A Method for Verifying Indicators of Journal Quality

Jere Odell
jdodell@iupui.edu

David Edward (Ted) Polley
dapolley@iupui.edu

IUPUI University Library
Center for Digital Scholarship
October 26, 2018
Outline

• Project Context
  • “Predatory” OA?
  • Dishonest APC journals at IUPUI?
  • Data analysis tasks at IUPUI University Library

• The Details: Digital Measures + Open Refine + R + APIs

• The Results

• What’s next for this project?
IUPUI Leads Indiana in the Publication of Articles in Legitimate OA journals

But what about the “predatory” journals?

Yes, “worthless” journals are a problem, but ...

Shen and Bjork, 2015

Bohannon, 2013
... part of the problem is: we don’t know what we’re talking about.
“SciELO and Redalcy aggregate scholarly content but do a poor job of distributing it or increasing its visibility, despite the fact that both services are open-access. Many North American scholars have never even heard of these meta-publishers or the journals they aggregate. Their content is largely hidden, the neighborhood remote and unfamiliar.”

“Thus, commercial publisher platforms are nice neighborhoods for scholarly publications. On the other hand, some open-access platforms are more like publication favelas.” J.B.

Xenophobic Turf Battle?

World scaled by number of documents with authors from each country in Web of Science

https://doi.org/10.6084/m9.figshare.7064771.v1
Evidence-based decisions?

But a scam is a scam.
And it might be a big one.

420,000 articles in 2014 – sample of 613 journals on B.’s list
$74 million
(Shen and Bjork, 2015)

- OR -

135,000 articles in 2014 – sample of 655 journals on B.’s list (Crawford, 2015)
$32 million
(assuming 2.5 million articles per year ... probably under counting)

5% - 15% of global production in disreputable journals
Who writes these articles?

Stratified sample: 200 articles from 47 journals

U.S. authors: 6% of sample

(Shen and Bjork, 2015)
Who writes these articles?

India: 27% of sample (> 10% of national output; Scopus)
USA: 15% of sample (< 1% of national output; Scopus)

But, of course, most of these journals are not indexed in Scopus.
Why do authors submit to fraudulent journals?

71% of respondents
unaware that the journal could not be trusted
Sample: 96 authors
(61% published in a journal from their country)

“I sent my paper to that journal because I checked its editorial pages and I found [that its] editors [were] from India. Since I am also from India, I thought this journal would understand my paper more and decide accordingly.”

(Kurt, 2018)
Authors from non-Western countries may be justified in their suspicions

“manuscripts were more likely to be accepted when reviewed by at least one gatekeeper with the same national affiliation as the corresponding author. Our results indicated that homogeneity between author and gatekeeper gender and nationality is associated with the outcomes of scientific peer review.”

Conundrum

1) How can we best advise authors and others in our universities, while not reinforcing a broken value system that favors wealthy?

   Short-term/local interests – vs – long-term/global well-being

2) If, reluctantly, we’re going reinforce a broken value system, is there some way that we can do so without relying on subjective and/or prejudiced measures?
But do we have a problem at IUPUI?
IUPUI Articles in “Predatory” Journals?


Sample size: 1,907 articles in 200 journals

IUPUI = 0
IUPUI articles = 6
(about 1 per year)

5 corresponding authors
(4 with tenure at date of publication)
IUPUI OA Fund Requests (Oct 2013-March 2018)

Total Requests for Support: 176
Requests Denied for Journal Quality: 3 (2%)
(OA Journal not included in the DOAJ)

2 = Oncotarget (http://www.oncotarget.com)
[Indexed in Web of Science 2017; IF 5.8]

1 = International Journal of Emerging Technology and Advanced Engineering (http://www.ijetae.com/)
Digital Measures Activity Insight – IUPUI 2017 (did not include IU School of Medicine)
• Do you or your colleagues know the journal?
• Can you easily identify and contact the publisher?
• Is the journal clear about the type of peer review it uses?
• Are articles indexed in services that you use?
• Is it clear what fees will be charged?
• Do you recognise the editorial board?
• Is the publisher a member of a recognized industry initiative? (DOAJ, OASPA, COPE, AJOL, etc)

http://thinkchecksubmit.org/
Verifiable Characteristics of Disreputable APC Journals

Filters
• Not indexed by the DOAJ: https://doaj.org/
• Not affiliated with a scholarly society or university

Faults
• The journal makes false claims about indexing in PubMed, DOAJ, etc.
• The journal makes false claims about affiliation with a scholarly society or university
• The journal makes false claims about location of publication (e.g., title says “American” but it is published in Nigeria)
Preparing the Data for Analysis – 1/4

• Entries in DMAI for “Publications / Scholarship of Discovery”: **2,882**
  • (Note: IU School of Medicine did not use DMAI in 2017)

