



STUDY PROTOCOL

**REVISED** **Altmetric coverage of health research in Ireland 2017-2023: a protocol for a cross-sectional analysis**

[version 3; peer review: 3 approved, 1 not approved]

Melissa K Sharp <sup>1</sup>, Patricia Logullo <sup>2</sup>, Pádraig Murphy <sup>3</sup>, Prativa Baral<sup>4</sup>, Sara Burke <sup>5</sup>, David Robert Grimes <sup>6</sup>, Máirín Ryan<sup>7,8</sup>, Barbara Clyne<sup>1</sup>

<sup>1</sup>Department of Public Health and Epidemiology, RCSI University of Medicine and Health Sciences, Dublin 2, Ireland

<sup>2</sup>Centre for Statistics in Medicine, EQUATOR Network UK Centre, University of Oxford, Oxford, England, UK

<sup>3</sup>School of Communications, Dublin City University, Dublin, Leinster, Ireland

<sup>4</sup>Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA

<sup>5</sup>Centre for Health Policy and Management Discipline of Public Health and Primary Care, The University of Dublin Trinity College, Dublin, Leinster, Ireland

<sup>6</sup>School of Medicine, The University of Dublin Trinity College, Dublin, Leinster, Ireland

<sup>7</sup>Health Information and Quality Authority, Dublin 7, Ireland

<sup>8</sup>Department of Pharmacology and Therapeutics, Trinity Health Sciences, The University of Dublin Trinity College, Dublin, Leinster, Ireland

**V3** **First published:** 18 Jun 2024, 7:36  
<https://doi.org/10.12688/hrbopenres.13895.1>  
**Second version:** 03 Sep 2024, 7:36  
<https://doi.org/10.12688/hrbopenres.13895.2>  
**Latest published:** 22 Oct 2024, 7:36  
<https://doi.org/10.12688/hrbopenres.13895.3>

**Abstract**

**Background**

Scientific publications have been growing exponentially, contributing to an oversaturated information environment. Quantifying a research output’s impact and reach cannot be solely measured by traditional metrics like citation counts as these have a lag time and are largely focused on an academic audience. There is increasing recognition to consider ‘alternative metrics’ or altmetrics to measure more immediate and broader impacts of research. Better understanding of altmetrics can help researchers better navigate evolving information environments and changing appetites for different types of research.

**Objectives**

Our study aims to: 1) analyse the amount and medium of Altmetric coverage of health research produced by Irish organisations (2017 – 2023), identifying changes over time and 2) investigate differences in the amount of coverage between clinical areas (e.g., nutrition vs.

**Open Peer Review**

**Approval Status**

	1	2	3	4
<b>version 3</b> (revision) 22 Oct 2024				
<b>version 2</b> (revision) 03 Sep 2024				
<b>version 1</b> 18 Jun 2024				

- Dimity Stephen** , German Centre for Higher Education Research and Science Studies (DZHW), Berlin, Germany
- Fei Yu** , University of North Carolina at Chapel Hill, North Carolina, USA
- Andreas Nishikawa-Pacher** , Vienna School of International Studies and University

neurology).

## Methods

Using Altmetric institutional access, we will gather data on research outputs published 1 January 2017 through 31 December 2023 from active Irish organisations with Research Organisation Registry (ROR) IDs. Outputs will be deduplicated and stratified by their Australian and New Zealand Standard Research Classification relating to  $\geq 1$  field of health research: Biological Sciences, Biomedical and Clinical Sciences, Chemical Sciences, Health Sciences, and Psychology. We will clean data using R and perform descriptive analyses, establishing counts and frequencies of coverage by clinical area and medium (e.g., traditional news, X, etc.); data will be plotted on a yearly and quarterly basis where appropriate.

## Results and Conclusions

Improved understanding of one's information environment can help researchers better navigate their local landscapes and identify pathways for more effective communication to the public. All R code will be made available open-source, allowing researchers to adapt it to evaluate their local landscapes.

## Keywords

Altmetric, science communication, media coverage, knowledge dissemination, health research

of Vienna, Vienna, Austria

TU Wien, Vienna, Austria

4. **Nicole Llewellyn**, Emory University, Georgia, USA

Any reports and responses or comments on the article can be found at the end of the article.

**Corresponding author:** Melissa K Sharp ([melissasharp@rcsi.ie](mailto:melissasharp@rcsi.ie))

**Author roles:** **Sharp MK:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Software, Supervision, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Logullo P:** Conceptualization, Funding Acquisition, Methodology, Writing – Review & Editing; **Murphy P:** Conceptualization, Funding Acquisition, Methodology, Writing – Review & Editing; **Baral P:** Conceptualization, Funding Acquisition, Methodology, Writing – Review & Editing; **Burke S:** Conceptualization, Funding Acquisition, Methodology, Writing – Review & Editing; **Grimes DR:** Conceptualization, Formal Analysis, Funding Acquisition, Methodology, Writing – Original Draft Preparation; **Ryan M:** Conceptualization, Funding Acquisition, Methodology, Writing – Review & Editing; **Clyne B:** Conceptualization, Funding Acquisition, Methodology, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work is supported by Health Research Board Ireland [ARPP-2023-010].

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2024 Sharp MK *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**How to cite this article:** Sharp MK, Logullo P, Murphy P *et al.* **Altmetric coverage of health research in Ireland 2017-2023: a protocol for a cross-sectional analysis [version 3; peer review: 3 approved, 1 not approved]** HRB Open Research 2024, 7:36 <https://doi.org/10.12688/hrbopenres.13895.3>

**First published:** 18 Jun 2024, 7:36 <https://doi.org/10.12688/hrbopenres.13895.1>

**REVISED Amendments from Version 2**

We have clarified the section relating to Open Alex topics.

**Any further responses from the reviewers can be found at the end of the article**

## Introduction

Scientific publications have grown exponentially in recent years, contributing to an ‘infodemic’ as seen during the Covid-19 pandemic<sup>1-3</sup>. With around 1.5 million new items being added to PubMed per year, or 2 papers per minute, there is a need to demonstrate impact of the research to ensure that the scientific community is not favouring quantity over quality<sup>4</sup>. For many years, impact measures were focused on citation counts which are primarily metrics of influence to other academics working in related fields. However, there have been shifts in the past decade away from traditional bibliometrics like citation counts as they can be poor predictors of quality and impact, have a large lag time, and are largely focused on academics<sup>5,6</sup>. Increasingly researchers, institutions, and funders are considering ‘alternative metrics’ or altmetrics to measure the real-time impacts of research on a broader population<sup>7</sup>.

Traditional bibliometrics and altmetrics are complementary measures which provide a more complete picture of impact<sup>8</sup>. Altmetric tracks the immediate online attention given to scientific publications (<https://www.altmetric.com/about-us/our-data/how-does-it-work/>), making it an invaluable tool in crowded information environments. It provides an ‘Altmetric-Attention Score’ (AAS) which factors in sources from social media (e.g., Facebook, X), YouTube videos, newspapers, policy documents, Wikipedia, question-and-answer sites (e.g., Stack Overflow), and more (<https://www.altmetric.com/about-us/our-data/our-sources/>). The broadness of this measure of impact offers the opportunity capture a more diverse picture of the impact of a piece of research<sup>9</sup>. The AAS has been found to be associated with citation counts<sup>7,10</sup>, journal impact factor<sup>11</sup>, and the likelihood of being cited in policy documents<sup>12</sup>. The attention score also has showed differences during the Covid-19 pandemic where Covid-19 related work had significantly higher AAS than for non-Covid-19 articles in 2020<sup>8</sup>. Despite broad criticism about how the AAS is calculated and its reproducibility, it remains one of the strongest proxies for social attention and is widely used in the health and social sciences<sup>13</sup>, particularly as it the health sciences often show the highest Altmetric data coverage and attention<sup>14-16</sup>. Previous research evaluating coverage of Web of Science documents indexed on [Altmetric.com](https://www.altmetric.com) has shown relatively high percentage of coverage for Ireland (68%), especially in comparison to other European countries<sup>16</sup>.

How research findings are presented through domestic news can influence behaviour and risk perceptions<sup>17-20</sup>. Therefore, it would be beneficial for health researchers and healthcare practitioners to better understand the influence that the dissemination of research publications and their subsequent

coverage can have on public behaviour. Analyses of research publication’s online impact can provide insights on effective communication strategies for research outputs produced during COVID-19, for future pandemics, and in generally oversaturated complex information environments<sup>21</sup>. Despite the shift to online global social media, countries still have their own unique landscape and conditions with varying rates of audience engagement and trust in their local and international news sources<sup>22</sup>.

According to the 2023 Reuters Digital News report<sup>22</sup>, Irish consumers have bucked international trends, with levels of trust in news remaining fairly high. Almost half (47%) agreed that they can trust most news most of the time. Ireland can also be considered an outlier in other ways, with 96% of adults having received the full primary COVID-19 vaccination course in 2022, compared to the EU average of 82%<sup>23,24</sup> and registering the fourth lowest rate of excess deaths among OECD countries during the Covid-19 pandemic (2020–2022). While the strong uptake of vaccination clearly had an impact, evidence-based public health messaging (e.g., <https://ihealth-facts.ie/>) and clear messaging in the mainstream media likely also contributed to beneficial behaviour changes<sup>25,26</sup>.

