

# Leveraging Big Data - Bibliometric Intelligence for Editors

Elizabeth Caley

COO, Meta



# Meta<sup>α</sup>

## Tools for Researchers and Editors to help with Information Overload

Founded in 2010 • Team of 25+ • Venture Backed

Toronto (HQ) • San Francisco

**MIT  
Technology  
Review**

**Bloomberg**

**WIRED**

**TheScientist**

**TECHVIBES**

 **OUTSELL®**

**YAHOO!**

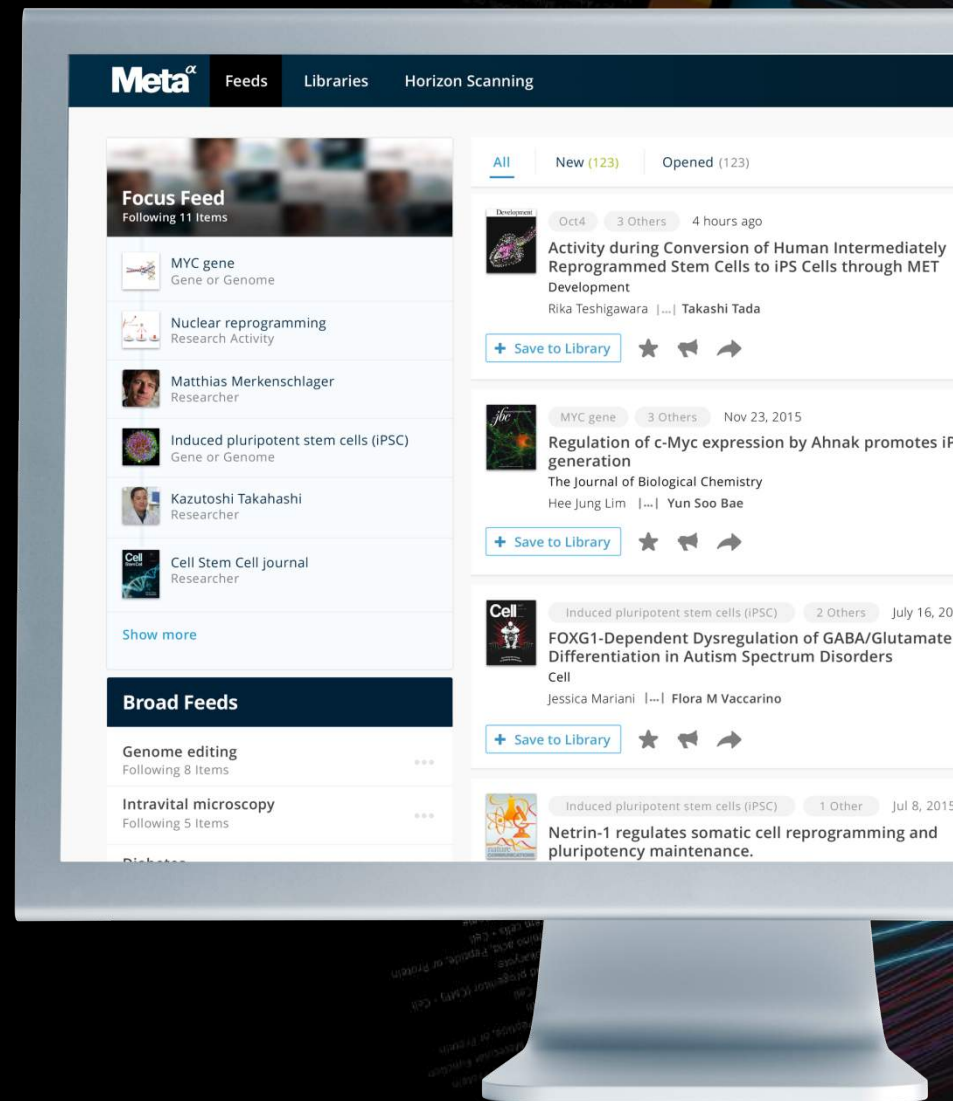
**VentureBeat**

 **TechCrunch**

Literature Discovery

# Meta Science

A free literature discovery site for researchers to stay apprised of the latest developments in their areas of research, explore the evolution of topics, and follow high-impact journals and authors.



