

Introduction to using Altmetric data for research

Stacy Konkiel
Director of Research Relations
Altmetric & Dimensions

Agenda

- ★ About the Program
- ★ How Altmetric collects data
- ★ Free data access options
- ★ How to query Altmetric's data
- ★ Examples of research done with Altmetric data
- ★ Q&A



Welcome!



Stacy Konkiel

Director of Research Relations

stacy@altmetric.com



About the Altmetric Researcher Data Access Program



Altmetric Researcher Data Access Program

Benefits

1. No-cost access for noncommercial research
2. Support for using our products in your research
3. Help understanding and interpreting our data
4. Connections with other researchers, grant opportunities, etc.

Provided through:

1. Email support
2. Educational webinars
3. Community listserv
4. Quarterly newsletters
5. Events (*coming soon*)
6. Office hours (*coming soon*)



Huge amounts of data!

12.6 million+
outputs w/ attention

1.7m
Wikipedia
mentions

1.4m
policy
references

9.7m
non-patent
citations

4.6 million
news mentions

98 million+
mentions since
2011



What you get

- Access to altmetrics for 12m+ research outputs
- 6 months access, with option to renew
- Technical support



Program requirements

- *Noncommercial* purpose
- Defined research question
- Data security provisions
- Technical expertise (in some cases)
- Acknowledgement of Altmetric



Legal stuff

Get our permission in writing before sharing data

Access is at our sole discretion

No warranty



Also:

We'd love it if you made your resulting presentations and publications Open Access!

(But we also understand that not everyone can.)



How Altmetric collects data



Scope of our data

What we track

Online discussions
around research
outputs of all kinds!

We do not track

Social media metrics
Usage statistics
Concepts*
Author names*



For research to be tracked, we need...



A research output
with a...



Persistent identifier
that is...



Mentioned in a
source we track



Research outputs



Articles & preprints

Books

Book chapters

Datasets

Clinical trial records

News stories

*...and countless other research formats
if a persistent identifier has been assigned!*



Video data on figshare: [10.6084/m9.figshare.5721088.v1](https://figshare.com/10.6084/m9.figshare.5721088.v1)



Identifiers



- DOIs
- PubMed IDs
- ISBNs
- Handles
- arXiv IDs
- ADS IDs
- URNs
- SSRN IDs
- RePEC IDs
- ClinicalTrials.gov records

URLs, URIs, and ORCID identifiers are also tracked, but can't really be studied at scale



Data sources



17 types of platforms, thousands of sites indexed

- Policy documents
- News
- Blogs
- Twitter
- Post-publication peer-reviews
- Facebook
- Sina Weibo
- Syllabi
- Wikipedia
- Google+
- LinkedIn
- Reddit
- Faculty1000
- Q&A (Stack Overflow)
- Youtube
- Pinterest
- Patents

Mendeley, Connotea, CiteULike, Dimensions citations are also tracked but do not contribute towards the research's Altmetric Attention Score



Altmetric attention score



Output-level Score
based on:

- Volume
- Sources
- Authors



What this means for your research

- Study research with PIDs
- Sources: Comparing apples and oranges
- Altmetric Attention Score
 - Don't create journal-level aggregates or averages



Free Data Access Options



Data Access Options

Altmetric Explorer

API

Metrics (rate-limits)

Metrics (w/out limits)

Database snapshot file (JSON)



Altmetric Explorer

Advanced search × CLOSE

Research outputs

KEYWORDS [WHAT IS THIS?](#)
Enter a title, author name, editor name, and/or journal

SUBJECTS (FOR CLASSIFICATION) [WHAT IS THIS?](#)
Enter one or more subjects (e.g. "0607" or "Plant Biology")

AFFILIATION (GRID) [WHAT IS THIS?](#)
Enter the name of an organization or a GRID ID.