Ireland has also recently made significant investment in health research and healthcare reforms through its Health Service Executive (HSE) Action Plan for Health Research (2019 – 2029)<sup>27</sup> and Sláintecare reform (initially launched in 2017)<sup>28,29</sup>. The Action Plan emphasizes that dissemination and implementation of research are essential to achieving impactful policy and practices that meet the needs of patients, the health service, and policy makers<sup>27</sup>. Furthermore, from 2020, the Irish Research e-Library (IREL) signed the first open access publishing agreements, providing researchers with easier access to open access publishing<sup>30</sup>. Within this context of healthcare and publication reform and the Covid-19 pandemic, we have proposed to include data prior to these changes and the pandemic, to provide some baseline proxy, as well as data throughout and ‘post’-pandemic. This supports our aim to map a piece of the complex local landscape of research in Ireland, using a cross-sectional analysis of Altmetric data (2017 – 2023) and see how it has evolved since before, during, and after the Covid-19 pandemic. A better understanding of the online impact of recent health research can help researchers, and the communication specialists who help disseminate their work, identify pathways for more effective communication to the public. Innovations and dissemination can be improved through better recognition of changing narratives and key players.

## Objectives

Our primary objective is to analyse the amount and type (i.e., medium) of online attention given to health research produced by Irish organisations in recent years (2017 – 2023). We aim to investigate differences over time to identify changing trends, particularly as the online coverage of health research may have been affected during the Covid-19 pandemic. Our secondary objective is to identify differences in the amount of coverage between areas (e.g., nutrition vs. neurology). Our main research questions are: how did research outputs

change over time (amount, open access status, clinical area prevalence, etc.) and what are the differences in Altmetric coverage of research outputs during this period? We are also interested in: the relationships between the Altmetric data (as indicated by the Altmetric Attention Score) and citation data

## Methods

This project will be a cross-sectional study as it is a snapshot of online attention given to research outputs (i.e., articles, books, and chapters) from one defined time period (2017 – 2023) with the main focus to provide descriptive prevalence insights and changing trajectories of online attention. Project findings will be reported according to the STROBE Statement<sup>31</sup>

### Dataset

Data for this study will be gathered from Altmetric institutional access. Altmetric (altmetric.com) tracks 4,000 global news outlets, X (formerly Twitter), YouTube, Reddit, Stack Overflow (Q&A), a curated list of public Facebook pages, blogs, public policy documents, IFI CLAIMS patents, Wikipedia, Mendeley, and Publons. It uses a unique identifier (e.g., DOI, PubMedID, arXiv ID, ISBN, etc.) to track online attention given to a specific research output. (<https://www.altmetric.com/about-us/our-data/how-does-it-work/>) Research outputs included are largely journal articles although book chapters and books will also be included. Of note, some items within Altmetric are classified as ‘news’ but can be considered articles as they are perspectives, commentaries, overviews, or hot topics.

We will use Altmetric Explorer to search for all research outputs published between 1 January 2017 and 31 December 2023 from Irish organisations that have Research Organisation Registry (ROR) IDs. ROR is a global registry of open persistent identifiers for research organisations which helps link researchers and their outputs to institutions across sectors (e.g., education, government, healthcare, non-profit, etc.) (<https://ror.org/about/>). Altmetric uses the predecessor system, the Global Research Identifier Database (GRID) (<https://www.grid.ac/>) which maps to ROR. As of 9 April 2024, there were 663 research organisations with Ireland listed as their country of address. We searched both active and inactive IDs in case an organisation had outputs during the time period but then decided to inactivate their indexing GRID (e.g., they published in 2017–2019 but then deactivated their ID). The lead author (MKS) and two medical students will use the Altmetric Explorer interface to search for and download research outputs published within our date frame from each individual organisation. Datasets will be tracked using a tracking log in Excel to record the downloader (e.g., MKS), total number of research outputs, number of outputs mentioned, filename, and download date. If an organisation has a least 1 output, we will download their data as a csv file using a standard naming notation (ID\_YYYY-MM-DD). These csv files will be stored in one folder, spot checked for completeness (MKS), then combined into one dataset which will include research output data from all organisations producing output from 2017 – 2023. Of note, the datasets have the same 46 variables, will be downloaded

in UTF-8 to account for non-English characters, and dates will be checked prior to stacking. This tracking log and our RMarkdown code detailing combining of datasets, cleaning, pre-processing, and more can be available on our Open Science Framework accompanying our results manuscript<sup>32</sup>. While the Altmetric API (<https://www.altmetric.com/solutions/altmetric-api/>) is also available for information retrieval, we did not find it suitable for pulling data based on an organisation’s GRID ID.

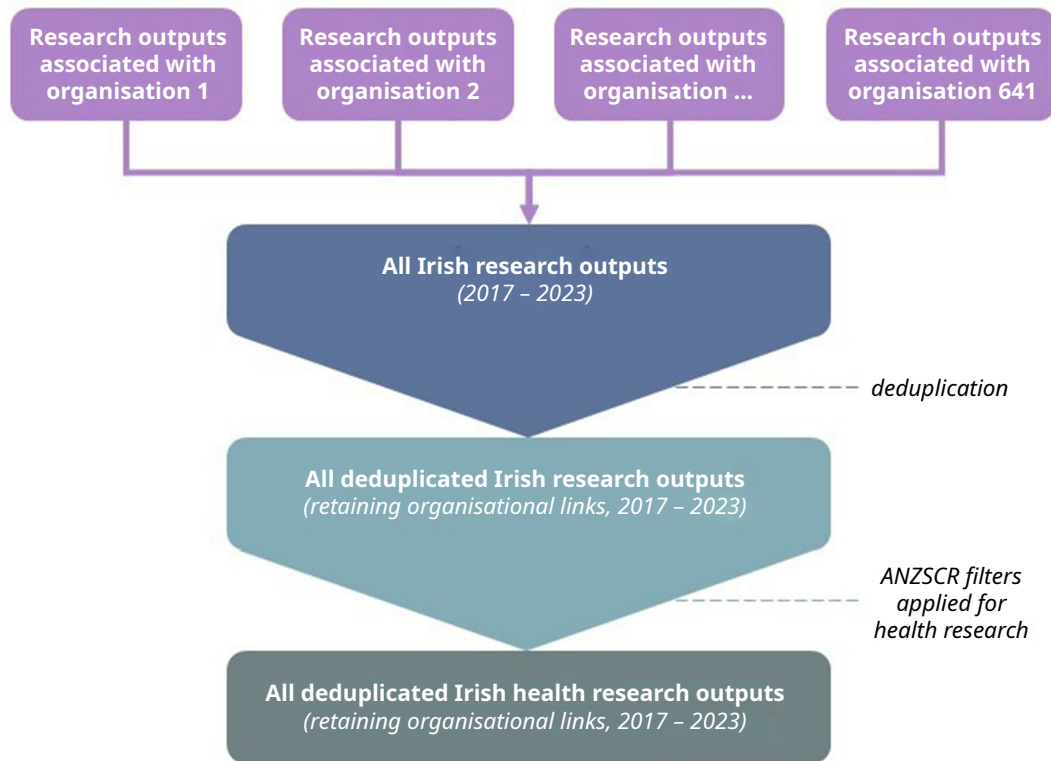
Each dataset (e.g., csv file) contains 46 standard columns or variables, one of which is the field of research. Each research output in the dataset is classified to at least one field of research (FoR) using the 2020 Australian and New Zealand Standard Research Classification (ANZSRC) system. The ANZSRC is a hierarchical system which contains divisions (broad subject areas or research disciplines) which are further detailed into subsets: groups and fields. ANZSRC was developed for use in the measurement and analysis of research and experimental development (R&D) statistics in Australia and New Zealand<sup>33</sup>. In instances where there is a lack of information and it cannot be classified at this level, the code is assigned based on a journal-level classification.

For the purposes of our project we are primarily interested in the following Divisions of biomedical research listed in [Table 1](#): Biological Sciences (31), Biomedical and Clinical Sciences (32), Chemical Sciences (34), Health Sciences (42), and Psychology (52). Excluded areas include: Agricultural, Veterinary and Food Sciences (30), Built Environment and Design (33), Commerce, Management, Tourism and Services (35), Creative Arts and Writing (36), Earth Sciences (37), Education (39), Engineering (40), Environmental Sciences (41), History, Heritage and Archaeology (43), Human Society (44), Indigenous Studies (45), Information and Computing Sciences (46), Language, Communication and Culture (46), Law and Legal Studies (48), Mathematical Sciences (49), Philosophy and Religious Studies (50), and Physical Sciences (51). As research outputs can be classified to several areas, as long as at least one of our included Divisions is in the field, the output will be included. For example, if a study is about the health impacts of environmental pollution and contamination, it could be classified under Biomedical and Clinical Sciences (32) and Environmental Sciences (41), thus it would still be retained.

All individual organisation datasets will be stacked and combined, deduplicated, and filtered to only contain research outputs pertaining to at least one field of health research as defined by the ANZSRC ([Figure 1](#)). Deduplication will be based on an output’s DOI and impact metrics should be identical across outputs with the same DOI. Of note, author affiliations with a research output do not depend on placement (e.g., first, corresponding). We will create a new variable to maintain links in the likely case where a research output is associated with multiple organisations. For example, if research output X was published by both author 1 at organisation A and author 2 at organisation B, the research output should contain the same DOI which is the information that is being used to track all the attention, therefore, that information should

**Table 1. Included divisions of research according to the Australian and New Zealand Standard Research Classification<sup>33</sup>.**

Included Divisions	31 BIOLOGICAL SCIENCES	32 BIOMEDICAL AND CLINICAL SCIENCES	34 CHEMICAL SCIENCES	42 HEALTH SCIENCES	52 PSYCHOLOGY
Included Groups	3101 Biochemistry and cell biology 3102 Bioinformatics and computational biology 3103 Ecology 3104 Evolutionary biology 3105 Genetics 3106 Industrial biotechnology 3107 Microbiology 3108 Plant biology 3109 Zoology 3199 Other biological sciences	3201 Cardiovascular medicine and haematology 3202 Clinical sciences 3203 Dentistry 3204 Immunology 3205 Medicalbiochemistry and metabolomics 3206 Medical biotechnology 3207 Medical microbiology 3208 Medical physiology 3209 Neurosciences 3210 Nutrition and dietetics 3211 Oncology and carcinogenesis 3212 Ophthalmology and optometry 3213 Paediatrics 3214 Pharmacology and pharmaceutical sciences 3215 Reproductive medicine 3299 Other biomedical and clinical sciences	3401 Analytical chemistry 3402 Inorganic chemistry 3403 Macromolecular and materials chemistry 3404 Medicinal and biomolecular chemistry 3405 Organic chemistry 3406 Physical chemistry 3407 Theoretical and computational chemistry 3499 Other chemical sciences	4201 Allied health and rehabilitation science 4202 Epidemiology 4203 Health services and systems 4204 Midwifery 4205 Nursing 4206 Public health 4207 Sports science and exercise 4208 Traditional, complementary and integrative medicine 4299 Other health sciences	5201 Applied and developmental psychology 5202 Biological psychology 5203 Clinical and health psychology 5204 Cognitive and computational psychology 5205 Social and personality psychology 5299 Other



**Figure 1.** Flow diagram of dataset creation.

not change and deduplication does not pose a threat to data loss.