• **Excluded entries by type:**
  • Article, non-peer review
  • Book Chapters
  • Book Reviews
  • Books
  • “Other”
  • Textbook
  • _blank_

• **Excluded entries by status:**
  • Accepted
  • Not accepted
  • Submitted
  • Resubmitted
  • Under revision
Preparing the Data for Analysis – 2/4

• Separated and created a new data set for “Conference proceedings”: **351 items**

• Excluded:
  • Presentations without published conference paper
  • Abstracts
  • Posters
  • Duplicate entries (by title)
  • Items published prior to 2017

• Eligible Conference Proceeding Items: **157 items**

• Retained data set for hand coding
Preparing the Data for Analysis – 3/4

• Duplicates result when co-authors enter the same item with different metadata (differences: author order, journal name, title)

• Removed duplicate entries from article data set by:
  • Title (inaccuracy in title entries made this difficult)
  • DOI (of this data set **56% of entries included author-supplied DOI**)
  • Retained entry of:
    • Corresponding author
      -- or, if no IUPUI corresponding --
    • Author with most complete metadata

• Total eligible items: **1,101** (944 articles + 157 conference papers)
Preparing the Data for Analysis – 4/4

Controlled Journal Name using OpenRefine’s Facet function

(Articles Only: 944 articles; ~703 unique journals)
Our Approach

• Looking for scalable way to reduce the labor involved in assessing the quality of venues for faculty research
• Increasing availability of faculty publication data through faculty reporting and research profiling systems
• Use the SHERPA/RoMEO and DOAJ APIs to supplement faculty reporting data to speed up decision-making
The Data

- We obtained data from faculty reporting system
- Data were predictably messy, but we had *title of contribution* and *journal title* consistently – occasionally faculty entered the *ISSN*
- What is needed to help assess journal quality:
  - Publisher
  - Indexed in DOAJ?
Pre-processing

- Journal titles were normalized using OpenRefine (http://openrefine.org/)
- Bare minimum data needed for this process:

<table>
<thead>
<tr>
<th>unique_ID</th>
<th>title_of_contribution</th>
<th>normalized_journal_title</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Format journal title as SHERPA/RoMEO API query in Excel:
  [link](http://www.sherpa.ac.uk/romeo/api29.php?jtitle=Academic%20Emergency%20Medicine%20Education%20and%20Training&qtype=exact)
Processing

• Load the data into RStudio (https://www rstudio.com/)

• R packages used:
  • dplyr ¹
  • httr ²
  • magrittr ³
  • XML ⁴

Step 1: Get publisher name and ISSN from SHERPA/RoMEO

<table>
<thead>
<tr>
<th>unique_ID</th>
<th>title_of_contribution</th>
<th>normalized_journal_title</th>
<th>issn*</th>
</tr>
</thead>
</table>

http://www.sherpa.ac.uk/romeo/apimanual.php?la=en&fIDnum=|&mode=simple

* In the data we obtained from faculty reporting, there were few ISSNs. If no record was found in SHERPA/RoMEO search and ISSN was present in faculty data, these values were merged to use for the next step.
Step 2: Check to see if journal is indexed in DOAJ

https://doaj.org/api/v1/docs
Results of R script

• R script processed 944 records in 4.333 minutes
• Reduced the number of records that required manual verification to 345 (313 records not in SHERPA/RoMEO and 32 returned multiple matches)
Limitations

• Journals that are not in SHERPA/RoMEO database that are indexed in DOAJ – our script misses these

• The multiple matches issue requires potentially unnecessary manual verification

• There are likely more comprehensive sources for ISSN and publisher information, such as ISSN API – but these are often not publicly accessible
### Hand Coding

#### Data Summary

<table>
<thead>
<tr>
<th>Data Summary</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annual Review entries</td>
<td>2882</td>
</tr>
<tr>
<td>Eligible items</td>
<td>1101</td>
</tr>
<tr>
<td>Articles</td>
<td>944</td>
</tr>
<tr>
<td>Conference papers</td>
<td>157</td>
</tr>
<tr>
<td>Unique journals</td>
<td>703</td>
</tr>
<tr>
<td>Articles to hand code</td>
<td>345</td>
</tr>
<tr>
<td>Conference papers to hand code</td>
<td>157</td>
</tr>
</tbody>
</table>

#### Trusted
- Pay-walled journal
- OA journal indexed in DOAJ
- Known publisher
- No-fee OA journal
- For-fee OA journal published by known university
- For-fee OA journal published by known scholarly society

#### Untrusted (if none of the above and ...)
- Journal makes false claims about one of the following:
  - Indexing
  - Scholarly society/university sponsorship
  - Origin or scope of literature ("American")
## Summary of Results

<table>
<thead>
<tr>
<th>Trusted?</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>1079</td>
</tr>
<tr>
<td>No (appears to be disreputable)</td>
<td>15 (14 articles + 1 conference)</td>
</tr>
<tr>
<td>Maybe</td>
<td>5</td>
</tr>
<tr>
<td>Ineligible</td>
<td>2 (preprints)</td>
</tr>
</tbody>
</table>

98% trusted publications (1079/1101)

OR

1.4% disreputable publications (15/1101)

No evidence of dishonesty, but below threshold for “trusted”
Difficult Decisions
(The following examples were not selected from the data set.)

