RESEARCH ARTICLE

Does the impact of medical publications vary by disease indication and publication type? An exploration using a novel, value-based, publication metric framework: the EMPIRE Index [version 5; peer review: 2 approved, 1 not approved]

Tomas Rees1, Avishek Pal2

1Oxford PharmaGenesis, Oxford, Oxfordshire, OX13 5QJ, UK
2Novartis Pharma AG, Basel, Switzerland

Abstract

Background

The EMPIRE (EMpirical Publication Impact and Reach Evaluation) Index is a value-based, multi-component metric framework to assess the impact of medical publications in terms of relevance to different stakeholders. It comprises three component scores (social, scholarly and societal impact), each incorporating related altmetrics that indicate a different aspect of engagement with the publication. Here, we present an exploratory investigation of whether publication types or disease indications influence EMPIRE Index scores.

Methods

Article-level metrics were extracted and EMPIRE Index scores were calculated for 5825 journal articles published from 1 May 2017 to 1 May 2018, representing 12 disease indications (chosen to reflect a wide variety of common and rare diseases with a variety of aetiologies) and five publication types.
Results

There were significant differences in scores between article types and disease indications. Median (95% CI) social and scholarly impact scores ranged from 1.2 (0.3–1.6) to 4.8 (3.1–6.6), respectively, for phase 3 clinical trials, and from 0.3 (0.3–0.4) to 2.3 (1.9–2.6), respectively, for observational studies. Social and scholarly impact scores were highest for multiple sclerosis publications and lowest for non-small cell lung cancer publications. Systematic reviews achieved greater impact than regular reviews. Median trends in the social impact of different disease areas matched the level of public interest as assessed through Google search interest. Although most articles did not register societal impact, mean societal impact scores were highest for migraine publications.

Conclusions

The EMPIRE Index successfully identified differences in impact by disease area and publication type, which supports the notion that the impact of each publication needs to be evaluated in the context of these factors, and potentially others. These findings should be considered when using the EMPIRE Index to assess publication impact.

Keywords
Altmetrics, bibliometrics, publication impact

This article is included in the Research on Research, Policy & Culture gateway.

This article is included in the Interactive Figures collection.
Article-level measures of publication impact (ALMs), including alternative metrics, play a crucial role in providing a more comprehensive understanding of the impact of research publications compared with traditional bibliometrics like the journal impact factor (JIF) and CiteScore. While JIF and similar metrics are useful for identifying journals with high readership, they fall short in assessing the quality and impact of individual research articles. This limitation arises from the fact that JIF primarily measures the average number of citations to articles in a journal, which may not reflect the significance or reach of a specific article within that journal. ALMs offer a more nuanced view by considering various factors beyond citations, such as social media mentions, downloads, and online discussions, providing insights into the broader impact of research across different audiences and contexts.

To date, most uses of alternative metrics to assess medical publications have looked at their correlation with conventional metrics, especially citations. For example, the correlation between alternative and traditional bibliometrics has been examined in publications in total joint arthroplasty research, neurology, COVID-19, critical care medicine and medical professionalism. Other uses of alternative metrics include the creation of an ‘altmetric impact factor’ for medical journals. They have also been used to identify trending topics in breast cancer, to explore the impact of multidisciplinary research groups, and to assess the convergence of altmetrics and societal impact assessments using data from the UK Research Excellence Framework. These studies have used either individual ALMs (Tweets, Mendeley saves, News mentions etc) or the Altmetrics Attention Score (AAS), an aggregated, weighted score that unifies multiple ALMs into a single metric. The AAS provides a convenient way of providing an summary of multiple article level metrics, but it excludes some key metrics (Mendeley saves and citations). The methodology by which it is calculated is not fully transparent and it is not open to adaptation by the users depending on their intended use.

We have previously described a novel and holistic approach to summarising ALMs, the EMPIRE (EMpirical Publication Impact and Reach Evaluation) Index, which uses weighted article-level metrics to assess the impact of medical publications in terms relevance to different stakeholders. The EMPIRE Index provides component scores for scholarly, social and societal impact, as well as a total impact score and predictive reach metrics (Figure 1). The scholarly component correlates weakly with CiteScore while the social score correlates closely with the AAS. The societal component, comprised of citations in guidelines, policy documents and patents, represents a distinct score.

The EMPIRE Index’s ability to correlate scholarly components with CiteScore and social components with AAS showcases its alignment with established metrics while providing a more nuanced evaluation of impact. This nuanced evaluation is particularly valuable in the medical field, where research impact transcends academic citations to societal influence and practical applications in healthcare settings.

Figure 1. Example of the EMPIRE Index score for a single publication. HCP, healthcare provider; NEJM, New England Journal of Medicine.
For researchers, the EMPIRE Index provides a detailed evaluation of publication value by offering component scores for scholarly, social, and societal impact, in addition to total impact scores and predictive reach metrics. This multifaceted assessment allows researchers to gain insights into the broader impact of their work beyond traditional citation-based metrics, enabling them to comprehend the reach and significance of their research across various dimensions.

Policymakers can derive value from the EMPIRE Index as it includes a societal impact component that takes into account citations in guidelines, policy documents, and patents. This feature is crucial for policymakers as it offers valuable insights into how research publications influence policy decisions and contribute to the development of guidelines and patents, assisting policymakers in making informed decisions based on the societal relevance and impact of research findings.

Healthcare providers can utilize the EMPIRE Index to evaluate the impact of medical publications in terms of societal implications and reach, allowing them to better grasp the practical implications of research in guiding clinical practice, policy development, and healthcare innovations. By considering societal impact metrics, healthcare providers can improve patient care and outcomes by staying abreast of impactful research in their field.

The EMPIRE Index uses randomized controlled clinical trials published in the NEJM as benchmark, which studies that are selected by the journal editors as likely to be of the highest impact for the practice of medicine. However, this standard cannot be uniformly applied to all types of studies, across different disease areas and development stages of a therapy. It is widely recognised that publication metrics vary by discipline, and to facilitate the comparison of publication impact across different fields, field-normalised citation impacts are frequently calculated. Although evidence is limited, some research suggests that ALMs can also vary by publication type. For example, review articles in pharmacology journals can receive twice as many citations as original articles and differences between quantitative and qualitative articles published in BMJ have been found for altmetric scores, but not citations.

Therefore, the nuances of each field and publication type should be considered in order to understand what constitutes a typical score. This additional context is an important consideration as it allows users to compare scores with a relevant frame of reference and so to accurately interpret and utilise publication metrics to derive meaningful insights.

It is likely that EMPIRE Index scores vary by disease area and by publication type. However, the scale and nature of these variations are unknown, which complicates efforts to compare scores of individual publications. This in turn limits the utility of the scale to identify ‘high impact’ publications, since what counts as an atypically high score will vary across therapy areas publication types.