*What's the topic?*

*How well does the topic fit my journal?*

*Who authored it?*

*What is their publication history?*

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

Choi et al.  
Functional impacts of amino acid variants  
Page 1

**Abstract**

As next generation sequencing projects generate massive genome-wide sequence variation data, bioinformatics tools are being developed to provide computational predictions on the functional effects of sequence variations and narrow down the search of casual variants for disease phenotypes. Different classes of sequence variations at the nucleotide level are involved in human diseases, including substitutions, insertions, deletions, frameshifts, and non-sense mutations. Frameshifts and non-sense mutations are likely to cause a negative effect on protein function. Existing prediction tools primarily focus on studying the deleterious effects of single amino acid substitutions through examining amino acid conservation at the position of interest among related sequences, an approach that is not directly applicable to insertions or deletions. Here, we introduce a versatile alignment-based score as a new metric to predict the damaging effects of variations not limited to single amino acid substitutions but also in-frame insertions, deletions and multiple amino acid substitutions. This alignment-based score measures the change in sequence similarity of a query sequence to a protein sequence homolog before and after the introduction of an amino acid variation to the query sequence. Our results showed that the scoring scheme performs well in separating disease-associated variants (n=21,662) from common polymorphisms (n=37,022) for UniProt human protein variations, and also in separating deleterious variants (n=15,179) from neutral variants (n=17,891) for UniProt non-human protein variations. In our approach, the area under the receiver operating characteristic curve (AUC) for the human and non-human protein variation datasets is ~0.85. We also observed that the alignment-based score correlates with the deleteriousness of a sequence variation. In summary, we have developed a new algorithm, PROVEAN (Protein Variation Effect Analyzer), which provides a generalized approach to predict the functional effects of protein sequence variations including single or multiple amino acid substitutions, and

*Are the findings novel?*

*How does it rank against other submissions?*

*Is this a highly citable paper?*

*Who could review it?*



**Evaluates** over 400 characteristics of the manuscript

Journal Matches  
**5**

Journal Matching Score  
**81%**

Projected Citation Count  
**23**

Forecasted Impact Score  
**0.81**

Top  
**10%**  
Manuscript Ranking














**New Editor Assignments - Stacey Middle Lavelle, PhD, MDd**

**Contents:** These are submissions that have been Assigned to the Editor. They require one of the following: another Editor assignment, Reviewer invitations, or Decision. Use arrows to change the sort order.

Page: 1 of 1 (1 total submissions)

Display 10 results per page.

Submission 	Manuscript Number 	Article Type 	Section/Category 	Article Title 	Author Name 	Initial Date Submitted 	Status Date 	Current Status 
<a href="#">View Submission</a> <a href="#">Predictive Bibliometrics Results (3/75%)</a> <a href="#">Details</a>  <a href="#">File Inventory</a> <a href="#">Assign Editor</a> <a href="#">Set Final Disposition</a> <a href="#">Submit Editor's Decision and Comm</a> <a href="#">Send E-mail</a>	11.0-46_335	Rapid Communication		test test	Stacey Middle Lavelle, PhD, MDd  ! #	May 10 2016 2:13:50:097PM	May 10 2016 2:20:25:877PM	With Editor

Predictive Bibliometrics Status Summary:  
 (/%) = Article Trajectory Score/Publication Match%  
 (...) = Predictive Bibliometrics analysis in process  
 (X) = Predictive Bibliometrics analysis error

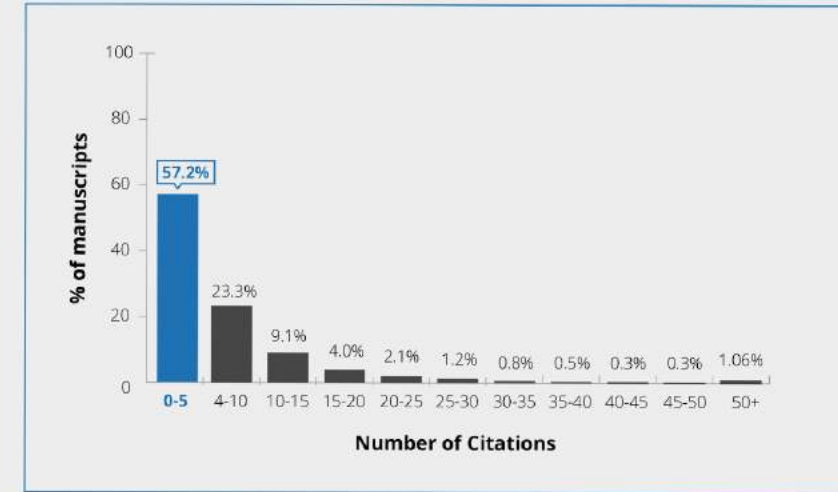
Page: 1 of 1 (1 total submissions)

Display 10 results per page.

