Amplifying the Impact of Open Access: Wikipedia and the Diffusion of Science

Misha Teplitskiy
*Department of Sociology and KnowledgeLab, University of Chicago, 5735 South Ellis Avenue, Chicago
IL 60637. E-mail: misha@uchicago.edu*

Grace Lu
*KnowledgeLab, University of Chicago, 5735 South Ellis Avenue, Chicago, IL 60637. E-mail:
gracelu@uchicago.edu*

Eamon Duede
*Computation Institute and KnowledgeLab, University of Chicago, 5735 South Ellis Avenue, Chicago, IL 60637.
E-mail: eduede@uchicago.edu*

With the rise of Wikipedia as a first-stop source for scientific information, it is important to understand whether Wikipedia draws upon the research that scientists value most. Here we identify the 250 most heavily used journals in each of 26 research fields (4,721 journals, 19.4M articles) indexed by the Scopus database, and test whether topic, academic status, and accessibility make articles from these journals more or less likely to be referenced on Wikipedia. We find that a journal's academic status (impact factor) and accessibility (open access policy) both strongly increase the probability of it being referenced on Wikipedia. Controlling for field and impact factor, the odds that an open access journal is referenced on the English Wikipedia are 47% higher compared to paywall journals. These findings provide evidence is that a major consequence of open access policies is to significantly amplify the diffusion of science, through an intermediary like Wikipedia, to a broad audience.

Introduction

Wikipedia, one of the most visited websites in the world,[1] has become a destination for information of all kinds, including information about science (Heilman & West, 2015; Laurent & Vickers, 2009; Okoli, Mehdi, Mesgari,

Nielsen, & Lanamäki, 2014; Spoerri, 2007). Given that so many people rely on Wikipedia for scientific information, it is important to ask whether Wikipedia's coverage of science is a balanced, high quality representation of the knowledge within the academic literature. One approach to asking this question involves looking at references used in Wikipedia articles. Wikipedia requires all claims to be substantiated by reliable references,[2] but what, in practice, are "reliable references?"

An intuitive approach is to compare the sources Wikipedia editors use to the sources scientists value most. In particular, within the scientific literature, a journal's status is often associated, albeit problematically (Seglen, 1997), with its impact factor. If status within the academic literature is taken as a "gold standard," Wikipedia's failure to cite high impact journals of certain fields would constitute a failure of coverage (Samoilenko & Yasseri, 2014), whereas a high correspondence between journals' impact factors and citations in Wikipedia would indicate that Wikipedia does indeed use reputable sources (Evans & Krauthammer, 2011; Nielsen, 2007; Shuai, Jiang, Liu, & Bollen, 2013).

Yet high impact journals often require expensive subscriptions (Björk & Solomon, 2012). The costs are, in fact, so prohibitive that even Harvard University has urged its faculty to "resign from publications that keep articles behind paywalls" because the library "can no longer afford the price hikes imposed by many large journal publishers" (Sample, 2012). Consequently, many within the scientific community advocate journals that provide free access to research - "open access" (OA) journals. (Van Noorden, 2013). A lively debate has arisen on the impact of OA on the scientific

literature, with some studies showing a citation advantage to publishing in OA versus paywall journals (Eysenbach, 2006a,b; Gargouri et al., 2010; "The Open Access Citation Advantage Service") whereas others find none (Davis, 2011; Davis, Lewenstein, Simon, Booth, & Connolly, 2008; Gaulé & Maystre, 2011; Moed, 2007).

Regardless of their impact on the scientific literature, OA journals may have a tremendous impact on the diffusion of scientific knowledge *beyond* this literature. To date, this potential of OA policies has been a matter chiefly of speculation (Heilman & West, 2015; Trench, 2008). Previous research has found that OA articles are downloaded from publishers' websites more often and by more people than closed access articles (Davis, 2010, 2011), but it is currently unclear by whom, and to what extent open access affects the use of science by the *general public* (Davis & Walters, 2011). We hypothesize that Wikipedia, with more than 8.5 million page views *per hour*,[3] is a new but crucial pathway through which the public consumes science and this diffusion of science may relate to its accessibility in two ways. By referencing findings from paywall journals that may be prohibitively expensive, Wikipedia distills and diffuses these findings to the general public. On the other hand, Wikipedia editors may be unable to access expensive paywall journals,[4] and consequently reference the easily accessible articles instead. For example, Luyt and Tan's (2010) study found accessibility to drive the selection of references in a sample of Wikipedia's history articles. In this case Wikipedia "amplifies" open access science by broadcasting its (already freely accessible) findings to millions. This "amplifier" effect may thus constitute one of the chief effects of open access.

Correspondence Between Academic and Wikipedia Statuses

This article tests both the distillation and amplifier hypotheses by evaluating which references Wikipedia editors use and do not use. In particular we study how a journal's status within the scientific community (impact factor) and accessibility (open access policy) relates to its status within Wikipedia (percent of its articles referenced on Wikipedia). It is important to note that an observed correspondence may be driven by a variety of mechanisms besides accessibility. First, the status ordering of academic journals as measured with impact factors may have only a tenuous relationship with the importance and notability—considerations of special relevance to Wikipedia[5]—of the published research. Citations, and therefore impact factors, are in part a function of the research field (Seglen, 1997), and may be affected by factors as circumstantial as whether a paper's title contains a colon (Jamali & Nikzad, 2011; Seglen, 1997). Second, the academic status ordering results from the objectives of millions of scientists and institutions, and may be irrelevant to the unique objectives of Wikipedia. Wikipedia's key objective is to serve as an encyclopedia, not a medium through which scientists communicate original research[6]. Relative to the decentralization of the scientific literature, Wikipedia is governed by explicit, if flexible, policies and a hierarchical power structure (Butler, Joyce, & Pike, 2008; Shaw & Hill, 2014). Apart from a remark that review papers serve Wikipedia's objectives better than primary research articles, Wikipedia's referencing policies generally pass no judgment over which items within the scientific literature constitute "the best" evidence in support of a claim[7]. Wikipedia's objectives and explicit, centrally accessible, policies differ from the decentralized decisions that produce status orderings within the scientific literature and do not imply that the two status orderings should correspond. Indeed, if editors are not scientists themselves they need not even be aware that journal impact factors exist.[8] On the other hand, despite the well-worn caveats, prestigious, high-impact journals may publish findings that are more important to both academics and Wikipedia's audience. In fact, a Wikipedia editor's *expectation* that the truly important research resides within high-impact journals may be enough to predispose them to want to reference such journals. Second, little is known about editors of science-related articles (West, Weber, & Castillo, 2012); they may be professional scientists with access to these high-impact journals, resulting in both the *motivation* and *opportunity* to reference them.

