



The over-representation of significant p values in abstracts compared to corresponding full texts: A systematic review of surgical randomized trials



Yusuf Assem^{a,c,*}, Sam Adie^{a,b,c}, Jason Tang^{a,c}, Ian A. Harris^{a,b,c}

^a University of New South Wales, South Western Sydney Clinical School, Liverpool Hospital, Liverpool, Australia

^b Whitlam Orthopaedic Research Centre, Ingham Institute for Applied Medical Research, Liverpool, Australia

^c South West Sydney Local Health District, Liverpool Hospital, Liverpool, Australia

ARTICLE INFO

Keywords:

P-Values

Significance

Bias

Spin

Surgical

Abstracts and statistical significance

ABSTRACT

Background: Abstracts are often the only read summaries of research findings, and it is essential that they accurately represent of the contents of the full text of the randomised control trial (RCT). We investigated whether outcomes in surgical trials were selectively reported in abstracts based on their statistical significance.

Objective: To compare the proportion of significant p-values reported in abstracts to their corresponding full texts in surgical RCTs.

Method: A Meta-analysis of 350 full text RCTs conducted on humans that compared a surgical intervention to any other intervention. An electronic search of MEDLINE, EMBASE, and the Cochrane Central Register of Controlled Trials (CENTRAL) was conducted. All outcomes were extracted from the abstract and the full text. Frequency histograms were used to plot the distribution of numerically reported p-values across the statistical significance spectrum. For each RCT, a 2×2 table was populated with that trial's outcomes and whether the outcome was statistically significant ($p < 0.05$). From each 2×2 table, an odds ratio (OR) was calculated describing the association between statistical significance, and reporting in the abstract. ORs were pooled in random effects meta-analysis for an overall estimate of the association.

Results: A total of 8258 reported outcomes were included. Outcomes reported in a surgical RCT abstract had three times the odds of being significant when compared to the corresponding full text (OR = 3.0, 95% confidence interval 2.5–3.6, $p < 0.001$). This finding was consistent and not subject to heterogeneity ($I^2 = 0\%$). Both histograms demonstrated a large drop in the frequency of reported p values between 0.04 and 0.05, and after the 0.06 thresholds.

Conclusions: Data presented in abstracts is biased to statistically significant outcomes. Clinicians and policy makers should do not rely solely on information presented in abstracts for their decision-making.

1. Introduction

Technological innovation and digital access to information has introduced an overwhelming volume of published studies in every medical discipline, in a growing number of online journals [1]. It is impractical for clinicians to keep up to date with the literature, even in their own subspecialty. This phenomenon is compounded, as half of scientific publications are limited to paid subscriptions, while abstracts are easily accessible free online via websites such as PubMed and Google Scholar [1–4]. Abstracts have become an appealing de facto source of evidence, digestible summaries with key information they are easily accessible at the point of care via mobile phones, and utilized to answer clinical questions [4]. Today abstracts are the most widely read, and often the only read, summaries of research findings [3,5].

In the scientific literature randomized controlled trials (RCTs) are the ‘gold standard’ in evaluating therapeutic interventions and thus are relied upon for the practice of evidenced based medical treatment [1]. Given abstracts are often relied upon, it is essential that they provide a concise and accurate representation of the contents of the full text of the RCT [1,4–6].

Several factors have been described in the literature to potentially contribute to inadequate reporting of clinical trial results in abstracts, including journal space constraints, an attempt to convey the most ‘clinically relevant’ results, selective reporting bias and ‘spin’ [2,5]. Boutroun et al. [7] defined spin as “use of specific reporting strategies, to highlight that the experimental treatment is beneficial, despite a statistically nonsignificant difference for the primary outcome.” Strategies include the authors’ use of language, selective reporting

* Corresponding author. University of New South Wales, South Western Sydney Clinical School, Liverpool Hospital, Liverpool, Australia.

E-mail addresses: yassem16@gmail.com (Y. Assem), sam.adie@gmail.com (S. Adie), Jason_tang1992@hotmail.com (J. Tang), iaharris1@gmail.com (I.A. Harris).