### Analysis

We will use R version 4.3.2 (<https://shiny.rstudio.com/>) to analyse data for this project. As data will be gathered using an institutional license for Altmetric Explorer, we will use R Markdown (<https://rmarkdown.rstudio.com/>) to create an html document to maintain privacy of the proprietary data but promote reproducible research practices. The .rmd file will also be available on GitHub (<https://github.com/sharpmel>) and the project will be registered on the Open Science Framework. (<https://osf.io/kfct6/>)

Data will be cleaned using R and descriptive analyses will be performed for general bibliometric information such as: the type of research output (i.e., article, book, chapter); open access status and type; 20 most frequent journals, funders, and organisations in our dataset; the prevalence of sectors (e.g., education, healthcare); and the five subject areas of health research and their subdivisions (yearly). To investigate trends over time, counts of frequencies will be plotted on a quarterly and yearly basis for the overall number of research outputs and yearly for subject and subdivision areas, sectors, and publishers.

For Altmetric analyses, we will report counts, and averages and medians per medium (e.g., X, Facebook, policy documents) for both the deduplicated dataset as a whole and for the deduplicated dataset with the zero-count AAS removed as there may be large amounts of zero counts in our dataset, potentially skewing measures of central tendency. We will also report the number of outputs with a score of 20 or above as Altmetric has indicated that this is a general score which can be considered as doing better than most of its ‘colleagues’<sup>34</sup>. All medium data will be plotted on a quarterly and yearly basis from 1 January 2017 through 31 December 2023. We will also segment data by subject areas and divisions. As the World Health Organisation (WHO) declared Covid-19 a pandemic on 11 March 2020<sup>35</sup> and no longer a public health emergency of international concern on 5 May 2023<sup>36</sup>, these cut points will be used for general discussions on pre-, during, and post-pandemic.

Assuming an adequately sized dataset, we will also investigate whether Altmetric Attention Scores correlate to traditional article-level citation metrics. Crossref lags significantly behind other major sources in terms of comprehensiveness<sup>37</sup>, therefore we will use OpenAlex to obtain citation data. OpenAlex<sup>38,39</sup> is an open-source catalogue which includes metadata for 209 million outputs like journal articles and

books and its average reference numbers are comparable to Web of Science and Scopus<sup>40</sup>. We will use zero-inflated negative binomial regression to account for the large number of AAS scores of zero. As articles in 2017 have had more time to accrue citations than those published in 2023, we will restrict our analyses and split our dataset by year. Previous research has indicated that a 3-year citation window is relatively stable so we will perform these analyses on the data from 2017–2020 only<sup>41,42</sup>.

**Field and clinical area.** Although the FoR classification system will give some insights as to content areas which are being covered, it does not provide more granular detail. OpenAlex provides a more thorough and in-depth view of academic research as it contains 4,500 topics and builds upon and fills the gap from the discontinued Microsoft Academic Graph (MAG)<sup>43,44</sup>. We will gather topic information by using the OpenAlexAPI to pull this information matched by the DOIs in our final dataset. OpenAlex uses a hierarchical system that organises topics into levels ranging from broad to more specific, from domain to field, subfield, and finally topic system. We will focus on lower level topics and will report the 20 most frequent topics per year.

**Limitations.** Our main limitation is the dataset itself and the quantitative focus of measuring impact. Firstly, Altmetric does not track certain platforms such as LinkedIn, TikTok, and Instagram, thus its generalisability is diminished. It has also been reported to have issues tracking publications with multiple versions (i.e., a pre- and post-print) and the replication of the data can sometimes be difficult due to constantly changing access agreements with data providers<sup>45</sup>.

The lack of a links between preprints (manuscripts uploaded to databases without peer review) and postprints is a larger issue (peer reviewed journal articles) within academic publishing<sup>46</sup> as each item is assigned a unique DOI and there are challenges indexing preprints alongside their peer-reviewed publication. Of note, in the Altmetric dataset, preprints are included and tracked, they just have their own individual altmetrics separate from the final publication. Preprints played an unprecedented role in disseminating Covid-19 research<sup>47</sup>, so we will include them in our dataset, account for this in statistical analyses, and explicitly disclose this preprint prevalence in our dataset, and frame our findings with this in mind.

Furthermore, although higher scores can be expected from newer papers as time since publication has shown to be associated with Altmetric scores<sup>11</sup>, even running searches one month apart resulted in changing numbers as old articles can be brought up for discussion at any time. Altmetric data has also been noted to be prone to manipulation and artificial inflation<sup>48</sup> and some sources are particularly unstable, with certain items ‘vanishing’<sup>13</sup>. We will try to address this by pulling the data in a discrete period of time and we have included a time buffer (i.e., the end of 2023). However, we do recognise that certain mediums may still be ‘incomplete’ as they have different trajectories of attention growth – for

example, Twitter (X) attention starts and ends quickly, Mendeley readers accumulate quickly but continue to grow over the years, and policy attention is the slowest form of impact to accumulate<sup>49</sup>.

The quantitative focus of the metrics in the dataset can also lose the context or tone of the coverage and does not account for ‘dose’, i.e., where a mention may be an extensive discussion or extremely brief. A high AAS does not indicate a high-quality piece of research<sup>50</sup> and the AAS is a measure of attention, not quality. However, recent work comparing altmetrics to norm-referenced peer review scores from the UK Research Excellence Framework 2021 found that Altmetric correlated more strongly with research quality than previously thought although there is large variability with the strength of correlations amongst mediums and between fields (e.g., stronger in health and physical sciences than in the arts and humanities)<sup>14</sup>.

Providing a broad overview of the online attention given to health research in Ireland in recent years is our primary objective. Our datasets are not meant for social listening purposes and audience metrics may be limited. However, our project’s results may provide a basis to build upon for future studies investigating *how* the media is actually covering the work. Altmetric may also be a flawed metric of impact on the public as previous research has shown that most tweets came from within academia with other academics interacting with them<sup>51</sup>. Notably, the data in our project likely will contain more health research produced from the academic sector as they primarily communicate via articles, books, and chapters, however, we have included pharmaceutical agencies, governmental health bodies, and hospitals (where they have an ROR) which may provide a broader overview of the online coverage of health research in Ireland. A recent bibliometric analysis of HRB supported publications<sup>52</sup> found the academic sector well-represented although our project is much broader in its scope.

## Discussion and implications

A better understanding of the amount and type of online attention given to health research involving Irish organisations can identify changing trends and gaps in attention. By looking at organisation-linked data and unique research outputs, we can provide insights to researchers and organisations (particularly universities) looking to evaluate the impact of their work and identify the strengths and weaknesses of their research portfolios. Results may be particularly useful for researchers and communication specialists who are aiming disseminate their research to the public and find ‘airtime’ in a particularly noisy information environment. Our use of open source coding also will offer a reproducible workflow for future monitoring and further investigations into the content of the health research coverage itself. Results from our ‘case study’ focused on the Irish landscape can also be used to compare and contrast with other countries. The time period chosen for our project may be of particular interest due to the variation in governmental responses to Covid-19 and differing rates of personal protective behaviours (e.g., vaccinations, masking, etc.). Overall, the project should provide us with

a piece of the puzzle of the landscape of online attention given to health research in Ireland.

## Data availability

No data are associated with this article.

