Buzzwords in Highly-Cited Manuscript Abstracts

A. Hall, H. Tanna, M. Besford, G. Allen, D. Sheard

Costello Medical, Cambridge, UK

Objectives

- To investigate the relationship between "buzzwords" used in *New England Journal of Medicine* article abstracts and their citation counts.
- To determine if any buzzwords related to good publication practice are more commonly used in abstracts of highly-cited articles than in abstracts of lowly-cited articles.

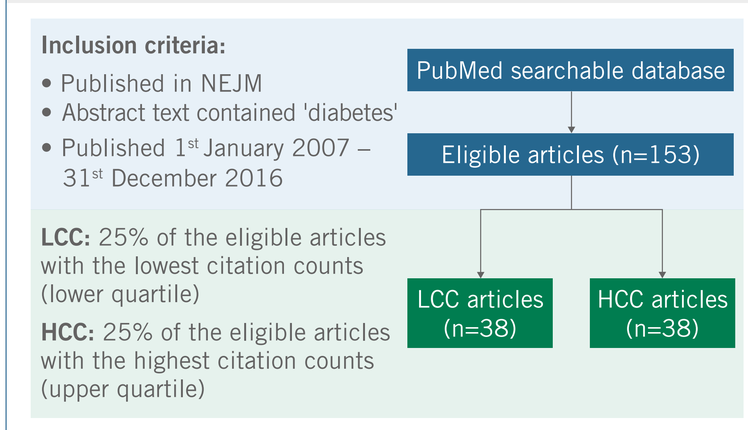
Background

- The Good Publication Practice guidelines (GPP3)[1] recommend that authors follow established reporting standards when publishing results.
- Commonly used reporting standards, including the Consolidated Standards of Reporting Trials (CONSORT)[2] for randomised controlled trials (RCTs), Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)[3] and Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)[4] guidelines have extensions for abstracts.[5–7]
- However, it is unclear whether closely adhering to such reporting guidelines affects the frequency with which articles are cited.

Methods

- A sample of abstracts were identified via a PubMed database search with the RISmed package in R.[8] Search terms were used to select for articles published in a single journal (*New England Journal of Medicine;* NEJM) and within a restricted timeframe (between 1st January 2007 and 31st December 2016) **(Figure 1)**.

Figure 1 | Abstract selection flow diagram



HCC: high citation count; LCC: low citation count; NEJM: New England Journal of Medicine.

- To reduce the risk of bias in the analysis, articles were selected from within a single disease area: diabetes.
- For each eligible article identified via PubMed, citation counts (CCs) were manually extracted from the NEJM website. High and low CC (HCC; LCC) cut-offs were defined as the upper and lower quartiles of the CCs for all eligible articles.
- Prior to the analysis, stop words (most common English words, e.g. 'and' or 'the') and the word 'diabetes' (common to all identified abstracts), were removed from the text of all abstracts.
- 10 buzzwords relevant to GPP3,[1] the CONSORT Checklist for Journal and Conference Abstracts,[5] PRISMA for Abstracts[6] and the STROBE Checklist for Conference Abstracts[7] were pre-selected **(Table 1)**.
- For each pre-selected buzzword, the number of HCC and LCC abstracts with ≥1 mention of the word were compared using the prop.test function in R,[9] as an implementation of the one-sided z-test with a Yates continuity correction. p≤0.001 indicates that a buzzword was featured in significantly more HCC abstracts than LCC abstracts.
- From the HCC and LCC abstracts, the article types were manually identified by one reviewer; abstracts were classified as either 'RCT' or 'Other'.