<http://dx.doi.org/10.1016/j.conctc.2017.07.007>

Received 3 January 2017; Received in revised form 2 July 2017; Accepted 20 July 2017

Available online 28 July 2017

2451-8654/ Crown Copyright © 2017 Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

emphasizing particular outcome results and omission, which have been described to exaggerate the effect of interventions in abstracts [3,4,8]. Additionally, it has been reported that in some high impact factor journals, abstracts often failed to report harm, despite reporting them in the full text [2,4,5]. Finally, authors may attempt to convey to the reader with the limited word count parameters, the results perceived to be of significant clinical difference, that may affect practice, encouraging them to read the entire article.

The primary objective of this study was to compare the proportion of significant p-values reported in abstracts, compared to their corresponding full texts in surgical RCTs. A secondary objective was to analyze the trends and impact of significance thresholds on the reporting of p values across the statistical significance spectrum in abstracts and full texts.

2. Method

2.1. Study design

We conducted a systematic review and meta-analysis of randomized trials that assessed a surgical intervention. This study was performed using data collected for a doctoral thesis. The protocol for the thesis was pre-approved and is available from the authors.

2.2. Inclusion of studies for the review

To be eligible, a study met the following criteria:

- i) A randomized controlled trial.
- ii) Published as a full text article in English. Studies published as abstracts or conference proceedings were excluded.
- iv) The primary/earliest publication from an investigation.
- v) Conducted on humans (not cadavers)
- vi) Compared a surgical intervention to any other intervention.

We defined a surgical intervention as any procedure that requires surgical training and performed by a surgeon of any subspecialty recognized by the Royal Australasian College of Surgeons. This included upper and lower gastrointestinal, transplant, cardiothoracic, neuro, ear nose and throat, paediatric, plastic and reconstructive, urology, vascular and orthopaedic surgery.

2.3. Electronic search strategy

A search on MEDLINE, EMBASE, and the Cochrane Central Register of Controlled Trials (CENTRAL) was executed in May 2009. Little is known about the appropriate sample size calculation for our hypotheses, but we aimed to include 350 trials to be comparable to similar methodological studies [2,5,9]. Trials were selected working backwards from May 2009, until the required number were included. The electronic search strategy was formulated in collaboration with two medical librarians and contained two filters. The first, a randomized trial filter, was based on the Cochrane highly sensitive search strategies for MEDLINE (Phase 1) and EMBASE. The second, a surgery filter, aimed to retrieve all studies of relevance to the surgical specialties of interest. The syntax of the search strategy is included in additional file 1, PRISMA checklist attached additional file 2.

2.4. Study identification method

Titles and abstracts of retrieved records were assessed according to the above inclusion criteria. Full texts of abstracts that appeared to meet the eligibility criteria were retrieved, and eligibility was assessed using the same process as for abstracts. Study identification methods were piloted by two authors to resolve any issues with the interpretation of the eligibility criteria and resulted in almost perfect agreement

among the two assessors ($n = 1000$, $\kappa = 0.85$, 95% confidence interval 0.77–0.93). Thereafter, one author, in an identical process, performed study identification.

2.5. Data extraction

Each trial's abstract and full text outcomes were extracted and the reported statistical significance was recorded. Statistical significance was determined by the reported p value or 95% confidence interval for each outcome. When these were not reported but a standard error or standard deviation was available, a p value was calculated. Outcomes that were not reported with any data regarding their statistical significance were unable to be included in these analyses.

2.6. Primary analysis

A meta-analysis was performed comparing the proportion of significant outcomes in abstracts vs. full text. For each trial, a contingency (2×2) table was populated with that trial's outcomes, describing whether each outcome was statistically significant vs. non-significant (contingency table columns), for the abstract and corresponding full text (contingency table rows). An exact p value was not required to determine statistical significance, such that when outcomes were reported as " $p < 0.05$ " or " $p > 0.05$ ", these were regarded as significant and non-significant, respectively. If the contingency table contained a single zero cell, or two diagonal zero cells, 0.5 was added to all four cells as per the default of standard meta-analysis statistical packages [10].