Previous Research

Wikipedia References and Academic Status

The first large-scale study of Wikipedia's scientific references was performed by Finn Arup Nielsen (Nielsen, 2007). Nielsen found that the number of Wikipedia references to the top 160 journals, extracted from the *cite-journal* citation templates, correlated modestly with the journal's *Journal Citation Reports* impact factor. This implication that Wikipedia preferentially cites high impact journals is delicate in part because the data used in the study included only a subset of journals with references that appear in Wikipedia, not journals that were and *were not* referenced. It is possible, albeit unlikely, that an even larger number of prestigious journals, made invisible by the methodology, are never referenced on Wikipedia at all, weakening the correlation to an unknown degree, or that the referenced journals are simply those that publish the most articles (see Nielsen 2007: Figure 1). Shuai et al. (2013) also found modest correlations when they investigated a possible correspondence between the academic rank of computer science papers, authors, and topics and their Wikipedia rank.

The altmetrics movement has also explored Wikipedia as a nonacademic venue on which academic literature makes an impact (ALM, Fenner, & Lin, ; "altmetrics,"; Priem, 2015). Evans and Krauthammer (2011) examined the use of Wikipedia as an alternative measure of the scholarly impact of biomedical research. The authors correlated scholarly metrics of biomedical articles, journals, and topics with Wikipedia citations and, in contrast to other studies, included in some of their analyses a random sample of
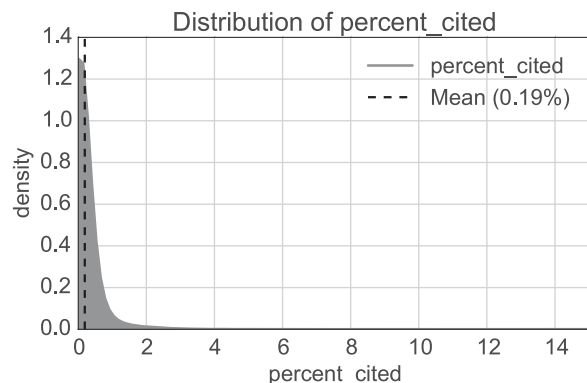
FIG. 1.   Distribution (kde) of percent_cited of 4,774 journals.

journals never referenced on Wikipedia. The authors also recorded a journal's OA policy but, unfortunately, do not appear to have used this information in analyses.

Open Access and the Web

The rather voluminous literature on OA has focused primarily on effects on the academic literature.[9] There is some debate on the size and direction of OA effects. Some evidence demonstrates that OA articles gain a citation advantage (Eysenbach, 2006a, 2006b; Gargouri et al., 2010; "The Open Access Citation Advantage Service"), whereas other evidence shows no such effect (Davis, 2011; Davis et al., 2008; Gaulé & Maystre, 2011; Moed, 2007). Regardless of the impacts on scientists in developed nations, increased accessibility through OA does yield benefits to scientists from developing nations (Davis & Walters, 2011; Evans & Reimer, 2009).

The promise of OA for disseminating scientific information to the world at large has gained much less attention (Davis & Walters, 2011; Trench, 2008; for an exception see Heilman & West, 2015). Yet, more and more of the world turns to the web for scientific information. For instance, as early as 1999 a full 20% of American adults sought medical and science information online (Miller, 2001). What's more, one who actively seeks such information within the academic literature will quickly discover that, despite the paywalls, many important and impactful research articles are made freely available by their authors or third parties (Björk, Laakso, Welling, & Paetau, 2014; Wren, 2005). This is to say nothing of the fact that science may also be disseminated through distillation of its findings into venues like Wikipedia or science-centric websites and blogs so that, here too, the impact of OA may be limited. Although full texts of the most impactful literature are, at least nominally, behind a paywall (Björk & Solomon, 2012), do Wikipedia's editors consult these texts? If they cite them in Wikipedia, have they consulted the full texts beyond a freely available abstract before referencing? If the academic literature is any guide, referenced material is sometimes consulted rather carelessly (Broadus, 1983; Rekdal, 2014). In short, the current understanding of the relationship between OA and the general public in the literature is limited at best (Davis & Walters, 2011).

Shortcomings and Our Contribution

In addition to the role of accessibility, a number of substantive and methodological shortcomings remain. First, it is unclear if professional scientists edit Wikipedia's science articles. As we will show, a preponderance of paywall references would suggest, albeit indirectly, this to be the case.[10] The scant existing evidence indicates that science articles are edited by people with general expertise, relative to the more narrow experts of popular culture articles (West et al., 2012). Second, most previous studies have completely ignored the articles that are never referenced on Wikipedia, thus sampling on the dependent variable. The only notable exception, (Evans & Krauthammer, 2011), treated the unreferenced articles outside the main analytic framework. Although the framework treated (referenced) articles or journals as the unit of analysis, the unreferenced articles and journals were treated as a homogeneous group.

This study extends existing work in three chief ways. First, it models the role of accessibility (OA status) on referencing. Second, it covers *all* major research areas of science by observing rates at which Wikipedia references nearly 5,000 journals, accounting for nearly 20 million articles. Third, it treats unreferenced articles in the same analytic framework as those referenced. Yet the study is not without its own limitations, which will be outlined more fully later in the article. Chief among these are that article-level characteristics are operationalized by the characteristics of the publishing journal. For example, the accessibility of articles is operationalized by their journal's OA policy, when, in fact, free access to many paywall articles exists through sanctioned or unsanctioned file-sharing (Björk et al., 2014; Wren, 2005). Thus, any observed advantage of OA referencing may be biased downward, that is, an underestimate of the true effect (see the Conclusion for a discussion of measurement error).