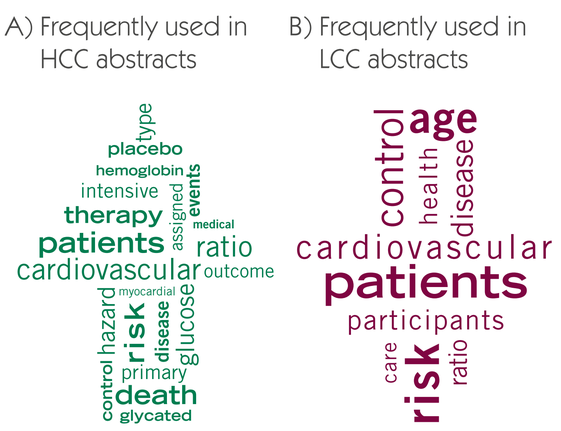
Results

- 153 abstracts met the inclusion criteria, with a median of 292 citations (min: 0, max: 3311). HCC abstracts had ≥540 citations and LCC abstracts had ≤133 citations.
- As may be expected for a medical journal, a much larger fraction of HCC abstracts than LCC abstracts were reporting RCTs of medical interventions (HCC: 73.7% [28/38]; LCC: 31.6% [12/38]).
- Abstracts contained 4440 different words between them. The most frequently used words are shown in **Figure 2**.
- The two words with the highest HCC/LCC frequency ratio were 'myocardial' and 'infarction' (6.2 [37/6] and 5.4 [38/7] respectively). **(Figure 3A)**. 'Outcome' was used 5.1 times more frequently in HCC abstracts (n=51) than in LCC (n=10) abstracts **(Figure 3A)**.
- 'Health' was used over 10 times more often in LCC (n=43) than HCC (n=4) abstracts **(Figure 3B)**. While 'patients' was used more frequently in HCC (n=166) than LCC (n=91) abstracts **(Figure 3A)**, 'participants' was used 4.9 times more frequently in LCC abstracts (n=44) than HCC abstracts (n=9, **Figure 3B**).
- Text related to 'ClinicalTrials.gov', 'death' and 'primary' were used significantly more in HCC abstracts than in LCC abstracts **(Table 2)**.

Table 1 | Pre-selected buzzwords relevant to publication guidelines

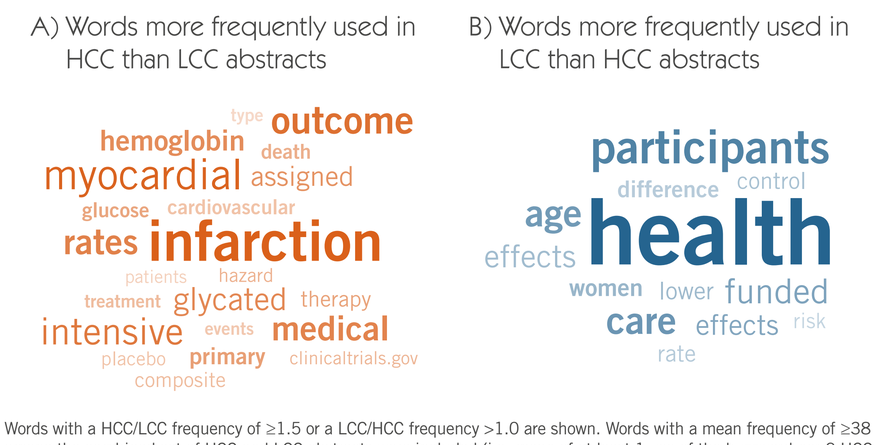
Items suggested for inclusion by publication guidelines		Pre-selected buzzwords
Item	Supporting text	
Study design	• "How participants were allocated to interventions"[5] • "Study and report characteristics used as criteria for inclusion"[6]	Randomisation
Statistical analysis	• "The estimated effect size and its precision"[5] • "Size of the effect in terms meaningful to clinicians and patients"[6] • "Describe statistical methods"[7]	Significance Statistic 0.001
Trial registration	• "Include the clinical trial identifier"[1] • "Registration number and name of trial register"[5] • "Registration number and registry name"[6]	ClinicalTrials.gov
Primary publications and primary endpoints	• "The primary publication or first presentation of any type of study should indicate this fact [of being the first published]"[1] • "Secondary publications should always reference the primary publication"[1] • "Clearly defined primary outcome"[5,7] • "Results for main outcomes"[6]	Primary
Safety analysis	• "Important adverse events or side effects"[5] • "Benefits and harms"[6] • "Consider translating estimates of relative risk into absolute risk"[7]	Death Safety Hazard
Efficacy results	• "For the primary outcome, a result for each group"[5] • "Direction of the effect (i.e. which group is favoured)"[6] • "Report estimates of associations"[7]	Efficacy

Figure 2 | Most frequently used words in HCC and LCC abstracts



Words with a mean frequency of ≥38 among all abstracts in the HCC or LCC subgroup (i.e. a mean of at least 1 use of the buzzword per abstract).