## References

- Borges do Nascimento IJ, Pizarro AB, Almeida JM, et al.: **Infodemics and health misinformation: a systematic review of reviews.** *Bull World Health Organ.* 2022; **100**(9): 544–61.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- The Lancet Infectious Diseases: **The COVID-19 infodemic.** *Lancet Infect Dis.* 2020; **20**(8): 875.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Calleja N, AbdAllah A, Abad N, et al.: **A public health research agenda for managing infodemics: methods and results of the first WHO infodemiology conference.** *JMIR Infodemiology.* 2021; **1**(1): e30979.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Novoa J, Chagoyen M, Benito C, et al.: **PMIDigest: interactive review of large collections of PubMed entries to distill relevant information.** *Genes (Base).* 2023; **14**(4): 942.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dougherty MR, Horne Z: **Citation counts and journal impact factors do not capture some indicators of research quality in the behavioural and brain sciences.** *R Soc Open Sci.* 2022; **9**(8): 220334.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Worrall JL, Cohn EG: **Citation data and analysis: limitations and shortcomings.** *J Contemp Crim Justice.* 2023; **39**(3): 327–40.  
[Publisher Full Text](#)
- Kolahi J, Khazaei S, Iranmanesh P, et al.: **Meta-analysis of correlations between altmetric attention score and citations in health sciences.** *Biomed Res Int.* 2021; **2021**: 6680764.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Peterson CJ, Anderson C, Nugent K: **Alternative publication metrics in the time of COVID-19.** *Proc (Bayl Univ Med Cent).* 2022; **35**(1): 43–5.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Fang Z, Costas R, Tian W, et al.: **An extensive analysis of the presence of altmetric data for web of science publications across subject fields and research topics.** *Scientometrics.* 2020; **124**(3): 2519–2549.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tornberg H, Moezinia C, Wei C, et al.: **Assessment of the dissemination of COVID-19-related articles across social media: altmetrics study.** *JMIR Form Res.* 2023; **7**(1): e41388.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Araujo AC, Vanin AA, Nascimento DP, et al.: **What are the variables associated with Altmetric scores?** *Syst Rev.* 2021; **10**(1): 193.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mullins CH, Boyd CJ, Ladowski JM: **The association between Altmetric Attention Scores and public engagement in the medical literature.** *J Surg Res.* 2023; **292**: 324–9.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Arroyo-Machado W, Torres-Salinas D: **Stranger things: the vanishing of the Altmetric Attention Score values in information and library science.** *Scientometrics.* 2024; [cited 2024 Jul 18].  
[Publisher Full Text](#)
- Thelwall M, Kousha K, Abdoli M, et al.: **Do altmetric scores reflect article quality? Evidence from the UK research excellence framework 2021.** *J Assoc Inf Sci Technol.* 2023; **74**(5): 582–93.  
[Publisher Full Text](#)
- Costas R, Zahedi Z, Wouters P: **Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective.** *J Assoc Inf Sci Technol.* 2015; **66**(10): 2003–19.  
[Publisher Full Text](#)
- Torres-Salinas D, Robinson-García N, Arroyo-Machado W: **Coverage and distribution of altmetric mentions in Spain: a cross-country comparison in 22 research fields.** *Profesional de la Información.* 2022; **31**(2).  
[Publisher Full Text](#)
- Qazi A, Qazi J, Naseer K, et al.: **Analyzing situational awareness through public opinion to predict adoption of social distancing amid pandemic COVID-19.** *J Med Virol.* 2020; **92**(7): 849–55.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sandell T, Sebar B, Harris N: **Framing risk: communication messages in the Australian and Swedish print media surrounding the 2009 H1N1 pandemic.** *Scand J Public Health.* 2013; **41**(8): 860–5.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Ogbodo JN, Onwe EC, Chukwu J, et al.: **Communicating health crisis: a content analysis of global media framing of COVID-19.** *Health Promot Perspect.* 2020; **10**(3): 257–69.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sallam M, Dababseh D, Yaseen A, et al.: **COVID-19 misinformation: mere harmless delusions or much more? A knowledge and attitude cross-sectional study among the general public residing in Jordan.** *medRxiv.* 2020; 2020.07.13.20152694.  
[Publisher Full Text](#)
- Raamkumar AS, Tan SG, Wee HL: **Measuring the outreach efforts of Public Health Authorities and the public response on facebook during the COVID-19 pandemic in early 2020: cross-country comparison.** *J Med Internet Res.* 2020; **22**(5): e19334.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Newman N, Fletcher R, Eddy K, et al.: **Reuters Institute digital news report 2023.** 2023; [cited 2024 Mar 7].  
[Reference Source](#)
- European Observatory on Health Systems and Policies: **State of Health in the EU Ireland Country Health Profile 2023.**  
[Reference Source](#)
- Report reveals that Ireland has the highest rate of reported good health in EU.** 2023; [cited 2024 Mar 7].  
[Reference Source](#)
- Sharp MK, Forde Z, McGeown C, et al.: **Irish media coverage of COVID-19 evidence-based research reports from one national agency.** *Int J Health Policy Manag.* 2022; **11**(11): 2464–2475.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Wheatley D: **Irish audiences and news information from official sources during COVID-19.** *Administration.* 2022; **70**(3): 7–32.  
[Publisher Full Text](#)
- Terrés AM: **HSE Action Plan for Health Research 2019–2029.** Health Service Executive; 2019; 36.  
[Reference Source](#)
- Delivering Sláintecare Reform.** 2019; [cited 2024 Mar 7].  
[Reference Source](#)
- Houses of the Oireachtas: **Committee on the future of healthcare sláintecare report.** 2017.  
[Reference Source](#)
- About IReL - IReL.** [cited 2024 Jul 24].  
[Reference Source](#)
- Vandenbroucke JP, von Elm E, Altman DG, et al.: **Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration.** *PLoS Med.* 2007; **4**(10): e297.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- OSF.** [cited 2023 Apr 3].  
[Reference Source](#)
- Australian and New Zealand Standard Research Classification (ANZSRC).** Australian Bureau of Statistics, 2020; [cited 2024 Jan 9].  
[Reference Source](#)
- Altmetric: **Altmetric attention score in context.** [cited 2024 Jul 18].  
[Reference Source](#)
- WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020.** [cited 2024 Feb 28].  
[Reference Source](#)
- Statement on the fifteenth meeting of the IHR: Emergency committee on the COVID-19 pandemic.** 2005; [cited 2024 Feb 28].  
[Reference Source](#)

37. Gerasimov I, Binita KC, Mehrabian A, *et al.*: **Comparison of datasets citation coverage in google scholar, web of science, scopus, crossref, and datacite.** *Scientometrics*. 2024; **129**: 3681–3704.  
[Publisher Full Text](#)
38. **OpenAlex.** About us. 2024; [cited 2024 Jul 30].  
[Reference Source](#)
39. Chawla DS: **Massive open index of scholarly papers launches.** *Nature*. 2022; [cited 2024 Jul 30].  
[PubMed Abstract](#) | [Publisher Full Text](#)
40. Culbert J, Hobert A, Jahn N, *et al.*: **Reference coverage analysis of openalex compared to web of science and scopus.** *arXiv*. 2024; [cited 2024 Jul 25].  
[Publisher Full Text](#)
41. Clermont M, Krolak J, Tunger D: **Does the citation period have any effect on the informative value of selected citation indicators in research evaluations?** *Scientometrics*. 2021; **126**(2): 1019–47.  
[Publisher Full Text](#)
42. Glänzel W: **Seven myths in bibliometrics about facts and fiction in quantitative science studies.** *COLLNET Journal of Scientometrics and Information Management*. 2008; **2**(1): 9–17.  
[Publisher Full Text](#)
43. Google Docs: **OpenAlex Topic Classification Whitepaper.docx.** [cited 2024 Oct 13].  
[Reference Source](#)
44. Google Docs: **OpenAlex\_topic\_mapping\_table.** [cited 2024 Aug 19].  
[Reference Source](#)
45. Bornmann L: **Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics.** *J Informetr*. 2014; **8**(4): 895–903.  
[Publisher Full Text](#)
46. Levchenko M, Parkin M, McEntyre J, *et al.*: **Enabling preprint discovery, evaluation, and analysis with Europe PMC.** *bioRxiv*. 2024; [cited 2024 Jul 24], 2024.04.19.590240.  
[Publisher Full Text](#)
47. Fraser N, Brierley L, Dey G, *et al.*: **The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape.** *PLoS Biol*. 2021; **19**(4): e3000959.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
48. Gupta S, Singh VK, Banshal SK: **Altmetric data quality analysis using Benford's law.** *Scientometrics*. 2024; **129**(1): 4597–4621.  
[Publisher Full Text](#)
49. Taylor M: **Slow, slow, quick, quick, slow: five altmetric sources observed over a decade show evolving trends, by research age, attention source maturity and open access status.** *Scientometrics*. 2023; **128**(4): 2175–200.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
50. Khatter A, Naughton M, Dambha-Miller H, *et al.*: **Is rapid scientific publication also high quality? Bibliometric analysis of highly disseminated COVID-19 research papers.** *Learn Publ*. 2021; **34**(4): 568–77.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
51. Zhang L, Gou Z, Fang Z, *et al.*: **Who tweets scientific publications? A large-scale study of tweeting audiences in all areas of research.** *J Assoc Inf Sci Technol*. 2023; **74**(13): 1485–97.  
[Publisher Full Text](#)
52. Draux H, Wastl J: **Bibliometric analysis of HRB supported publications.**  
[Reference Source](#)

## Open Peer Review

Current Peer Review Status:    

---

### Version 3

Reviewer Report 04 November 2024

<https://doi.org/10.21956/hrbopenres.15381.r42908>

© 2024 Nishikawa-Pacher A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Andreas Nishikawa-Pacher** 

<sup>1</sup> Vienna School of International Studies and University of Vienna, Vienna, Austria

<sup>2</sup> Bibliothek, TU Wien, Vienna, Austria

The minor issue about the potential confusion between 'concepts' and 'topics' in OpenAlex has been resolved.

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** bibliometrics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

### Version 2

Reviewer Report 30 September 2024

<https://doi.org/10.21956/hrbopenres.15326.r42289>

© 2024 Llewellyn N. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Nicole Llewellyn**

Georgia Clinical and Translational Science Alliance, Emory University, Georgia, USA

This article presents a protocol for an investigation of Altmetric attention to health related research conducted at Irish institutions from 2017-2023. The introduction makes solid arguments

for examining Altmetrics, which is a novel area of publication analysis that itself needs more attention in the literature. The methodology is thorough and well-explained. Suggest describing what STROBE is, adding the caveat that an AAS of 20+ was defined before COVID and may not apply the same way now, and suggest adding the relevant reference to Llewellyn & Nehl, 2022 [Ref 1] regarding AAS and adjusted citation scores.

