aiming for a TOFR of 1.0.[13,14] Train-of-four ratio should be normalised by baseline values when calibrated acceleromyography is used.[13] Second, the dosage of the reversal drug should be titrated according to the level of neuromuscular block, following recommendations and avoiding underdosage.[3]

Declarations of interest

MC has received payments for lectures from Merck Sharp & Dohme, Rome, Italy. The other authors declare no conflicts of interest.

References

1. Kirmeier E, Eriksson LI, Lewald H, et al. Post-anaesthesia pulmonary complications after use of muscle relaxants (POPULAR): a multicentre, prospective observational study. *Lancet Respir Med* 2019; **7**: 129–40
2. Blobner M, Hunter JM, Meistelman C, et al. Use of a train-of-four ratio of 0.95 versus 0.9 for tracheal extubation: an exploratory analysis of POPULAR data. *Br J Anaesth* 2020; **124**: 63–72
3. Brull SJ, Kopman AF. Current status of neuromuscular reversal and monitoring: challenges and opportunities. *Anesthesiology* 2017; **126**: 173–90
4. McCarthy GJ, Cooper R, Stanley JC, Mirakhur RK. Dose–response relationships for neostigmine antagonism of vecuronium-induced neuromuscular block in adults and the elderly. *Br J Anaesth* 1992; **69**: 281–3
5. Murphy GS, Szokol JW, Avram MJ, et al. Residual neuromuscular block in the elderly: incidence and clinical implications. *Anesthesiology* 2015; **123**: 1322–36
6. Kaufhold N, Schaller SJ, Stäuble CG, et al. Sugammadex and neostigmine dose-finding study for reversal of residual neuromuscular block at a train-of-four ratio of 0.2 (SUNDRO20). *Br J Anaesth* 2016; **116**: 233–40
7. Pongrácz A, Szatmári S, Nemes R, Fülesdi B, Tassonyi E. Reversal of neuromuscular blockade with sugammadex at the reappearance of four twitches to train-of-four stimulation. *Anesthesiology* 2013; **119**: 36–42
8. Muramatsu T, Isono S, Ishikawa T, et al. Differences of recovery from rocuronium-induced deep paralysis in response to small doses of sugammadex between elderly and nonelderly patients. *Anesthesiology* 2018; **129**: 901–11
9. Abd-Elfattah AEM. Reversal of rocuronium neuromuscular blockade in elderly patients undergoing elective open abdominal surgery: comparative study of sugammadex versus neostigmine. *Res Opin Anesth Intensive Care* 2019; **6**: 64–71
10. Oh CS, Rhee KY, Yoon TG, Woo NS, Hong SW, Kim SH. Postoperative delirium in elderly patients undergoing hip fracture surgery in the sugammadex era: a retrospective study. *Biomed Res Int* 2016; **2016**: 1054597
11. Williams 3rd WH, Cata JP, Lasala JD, Navai N, Feng L, Gottumukkala V. Effect of reversal of deep neuromuscular block with sugammadex or moderate block by neostigmine on shoulder pain in elderly patients undergoing robotic prostatectomy. *Br J Anaesth* 2020; **124**: 164–72
12. Togioka BM, Yanez D, Aziz MF, Higgins JR, Tekkali P, Treggiari MM. Randomised controlled trial of sugammadex or neostigmine for reversal of neuromuscular block on the incidence of pulmonary complications in older adults undergoing prolonged surgery. *Br J Anaesth* 2020; **124**: 553–61
13. Fuchs-Buder T, Brull SJ. Is less really more? A critical appraisal of a POPULAR study reanalysis. *Br J Anaesth* 2020; **124**: 12–4
14. Eikermann M, Vogt FM, Herbstreit F, et al. The predisposition to inspiratory upper airway collapse during partial neuromuscular blockade. *Am J Respir Crit Care Med* 2007; **175**: 9–15

doi: 10.1016/j.bja.2020.04.086
Advance Access Publication Date: 7 May 2020

Examining the correlation between Altmetric score and citation count in the anaesthesiology literature

Lisa Q. Rong*, Alexandra J. Lopes, Irbaz Hameed, Mario Gaudino and Mary E. Charlson

New York, NY, USA

*Corresponding author. E-mail: lir9065@med.cornell.edu

Keywords: altmetrics; anaesthesiology; bibliometrics; citation count; journal impact factor; social media

Editor—Traditional measures of impact of scientific research focus on article citation numbers and journal impact factor (IF).[1] With the increase of digital technology and use of social media platforms to discuss research, impact for these channels can also be assessed. Alternative-level metrics (altmetrics) are a new measure of the attention, dissemination, overall influence, and impact of scientific publications.[2] There are various altmetrics platforms being used, but the first and most popular of these, Altmetric, compiles the number of mentions of an article across the most commonly used social media platforms such as Twitter, Facebook, and blogs and research websites to generate a weighted score.[3] We explored the relationship between traditionally-used bibliometrics (citation counts) and altmetrics among highly cited articles in top anaesthesiology journals.