Data and Methods

Data Sample

Journal data.   Our analysis uses journal-level data from thousands of journals indexed by Scopus. Indexing over 21,000 peer-reviewed journals and with more than 2,800 classified as open access, Scopus is the world's largest database of scientific literature.[11] We obtained information on the 250 highest-impact journals within each of the following 26 major subject areas[12]: Agricultural Sciences, Arts and Humanities, Biochemistry and General Microbiology, Business Management and Accounting, Chemical Engineering, Chemistry, Computer Science, Decision Sciences, Earth and Planetary Sciences, Economics and Finance, Energy Sciences, Engineering, Environmental Sciences, Immunology and Microbiology, Materials Sciences, Mathematics, Medicine, Neurosciences, Nursing, Pharmacology, Physics, Psychology,

TABLE 1. 15 highest-impact journals within Scopus according to SCImago impact factor (2013).

Journal	Impact factor (SCImago2013)
CA - A Cancer Journal for Clinicians	45.894
Reviews of Modern Physics	34.830
Annual Review of Immunology	32.612
Cell	28.272
Annual Review of Biochemistry	27.902
Quarterly Journal of Economics	25.168
Nature Genetics	24.052
Nature Reviews Genetics	23.813
Nature Reviews Molecular Cell Biology	23.593
Chemical Reviews	23.543
Nature	21.323
Acta Crystallographica Section D: Biological Crystallography	20.717
Advances in Physics	20.349
Annual Review of Cell and Developmental Biology	19.686
Annual Review of Neuroscience	19.662

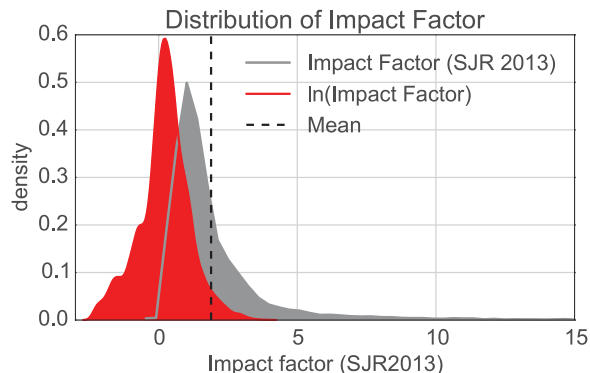FIG. 2. Distribution (kde) of impact factor and ln(impact factor). To aid visualization, the few journals with impact factor > 15 are not displayed.

Social Science, Veterinary Science, Dental, Health Professions. Assignment of journals to these broad subject areas is not exclusive; many journals fall into more than one category. As a result of cross listing, the list of candidate journals was less than 6,500. The final data consisted of 4,721 unique journals, 335 (7.1%) of which are categorized by the Directory of Open Access Journals (DOAJ) as "open access."

Journals were also categorized more narrowly using the more than 300 "All Science Journal Classification" (ASJC) subject codes,[13] for example, animal science and zoology, biophysics, etc. These narrow codes were used to identify journals that address similar topics and thus indicate whether the journal is at risk for reference *vis-à-vis* demonstrated demand. Journals with at least one narrow subject code in common were considered "neighbors" and if at least one of these neighboring journals has been referenced the original "ego" journal was considered to be at risk for reference as well. Journals with no demonstrable demand were excluded from analysis. As an example, consider the journal *Science*. It is listed under (ASJC) subject code 1000—general science. Other journals with this code—the "neighbors" of *Science*—are *Nature, PNAS*, and *Language Awareness*. *Language Awareness* is cross-listed under five other subject codes.

Impact factor was measured by the 2013 SCImago Journal Rank (SJR) impact factor. SJR correlates highly with the more conventional impact factor but takes into account self-citations and the diverse prestige of citing journals (González-Pereira, Guerrero-Bote, & Moya-Anegón, 2010; Leydesdorff, 2009). Table 1 displays the 15 highest SCImago impact journals, calculated with citation data available up to 2013.

English Wikipedia data. References in the English Wikipedia were extracted from the November 15, 2014 database dump of all articles. We parsed every page and following (Nielsen, 2007) extracted all references that use Wikipedia's

cite-journal template. Because it allows editors to easily include inline references that are automatically rendered into an end-of-article bibliography, this template is the recommended way for editors to reference scientific sources in Wikipedia.[14] In all, there were 311,947 "cite-journal" tags in the English Wikipedia. An exploratory analysis of the 49 largest non-English Wikipedias can be found in the Appendix.

Matching Scopus Journals to Wikipedia References

We checked each of the referenced journal names on Wikipedia against a list of Scopus-indexed journal names and common *Institute for Scientific Information (ISI; now Thomson Reuters)* journal name abbreviations. Of the 311,947 *cite-journal* tags, 203,536 could be linked to journals indexed by Scopus. Many of these references were nonunique, whereas our outcome of interest is whether articles from a journal are referenced on Wikipedia at all, not how many times. Therefore, to ensure that the counts for each journal included only unique articles, we distinguished articles by their DOIs and, if an article's DOI was not available, we used the article's title. Scopus's *coverage* of the output of various journals varies widely; our counts included only those articles published within the years of Scopus coverage.

In the end we matched 32,361 unique articles (and 55,262 total references) to our subset of Scopus journals (top 250 in each research field). Of the top Scopus journals, 2,005 are never referenced on the English Wikipedia. In most cases observed "journal names" that did not match to journals in Scopus were not academic journals but popular newspapers and magazines. Table A1 in the Appendix displays the 20 most frequently referenced sources that we were unable to link to Scopus. The top three nonscience

TABLE 2. Descriptive statistics of key variables.

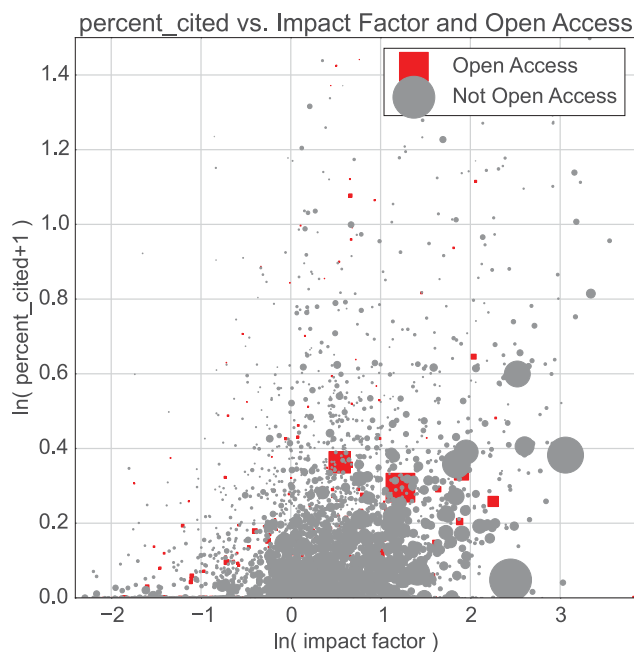| Variable name	Mean	Std.	Min	Max
percent_cited	0.193%	0.545	0%	14.7%
impact factor	1.89	2.47	0.100	45.9
ln(impact factor)	0.212	0.909	−2.30	3.83
open access	7.1% O.A.	—-	0	1

percent_cited vs. Impact Factor and Open Access

FIG. 3. Scatter plot of journals' percent_cited versus impact factor and open access policy. Markers are scaled by the total amount of articles published by each journal (and indexed by Scopus). OA journals are designated with red squares.