For each trial's 2×2 table, an odds ratio was calculated describing the association between reporting in the abstract, and statistical significance. Calculations were such that an odds ratio greater than one meant that a statistically significant outcome had a higher odds of being found in an abstract compared to the full text [11]. Odds ratios were then combined in random effects meta-analysis and a summary odds ratio (along with its 95% confidence interval and I^2 as a measure of heterogeneity) was calculated as an overall indicator of p value reporting in surgical RCTs.

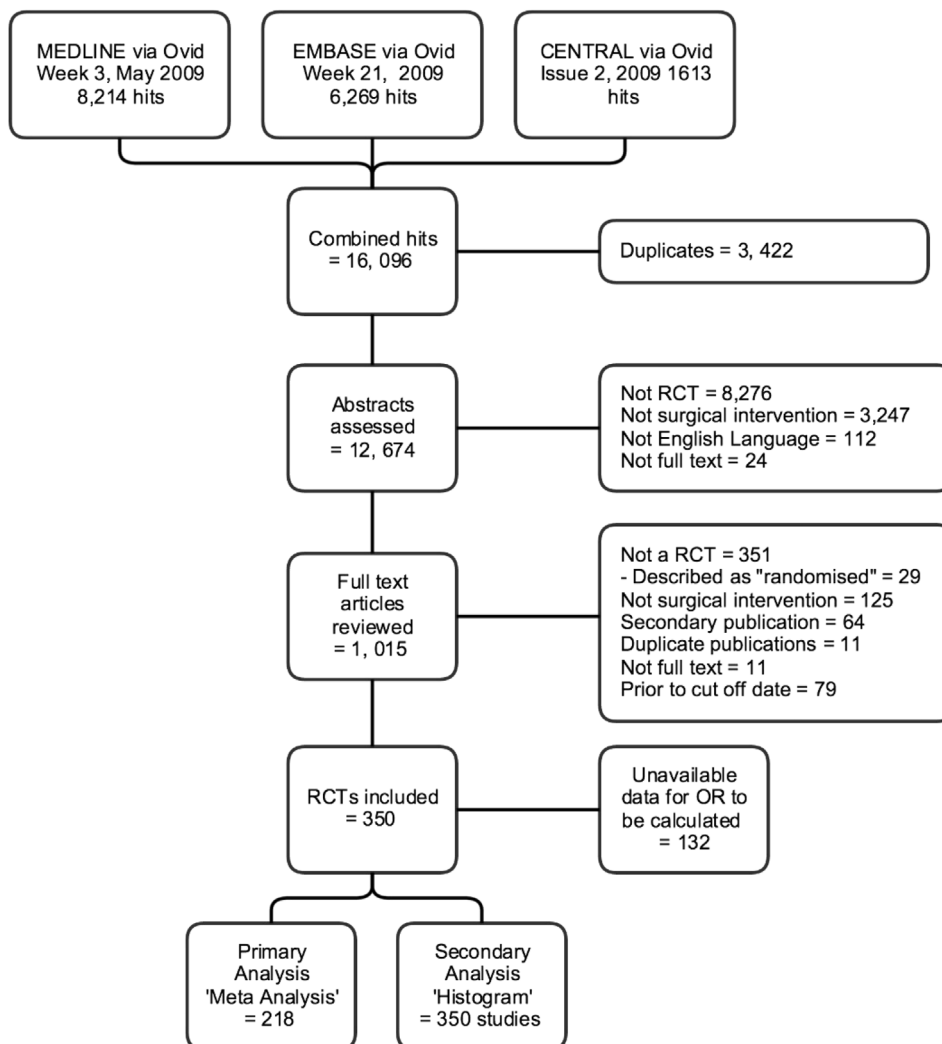
When a whole row or column in a trial's 2×2 table contained zero cells, then an odds ratio was incalculable for that trial and it was excluded from this analysis. This would occur when all the outcomes in a trial were non-significant, all the outcomes were significant, or when no outcomes were reported in either the abstract or the full text. This was unlikely to introduce a directional bias, but may have reduced the precision of our results since few studies were included [11,12].

A sensitivity analysis was also performed in order to assess whether the findings were robust to outcomes specified as primary or secondary. This was necessary as in theory, primary outcomes are more likely to be reported in the abstract, and are also more likely to be significant as trials are often powered for their primary outcome. We divided all outcomes into either primary or secondary, and repeated the above process for each of the two subgroups. Random effects metaregression was also performed to determine whether there were any differences between the primary and secondary outcome subgroups.