In this brief report we present an exploratory investigation of the interaction between disease indications and publication types and typical EMPIRE Index scores. We sought to explore the variations that may exist in typical (i.e. average and median) EMPIRE Index scores across different therapy areas and publication types, and to explore the magnitude of these variations. This will allow the metrics of individual publications to be placed in a richer context of potentially comparable publications to facilitate interpretation of individual publication scores.

We also sought to provide additional context for the observed typical scores in two ways. Citescore is a journal-level metric that is related to the average number of citations per paper in that journal and is an indicator of journal prominence in the academic community. Journal Citescore is modestly correlated with publication EMPIRE Index Scores, especially the Scholarly Impact components. We therefore compared mean EMPIRE Index scores across therapy areas and publication types with the mean Citescores of the journals in which the papers were published, to see if journal prominence could explain any variations seen.

Google search trends indicate the level search engine activity around a topic and have been used as a measure of the prominence of health topics among the general public. We therefore compared EMPIRE scores with search trends for each of the therapy areas we examined, to see if public interest in a therapy area was related to engagement with publications measured via article-level metrics, especially with regard to the EMPIRE Index Social Impact component.

**Methods**

This exploratory study investigated 12 disease indications, purposefully chosen to reflect a variety of common and rare diseases with a variety of aetiologies. Six of these were rare diseases, selected as a convenience sample of disease indications with which the authors were most familiar. No formal statistical power analysis was undertaken. However, we aimed for disease samples from approximately 1000 publications, which would enable publication type sub-analyses. The six rare disease samples were, therefore, pooled.

Relevant publications were identified for each disease by the appearance of the disease name in the publication title. We limited the search period to items with publication dates between 1 May 2017 and 1 May 2018, to give sufficient time for metrics to accumulate while also minimising the time-dependent variation in metrics.

The searches were conducted on PubMed between 22 June 2020 and 3 July 2020, using the following search string:

```
"(2017/05/01"[Date - Publication] : "2018/05/01"
[Date - Publication]) AND "<disease name>"[TI]
```

We chose five publication types that are commonly used in medical research and communication: clinical trials, phase 3 clinical trials, observational studies, reviews, and systematic reviews. Definitions of each publication type are provided by US National Library of Medicine. Clinical trials are experimental studies that test the efficacy and safety of new interventions...
or compare different interventions in human participants. Phase 3 clinical trials are the last stage of clinical research of new drug treatments and are typically large scale and provide evidence intended to guide treatment practice. Observational studies are non-experimental studies that examine the associations between exposures and outcomes in natural settings, without manipulating any variables. Reviews are articles that summarize and synthesize the existing literature on a specific topic, providing an overview of the current state of knowledge and identifying research gaps. Systematic reviews are a type of review that follow a rigorous and predefined methodology to search, select, appraise, and synthesize the evidence from multiple primary studies on a specific question.

For each disease, we conducted secondary searches for each publication type using PubMed tags for those of interest (i.e. the search string above and either "review", "systematic review", "clinical trial, phase iii", ‘clinical trial’ or "observational study"). PubMed publication types are metadata supplied by PubMed and derive originally from publisher submissions.

PubMed IDs were entered into the Altmetric Explorer and PlumX dashboards and ALMs downloaded over the period 23 June 2020 to 11 July 2020. ALMs were assumed to be zero for any publication for which Altmetric Explorer did not return a result. We also obtained the journal CiteScore for all publications.

EMPIRE Index scores were calculated for all publications as described previously. Briefly, selected ALMs that compose the EMPIRE Index were weighted and aggregated to form three component scores (social impact, scholarly impact and societal impact), which were then summed to form a total impact score.

Each disease area comprised a different mixture of publication types, which we expected could confound the analysis: multivariate analysis on such a homogenous, non-normal and zero-inflated data set is problematic. Therefore, we opted to create standardised samples through random polling. A sample was created for each disease area with a standardised mix of publication types chosen to maximise the total number of publications retained (the standardised publication types [SPT] set). First, the two least common publication types (phase 3 clinical trials and systematic reviews) were excluded because of the high variation between disease areas and because they are largely subsets of other publication types (clinical trials and reviews, respectively). Although the observational studies publication type was only slightly more common than systematic reviews, it was retained as it was considered to be functionally very different from clinical trials and reviews. The proportions of each of the remaining three publication types were calculated for each disease set, as well as for the overall set. Publications were then trimmed from each disease set by random sampling, as needed, to match the proportions in the overall set. The trimmed publication sets formed the SPT set.

Similarly, each publication type comprised a different mix of diseases. A standardised disease areas (SDA) set was created by random sampling using a similar approach that ensured each publication type included the same mix of diseases, while maximising the total number of publications retained.

We downloaded weekly Google Trends data on relative interest over time for the period of interest for these diseases (May 1 2017 to May 1 2018). Google Trends measures the relative frequency of searches for a given term on Google, normalized by the total number of searches in a given time and region. A score of 100 indicates the maximum interest in any week over the search period. We used each therapy area as the target term (e.g. asthma, multiple sclerosis) adding each to a comparative dataset so that the interest in each term is expressed relative to the maximum interest across any of the search terms of interest. The year averages presented here are expressed relative to that maximum score.

As these analyses were exploratory and primarily intended to assess the magnitude of impact scores observed for different publication types and disease areas, we primarily provide descriptive statistics, supported by some statistical analysis of comparative differences. Intra-group differences were assessed using Kruskal-Wallis one-way analysis of variance, a non-parametric test for equality of population means (a significant result indicates that at least one population median of one group is different from the population median of at least one other group).

Exploratory analyses were conducted using a two-part modelling approach, chosen to account for the zero-inflated and skewed nature of our impact measures. The model comprised a logistic regression to predict the probability of a publication having any impact (non-zero impact), and a generalized linear model (GLM) with a log-link function to analyse the magnitude of impact for publications with non-zero impact. Both components of the model included disease area, publication type, and CiteScore as predictor. Asthma and clinical trials were chosen arbitrarily as the reference categories for disease area and publication type, respectively.

To assess the relationship between public interest in a disease area and publication impact, we conducted correlation analyses at the disease area level. We calculated both simple correlations and partial correlations controlling for average CiteScore, to account for potential confounding effects of publication quality across disease areas.

All analyses were performed using Python (version 3.10.5) with the statsmodels library (version 0.14.4). Detailed methodology and code are provided in the supplementary materials.