TITLE OF OUTPUT [WHAT IS THIS?](#)
e.g., "Good vibrations: the role of music in Einstein's thinking"

TYPE OF OUTPUT [WHAT IS THIS?](#)
 All outputs
 Articles Books Book chapters
 Data sets Clinical trial records News stories

SCHOLARLY IDENTIFIERS [WHAT IS THIS?](#)
ADD SCHOLARLY IDENTIFIERS

ORCID [WHAT IS THIS?](#)
Enter a valid ORCID

PUBMED QUERY [WHAT IS THIS?](#)
ADD A PUBMED QUERY

Publishers, journals, and collections

PUBLISHER NAME [WHAT IS THIS?](#)
Enter one or more publisher names

DOI PREFIX [WHAT IS THIS?](#)
Enter one or more DOI prefixes, e.g., 10.6084

JOURNAL OR COLLECTION [WHAT IS THIS?](#)
e.g., Physics Letters, arXiv, figshare, 0028-0836
or **PASTE A LIST OF JOURNAL ISSNS**

HANDLE PREFIX [WHAT IS THIS?](#)
Enter one or more Handle prefixes.

Dates

PUBLICATION DATE [WHAT IS THIS?](#)
Between anytime and anytime

ALTMETRIC MENTIONS DURING [WHAT IS THIS?](#)
any time

CANCEL **CLEAR FIELDS** **RUN SEARCH**

- Journal name(s) or ISSN(s) (< 25,000)
- Output **identifier** (< 25,000)
- Keyword search in output title or abstract
- Publication date ranges
- Mention dates (fixed to previous 24h, day, three days, month, three months, six months, one year)
- Publisher name
- PubMed query
- Name of the outlet or social media user that has mentioned research
- Subject area (FoR code)
- Institutional affiliation



Search Capabilities

Advanced search includes:

- Keyword & subject search
- DOI or journal list
- Pubmed query builder
- Filter by date, journal, and more

The screenshot displays the 'Advanced search' interface with the following sections and fields:

- Research outputs:**
 - KEYWORDS:** Enter a title, author name, editor name, and/or journal.
 - SUBJECTS (FOR CLASSIFICATION):** Enter one or more subjects (e.g., "0607" or "Plant Biology").
 - AFFILIATION (GRID):** Enter the name of an organization or a GRID ID.
 - TITLE OF OUTPUT:** e.g., "Good vibrations: the role of music in Einstein's thinking".
 - TYPE OF OUTPUT:** Includes checkboxes for 'All outputs', 'Articles', 'Data sets', 'Books', 'Clinical trial records', 'Book chapters', and 'News stories'.
 - SCHOLARLY IDENTIFIERS:** Includes an 'ADD SCHOLARLY IDENTIFIERS' button.
 - ORCID:** Enter a valid ORCID.
 - PUBMED QUERY:** Includes an 'ADD A PUBMED QUERY' button.
- Publishers, journals, and collections:**
 - PUBLISHER NAME:** Enter one or more publisher names.
 - DOI PREFIX:** Enter one or more DOI prefixes, e.g., 10.6084.
 - JOURNAL OR COLLECTION:** Enter one or more journal/collection names (e.g., Physics Letters, arXiv, figshare, 0028-0836). Includes a 'PALETTE A LIST OF JOURNAL ISSUES' button.
 - HANDLE PREFIX:** Enter one or more Handle prefixes.
 - Dates:** Includes a 'PUBLICATION DATE' section with 'Between' and 'anytime' dropdowns, and an 'ALTMETRIC MENTIONS DURING' section with an 'anytime' dropdown.

At the bottom, there are 'CANCEL', 'CLEAR FIELDS', and 'RUN SEARCH' buttons.