sources are *Billboard*, *National Park Service*, and *Royal Gazette*. However, efforts to match Wikipedia references to Scopus were imperfect, and the list also includes a handful of academic journals, including *The Lancet*.

Journal Versus Article-Level Unit of Analysis

We chose to take journals instead of individual articles as our unit of analysis for several reasons. First and most important, accessibility of articles, the focal point of this study, was measured at the journal level by whether the journal is or is not OA. Switching the unit of analysis to individual articles would have simply assigned the same value of accessibility to all articles from a particular journal. Second, although article-level citations are an attractive, finely grained metric, a journal's impact factor is also designed to capture citation impact, albeit more coarsely. The general topic of any given article is also well captured by the host journal's Scopus -assigned topic(s). Lastly, the matching of Wikipedia journal title strings to Scopus required some manual matching and these efforts were more practical at the level of thousands of journals instead of hundreds of thousands of articles.

Percent_Cited and Other Variables

We present some of our results in terms of *percent_ cited*—the *percent of a journal's articles that are referenced*
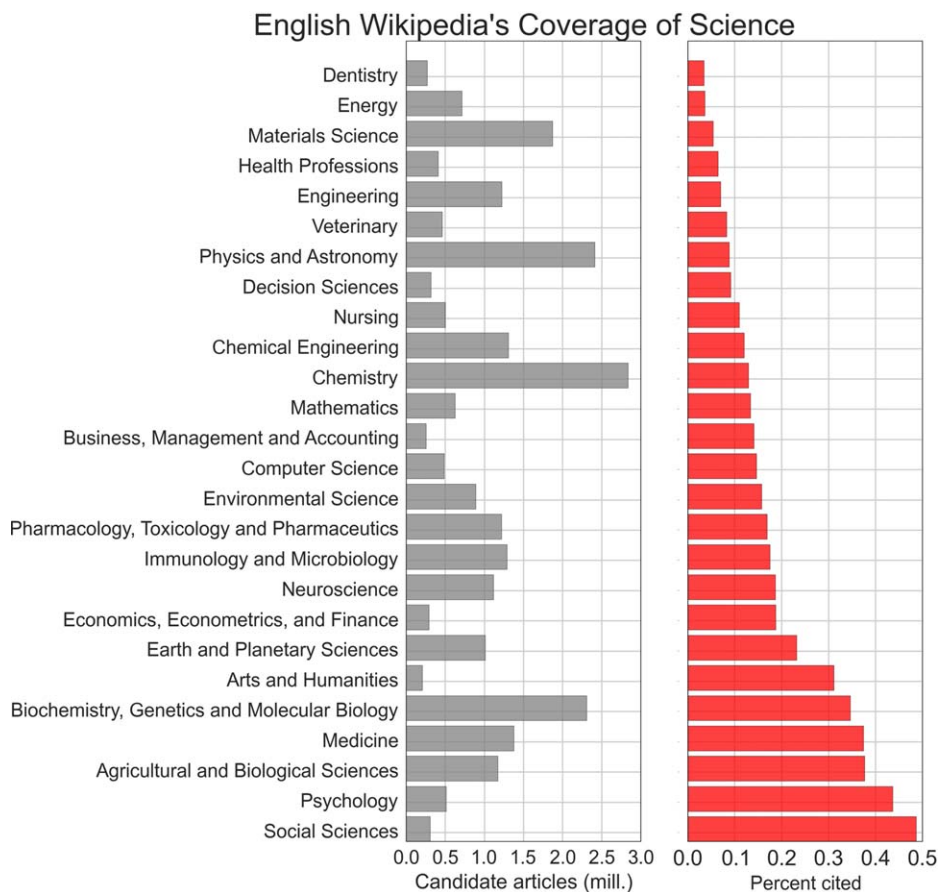

FIG. 4. English-language Wikipedia's coverage of academic research. Each field's candidate literature (left) consists of the articles published by the 250-highest impact journals within the field and indexed by Scopus. The right panel shows the percent of these articles referenced on Wikipedia. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 3. Estimates from the generalized linear model (GLM) estimated on English Wikipedia reference data. Variables with statistically insignificant odds ratios are not bolded.

Variable	Odds ratio	Std err of coeff	*p*-value
open_access	1.471	0.023	**0.000**
log_sjr2013	1.879	0.007	**0.000**
ag_bio_sciences	2.291	0.018	**0.000**
arts_hum	1.836	0.043	**0.000**
biochem_gen_mbio	1.059	0.014	**0.000**
bus_man_acct	0.714	0.057	**0.000**
chem	1.004	0.022	0.863
cheme	0.968	0.030	0.282
cs	0.991	0.041	0.831
decision_sci	0.957	0.064	0.489
dental	0.520	0.107	**0.000**
earth_plan_sci	1.515	0.024	**0.000**
econ_fin	1.106	0.046	**0.030**
energy	0.551	0.064	**0.000**
engineering	0.507	0.037	**0.000**
envi_sci	0.743	0.029	**0.000**
healthpro	0.787	0.063	**0.000**
immu_micro_bio	1.114	0.023	**0.000**
materials	0.640	0.035	**0.000**
math	0.716	0.039	**0.000**
medicine	0.660	0.014	**0.000**
neuro	1.033	0.023	0.168
nursing	1.206	0.045	**0.000**
pharm	1.481	0.025	**0.000**
phys	0.629	0.025	**0.000**
psyc	2.628	0.025	**0.000**
socialsci	1.357	0.029	**0.000**
vet	0.898	0.054	**0.048**

on Wikipedia. An equivalent interpretation of this journal-level metric is the *probability that a given article from a journal is referenced on Wikipedia*. Figure 2 illustrates the distribution (kde) of *percent_cited*.