Methods

We identified the top 10 most-cited articles in the five anaesthesiology journals with the highest Clarivate Analytics IF (*Anesthesiology*, *British Journal of Anaesthesia* [*BJA*], *European Journal of Anaesthesiology*, *Anaesthesia*, and *Anesthesia & Analgesia*) in 2016 and 2018.[4] Guidelines were excluded because of their disproportionally higher likelihood of being cited. For each journal, the following were recorded: the IF in 2016/2018, total number of tweets, and the 10 articles with the highest

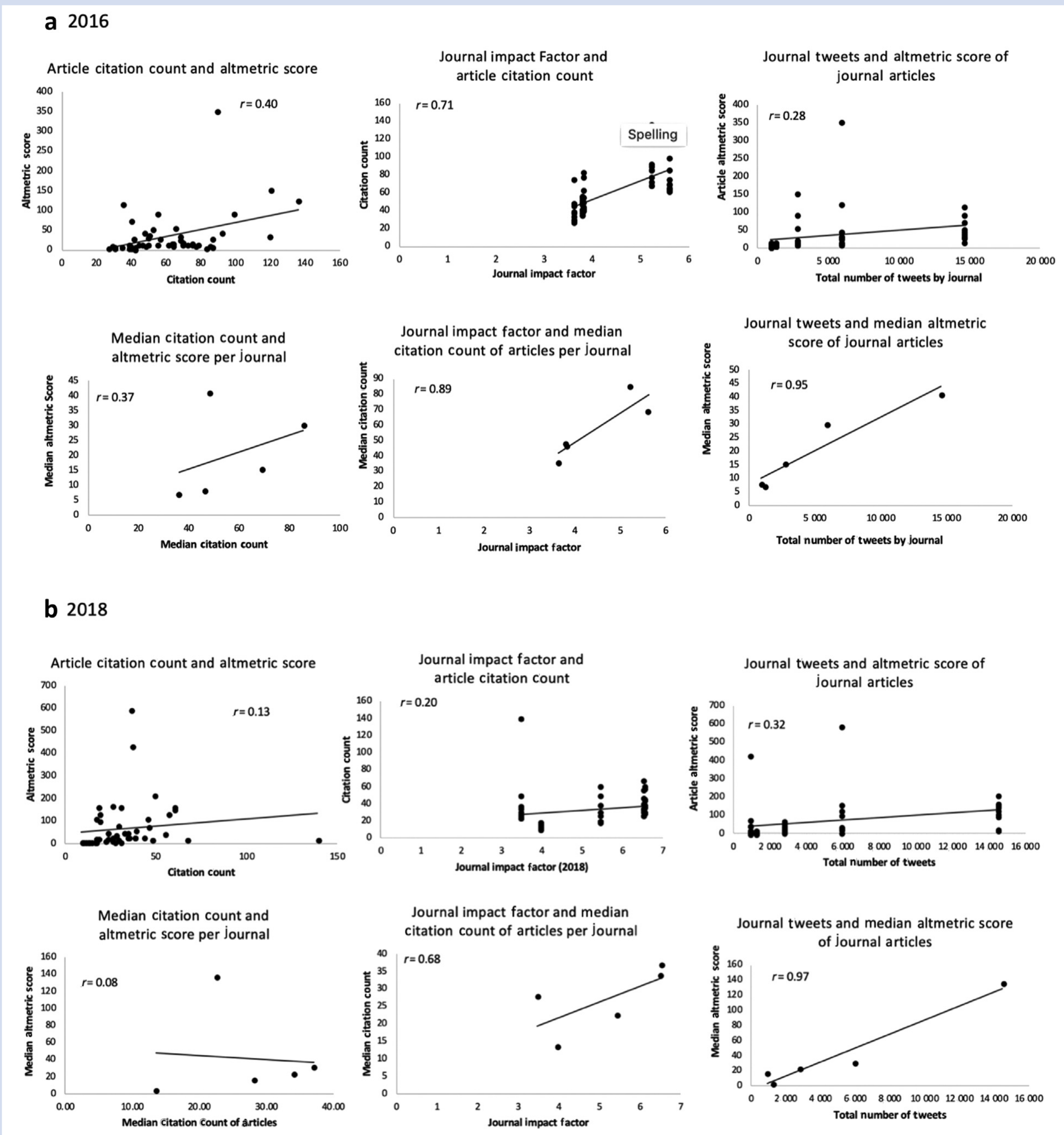

Fig. 1. Correlation between Altmetric score, citation count, journal impact factor, and overall journal tweets.

number of citations on Scopus in March, 2020.[5] For each article, the citation count and Altmetric score were recorded.[3]

After testing for normality, continuous variables such as citation count and Altmetric score were expressed as median and interquartile range (IQR). Descriptive statistics were performed and Pearson's correlation tests were used.