Search for research outputs...
+ ADVANCED SEARCH

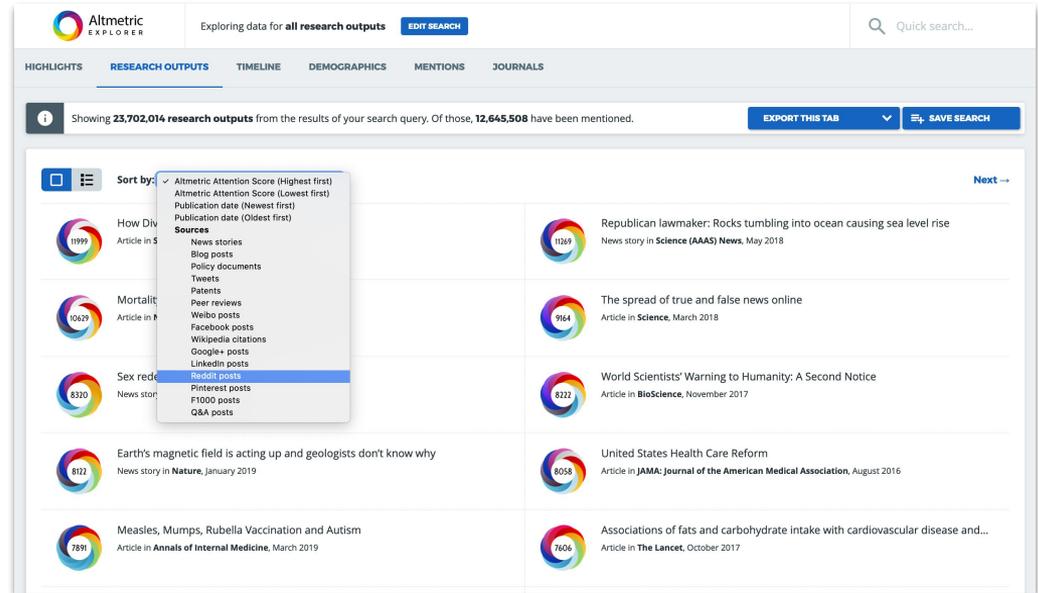


Exploring for trends

Sort by AAS

Sort by attention in particular online spaces (e.g. public policy documents)

Export to CSV



The screenshot displays the Altmetric Explorer interface. At the top, the logo and 'Exploring data for all research outputs' are visible. Below the navigation tabs (HIGHLIGHTS, RESEARCH OUTPUTS, TIMELINE, DEMOGRAPHICS, MENTIONS, JOURNALS), a summary bar indicates 'Showing 23,702,014 research outputs from the results of your search query. Of those, 12,645,508 have been mentioned.' A 'Sort by' dropdown menu is open, showing options: 'Altmetric Attention Score (Highest first)', 'Altmetric Attention Score (Lowest first)', 'Publication date (Newest first)', and 'Publication date (Oldest first)'. Under the 'Sources' section, 'Research posts' is highlighted. The main content area shows a grid of research outputs, each with a circular attention score icon and a brief description. Examples include 'Republican lawmaker: Rocks tumbling into ocean causing sea level rise' (AAS 11269) and 'Measles, Mumps, Rubella Vaccination and Autism' (AAS 7891).