As Figure 1 demonstrates, the vast majority of journals that scientists use are referenced on the English Wikipedia very little: on average 0.19% of a journal's articles are referenced.[15]

As mentioned, the academic status of journals was measured using (SCImago) impact factors. To limit the influence of the few journals with uncommonly high impact factors the impact factor variable was (natural) log-transformed when used in the models. Figure 2 displays the distribution of impact factor and log-impact factor; to aid visualization only journals with impact factor $<=15$ are shown.

Table 2 presents the summary statistics for key variables: percent_cited, impact factor, ln (impact factor), and open access. Additionally, analyses use dummies for the 26 subject categories, for example, psychology 0 or 1).

Lastly, Figure 3 displays a scatter plot of the key dependent variable, *percent_cited*, versus impact factor and open access.

The scatter plots appear to show a modest relationship between a journal's impact factor and *percent_cited*, the percent of its articles referenced on Wikipedia, especially when considering journal size (dot size). The next section analyzes these relationships statistically.

Results

We first present results of English Wikipedia's coverage. We ask, does Wikipedia draw equally on all branches of science? Next we focus on the role played by a journal's status and accessibility in predicting Wikipedia references. An exploratory analysis of references in the 49 largest *non-English* Wikipedias can be found in the Appendix.

Coverage

Figure 4 summarizes which branches of the scientific literature the English Wikipedia draws upon. The left panel shows the number of articles published by the top 250 journals in each field. The right panel shows the percent of those articles that are referenced at least once in the English Wikipedia.

Figure 4 indicates that the coverage of science, as measured by the use of references, is very uneven and limited across scientific fields (Samoilenko & Yasseri, 2014). The social sciences represent a relatively small candidate literature but a relatively large portion of this literature is referenced on the English Wikipedia (0.4–0.5%). At the other end of the spectrum, dentistry, also a relatively small literature, is rarely referenced ($< 0.05\%$). The ordering of disciplines by *percent_cited* does not engender a simple explanation. For example, such an ordering does not appear correlated with traditional distinctions like hard versus soft science, or basic versus applied. This finding is echoed by Nielsen (2007), who found that "computer and Internet-related journals do not get as many
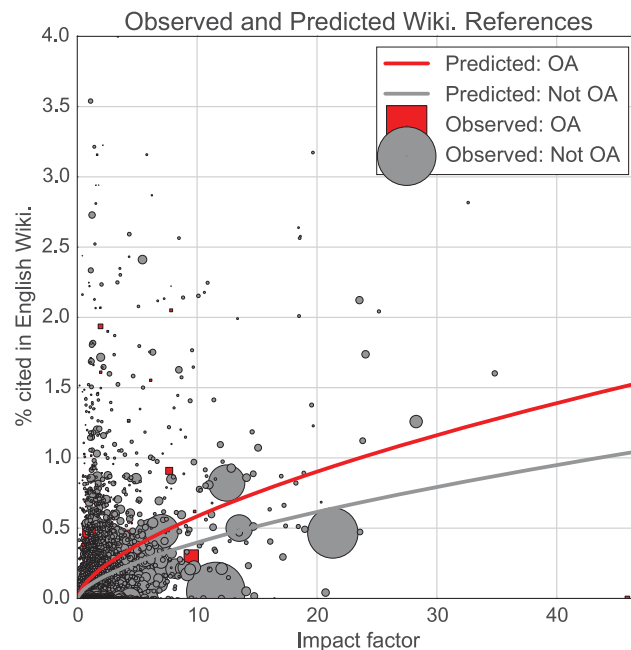
FIG. 5. Observed (dots) and predicted (solid lines) English Wikipedia references. Red squares designate OA journals. The marker size is proportional to the number of articles the journal published.

[references] as one would expect if Wikipedia showed bias towards fields for the 'Internet-savvy'." The highly uneven referencing across disciplines suggests that discipline should be controlled for in any statistical model, as is done below.

Status and Accessibility

We now present results from an intuitive statistical model that predicts the probability p that an article from a given journal will be referenced given that journal's characteristics. The data-generating process is assumed to be binomial each journal i publishes n_i articles and each of these articles is at risk p_i of being referenced in Wikipedia, where p_i depends on the journal. The probability that a journal i has k of its n_i articles referenced in Wikipedia is thus:

$Pr\left(y_i = k | n_i, p_i\right) \sim \binom{n_i}{k} p_{k_i}(1-p_i)^{n_i-k}$. p_i is assumed to be a (logistic) function of the journal characteristics x_i's (e.g., impact factor): $\ln(\frac{p}{1-p}) = x\beta$, where β are the parameters to be estimated. The model just described is commonly used for proportional outcomes; it embeds the familiar logistic regression within a binomial process and is known as a generalized linear model (GLM) of the binomial-logit family (Hardin & Hilbe, 2012: 153–4).

Table 3 displays estimates from this model of how journal characteristics are related to its p, probability of referencing, fitted to the English Wikipedia.

The column of odds ratios indicates how the odds of referencing change with unit changes in the independent variables. For indicator variables, for example, OA, these ratios are interpreted as the increase in odds when the indicator is true. For example, the odds that an article is referenced on Wikipedia increase by 47% if the article is published in an OA journal.

To interpret these results in terms of probabilities rather than odds ratios we must evaluate the model at particular values of the variables. Figure 5 displays the observed and predicted references for a range of values of impact factor and *open_access*. The indicator variables designating particular disciplines are set at their modes (0).

The figure demonstrates that a journal's impact factor has positive and asymptotic effect on the percent of its contents referenced in the English Wikipedia (*percent_cited*). Open access journals (red dots) are relatively uncommon, but these journals are referenced more often than paywall journals of similar impact factor. For example, in our sample of psychology journals, OA journals have an average impact factor of 1.59, whereas closed access journals have an average impact factor of 1.77. Yet in the English Wikipedia, editors reference an average of 0.49% of OA journals' articles but only 0.35% of closed access journals' articles, despite the higher impact factors.