### References

1. Llewellyn NM, Nehl EJ: Predicting citation impact from altmetric attention in clinical and translational research: Do big splashes lead to ripple effects?. *Clin Transl Sci.* 2022; **15** (6): 1387-1392 [PubMed Abstract](#) | [Publisher Full Text](#)

### Is the rationale for, and objectives of, the study clearly described?

Yes

### Is the study design appropriate for the research question?

Yes

### Are sufficient details of the methods provided to allow replication by others?

Yes

### Are the datasets clearly presented in a useable and accessible format?

Not applicable

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Evaluation, health publication analysis

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 19 September 2024

<https://doi.org/10.21956/hrbopenres.15326.r42245>

© 2024 Nishikawa-Pacher A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Andreas Nishikawa-Pacher** 

<sup>1</sup> Vienna School of International Studies and University of Vienna, Vienna, Austria

<sup>2</sup> Bibliothek, TU Wien, Vienna, Austria

The authors provided thoughtful responses, and the revised manuscript does seem to convey a more feasible approach.

I have only one comment left regarding the OpenAlex-based "topics". I am not sure whether the

authors may have mistaken, in parts, the OpenAlex-based "topics" with the OpenAlex-based "concepts", the latter of which has been deprecated (if I remember correctly). I can understand that one would mistake them given that they have a similar function, and given that the "topics" have only recently replaced the "concepts" in OpenAlex.

The (outdated) "concepts" used to have levels (like level 1, level 2), and perhaps keywords like the ones cited by the authors ("anxiety", "telemedicine").

In contrast, the (newer) "topics" do not have "levels" anymore (but a "topic" is part of a "subfield" which, in turn, is part of a "field" which, in turn, is part of a "domain").

In addition, the "topics" are more concrete. For example, instead of a concept called "anxiety", there are three topics that are named as such:

- 1.) "Cognitive Mechanisms of Anxiety and Depression"
- 2.) "The Relationship Between Music and Anxiety Management"
- 3.) "Dental Anxiety and Anesthetic Management in Dentistry"

Here is an API-link to the example of "Dental Anxiety ...": <https://openalex.org/topics/T12145>

The authors certainly looked at "topics" correctly (for instance, ref. 44 leads to a Google Sheets page enumerating all the "topics"), but I think that the paragraph they inserted on "topics" also contained, perhaps erroneously, some references to the outdated "concepts".

This is not a huge matter, but I thought I'd point it out in case it may be helpful.

Other than that, the draft seems very much enhanced!

**Is the rationale for, and objectives of, the study clearly described?**

Not applicable

**Is the study design appropriate for the research question?**

Not applicable

**Are sufficient details of the methods provided to allow replication by others?**

Not applicable

**Are the datasets clearly presented in a useable and accessible format?**

Not applicable

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** bibliometrics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 13 Oct 2024

**Melissa Sharp**

Thank you for looking at our new version and for your thoughtful comments and feedback. We agree that the text in our second version may have caused confusion between concepts and topics. As you noted, due to the deprecation of concepts, we will use topics. We have edited the section accordingly.

**Competing Interests:** No competing interests were disclosed.

---

**Version 1**

Reviewer Report 19 July 2024

<https://doi.org/10.21956/hrbopenres.15238.r40895>

© 2024 Nishikawa-Pacher A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Andreas Nishikawa-Pacher**

<sup>1</sup> Vienna School of International Studies and University of Vienna, Vienna, Austria

<sup>2</sup> Bibliothek, TU Wien, Vienna, Austria

Thank you for allowing me to review this study protocol which eyes to study the altmetric coverage of health research produced by Irish organisations, including with disaggregations regarding specific clinical areas and study types.

While I am no expert in the methodical approach as such (for instance, I cannot comment on the soundness of the use of latent Dirichlet allocation in this particular context), from the viewpoint of someone who engages in scientometric studies, the study protocol seems to be fine and worthwhile to be undertaken.

The reasons given for utility and the prospects of the study are well-argued. As regards the methods, the data sources should be good to use. From my experience, it should be possible to combine ROR/GRID with the Altmetric API without too many data losses (though there will certainly be a few omissions when affiliation names are not stored correctly in the publishers' metadata).

One aspect that I am uncertain about is whether the topic modelling approach for obtaining the research field and clinical area is really that helpful, especially since the topic modelling eyed would only look at the title of each publication. It would be interesting to find out whether the "topics" that are available as open data from OpenAlex could likewise serve as a good, if not better, equivalent regarding the clinical area, since OpenAlex' "topic"-assignment is based not only

on the title, but also on the abstract, the journal/source, and the references of each respective paper (see: <https://docs.openalex.org/api-entities/topics>).

With regards to the classification of study design, the approach sounds innovative; but I am also unsure about the success their approach will yield. It is certainly worthwhile to try it out, and it would be interesting to see the result. (And just as a sidenote, I wonder whether there is a good theoretical reason to believe that some study designs will structurally have a higher Altmetric Attention Score. For instance, would papers with 'randomized controlled trial' in the title or in the keywords get more attention in mass media because journalists may think that such studies are more reliable than others? Why would that be? Would a 'systematic review' catch more attention than a 'scoping review'? This is not to say that I'd demand from the authors that they come up with convincing theories; it is just an intriguing idea that was stimulated by the authors' study protocol.)

When it comes to the Discussion section, one future study that might be interesting would be to speculate whether a comparison between Ireland and another countries (e.g., one which, in the aggregate, was more sceptical regarding the COVID-19 vaccinations) would yield different results (given the reasons stated in the Introduction for the choice of research object).

Just as minor remark, I think the second paragraph should speak of an "Altmetric Attention Score" (rather than Altmetric-mentioned score).

Finally, I agree with the other peer-reviewer (Dimitry Stephan) that high Altmetric Attention Scores do not mean that a paper is really impactful & of a high quality. In fact, there are studies suggesting that high Altmetric Attention Scores can serve as a predictor of a retraction [Ref -1]. And there are enough "funny" or "strange" papers that lead to almost viral discussions in social media precisely because of their errors and flaws.

Anyway, the study protocol is well thought out and promises interesting results. I wish the authors all the best for their undertaking. The only change I would like to see as a reviewer is an explicit discussion on OpenAlex' "topics"; other than that, the study protocol seems perfect from me (notwithstanding my inability to judge on the exact statistical topic modelling method).

### References

1. Serghiou S, Marton RM, Ioannidis JPA: Media and social media attention to retracted articles according to Altmetric. *PLoS One*. 2021; **16** (5): e0248625 [PubMed Abstract](#) | [Publisher Full Text](#)

**Is the rationale for, and objectives of, the study clearly described?**

Yes

**Is the study design appropriate for the research question?**

Yes

**Are sufficient details of the methods provided to allow replication by others?**

Yes

**Are the datasets clearly presented in a useable and accessible format?**

Not applicable

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** bibliometrics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.**

Author Response 19 Aug 2024

**Melissa Sharp**

Thank you for allowing me to review this study protocol which eyes to study the altmetric coverage of health research produced by Irish organisations, including with disaggregations regarding specific clinical areas and study types.

While I am no expert in the methodical approach as such (for instance, I cannot comment on the soundness of the use of latent Dirichlet allocation in this particular context), from the viewpoint of someone who engages in scientometric studies, the study protocol seems to be fine and worthwhile to be undertaken.

The reasons given for utility and the prospects of the study are well-argued. As regards the methods, the data sources should be good to use. From my experience, it should be possible to combine ROR/GRID with the Altmetric API without too many data losses (though there will certainly be a few omissions when affiliation names are not stored correctly in the publishers' metadata).

**Response:** Thank you for the time spent reviewing our manuscript and for providing thoughtful and thorough feedback! The comment regarding the quality of the metadata is an important one (17). We will perform spot-checking, combining data sources and monitoring any retrieval issues, and make our R code open source to promote reproducibility and transparent reporting.

One aspect that I am uncertain about is whether the topic modelling approach for obtaining the research field and clinical area is really that helpful, especially since the topic modelling eyed would only look at the title of each publication. It would be interesting to find out whether the "topics" that are available as open data from OpenAlex could likewise serve as a good, if not better, equivalent regarding the clinical area, since OpenAlex' "topic"-assignment is based not only on the title, but also on the abstract, the journal/source, and the references of each respective paper (see: <https://docs.openalex.org/api-entities/topics>).

**Response:** Thank you for the suggestion of OpenAlex, we have added a section to the methods section to pivot to this approach: *"OpenAlex provides a more thorough and in-depth view of academic research as it contains 65,000 Wikidata concepts based on Microsoft Academic Graph (MAG)(9) and enhanced with machine learning, natural language processing, citation analysis, an expert feedback. We will gather the topic-, mid-, and lower- level concepts and subfield information by using the OpenAlexAPI to pull this information matched by the DOIs in our final dataset. Mid- and lower-level concepts could include things like 'pediatric oncology', 'telemedicine', or 'anxiety.' We will report the top 20 terms per year and if possible visualise this information using the VOS-Viewer free software. (11)"*

With regards to the classification of study design, the approach sounds innovative; but I am also unsure about the success their approach will yield. It is certainly worthwhile to try it out, and it would be interesting to see the result. (And just as a sidenote, I wonder whether there is a good theoretical reason to believe that some study designs will structurally have a higher Altmetric Attention Score. For instance, would papers with 'randomized controlled trial' in the title or in the keywords get more attention in mass media because journalists may think that such studies are more reliable than others? Why would that be? Would a 'systematic review' catch more attention than a 'scoping review'? This is not to say that I'd demand from the authors that they come up with convincing theories; it is just an intriguing idea that was stimulated by the authors' study protocol.)

**Response:** Thank you for this note. Anecdotally we have heard from colleagues in pre-clinical and evidence synthesis work that it is harder to get public interest and engagement in their work. This led us to the initial proposal. However, due to the final size of our dataset we no longer thought it was feasible to conduct the study design exploration aspect of our project and have taken this information out of our protocol (both study objectives and the study design section. However, we will still be creating the dataset which perhaps can be used to explore this in the future.