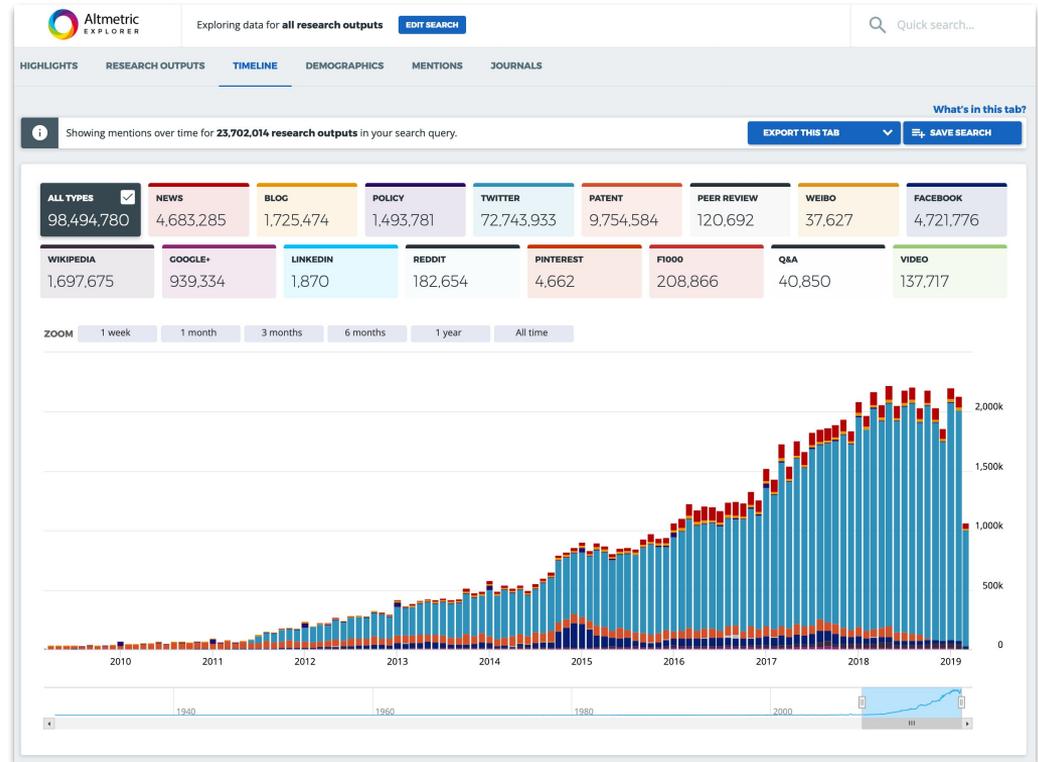


Navigating by Visualizations

Graph engagement over time to find trends, then dig into the data

Filter by source, time period

Export to CSV

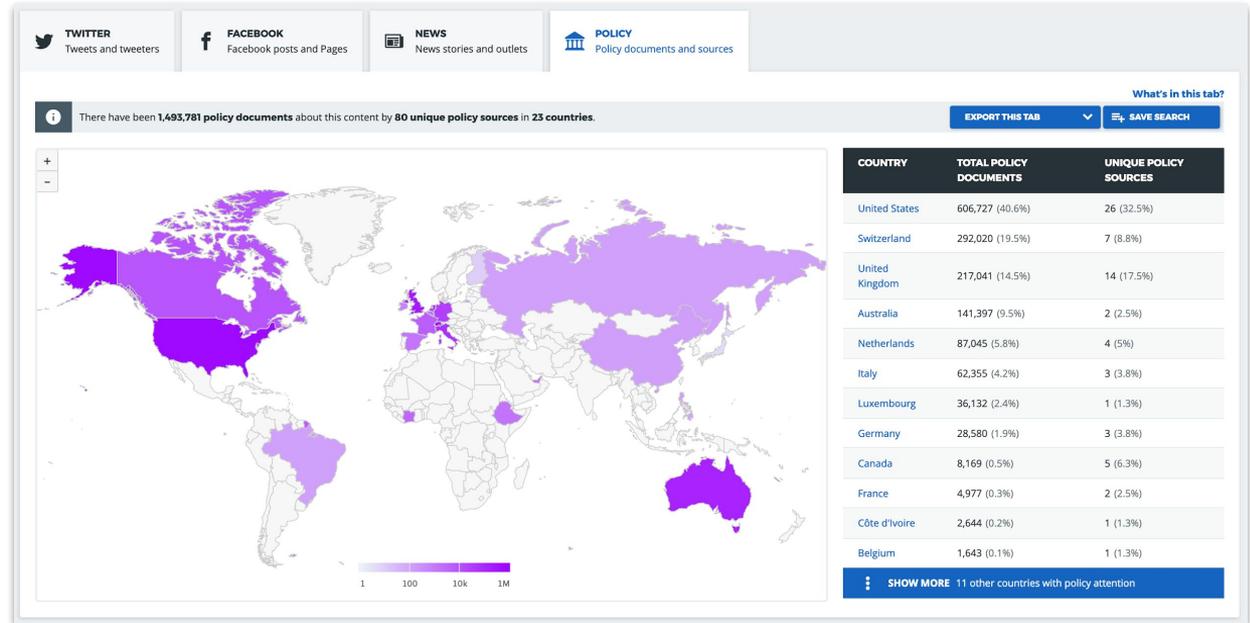


Navigating by Visualizations

Understand
global reach via:

- Twitter
- Facebook
- News media
- Public policy

Export to CSV



Finding Discussions

Filter by source,
time period

Search for influencers

The screenshot shows a search interface with tabs for 'ALL VIDEOS', 'ALL BLOG POSTS', and 'ALL POLICY DOCUMENTS'. A search bar contains 'fluoroquinolone-induced acute liver injury'. Below the search bar, there are filters for 'Show mentions from' and 'Search for a country'. The main content area displays a list of search results, including a research article titled 'Incidence, clinical features, and risk factors of fluoroquinolone-induced acute liver injury: a case-control study' and a Reddit post titled '[Journal] Theory of Bose condensation of light via laser cooling of atoms'. The interface also includes buttons for 'EXPORT THIS TAB' and 'SAVE SEARCH'.

Export to CSV

A tweet from World Bank (@WorldBank) with the text: "As economies develop, the composition of countries' wealth changes. Sustainable growth means looking #BeyondGDP to manage wealth as a portfolio of assets that includes human and #NaturalCapital: wrld.bg/nFai30i7iuY #BeyondGDP".

A document titled "More and Better Jobs through Socially Responsible Labour and Business Practices in the Electronics Sector of Viet Nam" with the ILO logo.



Mention CSV includes:

Mention metadata

Mention Type

Mention Date

Outlet or Author

Mention Title

Country

Mention URL

Output metadata

Research Output Title

Journal/Collection Title

Output Type

Publication Date

Altmetric Attention Score

Details Page URL

DOI, PMID, etc

Cannot export Mendeley readership, OSP syllabus mentions



API (metrics)

```
{  
  "title": "Protein hormone boosts memory",  
  "doi": "10.1038/news.2011.49",  
  "nlmid": "0410462",  
  "journal": "Nature News",  
  "altmetric id": 218594,  
  "cited by fbwalls count": 1,  
  "cited by posts count": 2,  
  "cited by rdts_count": 1,  
  "score": 0.5,  
  "history": {  
    "1d": 0,  
    ...  
    "1y": 0.5,  
    "at": 0.5  
  },  
  
  "url": "http://www.nature.com/news/2011/110126/full/  
news.2011.49.html",  
  "subjects": [  
    "science"  
  ],  
  "readers": {  
    "citeulike": "0",  
    "mendeley": "1",  
  }  
}
```

Unlimited no. of calls*

1. Firehose
 - a. Timeframe
 - b. "Mentioned in"
2. Output lookup by PID

Continuously updated

* Rate-limited = 1 call/second



Details Page API (metrics) includes:

Metrics

Counts for Facebook posts,
policy citations, news mentions,
Mendeley readers, etc

Altmetric Attention Score
(over time)

Demographics (Twitter
classifications)

Output metadata

Research Output Title

Journal/Collection Title

Publication Date

Journal subject (if provided)

Date added to Altmetric

URL, DOI, PMID, etc



API response

Metrics

```
"cited_by_fbwalls_count":1,  
"cited_by_posts_count":2,  
"cited_by_rdts_count":1,  
"score":0.5,  
"history":{  
  "1d":0,  
  ...  
  "1m":0,  
  "3m":0,  
  "1y":0.5,  
  "at":0.5  
},  
"readers":{  
  "mendeley":1  
}
```

Output metadata

```
"title":"Number of species on Earth tagged  
at 8.7 million",  
"doi":"10.1038/news.2011.498",  
"nlmid":"0410462",  
"journal":"Nature",  
"Altmetric_id":243700,  
"url":"http://www.nature.com/news/2011/110  
823/full/news.2011.498.html",  
  "added_on":1314135172,  
  "published_on":1314054000,  
  "subjects":[  
    "science"  
  ]
```



Database snapshot JSON file

```
"facebook": [  
  {  
    "title": "Stephen Hawking: 'There  
are no black holes'",  
    "url":  
"https://www.facebook.com/1288798742/post  
s/10201402755092686",  
    "posted_on":  
"2014-01-25T15:45:32+00:00",  
    "license": "public",  
    "author": {  
      "name": "Balram Bodhi",  
      "url":  
"https://www.facebook.com/1288798742",  
      "facebook_wall_name": false,  
      "image":  
"https://graph.facebook.com/1288798742/pi  
cture",  
      "id_on_source": "1288798742"  
    }  
  }  
]
```

10M+ records

Last update: July 2018

Metrics + links to
mentions

Query by PID, key, etc

Updated every 6-12 mo.