Conclusion

This article examined in unprecedented detail and scale how the English language Wikipedia references the scientific literature. (The appendix contains an exploratory analysis of scientific referencing in Wikipedias of 49 other languages.) Of central interest was the relationship between a journal's academic status and accessibility on its probability of being referenced in Wikipedia. Previous studies have focused only on the role of academic status on referencing in the English Wikipedia and have often ignored unreferenced articles. In contrast, we began with a large (4,721 journals, 19.4M articles) corpus of scientific literature that scientists use, and estimated a statistical model to identify the features of journals that predict how much Wikipedia editors use them as sources.

We found that a journal's academic status (impact factor) routinely predicts its appearance on Wikipedia. We also demonstrated, for the first time, that a journal's accessibility (OA policy) generally increases probability of referencing on Wikipedia as well, albeit less consistently than its impact factor. The odds that an OA journal is referenced on the English Wikipedia are about 47% higher compared to closed access, paywall journals. Moreover, of closed access journals, those with high impact factors are also significantly more likely to appear in the English Wikipedia. Therefore, editors of the English Wikipedia act as "distillers" of high quality science by interpreting and distributing otherwise closed access knowledge to a broad public audience, free of charge. Moreover, the English Wikipedia, as a platform, acts as an "amplifier" for the (already freely available) OA literature by preferentially broadcasting its findings to millions.

Limitations and Directions for Future Research

Our findings are not without limitations. First and foremost, it bears emphasis that this study did not investigate the nature of Wikipedia's sources as a whole (see Ford, Sen, Musicant, & Miller, 2013 for an excellent examination of sources). Only a fraction of Wikipedia's references use the scientific literature, and this is the subset on which we focused. Consequently the present study cannot address the concern expressed by others, for example, (Luyt & Tan, 2010), that sources outside the scientific literature are used too heavily in scientific articles. Second, the study was cross-sectional in nature; it is conceivable that OA articles differ from closed access, paywall articles in their relevance to Wikipedia. Future work can test the potential confounding factor of unmeasured relevance by observing reference rates for articles which have been experimentally assigned to open and closed-access statuses, as has been done by some psychology journals (Davis et al., 2008).

Third, the study measured accessibility of articles by the OA policy of the publishing journals. However, many articles in paywall journals are made freely available by their authors or third parties (Björk et al., 2014; Wren, 2005). The resulting error in the measurement of accessibility may bias the observed advantage of OA in either direction: if OA *articles* from paywall *journals*, erroneously

coded as closed access, are referenced at higher rates than the journals' truly closed access articles (Gargouri et al., 2010; Harnad & Brody, 2004), the *true* advantage of OA will be even higher than we observed. In the (unlikely) case that OA articles in paywall journals are referenced less than closed access articles, the observed open access advantage will be an overestimate. The academic status of articles is also operationalized by a journal characteristic—its impact factor. In fact, many articles out- or under-perform their journal's impact factor. Although this measurement error likely adds noise to the data, it probably does so without biasing the estimated effect of impact factor on referencing in one direction or another.

The Impact of Open Access Science

The chief finding of this study bears emphasis. We believe the existing discussion of open access has focused too narrowly on the academic literature. Early results showing that open access improves scientific outcomes such as citations have been tempered by newer experimental evidence showing small to null causal effects, and a lively debate has ensued. Our research shifts focus to diffusion, showing that open access policies have a tremendous impact on the diffusion of science to the broader general public through an intermediary like Wikipedia. This effect, previously a matter primarily of speculation, has empirical support. As millions of people search for scientific information on Wikipedia, the science they find distilled and referenced in its articles consists of a disproportionate quantity of open access science.

Acknowledgments

Endnotes

1. http://www.alexa.com/siteinfo/wikipedia.org
2. https://en.wikipedia.org/wiki/Wikipedia:Verifiability. Accessed 2015-06-15.
3. http://stats.wikimedia.org/EN/TablesPageViewsMonthly.htm. Accessed 2015-06-16.
4. Wikipedia has recently partnered with major publishers to provide editors access to some paywall literature: http://en.wikipedia.org/wiki/Wikipedia:The_Wikipedia_Library. Accessed 2015-09-02.
5. http://en.wikipedia.org/wiki/Wikipedia:Notability. Accessed 2015-06-11.
6. http://en.wikipedia.org/wiki/Wikipedia:Five_pillars. Accessed 2015-05-29.
7. http://en.wikipedia.org/wiki/Wikipedia:Identifying_reliable_sources#Scholarship. Accessed 2015-05-29.
8. Citation metrics often influence the ranking of academic search results and may thus promote high impact journals without searchers' knowledge.

9. This literature has grown to thousands of items and is impossible to summarize fully. See (Craig, Plume, McVeigh, Pringle, & Amin, 2007; Davis & Walters, 2011) for two reviews of parts of this literature.
10. As corroborating evidence consider the list of Wikipedia editors by [self-reported] degree lists more than 1000 users with PhDs: http://en.wikipedia.org/wiki/Category:Wikipedians_with_PhD_degrees. Accessed 2015-09-02.
11. http://www.elsevier.com/online-tools/scopus/content-overview. Accessed 2015-01-24.
12. The subject area "general" was excluded because it contained only four journals, all of which were cross-listed with other top-level categories.
13. http://info.sciencedirect.com/scopus/scopus-in-detail/content-coverage-guide/journalclassification. Accessed 2015-06-03.
14. Editors may also reference articles in other ways, for example by providing in-line links. We focus on the "cite-journal" template for three reasons. First, it shows clear intent to reference. Second, it has been used in previous research. Lastly, Ford and colleagues (Ford, Sen, Musicant, & Miller, 2013) found that "<ref>"tags were used most often to reference sources, and the "cite-journal" templates on which we focus are nested within such <ref> tags.
15. 2,005 (out of 4,721) journals are never referenced at all (*percent_cited* = 0).

