ISSI 2023 Proceedings, v2, pp 449–454
DOI: 10.5281/zenodo.8429008 https://cns-iu.github.io/workshops/2023-07-02_issi/

Confounding cross-correlation: A meta-analysis of the association between citations and altmetrics

Dimity Stephen[1] and Stephan Stahlschmidt[2]

[1]*stephen@dzhw.eu,* [2]*stahlschmidt@dzhw.eu*
German Centre for Higher Education Research and Science Studies (DZHW), Schützenstrasse 6A, 10117, Berlin, (Germany)

Abstract
This paper presents initial results from an ongoing meta-analysis of the correlation between altmetrics and citations. Since the Altmetric Manifesto was published in 2010, a large number of studies have examined altmetric-citation associations, resulting in a diverse array of observations. Very few studies have attempted to clarify this diversity with a meta-analysis. As such, we collect a large number of existing studies quantifying the correlation between altmetrics and citations, which for the first time will facilitate a meta-regression to establish how study characteristics such as discipline, non-zero values, and citation sources moderate the altmetric-citation relationship. To date, we have identified relevant literature examining altmetric-citation correlations in the Web of Science and then reviewed this literature against several inclusion criteria before extracting the correlation coefficient, sample size, and several other variables from each study. The final sample for the meta-analysis consists of 111 studies with a total of 914 correlation coefficients. Initial results show that the strength of the correlation with citations varies substantially by the source of altmetric data but also within each altmetric source. The meta-analysis strives to explain this variation and results of the moderating study characteristics will be presented during the conference.

Introduction
With the rise of social media and the digitalisation of science, alternative metrics – or altmetrics – emerged as a potential means to establish the impact of a publication in a much shorter timeframe than required by the conventional impact measure, citations (Priem et al. 2010). Altmetrics are generated from a range of sources, such as interactions on social media channels like Twitter and Facebook, mentions in blogs, mainstream media or Wikipedia, entries in reference managers, and usage metrics such as downloads and page views. Given the aim to supersede conventional metrics, much research has been undertaken to establish what exactly altmetrics measure and whether they "really reflect impact, or just empty buzz" (Priem et al. 2010). Consequently, many studies have investigated the correlation between altmetric indicators and citations. However, the results of these studies are exceedingly diverse. For instance, correlations between citations and Twitter mentions – one of the most thoroughly examined altmetrics – range from -0.20 to 0.78 depending on the sample used (e.g. Haustein et al. 2014; Malecki 2015; Xia et al. 2016). Further variability is observed between altmetrics indicators, with counts of Mendeley readers and usage metrics consistently more strongly associated with citations than other altmetric sources (e.g. Amath et al. 2017; Buttliere & Buder, 2017; Cho, 2021; Gorraiz, Blahous & Wieland, 2018).

The wealth of studies examining altmetric-citation correlations lends itself to a meta-analytical examination to assist in deriving clarity from the large pool of observations. However, very few meta-analyses appear to have been conducted. Bornmann (2015) examined the pooled correlation between citations and four altmetric data sources. He found weak correlations between citation counts and Twitter activity (pooled $r = 0.003$, n = 9) and mentions in blogs (pooled $r = 0.12$, n = 9), and stronger correlations between citations and CiteULike bookmarks (pooled $r = 0.23$, n = 19) and Mendeley reader counts (pooled $r = 0.51$, n = 27). Erdt et al.'s (2016) meta-analysis of 25 studies spanning nine altmetric sources replicated Bornmann's result for blogs ($r = 0.12$), but identified stronger correlations with Twitter mentions (0.11) and

CiteULike bookmarks (0.29), and a weaker correlation with Mendeley readers (0.37). In addition, Erdt et al. observed weak correlations between citations and mentions on Google+ (0.07), Delicious (0.07), Wikipedia (0.10), and Facebook (0.12), and F1000 ratings (0.23). Kolahi et al. (2021) also conducted a meta-analysis of 35 studies of correlations between citations and the Altmetric Attention Score (AAS) in health sciences, finding an overall pooled effect size of 0.19.