Difficult Decisions

http://www.jclindent.com/Information.html

The
Journal of
Clinical Dentistry®
THE INTERNATIONAL JOURNAL OF ORAL CARE PRODUCT RESEARCH
www.JClinDent.com

Instructions for Authors
Download Instructions

Mission Statement—The Journal is dedicated to the rapid publication of research and reviews focused on preventive oral healthcare products and new dental materials. Both laboratory and clinical research are accepted, as well as case reports. All scientific studies are blinded and peer-reviewed by an expert panel of researchers at independent academic institutions. The Journal also publishes issues dedicated to a single product and its research and development. These publications are designed to educate the dental professional on the safety and efficacy of these products so an informed and confident product recommendation can be made to patients. The Journal accepts no advertising. All papers accepted for publication will be assessed a placement fee of US$800 per published page.
Difficult “Decisions”

https://www.tandfonline.com/toc/uabr21/current
Difficult “Decisions”
Difficult “Decisions”

https://us.sagepub.com/en-us/nam/node/1008902
Difficult “Decisions”

Diagnostic usefulness of routine Lyme serology in patients with early inflammatory arthritis in nonendemic areas

Géraldine Muller *a, Anne Cherassé *, Jean Baptiste Bour *, Christian Tavernier *a, Jean Francis Maillefert *

* Department of Rheumatology, CRU Dijon, Dijon, France
a Department of Internal Medicine and Systemic Diseases, CRU Dijon, Dijon, France
* Rheumatology and Spondylitis Laboratory, CRU Dijon, Dijon, France

Received 5 March 2002, accepted 17 September 2002

Abstract

Objective. – To evaluate the diagnostic usefulness of routine Lyme serology in patients who live in nonendemic areas and present with early inflammatory joint disease.

Methods. – All patients admitted to a rheumatology department in a nonendemic area of France for evaluation of joint disease with onset within the last year. The evaluation included a medical history, a thorough physical examination, an electrocardiogram, and an ELSA for antibodies to Borrelia burgdorferi.

Results. – We included 90 patients, 51 women and 39 men, with a mean age of 48.1 ± 17.9 years. Mean duration of joint symptoms was 4.3 ± 4.3 months, with a median of 3 months. A patient (1.1%) reported a tick bite and no patients had a history of erythema migrans. Lyme serology was negative in all 90 patients.

Conclusion. – These results do not support routine Lyme serology in patients living in nonendemic areas and presenting with early inflammatory joint disease. However, Lyme serology remains appropriate in patients with features suggestive of Lyme disease. Given that Lyme disease is amenable to antibiotic treatment, a larger study is in order to confirm our findings.

Keywords: Arthritis, Polyarthropathy, Lyme disease; Borrelia burgdorferi; Diagnosis

Difficult? “Decisions”

Springer, IEEE withdrawing more than 120 nonsense papers

Two major publishers will remove more than 120 papers created with random paper generator SCIgen, according to Nature.

Richard van Noorden, who has the scoop, reports:

Over the past two years, computer scientist Cyril Labbé of Joseph Fourier University in Grenoble, France, has catalogued computer-generated papers that made it into more than 30 published conference proceedings between 2008 and 2013.

http://retractionwatch.com/2014/02/24/springer-ieee-withdrawing-more-than-120-nonsense-papers/
Be a Responsible Citizen of the Scholarly Ecosystem

Will publishing in this venue result in a return on your investment? (Do you own the work? Do you share in the profits?)

Will publishing in this venue contribute to your credibility as a scholar?

THINK

Are you submitting your research to a trusted journal? Is it the right journal for your work?

Will publishing in this venue result in a return on your community’s investment? (Does the university lose resources? Do researchers and students share the profits?)

Will publishing in this venue permit low-cost or no-cost dissemination to academic and community readers?
References

• Abrasco, Brazilian Public Health Association. (2015, August 2). Motion to repudiate Mr. Jeffrey Beall’s classist attack on SciELO. Retrieved from https://blog.scielo.org/en/2015/08/02/motion-to-repudiate-mr-jeffrey-bealls-classist-attack-on-scielo/

• Alperin, J. P. (2018, September 8). World scaled by number of documents with authors from each country in Web of Science: 2016. https://doi.org/10.6084/m9.figshare.7064771.v1