2.7. Secondary analysis

A secondary analysis utilized histograms to provide a graphical representation of the distribution of numerically reported p values and their equivalents (i.e. confidence intervals) across the statistical significance spectrum. This aimed to illustrate the effect of theoretical thresholds on the distribution pattern of precise numerical p values reported in abstracts and full texts. The X-axis plotted the range of p values from 0 to 1, at 0.01 increments, and the y-axis plotted the frequency of p values at each increment.

The histogram plotted numerical p values, and thus only outcomes that were reported with an exact p value, or could be converted into a



numerical *p* value were included. Therefore when a *p* value was missing, but a 95% confidence interval or standard error were reported, a *p* value was calculated instead. Outcomes that did not have any reported statistical significance were recorded, but were excluded from the primary and secondary analysis.

3. Results

Execution of the search strategy took place in May 2009. Three hundred and fifty RCTs were included as per the flow diagram presented in Fig. 1. A total of 8258 outcomes were reported in the included RCTs. On average, 24 outcomes were reported per trial (standard deviation = 22), with a range between 1 and 231 outcomes.

The meta-analysis comparing proportion of significant outcomes in abstracts vs. full texts is illustrated in Fig. 2. The pooled result showed an association between significance and the reporting of outcomes in abstracts (odds ratio = 3.0, 95% confidence interval 2.5–3.58, $p < 0.001$). Thus an outcome reported in an abstract had three times the odds of being statistically significant when compared to an outcome reported in the full text. This result was consistent and not subject to heterogeneity ($I^2 = 0\%$).

The findings were robust to the subgroup analysis assessing primary and secondary outcomes separately. The association remained when only primary outcomes were assessed (OR = 2.8, 95% CI 1.8–4.5, $p < 0.001$, $I^2 = 0$), and when only secondary outcomes were assessed (OR = 2.8, 95% CI 2.3–3.5, $p < 0.001$, $I^2 = 0$). In the metaregression

analysis, there was no difference between the primary and secondary outcome subgroups (OR = 1.0, 95% CI 0.6–1.7, $p = 0.9$).

There is an observable large drop in the frequency of reported *p* values between 0.04 and 0.05, and also after 0.06 in both the abstract (Fig. 3) and full text (Fig. 4) histograms. Both histograms illustrate the highest frequency of *p* value reporting between 0 and 0.01. Furthermore, there was a larger proportion of reported non-significant *p* values seen in the full text histogram, compared to the abstracts histogram.

4. Discussion

An outcome reported in an abstract had three times the odds of being significant when compared to the corresponding full text. This is the first study to show an association using meta-analysis of all reported results at the outcome level, and accurately quantifies the magnitude of the phenomenon. The findings were also robust to whether outcomes were specified as primary or secondary. There was no statistical heterogeneity, and this consistency enhances the generalizability of our findings across the body of surgical RCTs in the literature [13].

The disproportionate frequency of significant *p* values in abstracts is not a newly observed phenomenon and these results are consistent with those reported in the literature [1,3,5]. Journal space constraints may influence authors to favor inclusion of primary outcome measures, which are more likely to be significant or to attempt to convey the most 'clinically relevant' findings that will affect practice. However significance alone is of limited value, and does not necessarily correlate