These studies are helpful in providing an aggregated view of the relationship between citations and altmetrics. However, a significant number of studies of altmetrics have been undertaken since Bornmann's (2015) and Erdt et al.'s (2016) studies and Kolahi et al. (2021) examined only one altmetrics indicator in one discipline. As such, we carry out here a broader and more recent meta-analysis of the association between citations and several altmetric indicators. The substantially extended coverage of included studies allows not only to observe a medial value and the underlying distribution as in former work, but facilitates a meta-model by regressing several confounders onto the correlation to crystallise the association between citations and altmetrics irrespective of the variation initiated by these confounders. Hence, we examine a number of variables as potential moderators of the relationship between altmetrics and citations. This study seeks to answer two questions:

1. What is the strength and direction of the association between several altmetric measures and citations according to the existing literature?
2. Which characteristics moderate the association between altmetrics and citations, and to what extent?

Methods

The study is comprised of two phases; first a search and review of the literature to identify and extract data from existing studies that assessed the correlation between altmetrics and citations and are within the scope of our study, and a meta-analysis of these studies.

Literature search and review

The literature search and review conducted in this study was carried out over four steps. First, we searched the literature about altmetrics and read the titles and abstracts of these studies in order to identify the broadest set of keywords that could be used to search for relevant correlation studies. Based on the results from this step, we then searched the German Kompetenznetzwerk Bibliometrie's in-house version of the Web of Science (WoS) database for documents with abstracts that contained at least one term from each of the sets (*citation, cite, citing, traditional metric*), (*relation*, associat*, predict*, correlat**), and (*altmetric*, alternative metric*, twitter, facebook, mendeley, tweet, f1000, blog, social media*) to identify potentially relevant studies. The WoS database covered publications up to April 2022. No restrictions were placed on publication year or document type. This step identified 1,051 relevant documents. However, we excluded 849 documents as they were obviously irrelevant based on a review of their titles and/or abstracts.

In the third step, we sought to review the full-texts of the remaining 202 documents against our inclusion/exclusion criteria. We thus excluded 27 documents for which full-texts could not be found. In examining the reference lists of the remaining documents, we identified an additional 31 potentially relevant documents. As such, we reviewed the full-texts of 206 documents against the following inclusion criteria. Documents were included if they were empirical; examined the association between at least one altmetric indicator and citations; examined this association at the document level, i.e. not author, journal, etc; reported a Pearson, Spearman or

Kendall correlation or R^2 statistic; reported the sample size; and were written in English. We did not differentiate between Pearson and Spearman correlations as the statistics are comparable (Shen et al., 2021). We converted Kendall's correlation coefficients to Spearman's *r* based on the conversion table provided by Gilpin (1993).

As the final step of the literature search and review, we coded for each of the included studies the primary variables of interest, the correlation coefficient and the sample size. We also coded the following additional variables for each study to be used as moderators in the meta-analysis to identify the characteristics that influenced the altmetric-citation correlation: i) the study's publication year, ii) the publication years of documents in the study, iii) the altmetric data source used, iv) the citation data source used, v) the discipline of publications in the sample, vi) the reported discipline concorded to the OECD's Fields of Science and Technology (FOS), vii) a binary indicator of whether the sample included only articles that had non-zero altmetric values or citations, and viii) a binary indicator of whether the sample included only articles that were highly cited or had high altmetric values. The steps of the literature search, review and coding process and the number of studies included at each step are shown in Figure 1. Each study typically contained multiple samples as studies often sampled publications from different publication years, disciplines, or citation or altmetric data sources and reported a correlation coefficient for each sample. We concorded the disciplines used in each study from the native WoS and Scopus classifications to the FOS classification based on concordances provided by Clarivate Analytics and Elsevier, respectively.

