 (+/- 25.8) and MMSE of 20.1 (+/- 4.6) completed the trial. Both intervention A and control B resulted in improved NPSD symptoms and were equally effective in treating several NPI domains (the differences at 12 weeks intervention A: 16.7±15.6, control B: 17.9±16.3, p=0.06). However, control B showed a significant treatment advantage in the NPI domains irritation and agitation, F(1, 97) = 5.2, p=0.02. Intervention A also ameliorated cognitive functions where MMSE scores increased 2.8 points compared with baseline (95% CI: 1.96-3.52). No treament-related severe side effects occurred.

CONCLUSION: These results support that intervention A, with its benign safety profile, can be used as first-line treatment of NPSD symptoms, unless symptoms of irritation and agitation are prominent, where control B is more efficient.

(Without OS)

Intervention A was not more effective than control B in terms of neuropsychiatric symptoms in patients with dementia.

Control B is a generally used antipsychotics.

1000382

TITLE: Effects of intervention A for improving work functioning in major depressive disorder

BACKGROUND: Major depressive disorder is associated with significant impairment in occupational functioning and reduced productivity, which represents a large part of the overall burden of depression.

AIMS: To examine symptom-based and work functioning outcomes with intervention A treatment of major depressive disorder.

METHOD: Employed patients with a DSM-IV diagnosis of major depressive disorder were treated with escitalopram 10-20 mg/day and randomized to intervention A (n = 48) or control B (n = 51). Primary outcome was the Montgomery-Asberg Depression Rating Scale (MADRS), administered by masked evaluators via telephone. Secondary outcome was self-rated work functioning scales completed online.

RESULTS: After 12 weeks, there were no significant between-group differences in change in MADRS score [effect size (Cohen's d) 0.16, P=0.60] or in response /remission (response: ≥ 50% improvement in MADRS scores, remission: MADRS ≤ 12). However, participants in intervention A had significantly greater improvement on some measures of work functioning than the control B.

CONCLUSIONS: Intervention A with escitalopram significantly improved some self-reported work functioning outcomes, but not symptom-based outcomes, compared with escitalopram and control B.

(Without OS)
Intervention A with escitalopram was not more effective than control B with escitalopram in terms of depressive symptoms in patients with major depression.

Control B is the standard treatment for depression.

1000385

TITLE: Intervention A v. control as usual for common mental disorders: 8-month, cluster randomized controlled trial

AIMS: To evaluate the effectiveness of intervention A in the treatment of common mental disorders.

METHOD: An 8-month cluster randomized controlled trial comparing intervention A to control B. Primary outcomes were the percentage of patients responding to and remitting on Clinical Global Impression of Improvement Scale (CGI-I) after treatment.

RESULTS: Twenty general practitioners (GPs) and 8 psychiatric nurses were randomised to provide intervention A or control B. The GPs recruited 163 patients [intervention A (n=94)、treatment B (n=64)] of whom 85% completed the post-test measurements. At 4-month mid-test intervention A was superior to control B: 74.7% (n = 68) v. 50.8% (n = 31) responders (P = 0.003). At 8-month post-test and 12-month follow-up no significant differences were found as the patients in control B group improved as well [response at 8-month: 80.2% (n = 73) vs. 67.2% (n = 41), *P*=0.072; remission at 8 month: 58.9% (n = 53) vs. 51.7% (n = 31), *P*=0.383].

CONCLUSIONS: Intervention A resulted in an earlier treatment response compared with control B.

(Without OS)
Intervention A was not more effective than control B in terms of treatment response or remission in patients with common mental illness.

Control B is the standard treatment for common mental illness.

135

TITLE: Intervention A for elders with memory disorders: the pilot randomized trial

OBJECTIES: To assess whether intervention A delays time to transition from home (to a hospital or nursing home) and reduces unmet needs in elders with memory disorders.

DESIGN: 18-month randomized controlled trial of 303 community-living elders.
SETTING: 28 postal code areas of Baltimore, MD.
PARTICIPANTS: Age 70+, with a cognitive disorder, community-living, English-speaking, and having a study partner available.

INTERVENTION: 18-month intervention A. Care monitoring by an interdisciplinary team.

MEASUREMENTS: Primary outcomes were time to transfer from home and total percent of unmet care needs at 18 months (measured on Johns Hopkins Dementia Care Needs Assessment).