**Results**

**Sample characteristics**

In total, 20,577 publications were identified across the 12 disease areas, of which 5,825 (28%) were tagged with one of the publication types of interest (Table 1). Table 1 also shows the Google search interest for each of these diseases.
Table 1. Numbers of publications identified in the search. Google search interest is the average weekly interest across the search period and is a relative score 0–100 where 100 is the maximum score for any disease in any individual week.

| Disease       | All publications | Clinical trial | Phase 3 clinical trial | Observational study | Review | Systematic review | Any identifiable publication type | Google search interest |
|---------------|------------------|----------------|------------------------|---------------------|--------|-------------------|-----------------------------------|----------------------|
|               | n                | n              | n                      | n                   | n      | n                 | n                                 | n                    |
| Asthma        | 3487             | 225            | 6%                     | 21%                 | 1%     | 114               | 3%                                | 3%                   |
| Migraine      | 996              | 88             | 9%                     | 6%                  | 1%     | 36                | 4%                                | 17%                  |
| MS            | 2986             | 175            | 6%                     | 0%                  | 0%     | 80                | 3%                                | 548                  |
| NSCLC         | 3847             | 287            | 7%                     | 57%                 | 1%     | 42                | 1%                                | 417                  |
| Psoriasis     | 1654             | 139            | 8%                     | 53%                 | 3%     | 57                | 3%                                | 281                  |
| T2D           | 5341             | 644            | 12%                    | 66%                 | 1%     | 213               | 4%                                | 674                  |
| Rare diseases | 2266             | 82             | 4%                     | 18%                 | 1%     | 11                | 0%                                | 243                  |
| DLBLC         | 597              | 41             | 7%                     | 10%                 | 2%     | 5                 | 1%                                | 45                   |
| NASH          | 351              | 9              | 3%                     | 0%                  | 0%     | 1                 | 0%                                | 64                   |
| NET           | 187              | 3              | 2%                     | 1%                  | 1%     | 1                 | 0%                                | 70                   |
| SMA           | 196              | 8              | 4%                     | 2%                  | 1%     | 0                 | 0%                                | 42                   |
| TNBC          | 779              | 18             | 2%                     | 4%                  | 1%     | 2                 | 0%                                | 60                   |
| TSC           | 156              | 3              | 2%                     | 1%                  | 1%     | 2                 | 1%                                | 12                   |
| Total         | 20 577           | 1640           | 8%                     | 231%                | 1%     | 553               | 3%                                | 2907                 |

DLBLC, diffuse large B-cell lymphoma; MS, multiple sclerosis; NASH, non-alcoholic steatohepatitis; NET, neuroendocrine tumour; NSCLC, non-small cell lung cancer; SMA, spinal muscular atrophy; T2D, type 2 diabetes; TNBC, triple-negative breast cancer; TSC, tuberous sclerosis complex.

Analysis by publication type

The numbers of publications retained in the SDA set used for publication type comparisons (i.e. with the same disease indication composition for each publication type) are shown in Table 2.

Median EMPIRE Index scores and CiteScores for each disease in the SDA set are shown in Figure 2 and Table 3. Mean EMPIRE Index scores, shown in Figure 3, broadly reflect the median scores. Statistical analysis indicated that there was some significant variation in the medians of each component as well as the total impact score and journal CiteScore. In general, the ranking of publication type is relatively consistent across different types of impact. Notably, phase 3 clinical trials had the highest median and mean scores, while observational studies had the lowest. Systematic reviews had higher impact than reviews. Most articles across all publication types had no societal impact, and significant differences in societal impact were driven by outliers. Of note, eight of the ten publications with the highest societal impact were clinical trials, and six of those were in non-small cell lung cancer (NSCLC).

Results from the mixed model and regression analysis are presented in the supplemental material. Compared with non-phase 3 clinical trials, systematic reviews and phase 3 publications had higher social, scholarly, societal and total impact, although this was not consistent across logistic regression and linear models. In the logistic regression, systematic reviews consistently showed higher odds of having any impact across all impact types. Phase III clinical trials showed higher impact in the GLM analysis for scholarly and total impact, but this wasn’t consistent across other impact types. In the GLM, observational studies had lower social, scholarly and total impact.

Analysis by disease indication

The numbers of publications retained in the SPT set used for disease comparisons (i.e. with the same publication type composition for each disease indication) are shown in Table 4.

Median EMPIRE Index scores and journal CiteScores for each disease in the SPT set are shown in Figure 4 and Table 5. Kruskall–Wallis testing indicated at least one significant pairwise difference in the total scores, each component score and journal CiteScore. Migraine and multiple sclerosis (MS) had the highest impact across social and scholarly component scores as well as the total impact score, while NSCLC and psoriasis had the lowest. Most articles across all diseases had no societal impact, with significant differences in societal impact driven by outliers. The eight publications with the highest societal impact were all important clinical outcomes trials (three in type 2 diabetes, three in NSCLC and one each in migraine and asthma).

Mean EMPIRE Index scores for each disease in the SPT set are shown in Figure 5. The interactive version of Figure 5 (online publication only) also shows the mean EMPIRE Index.
Mean scores do not show clear trends for differences between disease indications, although societal impact appears to be lower for asthma and MS, and higher for migraine than other diseases. The high societal impact for migraine was driven by review articles; 16 of the 23 migraine articles with societal impact scores above zero were review articles. The scholarly impact for rare diseases appears to be higher than for other disease areas, albeit with low confidence owing to small numbers of publications included.

Results from the mixed model and regression analysis are presented in the supplemental material. Compared with the reference indication (asthma), other disease areas tended to have a lower probability of any total impact but, among publications with at least some impact, the average impact was higher with other diseases compared with asthma. This pattern was mostly consistent across the impact subscales (social, scholarly and societal).

Google search interest (public interest) did not correlate with any of the impact scores in the simple correlation. However, adjusted for CiteScore, the correlation between social impact and public interest was moderately strong and marginally significant ($r=0.69$, $p=0.0894$).

### Discussion

This analysis found that typical EMPIRE Index scores vary across both disease indications and publication types. These results provide valuable contextual information for interpreting EMPIRE Index scores and publication metric findings in general, for individual publications. For example, these findings can be used to help to understand whether a particular publication has notably high (or low) metrics.

We found considerable differences between disease areas, which broadly reflected public interest in the disease as assessed through Google search interest (Table 1). The variations in impact across disease areas highlight the complex nature of

---

**Table 2. Numbers of publications retained in the SDA set used for publication type comparisons.**

| Publication type          | Clinical trial | Phase 3 clinical trial | Observational study | Review | Systematic review | Proportion of total |
|---------------------------|----------------|------------------------|---------------------|--------|-------------------|---------------------|
| n                         | n              | n                      | n                   | n      | n                 | %                   |
| Asthma                    | 197            | 11                     | 51                  | 399    | 44                | 19%                 |
| Migraine                  | 66             | 4                      | 17                  | 133    | 15                | 6%                  |
| MS                        | 175            | 10                     | 45                  | 354    | 39                | 17%                 |
| NSCLC                     | 163            | 9                      | 42                  | 328    | 36                | 16%                 |
| Psoriasis                 | 104            | 6                      | 27                  | 210    | 23                | 10%                 |
| T2D                       | 334            | 19                     | 86                  | 674    | 74                | 32%                 |
| Total                     | 1039           | 59                     | 268                 | 2098   | 231               | 100%                |

MS, multiple sclerosis; NSCLC, non-small cell lung cancer; T2D, type 2 diabetes.