# Bibliometric Intelligence

- Determines topic and key concepts
- Pre-ranks incoming manuscripts based on deep predictive profiling
- Provides reviewer recommendations
- Can cascade manuscripts to more appropriate sister journals within a portfolio
- Integrated into Aries Systems' Editorial Manager v13.1

## Predicted 3-Year Citation Count



The predicted 3-year citation count for a manuscript is the projected number of citations that the manuscript will accrue over the next three years.

The predicted citation count for this manuscript is **4.064** over the next three years. Out of over 500K manuscripts published in the last six months, **23.3%** fall within this range.



## Predictive Analytics

The major research areas identified in this manuscript are:

 <b>Chronic Lymphocytic Leukaemia Refractory</b>	 <b>Alemtuzumab</b>
 <b>Therapy-related Myelodysplastic Syndrome</b>	 <b>Fludarabine</b>
 <b>Clinical Trial [Publication Type]</b>	 <b>Hematopoietic Stem Cells</b>
 <b>Sequencing</b>	 <b>Acute Myeloid Leukemia Pathway</b>
 <b>DNA Damage</b>	 <b>Cytopenia</b>






## Journal Matching Score

The Journal Matching score is a score representing how well the submitted manuscript matches the topics and scope of the desired journal, based on an analysis of the research areas covered in the manuscript.

The research areas identified in this manuscript are **58.4%** consistent with the research areas covered in the selected journal.

## Recommended Reviewers

Recommended Reviewers is an intelligently curated list of active researchers who would be best suited to review the manuscript, based on the topics identified in the manuscript. By comparing the topics within this manuscript to the five-year publication history of over seven million authors, the following active researchers (and their most relevant paper) have been recommended to review this manuscript.

	<b>Cheng-Hwai Tzeng</b> Clinicopathologic features and outcome of acute erythroid leukemia based on 2008 revised World Health Organization classification <i>Leukemia &amp; Lymphoma</i>
	<b>Gail J Roboz</b> Targeted Deletion of Autophagy Genes Atg5 Or Atg7 in The Chondrocytes Promotes Caspase-Dependent Cell Death And Leads To Mild Growth Retardation <i>Journal of Bone and Mineral Research</i>
	<b>Detlef Haase</b> New comprehensive cytogenetic scoring system for primary myelodysplastic syndromes (MDS) and oligoblastic acute myeloid leukemia after MDS derived from an international database merge <i>Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology</i>
	<b>Richard J B Wells</b> Initial transfusion intensity predicts survival in myelodysplastic syndrome <i>Leukemia &amp; Lymphoma</i>
	<b>Juan Gao</b> Correlation of chromosome karyotype with dyshaematopoiesis and reticulin in myelodysplastic syndrome <i>Zhongguo shi yan xue ye xue za zhi</i>