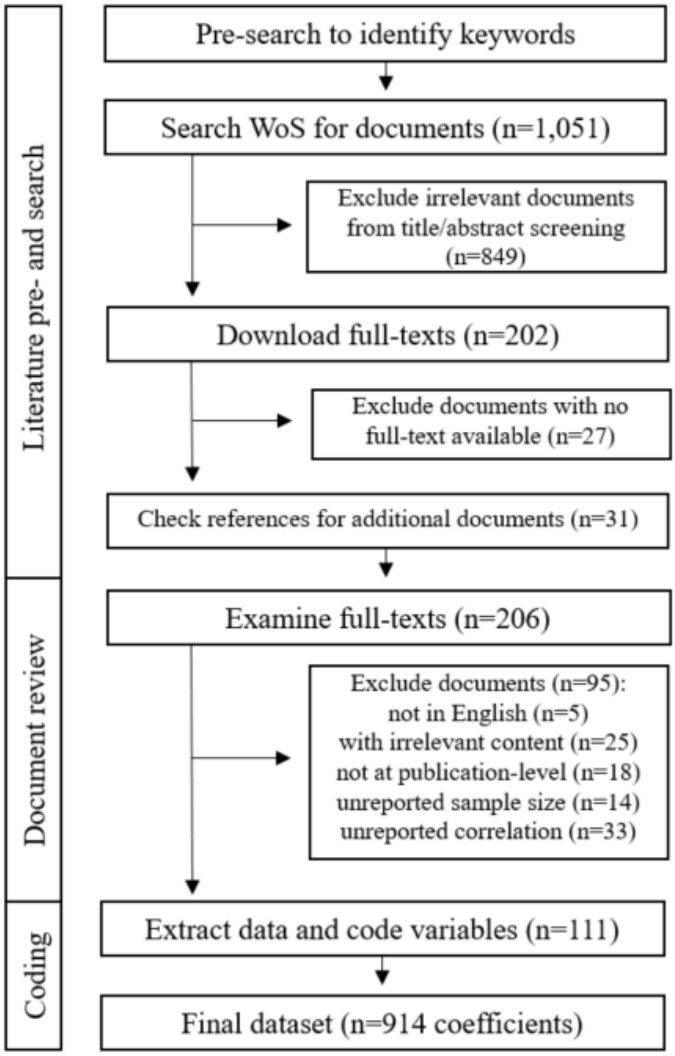

Figure 1. Flowchart of the steps taken for the literature search, review, and coding.

Meta-analytic method

The second phase of the study will be the meta-analysis to examine the strength of the pooled association between citations and several altmetric indicators. The multiple samples obtained from each study were usually independent from one another, however some samples were not. For instance, some studies provided correlations between altmetrics and citations for all publications in a sample and also only the publications in the sample that had non-zero altmetric values. To account for the lack of independence of samples, we will use multi-level random-effects meta-analyses to determine how particular characteristics, such as discipline, citation sources, and the use of non-zero values moderate the relationship between altmetrics and citations, while acknowledging the persisting uncertainty.

Preliminary results

The final number of studies included in our sample was 111 and the total number of correlation coefficients was 914. The studies included were published between 2004 and 2022. The number of publications included in the samples ranged from 3 to 3,808,747, with a mean of 41,665. The correlations observed ranged from -0.48 to 0.95, with a mean of 0.30. However, as shown in Figure 2, the distributions of the correlations observed varied substantially by the source of

altmetric data used. As this is research currently being undertaken, the results of the meta-analysis and moderator analyses are incomplete at this time and will be presented in a final version of this paper.

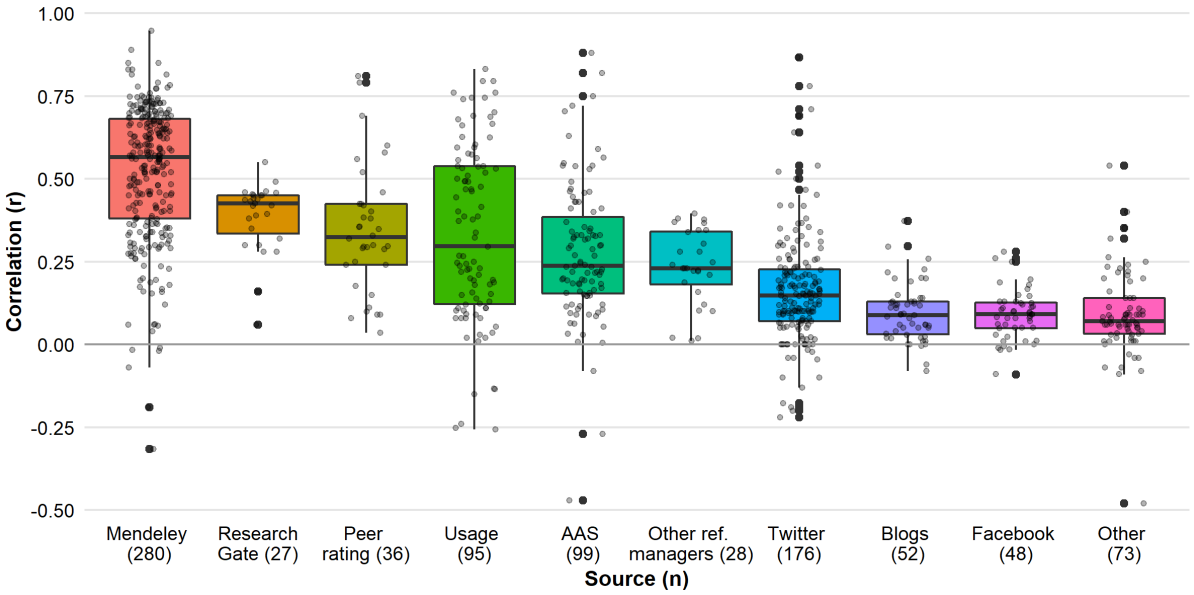

Figure 2. Distribution of observed altmetric-citation correlations by altmetric source.