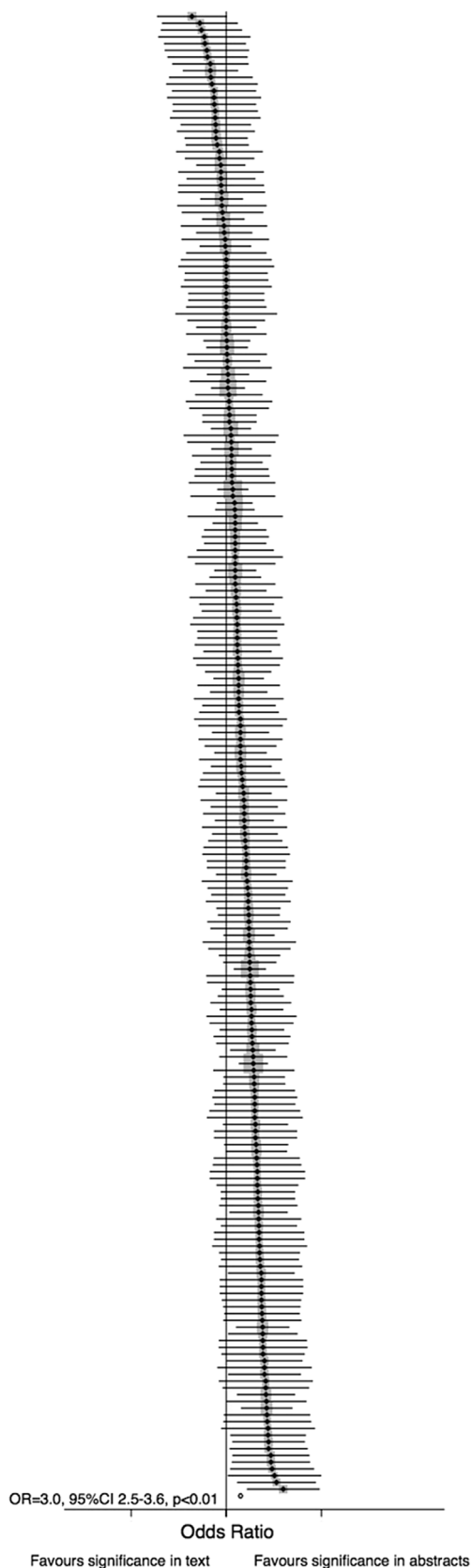


Fig. 2. Forest plot of the overall pooled analysis of 218 studies showing an association favoring the reporting of significant outcomes in abstracts.

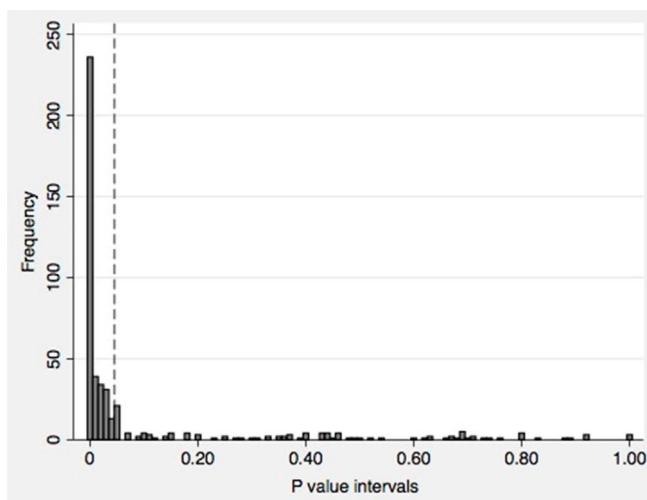


Fig. 3. Histogram depicting frequency of reported abstract outcomes according to exact p value. P values were grouped into 0.01 intervals. The dotted line represents a p value of 0.05.

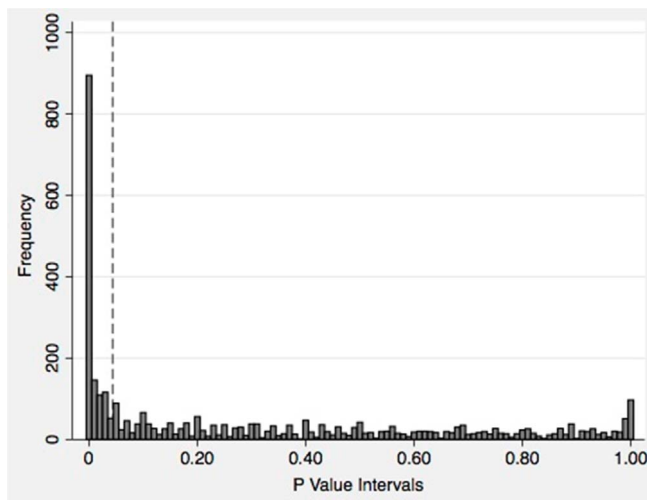


Fig. 4. Histogram depicting frequency of reported full text outcomes according to exact p value. P values were grouped into 0.01 intervals. The dotted line represents a p value of 0.05.

with clinical efficacy. Information regarding treatment effect i.e. magnitude or direction of difference and confidence intervals, are required to improve interpretation of clinical efficacy [14]. Alternatively authors may seek to incentivize readers or journal editors to select their article by portraying increased significant findings.