Publishing Partner Services

# Bibliometric Intelligence

Life Genomics 2016



## Analysis Summary

Analysis Summary

Contents

Eigenfactor

Citations

Rank

2,274

Articles

413

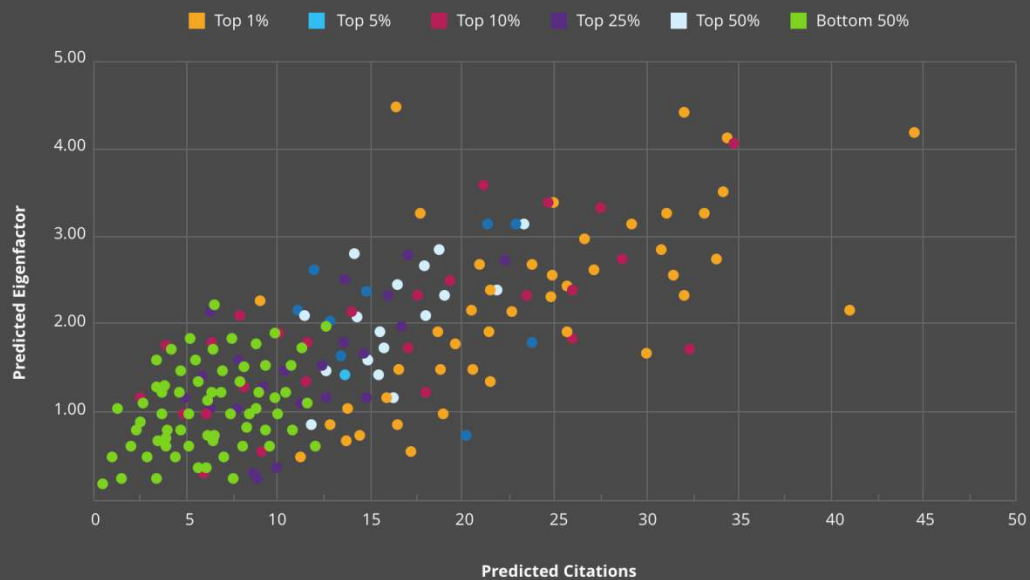
Journals

0.111

Average Predicted Eigenfactor

2

Average Predicted Citations



## NON-COMMUNICABLE DISEASES IN THE ASIA-PACIFIC REGION: PREVALENCE, RISK FACTORS AND COMMUNITY-BASED PREVENTION

WAH-YUN LOW, Yew-Kong Lee and ALEXANDER LOURDES SAMY

University of Malaya, Kuala Lumpur, Malaysia  
Faculty of Medicine

### Abstract

Non-communicable diseases (NCDs) are a substantial mortality and morbidity worldwide. The most common NCDs are cardiovascular diseases (CVDs), **Diabetes** and chronic respiratory diseases. With the rapid increase in NCD-related deaths in Asia Pacific countries, NCDs are one of the major causes of deaths and disease burden in the region. **NCD** poses a challenge to the achievement of the Millennium Development Goals (MDG). People in the low socioeconomic group are most affected by NCDs as they have poor access to policies, legislation, regulation and healthcare services meant to control NCDs. This results in loss of productivity by accumulating 100s of years with implications at the macroeconomic level. The 3 major NCDs in the Asia Pacific region are **CVD**, **Diabetes** and **Chronic respiratory diseases**. The increasing prevalence of NCDs and its associated risk factors, such as **Obesity** and **Hypertension**, are major public health concerns. Strategies to combat NCDs in the Asia Pacific region are as follows: population-based dietary salt reduction, health education, psychological interventions, i.e., cognitive behavioral therapy and motivational-interviewing, taxation and bans on tobacco-related advertisements, implementing smoke-free zones and surveillance by the World Health Organization. Control measures must focus on prevention and strengthening inter-sectoral collaboration.

### Key words:

Non-communicable diseases (NCDs), Urbanization, Hypertension, Obesity, Community-based prevention, Inter-sectoral collaboration

### INTRODUCTION

Non-communicable diseases (NCDs) are the pivotal cause of disease burden and mortality in the Asia Pacific region, claiming 55% of total life in the South East Asia region each year and 75% in the Western Pacific region [1,2]. The Asia Pacific region is experiencing a rapid increase in NCD-related deaths; the World Health Organization estimates that the highest worldwide increment in total

mortality in a 10-year time frame (2005–2015) will be observed in the South-East Asia and Western Pacific regions with 21% increase in the South-East Asia region [3] and 12.3 million deaths in the Western Pacific region [2]. This increase in NCDs presents a major barrier to global development, specifically to the achievement of the Millennium Development Goals [4] in low-and-middle income countries.