Figure 3 | Differences in the words most frequently used in HCC and LCC abstracts



Words with a HCC/LCC frequency of ≥1.5 or a LCC/HCC frequency >1.0 are shown. Words with a mean frequency of ≥38 among the combined set of HCC and LCC abstracts were included (i.e. mean of at least 1 use of the buzzword per 2 HCC or LCC abstracts). Words are scaled and coloured such that larger, darker words correspond to a higher HCC/LCC frequency ratio **(Figure 3A)** or higher LCC/HCC frequency ratio **(Figure 3B)**. Numbers were not counted as words in this analysis.

Table 2 | Instances of pre-selected buzzwords relevant to publication guidelines

Buzzword[a]	Text search term[b]	Abstracts with ≥1 instance of the text search term		p value[c]
		HCC (n=38)	LCC (n=38)	
Randomisation	randomi	15	9	0.109
Significance	significan	24	21	0.320
Hazard	hazard	19	9	0.016
ClinicalTrials.gov	clinicaltrials.gov	28	13	**<0.001**
0.001	0.001	23	13	0.019
Primary	primar	27	13	**0.001**
Death	death	29	11	**<0.001**
Safety	safe	3	6	NA[d]
Efficacy	efficac	2	5	NA[d]
Statistic	statistic	2	0	NA[d]

[a]Pre-selected buzzwords based on relevance to publications guidelines; [b]Exact term searched in abstracts; [c]Words featured in significantly more HCC abstracts than LCC abstracts **(p≤0.001)** are highlighted in **bold**; [d]No statistical test was run due to low counts in the HCC and/or LCC sample.

Conclusions

- Our results suggest that abstracts using buzzwords related to adherence with reporting guidelines, as recommended by GPP3, may be cited more frequently than abstracts that do not use these buzzwords.
- These results provide further motivation for following relevant reporting guidelines, such as including trial registration numbers and clear primary outcomes in abstracts.

Abstract

Objective

- To investigate the relationship between "buzzwords" in New England Journal of Medicine (NEJM) article abstracts and citation counts (CCs).

Research Design and Methods

- Abstracts mentioning "diabetes" from 2007–2016 NEJM articles were identified via PubMed and their CCs recorded from the NEJM website. High and low CC (HCC; LCC) cut-offs were the upper and lower quartiles of all CCs. 10 "buzzwords" were pre-selected. The numbers of HCC and LCC abstracts with ≥1 mention of each buzzword were compared using the prop.test function in R. Words frequently used in HCC and LCC abstracts were determined.

Results

- 153 abstracts, with 4440 unique words, met the inclusion criteria. "ClinicalTrials.gov", "death" and "primary" were used significantly more in HCC abstracts. "Outcome" and "placebo" were among the most common words used in abstracts of HCC, but not LCC, articles.

Conclusions

- Buzzwords associated with good publication practice, including "ClinicalTrials.gov", in manuscript abstracts may increase their citation likelihood.

References

1. Battisti WP. *et al.* Ann Intern Med 2015;163:461–4; **2.** Schulz KF, Altman DG, Moher D, for the CONSORT Group BMJ 2010;340:c332; **3.** Moher D. *et al.* BMJ 2009;339:b2535; **4.** von Elm E. *et al.* BMJ 2007;335(7624):806–8; **5.** Hopewell S. *et al.* PLoS Med 2008;5(1):e20. doi: 10.1371/journal.pmed.0050020; **6.** Beller EM. *et al.* PLoS Med 2013;10(4):e1001419. doi: 10.1371/journal.pmed.1001419; **7.** STROBE. STROBE Conference Abstracts Checklist. Available at: www.equator-network.org/reporting-guidelines/strobe-abstracts/ [Last accessed 22.12.17]; **8.** Stephanie Kovalchik (2017). R package version 2.1.7. Available at: CRAN.R-project.org/package=RISmed [Last accessed 22.12.17]; **9.** R Core Team (2017). R Foundation for Statistical Computing, Vienna, Austria. Available at: www.R-project.org/ [Last accessed 22.12.17].

Author Contributions

Substantial contributions to study conception/design, or acquisition/analysis/interpretation of data: AH, HT, MB, GA, DS; Drafting of the publication, or revising it critically for important intellectual content: AH, HT, MB, GA, DS; Final approval of the publication: AH, HT, MB, GA, DS.

Acknowledgements

The authors thank Kevin Racaza, Costello Medical, for graphic design assistance.

Presented at the 2018 European Meeting of the International Society for Medical Publication Professionals (ISMPP), 23–24 January 2018, London, UK

Copyright © Costello Medical Consulting Ltd