When it comes to the Discussion section, one future study that might be interesting would be to speculate whether a comparison between Ireland and another countries (e.g., one which, in the aggregate, was more sceptical regarding the COVID-19 vaccinations) would yield different results (given the reasons stated in the Introduction for the choice of research object).

**Response:** This is an important note. We have added related text to the discussion and implications section: *"Results from our 'case study' focused on the Irish landscape can also be used to compare and contrast with other countries. The time period chosen for our project may be of particular interest due to the variation in governmental responses to Covid-19 and differing rates of personal protective behaviours (e.g., vaccinations, masking, etc.)."*

Just as minor remark, I think the second paragraph should speak of an "Altmetric Attention Score" (rather than Altmetric-mentioned score).

**Response:** We have fixed this typo.

Finally, I agree with the other peer-reviewer (Dimitry Stephan) that high Altmetric Attention Scores do not mean that a paper is really impactful & of a high quality. In fact, there are studies suggesting that high Altmetric Attention Scores can serve as a predictor of a retraction [Ref -1]. And there are enough "funny" or "strange" papers that lead to almost viral discussions in social media precisely because of their errors and flaws.

**Response:** We agree that a high AAS should not be synonymous with good or indicate a high-quality piece of research. We have added text relating to this to the limitations section: *"A high AAS does not indicate a high-quality piece of research(26) and the AAS is a measure of attention, not quality. However, recent work comparing altmetrics to norm-referenced peer review scores from the UK Research Excellence Framework 2021 found that Altmetrics correlated more strongly with research quality than previously thought although there is large variability with the strength of correlations amongst mediums and between fields (e.g., stronger in health and physical sciences than in the arts and humanities). (3)"*

Anyway, the study protocol is well thought out and promises interesting results. I wish the authors all the best for their undertaking. The only change I would like to see as a reviewer is an explicit discussion on OpenAlex' "topics"; other than that, the study protocol seems perfect from me (notwithstanding my inability to judge on the exact statistical topic modelling method).

**Response:** Per our earlier response, we have pivoted our approach and removed the LDA topic modelling. Thank you for your thoughtful and positive comments.

**Competing Interests:** No competing interests were disclosed.

Reviewer Report 16 July 2024

<https://doi.org/10.21956/hrbopenres.15238.r40898>

© 2024 Yu F. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Fei Yu 

University of North Carolina at Chapel Hill, North Carolina, USA

The manuscript analyzes the online attention of health research outputs produced by Irish organizations from 2017 to 2023 using Altmetric data. It emphasizes using altmetrics as an alternative measure to address the limitations and biases of traditional bibliometric measures, such as raw citation counts. However, several areas could benefit from further clarification and improvement.

**Areas for Improvement:**

**1. Literature Review:** The introduction could include a more extensive review of relevant literature on applying altmetrics to assess online attention (digital impact) in the field of biomedical and health sciences. A simple search in PubMed will yield many relevant studies. This would provide a stronger foundation for the study's rationale and methodology.

**2. Research Questions:** Although the authors state their primary and secondary objectives, no specific research questions are proposed to guide the data search and analysis.

**3. Methodological Details:**

- It would be helpful to include more details on the data cleaning and preprocessing steps, particularly regarding how duplicates will be handled and how the dataset will be validated. For example, will the dataset be manually screened by two reviewers independently to ensure its validity and relevance?
- The manuscript should provide a rationale for the selected time frame (2017-2023) and explain any potential biases this period may introduce. This time range covers before, during, and after COVID-19. The authors need to explain what potential changes they expect to observe and why the changes are important.
- If Table 1 is based on the Altmetric help page of "field of research," the authors need to cite the source.

- Since there are no specific research questions or theories/models to guide the data extraction, the authors need to explain the topic mapping using titles. There are existing bibliometric tools for topic mapping (e.g., VOSviewer, Biblioshiny). The authors need to explain why they chose to use titles only rather than abstracts, author keywords, or MeSH terms.
- Additionally, the Altmetric platform supports PubMed search queries. The authors might want to explain why they prefer API for data retrieval instead of the user interface.
- Altmetric.com recommends using an Altmetric attention score of 20 as a benchmark. The authors did not mention this for quantitative analysis.

**4. Limitations:** The manuscript could benefit from addressing potential biases in Altmetric data collection and how these might affect the study's findings.

**Specific Comments:**

- The statement "With around 1.5 million new items being added to PubMed per year, or 2 papers per minute..." could include a reference to the source of this statistic for credibility.
- The sentence "Ireland can also be consider an outlier in other ways..." has a grammatical error ("consider" should be "considered").

Overall, the manuscript presents a potentially valuable contribution to understanding online attention to health research in Ireland. However, the research protocol needs improvement, particularly in the areas of method design and justification.

**Is the rationale for, and objectives of, the study clearly described?**

Yes

**Is the study design appropriate for the research question?**

Partly

**Are sufficient details of the methods provided to allow replication by others?**

Partly

**Are the datasets clearly presented in a useable and accessible format?**

No

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Health sciences research, bibliometrics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.**

Author Response 19 Aug 2024

**Melissa Sharp**

The manuscript analyzes the online attention of health research outputs produced by Irish organizations from 2017 to 2023 using Altmetric data. It emphasizes using altmetrics as an alternative measure to address the limitations and biases of traditional bibliometric

measures, such as raw citation counts. However, several areas could benefit from further clarification and improvement.

**Response:** Thank you for the time spent reviewing our protocol and for the thoughtful comments, questions, and suggestions.

Areas for Improvement:

1. Literature Review: The introduction could include a more extensive review of relevant literature on applying altmetrics to assess online attention (digital impact) in the field of biomedical and health sciences. A simple search in PubMed will yield many relevant studies. This would provide a stronger foundation for the study's rationale and methodology.

**Response:** We have edited and added additional information about the value of altmetrics/Altmetric as a measure of alternative impact of research to the paragraph describing Altmetric: *"The AAS has been found to be associated with citation counts, 7, 9 journal impact factor 10, and the likelihood of being cited in policy documents. 11 The attention score also has showed differences during the Covid-19 pandemic where Covid-19 related work had significantly higher AAS than for non-Covid-19 articles in 2020. (1) Despite broad criticism about how the AAS is calculated and its reproducibility, it remains one of the strongest proxies for social attention and is widely used in the health and social sciences (2), particularly as it the health sciences often show the highest Altmetric data coverage and attention. (3–5) Previous research evaluating coverage of Web of Science documents indexed on Altmetric.com has shown relatively high percentage of coverage for Ireland (68%), especially in comparison to other European countries. (5)"*

2. Research Questions: Although the authors state their primary and secondary objectives, no specific research questions are proposed to guide the data search and analysis.

**Response:** We have clarified our research questions in the objectives section and have removed part of our secondary objective (relating to study types): *"Our main research questions are: how did research outputs change over time (amount, open access status, clinical area prevalence, etc.) and what are the differences in Altmetric coverage of research outputs during this period? We are also interested in: how are the relationships between the Altmetric data (as indicate by the Altmetric Attention Score) and citation data?"*

3. Methodological Details:

It would be helpful to include more details on the data cleaning and preprocessing steps, particularly regarding how duplicates will be handled and how the dataset will be validated. For example, will the dataset be manually screened by two reviewers independently to ensure its validity and relevance?

**Response:** Thank you for this suggestion. We have clarified the information regarding duplicates and data processing and validation: *"We will use Altmetric Explorer to search for all research outputs published between 1 January 2017 and 31 December 2023 from Irish organisations that have Research Organisation Registry (ROR) IDs. ROR is a global registry of open persistent identifiers for research organisations which helps link researchers and their outputs to institutions across sectors (e.g., education, government, healthcare, non-profit, etc.) (<https://ror.org/about/>). Altmetric uses the predecessor system, the Global Research Identifier Database (GRID) (<https://www.grid.ac/>) which maps to ROR. As of 9 April 2024, there were 663 research organisations with Ireland listed as their country of address. We searched both active and inactive IDs in case an organisation had outputs during the time period but then decided to*

*inactivate their indexing GRID. (e.g., they published in 2017-2019 but then deactivated their ID). The lead author (MKS) and two medical students will use the Altmetric Explorer interface to search for and download research outputs published within our date frame from each individual organisation. Datasets will be tracked using a tracking log in Excel to record the downloader (e.g., MKS), total number of research outputs, number of outputs mentioned, filename, and download date. If an organisation has a least 1 output, we will download their data as a csv file using a standard naming notation (ID\_YYYY-MM-DD). These csv files will be stored in one folder, spot checked for completeness (MKS), then combined into one dataset which will include research output data from all organisations producing output from 2017 - 2023. Of note, the datasets have the same 46 variables, will be downloaded in UTF-8 to account for non-English characters, and dates will be checked prior to stacking. This tracking log and our RMarkdown code detailing combining of datasets, cleaning, pre-processing, and more can be available on our Open Science Framework accompanying our results manuscript.(6) While the Altmetric API (<https://www.altmetric.com/solutions/altmetric-api/>) is also available for information retrieval, we did not find it suitable for pulling data based on an organisation's GRID ID."*

The manuscript should provide a rationale for the selected time frame (2017-2023) and explain any potential biases this period may introduce. This time range covers before, during, and after COVID-19. The authors need to explain what potential changes they expect to observe and why the changes are important.

**Response:** We have added additional information regarding potential changes expected into the research questions in the objectives section which is provided in further detail in a later response. We have clarified and added information to the last paragraph of the introduction regarding the choice of the time frame and *"Ireland has also recently made significant investment in health research and healthcare reforms through its Health Service Executive (HSE) Action Plan for Health Research (2019 - 2029) <sup>23</sup> and Sláintecare reform (initially launched in 2017) (7)<sup>24</sup>."* ...