Snapshot JSON file includes:

Metrics

Counts for Facebook posts,
policy citations, news mentions,
Mendeley readers, etc

Altmetric Attention Score
(over time, in context)

Demographics (Twitter
classifications, Mendeley
readers, location)

Mention snippets/links

Author name

Mention URL

Snippets

Location

Output metadata (as before)



How to query our data



How to think about our data

1. What are the **outputs** related to your research question?
2. What are the **timeframes** you want to study?
3. What **attention sources** do you want?
4. **What do you want** from those sources?



That's because...

1. It's easiest to search Explorer and query our APIs by output identifiers (e.g. PMIDs or DOIs)
2. Proscribed *mention* time frames in API & Explorer
3. Attention sources aren't all alike
 - a. Twitter vs. blogs vs. Mendeley readers vs. syllabi
4. Full-text (and sometimes metrics!) availability varies



Example: SA of Reddit mentions

“Are conversations around neuroscience research on Reddit positive or negative in nature?”



Example: Pulling the data

1. What are the outputs or title/abstract keywords related to your research question?
All articles assigned a 'neuroscience' FoR code
2. What are the timeframes you want to study?
Mentioned during 'all time'
3. What attention sources do you want?
Reddit
4. What do you want from those sources?
Known Reddit links *and* conversation full-text



Script + snapshot file option

Scan all files in snapshot

Where “posts” -> “reddit” key exists:

Follow Reddit links

If comments appear, scrape comments then run through sentiment analysis script

If no comments appear, mark “No comments”



Studies done with Altmetric data

Do Altmetrics Work? Twitter and Ten Other Social Web Services

Mike Thelwall^{1*}, Stefanie Haustein², Vincent Larivière³, Cassidy R. Sugimoto⁴

1 School of Technology, University of Wolverhampton, Wolverhampton, United Kingdom, **2** École de bibliothécaire et de sciences de l'information, Université de Montréal, Montréal, Québec, Canada and Centre de recherche en éducation, Québec, École de bibliothécaire et de sciences de l'information, Université de Montréal, Montréal, Québec, Canada and Observatoire de données de la recherche, École de bibliothécaire et de sciences de l'information, Université de Montréal, Montréal, Québec, Canada, **3** School of Information and Library Studies, Indiana University Bloomington, Bloomington, Indiana, United States of America, **4** School of Information and Library Studies, Indiana University Bloomington, Bloomington, Indiana, United States of America

Abstract

Altmetric measurements derived from the social web are increasingly advocated and used as early indicators of article impact and usefulness. Nevertheless, there is a lack of systematic scientific evidence that altmetrics are valid proxies of either impact or utility although a few case studies have reported median correlations between specific altmetrics and citation rates for individual journals or fields. To fill this gap, this study compares 11 altmetrics with Web of Science citations for 79 to 206,739 PubMed articles with at least one altmetric mention in each case and up to 1497 journals per metric. It also introduces a simple sign test to overcome biases caused by different citation and usage windows. Statistically significant correlations were found between higher metric scores and higher citations for articles and higher altmetric scores for all cases with significant evidence (Twitter, Facebook wall posts, research4life, blogs, altmetric.com media and forums) and correlations that did differ across journal categories. Evidence was insufficient for LinkedIn, Pinterest, question and answer sites, and Reddit, and no correlations that did differ across journal categories were observed for any other metric. The strength of any correlation between altmetrics and citations. Nevertheless, comparisons between citations and metric values for articles published at different times, even within the same year, city, country or region. This association and its publication and socioeconomics should consider the effect of time when using altmetrics to rank articles. Finally, the coverage of all the altmetrics except for Twitter seems to be low and it is not clear if they are granular enough to be useful in practice.