The dip in frequency between 0.04 and 0.05 in the abstract and full text histograms (Figs. 3 and 4) may be explained by the clustering of specific numerical p values close to 0.05 under the significance threshold of $p < 0.05$. There has been a wide misuse of the arbitrary significance threshold ‘ $p < 0.05$ ’ to incorrectly equate to ‘proof’ of treatment difference. Authors’ tendency to cluster p values immediately below 0.05 with $p < 0.05$ thresholds, or simply report as ‘significant’ hinders effective scientific reporting [8,15]. Whereas the dip in frequency of reported outcomes after 0.06 may be due to tendency to report non-significant outcomes ‘close to significance’ numerically and others as simply “non-significant” or $p > 0.05$.

A gross increase in the frequency of reported p values is observed within the parameters of ‘significance’ i.e. $p < 0.5$, across the statistical significance spectrum, comparative to the reduced frequency of non-significant p values reported in both histograms. This illustrates that overall there is an increased prevalence of significant p values in

research findings, not a trend exclusive to abstracts.

The tendency for non-significant outcomes to be omitted from abstracts was similarly observed in a survey of 73 observational studies, which showed an excess of *p* values between 0.01 and 0.05 in abstracts, indicative of biased reporting or analysis [3]. Ginsel et al. concluded using a similar histogram distribution analysis that there is evidence of systematic error in reporting of significant *p* values in abstracts, potentially by methodological errors, publication bias or fraud [9].

A cross sectional study of 260 PubMed abstracts exposed an unexpected prevalence of significant results in abstracts and concluded that they should generally be disbelieved [16]. The finding was described as ‘unexpected’ as often the premise behind conducting RCTs is clinical equipoise, i.e. the null hypothesis of no ‘known’ difference between interventions is likely to be accepted [3,16]. Pocock et al. described authors’ tendency to emphasize more significant findings in abstracts, showing the odds were nine times higher for the reporting of significant results in the abstract in a survey of 19 clinical trials [8].

The disproportionate reporting of significant *p* values in abstracts demonstrated may be a consequence of the ‘spin’ phenomenon. Spin involves selective reporting to convince readers of the beneficial effect of an intervention, greater than indicated by the results [2]. Non-significance or omission of stating a primary outcome measure predisposes it to spin and selective reporting bias [2,4]. Unspecified primary outcomes facilitate the ‘cherry picking’ of particular results, including significant secondary outcomes, to create emphasis [2]. Lockyer et al. demonstrated that 63% of abstracts contained spin in wound related RCTs. They showed that abstracts reporting favorable treatment effects often presented no statistical analysis, non significant findings in support or only reported significant results [17].

Numerous studies caution readers on relying upon abstracts alone, due to the potential for misleading information and erroneous decision-making [2,4,8]. Marcelo et al. demonstrated that clinical decision by residents guided by full texts were more accurate than those by abstracts alone, particularly in the department of surgery (*p* = 0.016) [18]. Boutron et al. randomly assigned 150 clinicians’ abstracts with spin and a different 150 abstracts without spin. In the abstracts with spin, the experimental treatment was rated more beneficial (*p* = 0.03) and clinicians were more interested in reading the article (*p* = 0.029) [2]. Our findings substantiate that a reading of abstracts alone presents information skewed towards significance, which may provide an exaggerated perception of experimental treatment effect, or attempt to convey statistically significant outcomes and omit relevant non-significant findings in the process. Thus clinical decision-making should optimally be based upon a thorough reading and critical appraisal of the corresponding full texts.

Ioannidis proposes that research findings are less likely to be true when teams are working in scientific fields that chase statistical significance [19]. Gelman and Stern assert that differences in statistical significance are often not statistically significant, and that large variations in significance levels may correspond to minor non-significant variation in underlying quantities [20]. Thus there needs to be a shift in scientific culture: from the simple effort to dichotomize results into significant or not, to precision reporting of *p* values and confidence intervals that are less susceptible to chance [8,21,22].