*"Furthermore, from 2020, the Irish Research e-Library (IReL) signed the first open access publishing agreements, providing researchers with easier access to open access publishing.(8) Within this context of healthcare and publication reform and the Covid-19 pandemic, we have proposed to include data prior to these changes and the pandemic, to provide some baseline proxy, as well as data throughout and 'post'-pandemic. This supports our aim to map a piece of the complex local landscape of research in Ireland, using a cross-sectional analysis of Altmetric data (2017 - 2023) and see how it has evolved since before, during, and after the Covid-19 pandemic. A better understanding of the online impact of recent health research can help researchers, and the communication specialists who help disseminate their work, identify pathways for more effective communication to the public..."*

If Table 1 is based on the Altmetric help page of "field of research," the authors need to cite the source.

Response: We have added the reference to the ANZSCR system.

Since there are no specific research questions or theories/models to guide the data extraction, the authors need to explain the topic mapping using titles. There are existing bibliometric tools for topic mapping (e.g., VOSviewer, Biblioshiny). The authors need to explain why they chose to use titles only rather than abstracts, author keywords, or MeSH

terms.

**Response:** We have clarified our research question in a previous response and have provided further details on our planned descriptive analyses which includes the ANZSCR areas: *“Data will be cleaned using R and descriptive analyses will be performed for general bibliometric information such as: the type of research output (i.e., article, book, chapter); open access status and type; top 20 journals, funders, and organisations in our dataset; the prevalence of sectors (e.g., education, healthcare); and the five subject areas of health research and their subdivisions (yearly). To investigate trends over time, counts of frequencies will be plotted on a quarterly and yearly basis for: the overall number of research outputs and yearly for subject and subdivision areas, sectors, and publishers.”*

Following a suggestion from another reviewer regarding the OpenAlex API and its classification system of topics (developed over many years by experts in the field), we have decided to no longer use our initially proposed topic modelling approach and the Crossref API. We have added text regarding this and will explore using VOSviewer to visualise this (although there may be issues due to file formats and the complexity of our dataset): *“OpenAlex provides a more thorough and in-depth view of academic research as it contains 65,000 Wikidata concepts based on Microsoft Academic Graph (MAG)(9) and enhanced with machine learning, natural language processing, citation analysis, an expert feedback. We will gather topic information by using the OpenAlexAPI to pull this information matched by the DOIs in our final dataset. OpenAlex uses a hierarchical system that organises topics into levels ranging from broad to more specific. We will focus on lower level, more narrow fields (i.e., level 1 and beyond) that represent increasingly specific subfields within disciplines (e.g., ‘pediatric oncology’, ‘telemedicine’, ‘anxiety’). (10) We will report the 20 most frequent terms per year and if possible visualise this information using the VOS-Viewer free software. (11)”*

Additionally, the Altmetric platform supports PubMed search queries. The authors might want to explain why they prefer API for data retrieval instead of the user interface.

**Response:** To clarify, we did use the user interface. The initial phrasing relating to the API was to allow us to explore if this was a good approach for our purposes. Previous literature has used the API to pull based on article-level information such as DOI or keywords but after discussions with Altmetric, it was confirmed that pulling data based on an organisations GRID was not feasible with our licensing access. As the focus of our project was on the organisation level, PubMed search queries would not be appropriate in our circumstance. We have edited and added text to the dataset section to clarify this: *“We will use Altmetric Explorer to search for all research outputs published between 1 January 2017 and 31 December 2023 from active Irish organisations that have Research Organisation Registry (ROR) IDs. ROR is a global registry of open persistent identifiers for research organisations which helps link researchers and their outputs to institutions across sectors (e.g., education, government, healthcare, non-profit, etc.) (<https://ror.org/about/>). Altmetric uses the prior system, the Global Research Identifier Database (GRID) (<https://www.grid.ac/>) which maps to ROR. As of 9 April 2024, there were 663 active research organisations with Ireland listed as their country of address. We searched both active and inactive IDs in case an organisation had outputs during the time period but then decided to inactivate their indexing GRID (e.g., they published in 2017-2019 but then deactivated their ID). The lead author (MKS) and two medical students will use the Altmetric Explorer interface to search for research outputs published within our date frame from each individual organisation. Datasets will be tracked using a tracking log in Excel to record the downloader (initials), total number of research outputs, number of outputs mentioned, filename,*

and download date. If an organisation has a least 1 output, we will download their data Altmetric uses the prior system, the Global Research Identifier Database (GRID) (<https://www.grid.ac/>) which maps to ROR. If an organisation has a least 1 output, we will download their data as a csv file using a standard naming notation (ID\_YYYY-MM-DD). These csv files will be stored in one folder, spot checked for completeness (MKS), then combined into one dataset which will include research output data from all organisations producing output from 2017 - 2023. Of note, the datasets have the same 46 variables, will be downloaded in UTF-8 to account for non-English characters, and dates will be checked prior to stacking. This tracking log and our RMarkdown code detailing combining of datasets, cleaning, pre-processing, and more can be available on our Open Science Framework accompanying our results manuscript.(10) While the Altmetric API (<https://www.altmetric.com/solutions/altmetric-api/>) is also available for information retrieval, we did not find it suitable for pulling data based on an organisation's GRID ID."

Altmetric.com recommends using an Altmetric attention score of 20 as a benchmark. The authors did not mention this for quantitative analysis.

**Response:** We have added a comment within the analysis section to address this: "We will also report the number of outputs with a score of 20 or above as Altmetric has indicated that this is a general score which can be considered as doing better than most of its 'colleagues.' (12)" As this score can vary by field and is not necessarily an indicator of quality, we do not want to relay or overemphasize its importance by integrating it into more detailed analyses.

4. Limitations: The manuscript could benefit from addressing potential biases in Altmetric data collection and how these might affect the study's findings.

**Response:** We have elaborated upon the limitations, potential biases, and our efforts to address them.

"The lack of a links between preprints (manuscripts uploaded to databases without peer review) and postprints is a larger issue (peer reviewed journal articles) within academic publishing (13) as each item is assigned a unique DOI and there are challenges indexing pre-prints alongside their peer-reviewed publication. Of note, in the Altmetric dataset, pre-prints are included and tracked, they just have their own individual altmetrics separate from the final publication. Pre-prints played an unprecedented role in disseminating Covid-19 research, (14) so we will include them in our dataset, account for this in statistical analyses, and explicitly disclose this pre-print prevalence in our dataset, and frame our findings with this in mind." ...

"Altmetric data has also been noted to be prone to manipulation and artificial inflation (16) and some sources are particularly unstable, with certain items 'vanishing'. (3) We will try to address this by pulling the data in a discrete period of time and we have included a time buffer (i.e., the end of 2023). However, we do recognise that certain mediums may still be 'incomplete' as they have different trajectories of attention growth – for example, Twitter attention starts and ends quickly, Mendeley readers accumulate quickly but continue to grow over the years, and policy attention is the slowest form of impact to accumulate. (17)"

"Furthermore, a high AAS does not necessarily indicate a high-quality piece of research.(15) Retracted publications have been widely shared online and certain topics (e.g., applied research, lifestyle behaviours, pop psychology, etc.) often simply receive more attention.(16) However, recent work comparing altmetrics to norm-referenced peer review scores from the UK Research Excellence Framework 2021 found that Altmetric correlated more strongly with research quality than previously thought although there is large variability with the strength of correlations amongst mediums and between fields (e.g., stronger in health and physical sciences than in the

*arts and humanities). (3)"*

Specific Comments:

The statement "With around 1.5 million new items being added to PubMed per year, or 2 papers per minute..." could include a reference to the source of this statistic for credibility.

**Response:** The reference for this statement is at the end of the sentence. We have moved it to clarify that it is the source of the information.

The sentence "Ireland can also be consider an outlier in other ways..." has a grammatical error ("consider" should be "considered").

**Response:** We have fixed this typo.

Overall, the manuscript presents a potentially valuable contribution to understanding online attention to health research in Ireland. However, the research protocol needs improvement, particularly in the areas of method design and justification.

**Response:** Thank you for this comment. We have attempted to address the methodological concerns in the response to the prompts above. Thank you for your helpful suggestions!

**Competing Interests:** No competing interests were disclosed.

Reviewer Report 27 June 2024

<https://doi.org/10.21956/hrbopenres.15238.r40897>

© 2024 Stephen D. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Dimity Stephen** 

German Centre for Higher Education Research and Science Studies (DZHW), Berlin, Germany

The protocol presents the planned method for conducting a cross-sectional study that explores the online attention received by medical/health-related research published by institutes in Ireland in 2017-2023, particularly in the context of the COVID pandemic. The study seeks to identify how much attention is received and via which channels, disambiguated by sector, research institute, field, research topic, and study design type over time. The study will also assess the correlation between altmetric attention and citations. The methodology is sound and the sources of data are reputable and reliable. The authors also thoughtfully outline the limitations of their study and its potential findings. However, further details about specific aspects of the methodology could be given and attention should be given to citation windows (detailed below).

Specific comments:

Are you planning to restrict the content to a particular document type, e.g., journal article, or any published item is within scope?

The data collection will produce a very large amount of information -- 7 years of daily counts of altmetric events for thousands of publications from over 600 institutions. You could provide more information about how exactly you will analyse this data to identify meaningful trends in attention between channels, fields, institutions, etc, and what statistical comparison will be made. A key point here is that altmetric data are rife with zero counts. Any statistical modelling will need to take this into account.