**Figure 2.** Median (95% CI) total impact scores for each publication type (standardised set). CI, confidence interval.
Table 3. Median and maximum scores for each EMPIRE Index component and CiteScore by publication type (SDA set).

| Publication Type          | n    | Median | 95% CI | Max  | Median | 95% CI | Max  | Median | 95% CI | Max  | Median | 95% CI | Max  | Median | 95% CI | Max  |
|---------------------------|------|--------|--------|------|--------|--------|------|--------|--------|------|--------|--------|------|--------|--------|------|
| Social                    |      |        |        |      |        |        |      |        |        |      |        |        |      |        |        |      |
| Clinical trial            | 1039 | 0.4    | 0.4–0.6 | 411.9 | 3.0    | 2.8–3.2 | 303.8 | 0.0    | 0.0–0.0 | 267.0 | 1.4    | 1.2–1.5 | 274.4 | 2.9    | 2.9–2.9 | 19.1 |
| Phase 3 clinical trial    | 59   | 1.2    | 0.3–1.6 | 222.4 | 4.8    | 3.1–6.6 | 68.3  | 0.0    | 0.0–0.0 | 192.8 | 2.3    | 1.4–3.2 | 156.9 | 3.2    | 2.9–4.9 | 16.1 |
| Observational study       | 268  | 0.3    | 0.3–0.4 | 279  | 2.3    | 1.9–2.6 | 38.7  | 0.0    | 0.0–0.0 | 89.0  | 1.0    | 0.8–1.2 | 32.5  | 2.5    | 2.3–2.6 | 10.5 |
| Review                    | 2098 | 0.4    | 0.3–0.4 | 176.5 | 3.3    | 3.1–3.5 | 183.1 | 0.0    | 0.0–0.0 | 178.0 | 1.4    | 1.3–1.5 | 106.6 | 2.9    | 2.7–2.9 | 23.2 |
| Systematic review         | 231  | 0.7    | 0.6–0.9 | 176.5 | 4.1    | 3.6–4.4 | 53.6  | 0.0    | 0.0–0.0 | 178.0 | 1.9    | 1.6–2.1 | 71.5  | 2.9    | 2.7–3.0 | 8.7  |
| p value                   |      | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 | < 0.0001 |

CI, confidence interval.
Table 4. Numbers of publications retained in the SPT set used for disease comparisons.

| Disease          | Asthma | Migraine | MS | NSCLC | Psoriasis | Rare disease | T2D | Proportion of total |
|------------------|--------|----------|----|-------|-----------|--------------|-----|---------------------|
| Clinical trial   | 225    | 88       | 175| 287   | 139       | 32           | 394 | 11%                 |
| Observational study | 78    | 31       | 61 | 42    | 48        | 11           | 137 | 11%                 |
| Review           | 385    | 150      | 299| 417   | 238       | 54           | 674 | 56%                 |
| Total            | 688    | 269      | 535| 746   | 425       | 97           | 1205| 100%                |

MS, multiple sclerosis; NSCLC, non-small cell lung cancer; SPT, standardised publication types; T2D, type 2 diabetes.

Figure 3. Mean EMPIRE Index scores for each publication type (SDA set). The interactive version (online only, accessible here: https://s3.eu-west-2.amazonaws.com/ox.em/webflow/p29ieu21/chart1.html) also shows mean EMPIRE Index scores for each disease by publication type (full set).

Figure 4. Median (95% CI) total impact scores for each disease (standardised set). MS, multiple sclerosis; NSCLC, non-small cell lung cancer; T2D, type 2 diabetes.
|               | Social |                     | Scholarly |                     | Societal |                     | Total |                     | CiteScore |                     |
|---------------|--------|---------------------|-----------|---------------------|----------|---------------------|-------|---------------------|-----------|---------------------|
|               |        | n | Median | 95% CI | Max | Median | 95% CI | Max | Median | 95% CI | Max | Median | 95% CI | Max | Median | 95% CI | Max |
| Asthma        |        | 688 | 0.6   | 0.4–0.7 | 122.5 | 3.0 | 2.8–3.2 | 70.5 | 0.0 | 0.0–0.0 | 178.0 | 1.3 | 1.2–1.5 | 88.0 | 2.9 | 2.7–3.0 | 16.1 |
| Migraine      |        | 269 | 0.6   | 0.4–0.9 | 222.4 | 3.3 | 2.8–3.9 | 68.3 | 0.0 | 0.0–0.0 | 192.8 | 1.7 | 1.3–2.1 | 156.9 | 2.9 | 2.4–2.9 | 16.1 |
| MS            |        | 535 | 0.9   | 0.7–1.0 | 142.1 | 3.8 | 3.4–4.2 | 267.4 | 0.0 | 0.0–0.0 | 133.5 | 1.8 | 1.6–1.9 | 150.0 | 2.8 | 2.6–2.9 | 16.1 |
| NSCLC         |        | 369 | 0.1   | 0.1–0.1 | 256.1 | 2.4 | 1.9–2.8 | 259.7 | 0.0 | 0.0–0.0 | 267.0 | 0.9 | 0.8–1.1 | 186.7 | 3.2 | 2.9–3.6 | 23.2 |
| Psoriasis     |        | 425 | 0.1   | 0.1–0.3 | 60.7  | 2.4 | 2.0–2.8 | 85.8 | 0.0 | 0.0–0.0 | 133.5 | 1.0 | 0.8–1.1 | 51.3 | 2.4 | 2.2–2.4 | 10.3 |
| T2D           |        | 1205 | 0.4  | 0.3–0.4 | 833.4 | 3.4 | 3.1–3.7 | 485.5 | 0.0 | 0.0–0.0 | 548.8 | 1.5 | 1.4–1.6 | 532.5 | 2.8 | 2.8–3.0 | 19.1 |
| Rare diseases |        | 97 | 0.3   | 0.1–0.6 | 236.1 | 3.2 | 2.2–4.2 | 143.4 | 0.0 | 0.0–0.0 | 89.0  | 1.3 | 0.9–1.9 | 136.4 | 3.3 | 2.5–3.7 | 16.1 |
| p value       | < 0.0001 < 0.0001                      0.0002 < 0.0001 < 0.0001 |

CI, confidence interval; MS, multiple sclerosis; NSCLC, non-small cell lung cancer; T2D, type 2 diabetes.
research dissemination and engagement. While publications in areas other than asthma tended to have lower odds of any impact, those that did have impact often showed higher magnitudes. This suggests that certain disease areas may have fewer but more impactful publications, possibly due to factors such as disease prevalence, research funding, or public health priorities.