Acknowledgements

This study was funded by the Federal Ministry of Education and Research of Germany (BMBF) under grant number 16PU17024. This study and the simultaneously submitted study "Mind the gap: Nowcasting the citation impact of research institutions", are studies in the same project.

References

Amath, A., Ambacher, K., Leddy, J. J., Wood, T. J., & Ramnana, C. J. (2017). Comparing alternative and traditional dissemination metrics in medical education. *Medical Education, 51*(9), 935-941, DOI: 10.1111/medu.13359.

Bornmann, L. (2015). Alternative metrics in scientometrics: A meta-analysis of research into three altmetrics. *Scientometrics, 103*, 1123-1144. DOI: 10.1007/s11192-015-1565-y.

Buttliere B., & Buder, J. (2017). Personalizing papers using Altmetrics: Comparing paper 'Quality' or 'Impact' to person 'Intelligence' or 'Personality'. *Scientometrics, 111*, 219-239. DOI: 10.1007/s11192-017-2246-9.

Cho, J. (2021). Altmetrics analysis of highly cited academic papers in the field of library and information science. *Scientometrics, 126*, 7623-7635. DOI: 10.1007/s11192-021-04084-w.

Erdt, M., Nagarajan, A., Sin, S-C. J., & Theng, Y-L. (2016). Altmetrics: An analysis of the state-of-the-art in measuring research impact on social media. *Scientometrics, 109*, 1117-1166. DOI: 10.1007/s11192-016-2077-0.

Kolahi, J., Khazaei, S., Iranmanesh, P., Kim, J., Bang, H., & Khademi, A. (2021). Meta-analysis of correlations between Altmetric Attention Score and citations in health sciences. *BioMed Research International.* DOI: 10.1155/2021/6680764.

Gilpin, A. R. (1993). Table for conversion of Kendall's tau to Spearman's rho within the context of measures of magnitude of effect for meta-analysis. *Educational and Psychological Measurement, 53*(1), 87–92. DOI: 10.1177/0013164493053001007.

Gorraiz, J., Blahous, B., & Wieland, M. (2018). Monitoring the broader impact of the journal publication output on country level: A case study for Austria. In: M. Erdt, A. Sesagiri Raamkumar, E. Rasmussen & YL. Theng (Eds.), *Altmetrics for Research Outputs Measurement and Scholarly*

Information Management. AROSIM 2018. Communications in Computer and Information Science, (pp. 39-62) Springer, Singapore. DOI: 10.1007/978-981-13-1053-9_4.

Haustein, S., Peters, I., Sugimoto, C., Thelwall, M., & Larivière, V. (2014). Tweeting biomedicine: An analysis of tweets and citations in the biomedical literature. *Journal of the Association for Information Science and Technology, 65*(4), 656-669, DOI: 10.1002/asi.23101.

Malecki, A. (2015). PubMed and arXiv vs. Gold Open Access: Citation, Mendeley, and Twitter uptake of academic articles of Iran. In: A. A. Salah, Y. Tonta, A.A. Akdag Salah, C. Sugimoto & U. Al (Eds.), *Proceedings of ISSI 2015 Istanbul: 15th International Society of Scientometrics and Informetrics Conference* (pp. 46-58). Bogaziçi University Printhouse. https://www.issi-society.org/proceedings/issi_2015/0046.pdf.

Priem, J, Taraborelli, D., Groth, P., & Neylon, C. (2010). Altmetrics: A manifesto. Accessed 8 November 2022 from http://altmetrics.org/manifesto.

Shen, H., Xie, J., Li, J., & Cheng, Y. (2021). The correlation between scientific collaboration and citation count at the paper level: a meta-analysis. *Scientometrics, 126*, 3443-3470, DOI: 10.1007/s11192-021-03888-0.

Xia, F., Su, X., Wang, W., Zhang, C., Ning, Z., & Lee, I. (2016). Bibliographic analysis of Nature based on Twitter and Facebook altmetrics data. *PLoS ONE, 11*(12): e0165997. DOI:10.1371/journal.pone.0165997.