RESULTS: Intervention participants had a significant delay in time to all-cause transition from home and the adjusted hazard of leaving the home was decreased by 37% (HR = 0.63, 95% CI 0.42 to 0.94) compared to control participants. While there was no significant group difference in reduction of total percent of unmet needs from baseline to 18 months (p=0.054), the intervention group had significant reductions in the proportion of unmet needs in safety and legal/advance care domains relative to controls. Participants in intervention A group had a significant improvement in self-reported quality of life (QOL) relative to control participants. No group differences were found in proxy- rated QOL, neuropsychiatric symptoms, or depression.

Conclusions—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.

(Without OS)
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.

Appendix 3-The results of each abstract

abstract	overstatement		q1	q2	q3
Sternfeld 2014	without OS (N=56)	average	2.98	5.64	3.54
		SD	2.385	2.393	2.703
	with OS (N=60)	average	3.4	3.12	2.78
		SD	1.942	2.415	2.300
Levi 2014	without OS (N=58)	average	4	4.09	3.47
		SD	2.362	2.430	2.773
	with OS (N=51)	average	3.94	3.29	3.18
		SD	2.240	2.452	2.613
Lam 2013	without OS (N=57)	average	4.37	4.54	3.3
		SD	2.143	2.646	2.464
	with OS (N=58)	average	3.97	4.36	3.19
		SD	2.255	2.375	2.806
Oosterbaan 2013	without OS (N=53)	average	3.96	4.58	3.02
		SD	2.038	2.365	2.162
	with OS (N=60)	average	3.92	3.53	3.4
		SD	2.149	2.174	2.402
Samus 2014	without OS (N=57)	average	5.56	5.37	4.26
		SD	1.711	1.789	2.489
	with OS (N=57)	average	5.3	5.07	4.53
		SD	1.861	1.850	2.726

CONSORT 2010 checklist of information to include when reporting a randomised trial*

Section/Topic	Item No	Checklist item	Reported on page No
Title and abstract			
	1a	Identification as a randomised trial in the title	P1
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	P2
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	P4
	2b	Specific objectives or hypotheses	P5
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	P6 Randomisation P5 Setting and design
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	NA
Participants	4a	Eligibility criteria for participants	P5 Participants and recruiting
	4b	Settings and locations where the data were collected	P5 Setting and design
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	P6 Selecting abstracts with overstatements P7 Constructing abstracts without overstatements
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	P7 Outcomes
	6b	Any changes to trial outcomes after the trial commenced, with reasons	P8 Statistical analysis
Sample size	7a	How sample size was determined	P8 Sample size

	7b	When applicable, explanation of any interim analyses and stopping guidelines	NA
Randomisation:			
Sequence	8a	Method used to generate the random allocation sequence	P6 Randomisation
generation	8b	Type of randomisation; details of any restriction (such as blocking and block size)	P6 Randomisation
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	P6 Randomisation
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	P6 Randomisation
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	P6 Blinding P8 Blinded data interpretation
	11b	If relevant, description of the similarity of interventions	P7 Constructing abstracts without overstatements
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	P8 Statistical analysis
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	P8 Statistical analysis
Results			
Participant flow (a diagram is strongly	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	P9 Results, Fig1
recommended)	13b	For each group, losses and exclusions after randomisation, together with reasons	P9 Results, Fig1
Recruitment	14a	Dates defining the periods of recruitment and follow-up	P5 Setting and design
	14b	Why the trial ended or was stopped	P9 Results
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	Table 3
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	Fig1
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	P9 Primary outcome, Secondary outcomes and subgroup analyses, Table 4

	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	NA
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	P9 Secondary outcomes and subgroup analyses
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)	NA
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	P11 Limitations and strengths
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	P11 Limitations and strengths
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	P10 Discussion
Other information			
Registration	23	Registration number and name of trial registry	P1
Protocol	24	Where the full trial protocol can be accessed, if available	NA
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	P12 Funding

*We strongly recommend reading this statement in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items. If relevant, we also recommend reading CONSORT extensions for cluster randomised trials, non-inferiority and equivalence trials, non-pharmacological treatments, herbal interventions, and pragmatic trials. Additional extensions are forthcoming: for those and for up to date references relevant to this checklist, see www.consort-statement.org.