Results

A total of 100 articles were evaluated. For articles published in 2016, overall Altmetric scores were weakly correlated with citation count (r=0.40) but not journal IF (r=0.25). There was strong correlation between journal IF and overall citation count (r=0.71), and journal IF and median citation count for that journal (r=0.89). There was strong correlation between the number of journal tweets and the median journal Altmetric score (r=0.95) (Fig 1). The largest Altmetric score for any article was 351 (median 13.0; IQR [7.5, 33.0]) and the largest citation count was 136 (median 55.5; IQR [41.5, 75.0]). *Anaesthesia* had the largest median article Altmetric score (41) and largest number of journal tweets (14 600), while the *BJA* had the largest IF (5.62), and *Anesthesiology* had the largest median number of citations (85.5).

For articles published in 2018, Altmetric scores were not correlated with citation count (r=0.13) or journal IF (r=0.15). There was a moderate correlation between journal IF and median citation count (r=0.68) but no correlation between journal IF and overall citation count for that journal (r=0.20). There was a strong correlation between the number of journal tweets and the median journal Altmetric score (r=0.97). The highest Altmetric score for any article was 589 (median 23.5; IQR [7, 95]) and the highest citation count was 139 (median 28.5; IQR [18, 38]). *Anesthesiology* had the highest IF (6.5) and the highest median number of citations (37). *Anaesthesia* had the highest number of journal tweets (14,600) and largest median Altmetric score (136).

Discussion

To our knowledge, this is the first study evaluating the relationship between altmetrics and traditionally used bibliometrics for the top-cited anaesthesiology articles. We found that for articles published in 2016, the Altmetric score was weakly correlated with citation count and its median score was strongly correlated with the overall number of journal tweets. In 2018, the median citation counts (29 *vs* 53) were less, and Altmetric scores (24 *vs* 13) were greater than in 2016. The overall highest individual citation count and Altmetric score was found in 2018. While there was no correlation between citation count (139) and Altmetric score (589), there was a moderate correlation between journal IF and median citation count, a weak correlation between journal tweets and IF, and a strong correlation between journal tweets and Altmetric score.

Our results demonstrate that journal social media activity may influence Altmetric scores but that there is a limited correlation between citation and Altmetric scores.

Top anaesthesiology journals routinely report altmetrics online. For example, *Anesthesiology* articles link to Altmetrics while the *BJA* articles link to Plum Analytics.[6] Researchers in urology, emergency medicine, and paediatric surgery have studied the effect of social media on academic influence and have shown a weak correlation between Altmetric score and citation counts.[1,2,7] Other studies have found that journals with social media accounts had significantly higher Altmetric scores than those without accounts[8] and that tweets can predict highly cited articles within the first 3 days of article publication.[9] Another recent study demonstrated a weak correlation between Altmetric score and citations for a percentage of articles published in top medicine journals.[10]

Altmetrics and traditional bibliometrics such as citation count both provide insight into the impact and influence of research. However, while an article needs time to accumulate citations and influence, altmetrics provide rapid feedback on the ripple effect of research which may not be sustained.[2] It has been shown that there is an initial 3 month spike in social media interest that often abates.[2] This short-lived impact may reflect a different measure of influence compared with citations which take time to accrue.

Our study has important limitations. We only analysed the top five anaesthesiology journals, selected by IF in 2016 and 2018. It is unknown whether journals of lower IF would have shown similar results. We also analysed only the top 10 most cited articles of each journal, and our findings may not be consistent with respect to less cited articles. For 2018 data, there may not have been enough time for citations to accrue. It is also possible that our results may not reflect the impact of social media use in 2020. Finally, we only used each journal's Twitter activity as a measure of the journal's social media presence.

As the use of social media for research dissemination grows, further research is needed to understand the relationship between traditional bibliometrics and altmetrics over time, qualitative aspects of articles with high citations, almetrics, or both, and the impact of social media exposure on articles and journals, both short-term and long-term.

Author contributions

Study concept and design: LQR, IH, MFLG, MEC
Data extraction: LQR
Data analysis: LQR, AL
Interpretation of data: all authors
Preparation of first draft of manuscript, revision of manuscript: LQR, AL
Preparation of manuscript: all authors

Declarations of interest

The authors declare that they have no conflicts of interest.