We found that the three diseases with the highest median EMPIRE Index scores, especially social impact, were migraine, MS and asthma (they also tended to have the greatest effect sizes in the model). These also had the highest public interest. These differences were not observed in journal CiteScores, meaning that the disease areas with higher EMPIRE Index impact were not necessarily published in ‘high impact’ journals. NSCLC had low public interest (‘lung cancer’ as a general term was higher, but still lower than any of the other five major disease areas examined). Publications in NSCLC also had low median total impact scores, particularly in terms of social impact scores, despite being published in journals with higher median CiteScores. Of all the three impact scores, only social impact correlated meaningfully with public interest. Overall, this suggests that the reason publications in some disease areas attract higher social impact scores (driven by citations in news articles and social media) is that these disease areas are of greater interest to the general public. Although this suggests distinct differences between diseases in terms of publication impact, it should be noted that the period of interest was only a single year. The findings could therefore have been influenced by the completion of important clinical studies, which can vary from year to year across disease areas.

A clear picture is seen for publication types, with phase 3 trials demonstrating much higher metrics than other types. Phase 3 clinical trials, the last stage of clinical research of new drug treatments, are typically large scale and provide evidence intended to guide treatment practice. The higher impact observed for this publication type likely relates to higher public interest, higher scholarly interest, and greater likelihood of citations in guidelines and policy documents. Systematic reviews had higher impact than general reviews; interestingly, this was despite being published in journals with similar median CiteScores. This likely reflects that the methodological approach to synthesising systematic literature reviews makes them more impactful. Observational studies had the lowest impact, suggesting observational analyses are still generally regarded as having lower interest, despite the fact that they are increasingly regarded as valuable complementary research to interventional studies. These findings from simple comparisons were also supported by results from the exploratory model.

In general, across both publication types and disease indications, median scores were higher for scholarly impact than for social or societal impact, while mean and maximal scores were broadly similar (or lower). This suggests that score distribution is more skewed for social and societal impact, with many papers generating little interest despite some scholarly impact.

A key strength of this study is the use of an automated approach to identify a large pool of publications for analysis. However, the automated process used depends on the reliability of the underlying data. For example, disease areas were identified through a PubMed search on article titles, which may have excluded some relevant articles or included irrelevant ones. The PubMed search engine uses automatic term mapping, which usually makes the search more inclusive but can introduce inconsistencies. Publication types were identified by metadata tags, but these can often be inconsistently applied.
or missing. It can also result in duplication; for example, some phase 3 clinical trial publications in our sample were also classified as clinical trials.

A further limitation of this study is that it only includes publications from 2019, which may not reflect the long-term impact of publications over time. We acknowledge that some of the publications may have been published towards the end of the period, which could affect the relationship with Google Trends search data. Moreover, the EMPIRE Index relies on data sources that may have incomplete or inaccurate information, such as missing citations, delayed updates, or misclassified publications. Furthermore, the definition or interpretation of publication impact varies based on the stakeholder, intended audience and channel of dissemination, so other aspects such as clinical utility, policy influence, or educational value, may require more qualitative or specific indicators. Therefore, the findings of this study should be complemented by other methods of publication evaluation. Despite these potential limitations, it is important to remember that the original aim of the EMPIRE Index was to allow for flexibility to the users to define publication impact in terms of which individual ALM they would like to emphasise on through more weighting, which then customizes the measurement in the context of the intended pre-defined objectives.

The findings of this study have several implications for publication planning and evaluation in academic and also the pharmaceutical industry. First, they highlight the importance of tailoring communication strategies to different disease areas and publication types, as different audiences may have different preferences and expectations for the content and format of publications. For example, publications on rare diseases may benefit from more sharing on social media or patient engagement, while publications on oncology may benefit from explication of study results in clinical contexts.

Second, they demonstrate the value of using a multidimensional index that captures various aspects of publication impact, rather than relying on a single metric such as citations or Altmetric Attention Score. The EMPIRE Index provides a comprehensive and balanced view of how publications perform across scholarly, social and societal domains, and can help identify strengths and weaknesses of different publications and inform future improvement.

Third, they suggest that publication impact is not solely determined by the quality and novelty of the research, but also by the relevance and timeliness of the topic, the appropriateness and effectiveness of the dissemination channels, and the engagement and influence of the stakeholders. Therefore, publication planning and evaluation should consider not only the scientific merit of the publication, but also the context and the audience of the publication.

In conclusion, the EMPIRE Index successfully identified differences in impact by disease indication and publication type. This supports the notion that there is no universal gold standard metric for publications, and instead the impact of each publication needs to be evaluated in the context of the type of publication, disease area and potentially other factors. These findings should be considered when using the EMPIRE Index to assess publication impact.

Data availability
Figshare: EMPIRE Index disease and publication type analysis. https://doi.org/10.6084/m9.figshare.17072435.v2

This project contains the following underlying data:
- SMA metrics unlinked 11Jul20.xlsx
- Psoriasis metrics unlinked 11Jul20.xlsx
- NSCLC metrics unlinked 5Jul20.xlsx
- NET metrics unlinked 11Jul20.xlsx
- NASH metrics unlinked 11Jul20.xlsx
- MS metrics unlinked 5Jul20.xlsx
- Migraine metrics unlinked 5Jul20.xlsx
- Google search interest (30Jul21).xlsx
- DLBCL metrics unlinked 11Jul20.xlsx
- Asthma metrics unlinked 5Jul20.xlsx
- TSC metrics unlinked 11Jul20.xlsx
- TNBC metrics unlinked 11Jul20.xlsx
- T2DM metrics unlinked 5Jul20.xlsx
- Mixed mode analysis supplemental.docx
- All_data_integrated_7Oct24_deduplicated.csv