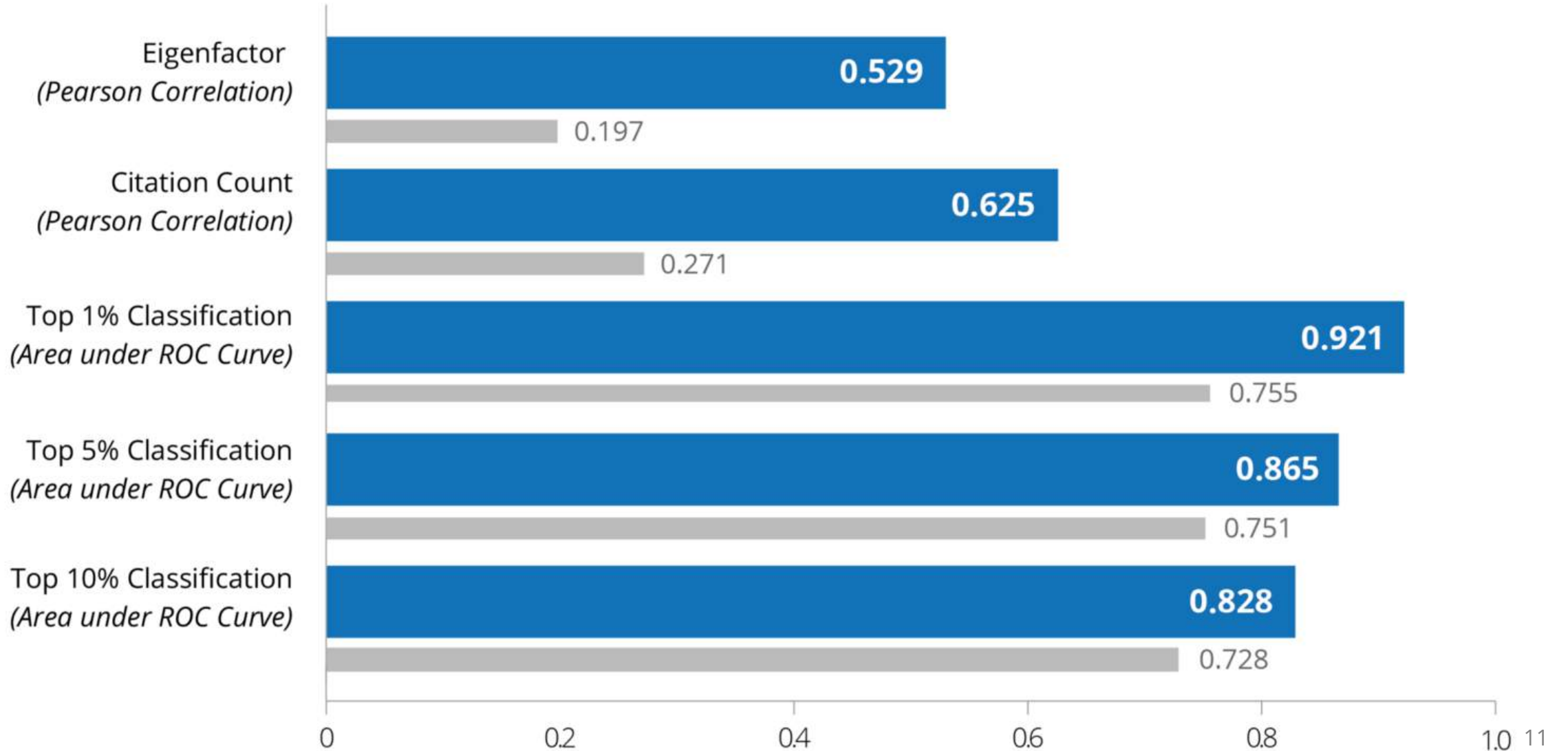
This paper was presented at the 6th ICQH International Conference on Work Environment and Cardiovascular Diseases, 2013 Mar 27–30; Tokyo, Japan. Received: May 9, 2014; Accepted: September 2, 2014.  
Corresponding author: W.-Y. Low, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia (e-mail: lowwy@um.edu.my).

Bibliometric Intelligence

**Over 1M Biomedical manuscripts  
analyzed over 3 years  
Examines over 400 characteristics**

# Bibliometric Intelligence Performance

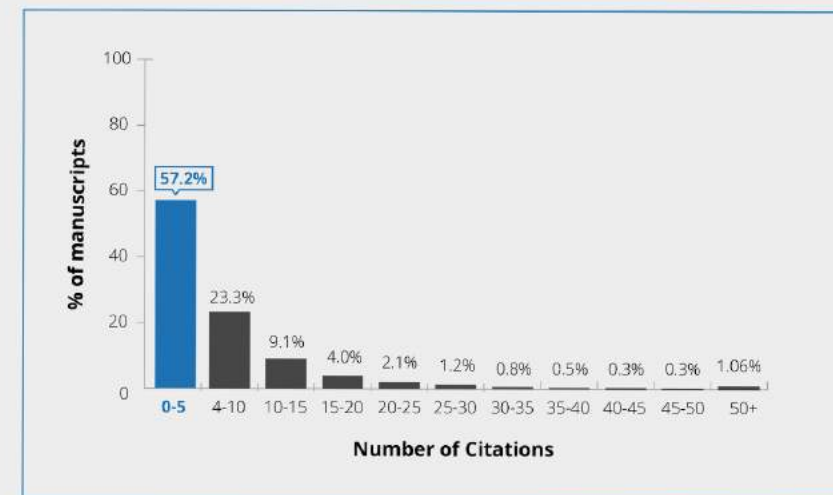
■ Predicted Impact   ■ Baseline



# Bibliometric Intelligence

A set of quantitative data for editors to use that complements your expertise in evaluating manuscripts.

## Predicted 3-Year Citation Count



The predicted 3-year citation count for a manuscript is the projected number of citations that the manuscript will accrue over the next three years.

The predicted citation count for this manuscript is **4.064** over the next three years. Out of over 500K manuscripts published in the last six months, **23.3%** fall within this range.



# Thank You

For more information visit [meta.com](https://meta.com)

Elizabeth Caley · [ec@meta.com](mailto:ec@meta.com)