References

Altmetrics: a manifesto - altmetrics.org. Retrieved from http://altmetrics.org/manifesto

Björk, B.-C., Laakso, M., Welling, P., & Paetau, P. (2014). Anatomy of green open access. Journal of the Association for Information Science and Technology, 65(2), 237–250. http://doi.org/10.1002/asi.22963

Björk, B.-C., & Solomon, D. (2012). Open access versus subscription journals: A comparison of scientific impact. BMC Medicine, 10(1), 73. http://doi.org/10.1186/1741-7015-10-73

Broadus, R.N. (1983). An investigation of the validity of bibliographic citations. Journal of the American Society for Information Science, 34(2), 132–135. http://doi.org/10.1002/asi.4630340206

Butler, B., Joyce, E., & Pike, J. (2008). Don't look now, but we've created a bureaucracy: The nature and roles of policies and rules in Wikipedia. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1101–1110). New York, NY: ACM. http://doi.org/10.1145/1357054.1357227

Davis, P.M. (2010). Does open access lead to increased readership and citations? A randomized controlled trial of articles published in APS journals. The Physiologist, 53(6), 197, 200–201.

Davis, P.M. (2011). Open access, readership, citations: A randomized controlled trial of scientific journal publishing. The FASEB Journal, 25(7), 2129–2134. http://doi.org/10.1096/fj.11-183988

Davis, P.M., Lewenstein, B.V., Simon, D.H., Booth, J.G., & Connolly, M.J.L. (2008). Open access publishing, article downloads, and citations: Randomised controlled trial. BMJ, 337, a568. http://doi.org/10.1136/bmj.a568

Davis, P.M., & Walters, W.H. (2011). The impact of free access to the scientific literature: A review of recent research. Journal of the Medical Library Association: JMLA, 99(3), 208–217. http://doi.org/10.3163/1536-5050.99.3.008

Evans, P., & Krauthammer, M. (2011). Exploring the use of social media to measure journal article impact. AMIA Annual Symposium Proceedings, 2011, 374–381.

Evans, J.A., & Reimer, J. (2009). Open access and global participation in science. Science, 323, 1025.

Eysenbach, G. (2006a). Citation advantage of open access articles. PLoS Biol, 4(5), e157. http://doi.org/10.1371/journal.pbio.0040157

Eysenbach, G. (2006b). The open access advantage. Journal of Medical Internet Research, 8(2), e8. http://doi.org/10.2196/jmir.8.2.e8

Ford, H., Sen, S., Musicant, D.R., & Miller, N. (2013). Getting to the source: Where does Wikipedia get its information from? In

Proceedings of the ninth International Symposium on Open Collaboration (pp. 9:1–9:10). New York, NY: ACM. http://doi.org/10.1145/2491055.2491064

Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., & Harnad, S. (2010). Self-selected or mandated, open access increases citation impact for higher quality research. PLoS One, 5(10), e13636. http://doi.org/10.1371/journal.pone.0013636

Gaulé, P., & Maystre, N. (2011). Getting cited: Does open access help? Research Policy, 40(10), 1332–1338. http://doi.org/10.1016/j.respol.2011.05.025

González-Pereira, B., Guerrero-Bote, V.P., & Moya-Anegón, F. (2010). A new approach to the metric of journals' scientific prestige: The SJR indicator. Journal of Informetrics, 4(3), 379–391. http://doi.org/10.1016/j.joi.2010.03.002

Hardin, J.W. & Hilbe, J.M. (2012). Generalized Linear Models and Extensions. 3rd ed. College Station: Stata Press.

Harnad, S., & Brody, T. (2004). Comparing the impact of open access (OA) vs. non-OA articles in the same journals. D-Lib Magazine, 10(6). Retrieved from http://eprints.soton.ac.uk/260207/

Heilman, J.M., & West, A.G. (2015). Wikipedia and medicine: Quantifying readership, editors, and the significance of natural language. Journal of Medical Internet Research, 17(3), e62. http://doi.org/10.2196/jmir.4069

Jamali, H.R., & Nikzad, M. (2011). Article title type and its relation with the number of downloads and citations. Scientometrics, 88(2), 653–661. http://doi.org/10.1007/s11192-011-0412-z

Jervell, E.E. (2014). For this author, 10,000 Wikipedia articles is a good day's work. Wall Street Journal. Retrieved from http://www.wsj.com/articles/for-this-author-10-000-wikipedia-articles-is-a-good-days-work-1405305001. Published 2014-6-13. Last accessed 2016-1-20.

Joe, W. (2015). Real-time stream of DOIs being cited in Wikipedia. Retrieved from http://crosstech.crossref.org/2015/03/real-time-stream-of-dois-being-cited-in-the-wikipedia.html Published 2015-3-3. Last accessed 2016-1-20.

Laurent, M.R., & Vickers, T.J. (2009). Seeking health information online: Does Wikipedia matter? Journal of the American Medical Informatics Association, 16(4), 471–479. http://doi.org/10.1197/jamia.M3059

Leydesdorff, L. (2009). How are new citation-based journal indicators adding to the bibliometric toolbox? Journal of the American Society for Information Science and Technology, 60(7), 1327–1336. http://doi.org/10.1002/asi.21024

Lin, J. & Fenner, M., (2014). An analysis of Wikipedia references across PLOS publications. Retrieved from http://figshare.com/articles/An_analysis_of_Wikipedia_references_across_PLOS_publications/1048991

Luyt, B., & Tan, D. (2010). Improving Wikipedia's credibility: References and citations in a sample of history articles. Journal of the American Society for Information Science and Technology, 61(4), 715–722. http://doi.org/10.1002/asi.21304

Mesgari, M., Okoli, C., Mehdi, M., Nielsen, F.Å., & Lanamäki, A. (2015). "The sum of all human knowledge": A systematic review of scholarly research on the content of Wikipedia. Journal of the Association for Information Science and Technology, 66(2), 219–245.

Miller, J.D. (2001). Who is using the web for science and health information? Science Communication, 22(3), 256–273. http://doi.org/10.1177/1075547001022003003

Moed, H.F. (2007). The effect of "open access" on citation impact: An analysis of ArXiv's condensed matter section. Journal of the American Society for Information Science and Technology, 58(13), 2047–2054. http://doi.org/10.1002/asi.20663

Nielsen, F.A. (2007). Scientific citations in Wikipedia. First Monday, 12(8). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/1997

Okoli, C., Mehdi, M., Mesgari, M., Nielsen, F.Å., & Lanamäki, A. (2014). Wikipedia in the eyes of its beholders: A systematic review of scholarly research on Wikipedia readers and readership. Journal of the Association for Information Science and Technology, 65(12), 2381–2403. http://doi.org/10.1002/asi.23162

Priem, J. (2015). Altmetrics (Chapter from Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact). arXiv: 1507.01328. Retrieved from http://arxiv.org/abs/1507.01328