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

References
1. Haustein S, Larivière V: The use of bibliometrics for assessing research: possibilities, limitations and adverse effects. In Incentives and performance: governance of research organizations. (eds. Welpe, I., Wollersheim, J., Ringelhan, S. & Osterloh, M.) Springer, Cham, 2015; 121–139. Publisher Full Text
2. Jeong S: Altmetrics: a practical guide for librarians, researchers and academics. Technical Services Quarterly. 2017; 34(2): 230-231. Publisher Full Text
3. Raff JW: The San Francisco declaration on research assessment. Biol Open. 2013; 2(6): 533-534. PubMed Abstract | Publisher Full Text | Free Full Text
4. Bornmann L, Marx W, Gasparyan A, et al.: Diversity, value and limitations of
the journal impact factor and alternative metrics. Rheumatol Int. 2012; 32(7): 1861–1867.
PubMed Abstract | Publisher Full Text
5. Ramamurti P, Gu A, Fassihi SC, et al.: Correlation between Altmetric score and traditional bibliometrics in total joint arthroplasty research. Arthroplasty Today. 2021; 7: 225–229. PubMed Abstract | Publisher Full Text | Free Full Text
6. Kim Y, Kim J, Lee S, et al.: Analysis of altmetrics in social recognition of neurology and neurological disorders. Healthcare (Basel). 2020; 8(4): 367. PubMed Abstract | Publisher Full Text | Free Full Text
7. Edakar NMAM, Shehata AMK: Measuring the impact of COVID-19 papers on the social web: an altmetric study. Glob Knowl Mem Commun. 2022; 71(1/2): 1–26. Publisher Full Text
8. Lehane DJ, Black C: Can altmetrics predict future citation counts in critical care medicine publications? J Intensive Care Soc. 2021; 22(1): 60–66. PubMed Abstract | Publisher Full Text | Free Full Text
9. Azer SA, Azer S: Top-cited articles in medical professionalism: a bibliometric analysis versus altmetric scores. BMJ Open. 2019; 9(7): e029433. PubMed Abstract | Publisher Full Text | Free Full Text
10. Fassoulaki A, Tsaroucha A, Mitha G, et al.: Bibliometric analysis of alternative performance metrics for peri-operative, critical care and pain medicine journals. Anesthesia. 2021; 76(6): 856–857. PubMed Abstract | Publisher Full Text
11. Bayar B, Pekoz R: Analysis of trending topics in breast cancer articles from an Altmetric perspective. Cureus. 2022; 14(7): e26565. PubMed Abstract | Publisher Full Text | Free Full Text
12. Gregori MD, Scotti V, De Silvestri A, et al.: Does a research group increase impact on the scientific community or general public discussion? alternative metric-based evaluation. J Pain Res. 2016; 9: 391–5. PubMed Abstract | Publisher Full Text | Free Full Text
13. Bornmann L, Haunschild R, Adams J: Do altmetrics assess societal impact in a comparable way to case studies? An empirical test of the convergent validity of Altmetrics based on data from the UK Research Excellence Framework (REF). J Informetr. 2019; 13(1): 325–340. Publisher Full Text
14. Pal A, Rees T: Introducing the EMPIRE Index: a novel, value-based metric framework to measure the impact of medical publications. PloS One. 2022; 17(4): e0265381. PubMed Abstract | Publisher Full Text | Free Full Text
15. Ruiz-Castillo J, Waltman L: Field-normalized citation impact indicators using algorithmically constructed classification systems of science. J Informetr. 2015; 9(1): 102–117. Publisher Full Text
16. Amiri M, Michel MC: Do review articles boost journal impact factors? A longitudinal analysis for five pharmacology journals. Naunyn Schmiedebergs Arch Pharmacol. 2018; 391(9): 1033–1035. PubMed Abstract | Publisher Full Text
17. Retrouvey H, Webster F, Zhong T, et al.: Cross-sectional analysis of bibliometrics and altmetrics: comparing the impact of qualitative and quantitative articles in the British Medical Journal. BMJ Open. 2020; 10(10): e040950. PubMed Abstract | Publisher Full Text | Free Full Text
18. Scopus Citescore. Reference Source
19. Hong Y, Lawrence J, Williams D Jr, et al.: Population-level interest and telehealth capacity of US Hospitals in response to COVID-19: cross-sectional analysis of google search and national hospital survey data. JIR Public Health Surveill. 2020; 6(2): e18961. PubMed Abstract | Publisher Full Text | Free Full Text
20. Publication Characteristics (Publication Types) with Scope Notes. Reference Source
21. Rees T: EMPIRE Index disease and publication type analysis. figshare. Dataset. 2021. http://www.doi.org/10.6084/m9.figshare.17072435.v2
22. US Food and Drug Administration: Step 3: clinical research. Reference Source
23. Bosdriesz JR, Stel VS, van Diepen M, et al.: Evidence-based medicine-when observational studies are better than randomized controlled trials. Nephrolody (Carlton). 2020; 35(10): 737–743. PubMed Abstract | Publisher Full Text | Free Full Text
24. Kang P, Kalloniatis M, Doig GS: Using updated pubmed: new features and functions to enhance literature searches. JAMA. 2021; 326(6): 479–480. PubMed Abstract | Publisher Full Text
Open Peer Review

Current Peer Review Status: ✗ ✓ ✓

-- Version 5 --

Reviewer Report 08 November 2024

https://doi.org/10.5256/f1000research.173716.r336422

© 2024 Saleh 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.

✔ Arman Yurisaldi Saleh

Universitas Pembangunan Nasional Veteran Jakarta, Jakarta, Special Capital Region of Jakarta, Indonesia

The author has answered my feedback well and performed statistical regression, also explaining terms that were previously unclear.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Neurorestoration, neurobehavior, neuroprotective agent, bibliometric

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 4 --

Reviewer Report 10 October 2024

https://doi.org/10.5256/f1000research.168629.r324069

© 2024 Aviv-Reuven S. 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.

✔ Shir Aviv-Reuven

Bar Ilan University, Ramat Gan, Israel

The authors have responded to all my comments and the manuscript, I have no further comments and I now Approve it without revisions.
**Competing Interests:** No competing interests were disclosed.

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 30 September 2024**

https://doi.org/10.5256/f1000research.168629.r327843

© 2024 Saleh 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.

**Arman Yusalsal Saleh**

Universitas Pembangunan Nasional Veteran Jakarta, Jakarta, Special Capital Region of Jakarta, Indonesia

This article has strengths when viewed from the amount of data used, which is quite a lot in the form of articles, so that it provides a fairly clear picture when analyzed.

Actually, this article would be more convincing if the right statistical test was used, but no statistical test was carried out in the manuscript. For example, a statistical test can be added whose results are a picture of the scientific score and social impact score for various articles in the form of systematic reviews, phase 3 clinical trials and observational. The difference in impact scores between the two groups can be analyzed using the t-test in statistical tests. So this is the main suggestion for improving the article.

In the manuscript, it can be seen that the dependent variables are, for example, the social and scientific impact scores. While the independent variables are public interest, disease indications, number of article citations, type of publication. There are shortcomings in terms of statistical analysis, related to the study of the relationship between the variables above, for example, what if you want to know the relationship between the social and scientific impact scores and public interest, and this is not found in the manuscript. A small point to note is that a linear regression analysis was not included. I kindly suggest that statistical calculations in the form of linear regression be added. I suggest that the author be clearer in explaining Cite Score, Altimetrik and Google Trends. A more detailed description is still needed regarding the 3 terms above so that the reader is clearer.

My conclusion is that the author has responded well to the reviewers’ input, and I provide suggestions for improvement in terms of the use of statistical tests and more detailed descriptions. The article is worthy of being indexed.

**Is the work clearly and accurately presented and does it cite the current literature?**
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
No

If applicable, is the statistical analysis and its interpretation appropriate?
No

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

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

**Reviewer Expertise:** Neurorestoration, neurobehavior, neuroprotective agent, bibliometric

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 18 Oct 2024**

**Tomas Rees**

We thank the reviewer for their comments. The initial version of this analysis provided only simple statistical analysis (a non-parametric test of between-group differences) because we were primarily interested in understanding the magnitude of the scores that could be expected for different kind of publications in different disease areas. Furthermore, the data are complicated to analyse because they are zero inflated (many publications have a score of 0), highly skewed (a few publications score very highly) and partially nested (some, but not all, 'phase 3 clinical trials' are also tagged as 'clinical trials', for example).

However, we do agree that statistical analysis can provide useful and interesting perspective on the relationship between our independent variables and the EMPIRE Index scores. Therefore, we have provided a mixed model analysis as well as a regression analysis versus Google search interest. The results of these analyses largely accord with our original conclusions, but provide rich detail on these interactions.