The strengths of this study include a systematic review design, the meta-analysis facilitating a statistical comparison and filters allowing generalizability to surgical practice. Given we limited our inclusion criteria to English language trials, the results are only generalizable to RCTs published in English literature. There is no evidence that language restriction results in different results for clinical meta-analyses [23], and is even less likely to be the case for a non-clinical research question such as those explored in this study. Another weakness was the pattern of outcome reporting in abstracts, where primary outcomes are more likely to be reported, and are in theory more likely to be statistically significant. This was a potential confounder of our primary analysis. However, we performed a sensitivity analysis and found that when only

primary outcomes (or secondary outcomes) were considered, the association remained. Therefore primary outcomes were also more likely to be reported in the abstract when they were statistically significant, compared to the full text. The same finding held for secondary outcomes and importantly, there was no significant difference in the effect size of the primary and secondary outcome subgroups. It is therefore unlikely that primary/secondary outcome status was a confounder of the results.

5. Conclusion

In conclusion, we found that an outcome reported in a surgical RCT abstract has three times the odds of being significant when compared to the corresponding full text. We also found a clustering of reported *p* values around the 0.05 cut-off - a concerning finding that suggests data is selectively analyzed and reported to achieve a statistically significant status. It is imperative that clinicians and policy makers do not rely solely on information presented in abstracts for their decision-making. Guidelines for the reporting of abstracts and full texts, such as those developed by the CONSORT group, should be promoted and adhered to.

Ethical Approval and Consent to participate

N/A.

Consent for publication

All authors have consented to the publication of this paper.

Availability of supporting data

This study was performed using data from a doctoral thesis. The protocol for the thesis was pre-approved and is available from one of the authors, Dr Sam Adie.

Competing interests

There were no competing or conflicts of interest.

Funding

SA was supported by scholarship grants from the National Health and Medical Research Council of Australia (630761) (Biomedical Postgraduate Scholarship), and the Royal Australasian College of Surgeons (165214) (Sir Roy McCaughey Research Fellowship). The funders had no role in the design, data collection or analysis of this study.

Authors' contributions

All the authors listed have significantly contributed to this paper in order of authorship. All authors read and approved the final manuscript.

1. Yusuf Assem – Research, data extraction, analysis and preparation of the first draft of the manuscript.
2. Dr Sam Adie – Formation of research question and concept, data analysis and preparation of manuscript.
3. Jason Tang – Research, data extraction and preparation of manuscript.
4. Prof Ian Harris - Formation of research question and concept, research and preparation of manuscript.

Authors' information

Mr Yusuf Assem – BMed, MD (Candidate)

Dr Sam Adie - BSc(Med) MBBS(Hons) MSpMed MPH PhD FRACS (Ortho)

- Department of Orthopaedic Surgery
- Conjoint associate lecturer at South West Sydney Clinical School, UNSW

Mr Jason Tang – BMed, MD (Candidate)

Professor Ian Harris MBBS, MMed (Clin Epi), PhD, FRACS(Orth), FAOrthA

- Professor of Orthopaedic Surgery, UNSW
- Director, Whitlam Orthopaedic Research Centre
- Director, Injury and Rehabilitation Research Stream, Ingham Institute for
- Applied Medical Research
- Director, Surgical Specialties Stream, SWSLHD
- Deputy Director, AOA NJRR (National Joint Replacement Registry)
- Chair, ACORN (Arthroplasty Clinical Outcomes Registry)
- Co-chair, ANZHFR (ANZ Hip Fracture Registry)

Acknowledgements

N/A.

Abbreviations

RCT Randomised Control Trial
 Consort Consolidated Standards of Reporting Trials
 CENTRAL Cochrane Central Register of Controlled Trials

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.conctc.2017.07.007>.