Funding

LQR is supported in part by the Foundation for Anesthesia Research and Education Training Grant (FAER MTRG-CT-08-15-2018-Rong). LQR is a fellow with the *British Journal of Anaesthesia* Peer Review Fellowship Programme.

References

1. Barbic D, Tubman M, Lam H, Barbic S. An analysis of altmetrics in emergency medicine. *Acad Emerg Med* 2016; **23**: 251–68
2. Nocera AP, Boyd CJ, Boudreau H, Hakim O, Rais-Bahrami S. Examining the correlation between altmetric score and citations in the urology literature. *Urology* 2019; **134**: 45–50

3. *Bookmarklet - altmetric* 2020. Available from: https://www.altmetric.com/products/free-tools/bookmarklet. [Accessed 20 December 2019]
4. *Journal citation reports, clarivate Analytics*. Available from: https://incites-clarivate-com.ezproxy.med.cornell.edu. [Accessed 20 December 2019]
5. Web of Science - Scopus. Available from: http://www.scopus.com. [Accessed 20 December 2019]
6. *PlumX metrics*. Available from: https://plumanalytics.com/learn/about-metrics/. [Accessed 20 December 2019]
7. Chang J, Desai N, Gosain A. Correlation between altmetric score and citations in pediatric surgery core journals. *J Surg Res* 2019; **243**: 52–8
8. Wang J, Alotaibi NM, Ibrahim GM, Kulkarni AV, Lozano AM. The spectrum of altmetrics in neurosurgery: the top 100 "trending" articles in neurosurgical journals. *World Neurosurg* 2017; **103**. 883–95.e1
9. Eysenbach G. Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *J Med Internet Res* 2011; **13**: e123
10. Barakat AF, Nimri N, Shokr M, et al. Correlation of altmetric attention score and citations for high-impact general medicine journals: a cross-sectional study. *J Gen Intern Med* 2019; **34**: 825–7

doi: 10.1016/j.bja.2020.04.086
Advance Access Publication Date: 20 June 2020

Should we accept a higher risk of type I errors in some trials?

Jens Laigaard

Copenhagen, Denmark

E-mail: xrw218@alumni.ku.dk

Keywords: clinical trial; power analysis; sample size; statistics; type I error; type II error

Editor—The well-performed sample size calculation is key when conducting a properly powered trial. In the sample size calculation, we make considerations about the least clinically important difference between the groups to be compared, that is effect size (δ). We choose the risk of type I (α) and type II (β) errors and make assumptions on the variability (σ) of the outcome measure in each group. These factors affect the sample size and thereby the cost of the trial. Sample sizes are relatively small in trials searching for large differences, accepting high risk of false-positive and false-negative findings. On the contrary, trials searching for small differences, with low risk of false-positive and false-negative findings, require larger sample sizes.

The most frequently used α value is 0.05. Accordingly, the risk of finding a statistically significant difference between groups in the sample that does not exist in the population is 5% (when ignoring Bayesian thinking). The statistical power is often 0.80 or 0.90. A power of 0.80 results in 20% risk for acceptance of a false null hypothesis – that is a false negative.

So, why do we accept a relatively higher risk of extrapolating non-existing differences to the population than of not finding existing differences to the population? For treatments that in some way require a large amount of resources, we would rather risk not introducing a beneficial treatment, than introduce an indifferent or potentially harmful treatment. This is the ‘first, do not harm’ principle. However, although this is perfectly rational when testing interventions against placebo, current clinical practice, a cheaper treatment, a lower dose or likewise, sometimes this is not the case.

Sometimes we compare interventions that are equal *a priori*, for example requiring equal resources and with equal risk of side-effects. In these cases, false positives (e.g. finding differences that do not exist) are not worse than false negatives (e.g. not finding differences that do exist). In other words, when we do not have a favourite between intervention arms, we should focus on minimising the overall risk of error. In trials where the intervention arms seem equal *a priori*, we should accept equal risk of type I and type II errors to minimise the combined risk of error for a given sample size.

One example is high *vs* low arterial oxygen fraction in critically ill patients as tested in the Handling Oxygenation Targets in the Intensive Care Unit (HOT-ICU) trial.[1] The trial tests whether a target of 8 or 12 kPa oxygen in arterial blood gas samples is preferable in critically ill hypoxaemic patients. With either oxygenation target, the same amount of time and effort is needed. Also, it is unlikely that the volume of oxygen used will have an impact on the health economic analysis. In this trial, an α of 0.05 and a β of 0.10 was chosen for the primary outcome of 90 day mortality with a sample size of 2928 patients to find a 20% relative risk reduction between groups. In the HOT-ICU trial, if they instead set the α at 0.075, a power of 0.93 could be maintained, while keeping the same sample size. This lowers the total risk of error from 15% to 14.5%. A small