**Competing Interests:** No competing interests were disclosed.
This manuscript aims to identify impact differences between medical publications according to diseases and publication types as well as observe variations in public interest by comparing the EMPIRE index with Google search trends.

In general I find this paper interesting, but I found it lacking in aspects which the previous reviewer mentioned. Further more, the references to related literature is lacking and there is no discussion related to other possible article level metrics and why the EMPIRE index was selected.

Introduction
1. I agree with previous reviewer that Introduction is lacking, not only in presentation of the research problem, but also in the significance of this study.
2. I further agree that Literature review section is missing. it should either be added or the introduction further expanded to address related studies utilizing this indicator as well those referring to analysis of medical publications in general and specifically those who refer to the 12 diseases under analysis in this paper.
3. last paragraph of Introduction should be rephrased as it is unclear, particularly this - "We also look to see if there are similar variations in public interest in therapy areas we explore, with particular interest in comparing this with the EMPIRE Index includes a social score."

Methods
4. The authors do not explain why the different types were selected for analysis and what each of these types presents. These should be detailed.
5. The authors do not give detailed the usage of Google Trends. what data was downloaded? was it filtered to specific areas?. Further more, as some of the papers may have been published towards the end of that period, it may have been more appropriate a longer period for the trends.

Results
1. The usage of CiteScore is unclear to me, as CiteScore aims to measure journal impact.

Discussion
The paragraph detailing the findings related to Google trends should be moved to results and a plot showing these trends should be added.
Discussion of the results seems short and in some cases state the obvious. for example "Observational studies had the lowest impact, suggesting observational analyses are still generally regarded as having lower interest."
A general discussion of possible significance of the findings should be added.

Is the work clearly and accurately presented and does it cite the current literature?
Partly
Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

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

**Reviewer Expertise:** Bibliometrics, Scientometrics, academic search, journals impact

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 28 Jun 2024**

**Tomas Rees**

We thank the Dr Aviv-Reuven for their constructive comments and suggestions on our manuscript. We have revised the manuscript accordingly and addressed each comment in detail below.

**Reviewer comments and responses**

**Introduction**

1. I agree with previous reviewer that Introduction is lacking, not only in presentation of the research problem, but also in the significance of this study.

**Response:** We have revised the introduction to provide more background on the research problem and the significance of this study. We have added a paragraph that explains the limitations of conventional metrics such as journal impact factor and CiteScore, and the advantages of using article-level metrics such as altmetrics to assess the impact of medical publications. We have also added a paragraph that describes the EMPIRE Index. We have highlighted the novelty and relevance of our study, which is the first to explore the variations in EMPIRE Index scores across different disease indications and publication types.

2. I further agree that Literature review section is missing. it should either be added or the introduction further expanded to address related studies utilizing this indicator as well those referring to analysis of medical publications in general and specifically those who refer to the 12 diseases under analysis in this paper.
Response: We have added a literature review where we discuss the previous studies that have used altmetrics to assess the impact of medical publications in various fields and contexts.

3. last paragraph of Introduction should be rephrased as it is unclear, particularly this - "We also look to see if there are similar variations in public interest in therapy areas we explore, with particular interest in comparing this with the EMPIRE Index includes a social score."
Response: We have rephrased the last paragraph of the introduction to clarify our research objectives and hypotheses.

Methods
4. The authors do not explain why the different types were selected for analysis and what each of these types presents. These should be detailed.
Response: We have added a paragraph in the methods section that explains the rationale for selecting the different publication types for analysis and what each of these types represents.

5. The authors do not give detailed the usage of Google Trends. what data was downloaded? was it filtered to specific areas?. Further more, as some of the papers may have been published towards the end of that period, it may have been more appropriate a longer period for the trends.
Response: We have added more details on the usage of Google Trends in the methods section.

Results
The usage of CiteScore is unclear to me, as CiteScore aims to measure journal impact.
Response: We have clarified the rationale for the comparison with of CiteScore in the introduction.

Discussion
The paragraph detailing the findings related to Google trends should be moved to results and a plot showing these trends should be added.
Response: We have clarified that the data are presented in Table 1 as averages from 2019. We have not included a plot of the weekly data for 2019 as we do not refer to these and they are not relevant to our discussion. However, they are available in the supplementary material in excel form.

Discussion of the results seems short and in some cases state the obvious. for example "Observational studies had the lowest impact, suggesting observational analyses are still generally regarded as having lower interest."
Response: We have clarified the value of observational studies.

A general discussion of possible significance of the findings should be added.
Response: We have added a discussion of potential significance, as well as expanding the limitations.
Competing Interests: No competing interests were disclosed.

José Luis Ortega
Institute for Advanced Social Studies (IESA-CSIC), Córdoba, Spain

The text is the same as the version 1. Only references were updated.

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bibliometrics, altmetrics, academic search engines, scholarly social networks

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.
Tomas Rees

We thank the reviewer for the commentary and perspective, and we have made numerous clarifications and additions to the manuscript to address these. One important caveat is that this was intended as a brief report. It follows on from the full publication detailing the background and rationale of the EMPIRE Index published in PLOS ONE, which covers how it was developed and what the component scores signify. The current report provides some useful additional investigation of the properties of the EMPIRE index and its underlying article-level metrics. It is for these reasons that this report does not include a full literature review and does not go into the details of the EMPIRE Index that are covered in the PLOS ONE paper. We originally referenced the preprint in the current submission (the primary publication had been approved but publication was delayed) but have updated the reference now that the original article is fully published.

We have addressed the reviewer’s other comments, as follows:

We expanded the introduction to provide greater clarity of the objectives and purpose of this analysis – including why we have used CiteScore and Google Trends. We have added information on data extraction from Altmetrics Explorer and PlumX. We have also added information on the definition of the publication types (these are defined by PubMed).

While most of the component metrics of the EMPIRE Index are altmetrics, article citations are not and we agree that using altmetrics as an umbrella term is therefore inappropriate: we have changed to article-level metrics (ALMs) throughout.

We agree that computing an average CiteScore would be inappropriate, because such an average would not indicate the average number of citations per paper across the set of journals and so could be misleading. However, in our analysis we use the median, which is robust in this case since it does not transform the underlying data.

We do not believe that it is obvious whether the metrics for a given paper are notably high or low. This can only be understood by understanding the typical metrics of comparable publications, which is the objective of this report.

We note in the discussion that the higher scores of migraine publications, in particular social scores, may be linked to higher public interest as evidence by higher Google Trends interest. However, we acknowledge that the analyses presented do not and cannot explain the differences in typical scores seen across therapy areas and publication types. Therefore, any such discussion must necessarily be speculative and as such we keep it to a minimum.