References

- [1] J.A. Lehmen, R.M. Deering, A.K. Simpson, et al., Inconsistencies between abstracts and manuscripts in published studies about lumbar spine surgery, *Spine* 39 (10) (2014) 841–845.
- [2] I. Boutron, D.G. Altman, S. Hopewell, et al., Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: the SPIIN randomized controlled trial, *J. Clin. Oncol.* 32 (36) (2014) 4120–4126.
- [3] P.C. Gøtzsche, Believability of relative risks and odds ratios in abstracts: cross sectional study, *BMJ* 333 (7561) (2006) 231–234.
- [4] P. Fontelo, A. Gavino, R.F. Sarmiento, Comparing data accuracy between structured abstracts and full-text journal articles: implications in their use for informing clinical decisions, *Evid. Based Med.* 18 (6) (2013) 207–211.
- [5] A.K. Altwaigi, C.M. Booth, W.M. Hopman, et al., Discordance between conclusions stated in the abstract and conclusions in the article: analysis of published randomized controlled trials of systemic therapy in lung cancer, *J. Clin. Oncol.* (2012) 3552–3557.
- [6] R. Dal-Ré, M. Castell, J. García-Puig, If the results of an article are noteworthy, read the entire article; do not rely on the abstract alone, *Rev. Clínica Española Engl. Ed.* 215 (8) (2015) 454–457.
- [7] I. Boutron, S. Dutton, P. Ravaud, et al., Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes, *Jama* 303 (20) (2010) 2058–2064.
- [8] S.J. Pocock, M.D. Hughes, R.J. Lee, Statistical problems in the reporting of clinical trials, *N. Engl. J. Med.* 317 (7) (1987) 426–432.
- [9] B. Ginsel, A. Aggarwal, W. Xuan, et al., The distribution of probability values in medical abstracts: an observational study, *BMC Res. Notes* 8 (1) (2015) 721.
- [10] M.J. Bradburn, J.J. Deeks, J.A. Berlin, et al., Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events, *Stat. Med.* 26 (1) (2007) 53–77.
- [11] A.-W. Chan, D.G. Altman, Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors, *BMJ* 330 (7494) (2005) 753.
- [12] A.-W. Chan, A. Hróbjartsson, M.T. Haahr, et al., Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles, *Jama* 291 (20) (2004) 2457–2465.
- [13] J.P. Higgins, S.G. Thompson, J.J. Deeks, et al., Measuring inconsistency in meta-analyses, *BMJ* 327 (7414) (2003) 557–560.
- [14] P. Page, Beyond statistical significance: clinical interpretation of rehabilitation research literature, *Int. J. Sports Phys. Ther.* 9 (5) (2014) 726.
- [15] M. Cowles, C. Davis, On the origins of the .05 level of statistical significance, *Am. Psychol.* 37 (5) (1982) 553.
- [16] C. Cook, C. Sheets, Clinical equipoise and personal equipoise: two necessary ingredients for reducing bias in manual therapy trials, *J. Man. Manip. Ther.* 19 (1) (2011) 55–57.
- [17] S. Lockyer, R. Hodgson, J.C. Dumville, et al., “Spin” in wound care research: the reporting and interpretation of randomized controlled trials with statistically nonsignificant primary outcome results or unspecified primary outcomes, *Trials* 14 (1) (2013) 371.
- [18] A. Marcelo, A. Gavino, I.T. Isip-Tan, et al., A comparison of the accuracy of clinical decisions based on full-text articles and on journal abstracts alone: a study among residents in a tertiary care hospital, *Evid. Based Med.* 18 (2) (2013) 48–53.
- [19] J.P. Ioannidis, Why most published research findings are false, *PLoS Med.* 2 (8) (2005) e124.
- [20] A. Gelman, H. Stern, The difference between “significant” and “not significant” is not itself statistically significant, *Am. Stat.* 60 (4) (2006) 328–331.
- [21] C. Poole, Low P-values or narrow confidence intervals: which are more durable? *Epidemiology* 12 (3) (2001) 291–294.
- [22] C.R. Weinberg, It's time to rehabilitate the P-value, *Epidemiology* 12 (3) (2001) 288–290.
- [23] D. Moher, T.P. Klassen, K.F. Schulz, et al., What contributions do languages other than English make on the results of meta-analyses? *J. Clin. Epidemiol.* 53 (9) (2000) 964–972.