"Assuming an adequately sized dataset, we will also like to investigate whether Altmetric Attention Scores correlate to traditional article-level citation metrics, we will use Crossref's metadata, the rccrossref package, and the Crossref API (<https://api.crossref.org/>) to match outputs by DOI to obtain data to run Pearson's correlation tests on the data." Please keep in mind that you need to allow a citation window during which point citations accrue to a level that represents the articles' likely long-term impact. This is usually 3 years, so it would be best to analyse only articles published 2017-2020. Similarly, this window should be applied to each publishing year, e.g., for items published in 2017, citations that occur within 2017-2019 are included, for 2018, citations 2018-2020, and so on. Otherwise articles published in 2017 cannot be reliably compared with articles published in 2022, because 2017 articles have had sufficiently longer to accrue citations. Applying a stable citation windows means citation counts between publication years are comparable. This is less important for Altmetric Attention Scores, because most attention occurs shortly after publication.

"Study design. Given an appropriate size of the dataset, we will first use the metadata available in Crossref dataset to pull the research output's abstract and author keywords which will then undergo classification. We will create a list of key terms, in consultation with our project's steering committee and medical librarians and to classify research outputs within the umbrella categories of: intervention studies (e.g., trials), observational studies (cohort, cross-sectional, case control, genetic association, DTA, etc.), qualitative research (focus groups, interviews), protocols, case reports, evidence syntheses (e.g., systematic and scoping reviews), editorials or opinions, and other. These likely will be linked to the stopwords created during topic modelling. At least one reviewer will check the classifications after review."

I'm a little confused about this section. It initially sounded as if you intended to manually classify the documents into categories. However, reading again, it sounds like the key terms will be developed by the steering committee and librarians and then perhaps some kind of automated classification process will be applied here. Could you please clarify how these documents will be classified? If this is to be done manually, this could consist of classifying thousands of documents and is quite a heavy workload. Also, in the last line "These likely will be linked to the stopwords created during topic modelling", do you perhaps means topics rather than stopwords?

Finally, care should be given in interpreting these data and the weight attributed to altmetric attention, particularly in the context of research assessment. Applied research and particularly "hot" topics that are trending or interesting to a general population, such as pop psychology and diet- and sleep-related topics, may draw more attention than topics from base research that are less accessible and applicable to a general public. However, that does not necessarily mean they are more valuable. Please consider this when interpreting the results and making recommendations.

A couple of formatting/language things:

Abstract: "reach can cannot be solely"

Introduction: "(e.g., Stack Overflow), and more. (<https://www.altmetric.com/about-us/our-data/our-sources/>)" -> fullstop after the closing bracket of the reference

Methods: "The ANZSRC is a hierarchical system ... statistics in Australia and New Zealand.(19)" Is 19 a reference? It doesn't appear to match the content for reference 19.

Analysis: "a daily quarterly, and yearly basis" -> daily,

**Is the rationale for, and objectives of, the study clearly described?**

Yes

**Is the study design appropriate for the research question?**

Yes

**Are sufficient details of the methods provided to allow replication by others?**

Partly

**Are the datasets clearly presented in a useable and accessible format?**

Not applicable

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Scientometrics, bibliometrics

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 19 Aug 2024

**Melissa Sharp**

The protocol presents the planned method for conducting a cross-sectional study that explores the online attention received by medical/health-related research published by institutes in Ireland in 2017-2023, particularly in the context of the COVID pandemic. The study seeks to identify how much attention is received and via which channels, disambiguated by sector, research institute, field, research topic, and study design type over time. The study will also assess the correlation between altmetric attention and citations. The methodology is sound and the sources of data are reputable and reliable. The authors also thoughtfully outline the limitations of their study and its potential findings. However, further details about specific aspects of the methodology could be given and attention should be given to citation windows (detailed below).

**Response:** Thank you for the time and energy spent reviewing our protocol and for asking thoughtful questions and giving constructive feedback. We have attempted to clarify and elaborate upon our methodological and statistical approaches through our response below. We have also taken your feedback regarding the citation windows on board and have addressed it in the in further detail below.

Specific comments:

Are you planning to restrict the content to a particular document type, e.g., journal article,

or any published item is within scope?

**Response:** We have added additional information regarding included 'research outputs'/document types within the dataset section. An overwhelming majority of research outputs will be journal articles but we will not exclude other outputs (i.e., book chapters or books). *"Research outputs included are largely journal articles although book chapters and books will also be included. Of note, some items within Altmetric are classified as 'news' but can be considered articles as they are perspectives, commentaries, overviews, hot topics."*

The data collection will produce a very large amount of information -- 7 years of daily counts of altmetric events for thousands of publications from over 600 institutions. You could provide more information about how exactly you will analyse this data to identify meaningful trends in attention between channels, fields, institutions, etc, and what statistical comparison will be made. A key point here is that altmetric data are rife with zero counts. Any statistical modelling will need to take this into account.

**Response:** The note about a potentially large prevalence of zero count scores is important to consider, particularly for the Altmetric Attention Score (AAS) and mediums. While a majority of our analyses are descriptive in nature (i.e., counts and percentages), a large number of 0 counts could influence things like measures of central tendency so we will run analysis on both the full (unique deduplicated) dataset as well as on a dataset which excludes the 0 counts to provide a more complete view of our dataset. We have added text regarding this in the analysis section: *"For Altmetric analyses, we will report counts, and averages and medians per medium (e.g., X, Facebook, policy documents) for both the deduplicated dataset as a whole and for the deduplicated dataset with the zero-count AAS removed as there may be large amounts of zero counts in our dataset, potentially skewing measures of central tendency."*

For any statistical comparisons, we will use zero-inflated negative binomial regression to account for metrics with excessive zeros and likely a quite dispersed outcome variable (in this case citation counts). We have added information in the analysis section addressing this: *"We will use zero-inflated negative binomial regression to account for the large number of AAS scores of zero."*

"Assuming an adequately sized dataset, we will also like to investigate whether Altmetric Attention Scores correlate to traditional article-level citation metrics, we will use Crossref's metadata, the rcrossref package, and the Crossref API (<https://api.crossref.org/>) to match outputs by DOI to obtain data to run Pearson's correlation tests on the data." Please keep in mind that you need to allow a citation window during which point citations accrue to a level that represents the articles' likely long-term impact. This is usually 3 years, so it would be best to analyse only articles published 2017-2020. Similarly, this window should be applied to each publishing year, e.g., for items published in 2017, citations that occur within 2017-2019 are included, for 2018, citations 2018-2020, and so on. Otherwise articles published in 2017 cannot be reliably compared with articles published in 2022, because 2017 articles have had sufficiently longer to accrue citations. Applying a stable citation windows means citation counts between publication years are comparable. This is less important for Altmetric Attention Scores, because most attention occurs shortly after publication.

**Response:** Thank you for raising this important concern, we will only have whole citation count data (not split by citations per year), so to account for this we will split our dataset by year and only perform analyses for the 2017-2020 years. We have edited the text

accordingly to describe these changes: *“As articles in 2017 have had more time to accrue citations than those published in 2023, we will restrict our analyses and split our dataset by year. Previous research has indicated that a 3-year citation window is relatively stable so we will perform these analyses on the data from 2017-2020 only. (17,18)”*

"Study design. Given an appropriate size of the dataset, we will first use the metadata available in Crossref dataset to pull the research output's abstract and author keywords which will then undergo classification. We will create a list of key terms, in consultation with our project's steering committee and medical librarians and to classify research outputs within the umbrella categories of: intervention studies (e.g., trials), observational studies (cohort, cross-sectional, case control, genetic association, DTA, etc.), qualitative research (focus groups, interviews), protocols, case reports, evidence syntheses (e.g., systematic and scoping reviews), editorials or opinions, and other. These likely will be linked to the stopwords created during topic modelling. At least one reviewer will check the classifications after review."

I'm a little confused about this section. It initially sounded as if you intended to manually classify the documents into categories. However, reading again, it sounds like the key terms will be developed by the steering committee and librarians and then perhaps some kind of automated classification process will be applied here. Could you please clarify how these documents will be classified? If this is to be done manually, this could consist of classifying thousands of documents and is quite a heavy workload. Also, in the last line "These likely will be linked to the stopwords created during topic modelling", do you perhaps means topics rather than stopwords?

**Response:** Thank you for this query. Due to the resource constraints noted, the size of the dataset, and the quality of keywords/tagging, we have removed the section relating to study design part of the secondary objective and the objective text relating to that approach.

Finally, care should be given in interpreting these data and the weight attributed to altmetric attention, particularly in the context of research assessment. Applied research and particularly "hot" topics that are trending or interesting to a general population, such as pop psychology and diet- and sleep-related topics, may draw more attention than topics from base research that are less accessible and applicable to a general public. However, that does not necessarily mean they are more valuable. Please consider this when interpreting the results and making recommendations.

**Response:** This is an important distinction which we have aimed to elaborate on in further detail in the Limitations section: *“Furthermore, a high AAS does not necessarily indicate a high-quality piece of research.(23) Retracted publications have been widely shared online and certain topics (e.g., applied research, lifestyle behaviours, pop psychology, etc.) often simply receive more attention.(24) However, recent work comparing altmetrics to norm-referenced peer review scores from the UK Research Excellence Framework 2021 found that Altmetric correlated more strongly with research quality than previously thought although there is large variability with the strength of correlations amongst mediums and between fields (e.g., stronger in health and physical sciences than in the arts and humanities). (5)”* We will also keep this distinction in mind for our results manuscript.

A couple of formatting/language things:

Abstract: "reach can cannot be solely"

Introduction: "(e.g., Stack Overflow), and more. (<https://www.altmetric.com/about-us/our-data/our-sources/>)" -> fullstop after the closing bracket of the reference

Methods: "The ANZSRC is a hierarchical system ... statistics in Australia and New Zealand.(19)" Is 19 a reference? It doesn't appear to match the content for reference 19.

Analysis: "a daily quarterly, and yearly basis" -> daily,

**Response:** Thank you for pointing out these errors. We have addressed and fixed all noted problems. We have removed 'daily' as due to the size of the dataset this granularity was deemed unnecessary as it did not add much additional information.

**Competing Interests:** No competing interests were disclosed.

---