The sole purpose of this report is to demonstrate that these variations exist and to explore the magnitude of these variations. The analyses presented are robust and do provide meaningful results that address the objective. We therefore believe that the current, amended report is worthy of publication despite its brief nature.
José Luis Ortega
Institute for Advanced Social Studies (IESA-CSIC), Córdoba, Spain

This manuscript attempts to find impact differences between medical publications according to diseases and publication types. These differences are calculated according to a new indicator, EMPIRE Index. More than 20k publications were retrieved from PubMed about 12 diseases areas. The resulting publications were also grouped by publication type. Using the Kruskal-Wallis test, authors found that the EMPIRE Indicator produces some differences between publications according to diseases and type of documents. These results allow them to justify the utility of this indicator and its power to identify different types of impacts.

However, the paper presents several weaknesses that impede my recommendation for approval. Next, I detail some problem points by section:

**Introduction**

The introduction is poor because it does not introduce the research problem properly. An introduction should explain why the research is necessary and what the problem is that the paper wants to address. Why is important to know that impact varies by disease or by publication type? Why is an indicator such as EMPIRE Index necessary?

Authors confound article-level with altmetrics. Many altmetric measures could be article-level metrics, but there are other article-level metrics (i.e. citations) that are not altmetrics. Author should revise the text and correct these statements.

I do not see a key section in a scientific article: a literature review. Authors should include a section where they contextualize the research, explaining previous studies that had deal with this problem. Studies about new indicators, differences in the impact by subject or document type, different types of impacts, etc. Overall, authors should present other studies in the introduction that show why is important to know that a disease has more social or scholarly impact, or why some document types attract more citations or tweets. In short, more conceptual background.
The paper includes only eight references, which illustrates the poor context of this research.

**Objective**

I think this section is the most important in a paper and it is missing in this manuscript. If there are not objectives, there are not research goals and a paper lacks sense. The final line of Introduction says: “Here, we present an exploratory investigation of whether disease indications and publication types influence the average EMPIRE Index scores.” Why do you want to know this? Why is important to know this? Authors should introduce a section for objectives where they explain the main objectives and secondary ones. It could be interesting to include some research questions that specify the aim of the paper.

**Methods**

Authors should include more detail about the data extraction process in Altmetric and PlumX. Authors claim that they used Altmetric Explorer to retrieve 20k publications. How did you exactly retrieve the data from this site? Do you use the API endpoint? How do you obtain the data from PlumX?

A key element in the paper is to test differences according to diseases and document types, but there is not any explanation about these typologies. For example, and about document types, what is the difference between “review” and “systematic review”? This part is key because the definition of groups and types influences the results, so it is very important to define the diseases and the typology of documents.

All the analysis relies on a new indicator, EMPIRE Index. In spite of the importance of this indicator in the paper, there is not any information about how it is calculated and conceptualized. We have to read a non-peer reviewed pre-print deposited in a repository to know what it is. I do not see this way to do science as correct. Authors have to explain how this indicator is calculated in this paper; how the three component scores are defined; what is the difference between “social impact” and “societal impact”, mainly from a conceptual point of view; likewise how do you find three components and not two or four; which metrics take part in each component and why?

CiteScore is an indicator for journals, not for articles. How it is used and why? The same for Google Trends - how they are used in the study, and why?

**Results**

Tables 3 and 5 use CiteScore median for valuing articles by diseases. I recall that it is not correct to use journal indicators for research articles, and the use of median or mean from a ratio (CiteScore is the ratio of citations by publications) is a mathematical artifact, and the results are spurious.

**Discussion**

This paper lacks of any substantial discussion about the results. Sometimes the interpretation is obvious and it does not contribute valuable information. For example, the first paragraph ends: “For example, these findings can be used to help to understand whether a particular publication has notably high (or low) metrics.” This is obvious and It is not necessary to do this study to reach
that conclusion. Every publication has high or low metrics. Another example: “Observational studies had the lowest impact, suggesting observational analyses are still generally regarded as having lower interest.” If they have low impact, then they have low interest.

It is interesting that the results from the EMPIRE index are not interpreted. For example, migraine, MS and asthma are the diseases with more social impact. Why? Is this positive or negative? Is there any evidence in other studies that confirm that these pathologies are more interesting for public audiences? For example, is there any correlation between these scores and Google Trends? Is there any survey that explore the public opinion about diseases? Etc.

This part should include the interpretation of the results according to the previous knowledge. It is interesting that this section only includes one citation (about using PubMed), but there is not any reference to previous results to discuss the validity of the findings or to explain the meaning of the results. For example, “The high impact of phase 3 clinical trials is to be expected, given that they are intended to provide practice-changing information.” Why is it that papers which provide practice-changing information should have more impact?

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
No

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
No

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Bibliometrics, altmetrics, academic search engines, scholarly social networks

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 28 Feb 2023

Tomas Rees
We thank the reviewer for the commentary and perspective, and we have made numerous clarifications and additions to the manuscript to address these. One important caveat is that this is a F1000 Brief Report (https://f1000research.com/for-authors/article-guidelines/brief-report), and so is limited by the journal style template (no ‘Objectives’ section) and word count limit (2,500 words). It follows on from the full publication detailing the background and rationale of the EMPIRE Index published in PLOS ONE, which covers how it was developed and what the component scores signify. The current report is intended to provide some useful additional investigation of the properties of the EMPIRE index and its underlying article-level metrics. It is for these reasons that this report does not include a full literature review and does not go into the details of the EMPIRE Index that are covered in the PLOS ONE paper. We originally referenced the preprint in the current submission (the primary publication had been approved but publication was delayed) but have updated the reference now that the original article is fully published.

We have addressed the reviewer’s other comments, as follows:

We expanded the introduction to provide greater clarity of the objectives and purpose of this analysis – including why we have used CiteScore and Google Trends. We have added information on data extraction from Altmetrics Explorer and PlumX. We have also added information on the definition of the publication types (these are defined by PubMed).

While most of the component metrics of the EMPIRE Index are altmetrics, article citations are not and we agree that using altmetrics as an umbrella term is therefore inappropriate: we have changed to article-level metrics (ALMs) throughout.

We agree that computing an average CiteScore would be inappropriate, because such an average would not indicate the average number of citations per paper across the set of journals and so could be misleading. However, in our analysis we use the median, which is robust in this case since it does not transform the underlying data.

We do not believe that it is obvious whether the metrics for a given paper are notably high or low. This can only be understood by understanding the typical metrics of comparable publications, which is the objective of this report.

We note in the discussion that the higher scores of migraine publications, in particular social scores, may be linked to higher public interest as evidence by higher Google Trends interest. However, we acknowledge that the analyses presented do not and cannot explain the differences in typical scores seen across therapy areas and publication types. Therefore, any such discussion must necessarily be speculative and as such we keep it to a minimum.

The sole purpose of this report is to demonstrate that these variations exist and to explore the magnitude of these variations. The analyses presented are robust and do provide meaningful results that address the objective. We therefore believe that the current, amended report is worthy of publication despite its brief nature.

**Competing Interests:** No competing interests were disclosed.
The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com