Citation: Thelwall M, Haustein S, Larivière V, Sugimoto CR (2013) Do Altmetrics Work? Twitter and Ten Other Social Web Services. PLOS ONE 8(2): e48411. doi:10.1371/journal.pone.0064841

Editor: Lisa Berenson, New York State, SENIOR

Received: February 18, 2013; **Accepted:** April 18, 2013; **Published:** May 28, 2013

Copyright: © 2013 Thelwall 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 author and source are credited.

Funding: This research was part of the International Digital Research Infrastructure for Health and Healthcare Research Canada/Economic and Social Research Council of the United Kingdom, Horizon Research, Social Sciences and Humanities Research Council Canada, and the National Science Foundation (award number: grant #0908207). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* m.thelwall@wlv.ac.uk

Introduction

Although scholars may traditionally have found relevant articles by browsing journals, attending meetings and checking over

questionnaire with peers, the use of digital sources that may only

specify relevant articles or online browsing means. While desktop

access to numerous articles and online provides potential

ways to identify the most relevant articles from amongst large

sets. In response, Google Scholar offers search results as

approximately 400 million items of digital, peer-reviewed

information that more highly cited articles are more

important or useful. Digital libraries with search facilities

offer the same service (e.g., ACM, IEEE). In addition

libraries regularly offer access to their fields and can

be used to compare the results to a specific year. Previews

of relevant articles in their fields can be

interested in recent articles. However, given that the

time to access them are not the best indicators of their

work. In response, some publishers have turned to blogs

which are means of citations for researchers to specific

articles, because they can appear more rapidly than citations. For

example, it would be reasonable to expect a typical article to be

most noticed on its publication day and most blogged within a

month of publication. Hence, social media venues have become

available marketing tools for publishers trying to promote current

high impact articles and there are also a number of altmetric

tracking websites that offer free and paid services (e.g.,

altmetric.com, impactstory.com, and altmetrics.org).

The fact that citations take time to accumulate also has an

impact on research evaluation, as a work of a few years after

Table 3. Correlations between metric values and citations (excluding self-citations) for all articles with non-zero scores on each altmetric.

Metric	Spearman	Articles (>0)	Metric total
Tweets	-0.190**	135,331	359,176
FbWalls	0.050**	24,822	35,317
RH	0.373**	23,980	35,365
Blogs	0.201**	13,325	17,699
Google+	0.034**	3,440	5,531
MSM	0.088**	2,402	3,209
Reddits	0.062**	1,516	1,766
Forums	0.033**	82	121
Q&A	0.048**	335	372
Pinners	0.005**	301	324
LinkedIn	0.009**	171	174

* Significant at $p=0.05$. ** Significant at $p=0.01$; both Bonferroni corrected for $n=11$.

doi:10.1371/journal.pone.0064841.t003

Data needed:
Papers with PMIDs;
Altmetric data for papers
 with AAS > 0

Search Explorer to get PMIDs for
 articles with AAS> 0, plus their
 Scores (CSV export)



Get citation data from Web of Science



JCE
JOURNAL OF CLINICAL EPIDEMIOLOGY

Articles and Issues ▾ Collections For Authors ▾ Journal Info ▾ Subscribe More Periodicals ▾

All Content Search Advanced Search

< Previous Article **May 2017** Volume 85, Pages 32-36 Next Article >

To read this article in full, please review your options for gaining access at the bottom of the page.

Health care articles with simple and declarative titles were more likely to be in the Altmetric Top 100

Nicola Di Girolamo^{a,b,*} Reint Meursinge Reynders^{c,d}

PlumX Metrics

DOI: <https://doi.org/10.1016/j.jclinepi.2016.11.018> | Check for updates

Article Info

Abstract Full Text Images References

Abstract

Objectives

The aim of this study was to assess whether specific title characteristics could influence the likelihood of being included in the "Altmetric Top 100."

Methods

We conducted a 1:2 matched case-control study with the cases being the health care articles included in the "Altmetric Top 100" lists (2013-2015) matched through a random computerized procedure with two health care articles published in the same journal and year. For each title, we extracted the number of characters in the title, the number of uncommon words, and whether the title was declarative. Conditional logistic regression was used to estimate odds ratio (OR) with 95% confidence intervals adjusted for a prespecified baseline confounder (open access).

Results

One hundred eight "Medical and health sciences" articles were retrieved in the 2013-2015 "Altmetric Top 100" and matched to 216 control articles. Titles of the "Altmetric Top 100" articles were 102.6 characters (s42) long, included 3.4 (s2.0) uncommon words, and 29.6% (32/108) were "declarative." Titles of the matched articles were 109.3 characters (s37.1) long, included 4.7 (s2.4) uncommon words, and 21.8% (47/216) were "declarative." After multivariate adjustment, declarative titles with a lower number of uncommon words were significantly more represented in the Altmetric list, with declarative titles having 2.8 times the odds of being in the top list (OR: 2.8; 95% confidence interval: 1.2, 6.4). For each additional uncommon word in the title, there was a 1.4 increase in the odds of being a non-Altmetric Top 100 article (1.4; 1.2-1.6).

Conclusion

An easy-to-understand, informative title may help bridge the gap between scholar and social media dissemination.

Data needed:
Altmetric Top 100 articles,
titles, and subject
categorization;
Control articles

Extract titles for Altmetric Top 100
articles in Medical Sciences from
Altmetric data archived on
Figshare



Compare titles and their
characteristics with set of control
articles

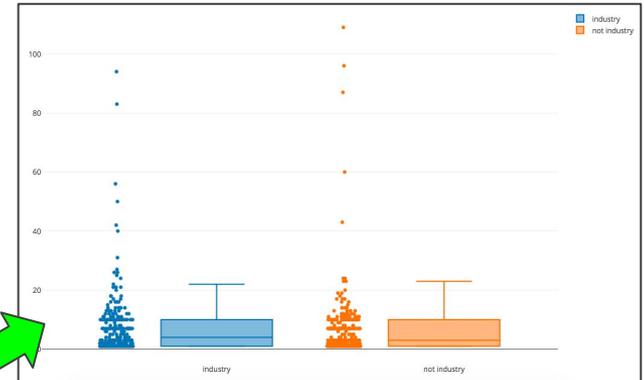


“Do industry-sponsored oncology clinical trials get more attention online?”

Data needed:
DOIs;
AASs for papers;
Acknowledgements/Col
text for papers

Search Explorer to get DOIs for
articles with attention and their
AASs (CSV export)

Article webpages to identify
industry-sponsored clinical trials



Resources



Resources

help.altmetric.com

Explorer, API

api.altmetric.com

API

3rd party API wrappers

API

rAltmetric

pyAltmetric

Stack Overflow

API, JSON

Researcher Resources

Emailed



Resources

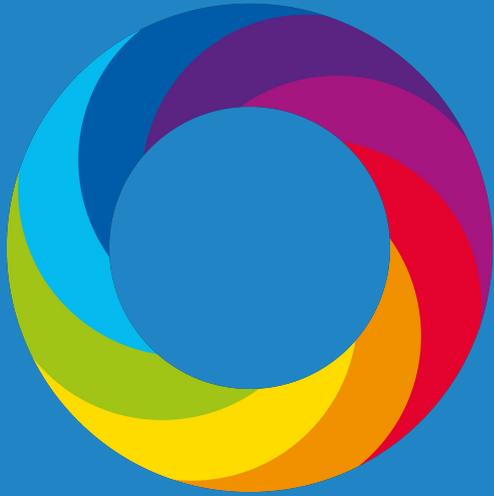


Stacy Konkiel

Director of Research Relations

info@altmetric.com





Thank you!
Questions?

info@altmetric.com