Rekdal, O.B. (2014). Academic urban legends. Social Studies of Science, 44(4), 638–654. http://doi.org/10.1177/0306312714535679

Samoilenko, A., & Yasseri, T. (2014). The distorted mirror of Wikipedia: A quantitative analysis of Wikipedia coverage of academics. EPJ Data Science, 3(1). http://doi.org/10.1140/epjds20

Sample, I. (2012). Harvard University says it can't afford journal publishers' prices. Retrieved from http://www.theguardian.com/science/2012/apr/24/harvard-university-journal-publishers-prices

Schroeder, R., & Taylor, L. (2015). Big data and Wikipedia research: Social science knowledge across disciplinary divides. Information, Communication & Society, 18(9), 1039–1056. http://doi.org/10.1080/1369118X.2015.1008538

Seglen, P.O. (1997). Why the impact factor of journals should not be used for evaluating research. BMJ, 314(7079), 497. http://doi.org/10.1136/bmj.314.7079.497

Shaw, A., & Hill, B.M. (2014). Laboratories of oligarchy? How the iron law extends to peer production. Journal of Communication, 64(2), 215–238. http://doi.org/10.1111/jcom.12082

Shuai, X., Jiang, Z., Liu, X., & Bollen, J. (2013). A comparative study of academic and Wikipedia ranking. In Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries (pp. 25–28). New York, NY: ACM. http://doi.org/10.1145/2467696.2467746

Spoerri, A. (2007). What is popular on Wikipedia and why? First Monday, 12(4). http://doi.org/10.5210/fm.v12i4.1765

Steiner, T. (2014). Bots vs. Wikipedians, Anons vs. Logged-Ins (Redux): A global study of edit activity on Wikipedia and Wikidata. In Proceedings of the International Symposium on Open Collaboration (pp. 25:1–25:7). New York, NY: ACM. http://doi.org/10.1145/2641580.2641613

The Open Access Citation Advantage Service. (n.d.). Retrieved from http://sparceurope.org/oaca/

Trench, B. (2008). Internet: Turning science communication inside-out? In M. Bucchi & B. Trench (Eds.), Handbook of public communication of science and technology. London and New York: Routledge. Retrieved from http://www.routledgesociology.com/books/Handbook-of-Public-Communication-of-Science-and-Technology-isbn9780415386173

Van Noorden, R. (2013). Open access: The true cost of science publishing. Nature, 495(7442), 426–429. http://doi.org/10.1038/495426a

West, R., Weber, I., & Castillo, C. (2012). Drawing a data-driven portrait of Wikipedia editors. In Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration (pp. 3:1–3:10). New York, NY: ACM. http://doi.org/10.1145/2462932.2462937

Wren, J.D. (2005). Open access and openly accessible: A study of scientific publications shared via the internet. BMJ, 330(7500), 1128. http://doi.org/10.1136/bmj.38422.611736.E0

Appendix: Non-English Wikipedias

Non-English Wikipedias have been noticeably neglected by the research community (Mesgari, Okoli, Mehdi, Nielsen, & Lanamäki, 2015; Schroeder & Taylor, 2015). It is thus important to test whether any of the findings of this article extend to the millions of articles in non-English Wikipedias. Here, we present an exploratory analysis of scientific references in the 49 largest non-English Wikipedias.

Data

Database dumps of the 49 largest non-English Wikipedias were downloaded on October 5, 2015. For each of

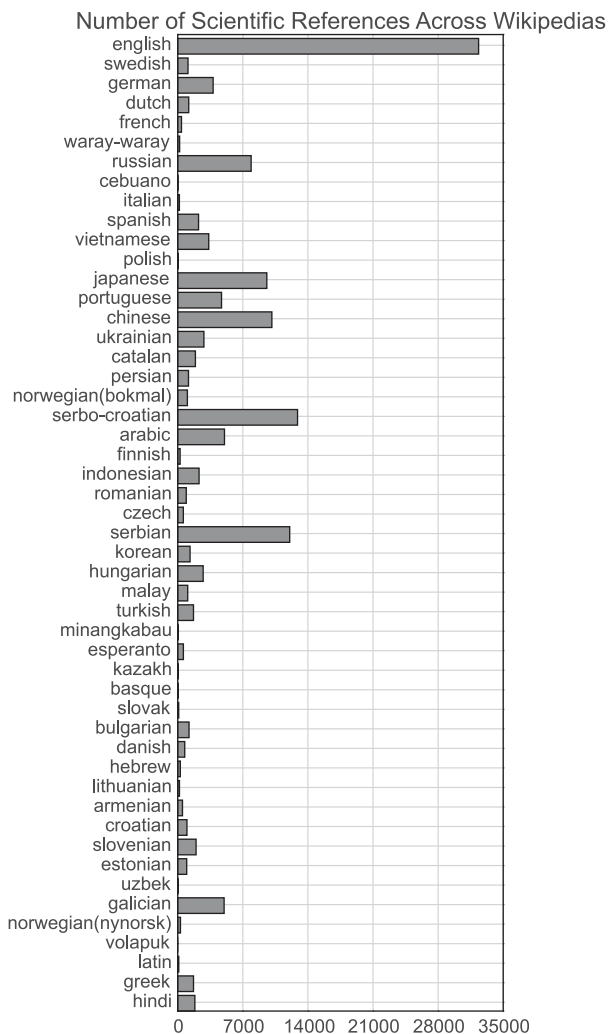## Number of Scientific References Across Wikipedias

FIG. A1. Number of unique scientific articles referenced on the 50 largest Wikipedias.

these, we extracted tags containing "journal" or "doi." Thus the process for obtaining scientific references in non-English Wikipedias did not take into account language-specific tags. Non-English Wikipedias may also reference domestic scientific journals that are not indexed by Scopus. Thus, this exploratory approach surely under-counts scientific references to non-English Wikipedias.

The English Wikipedia referenced by far the greatest number of unique articles. Figure 1 displays the number of unique articles referenced in other Wikipedias, sorted by size (total articles).

Empirical Strategy

Certainly not all findings published in the academic literature belong on Wikipedia. Only small subsets of published findings are important and notable enough to be referenced in Wikipedia. Ideally, studies of how Wikipedia editors reference sources should explain which items in this smaller subset are and are not referenced. Nevertheless, previous studies have struggled to distin-
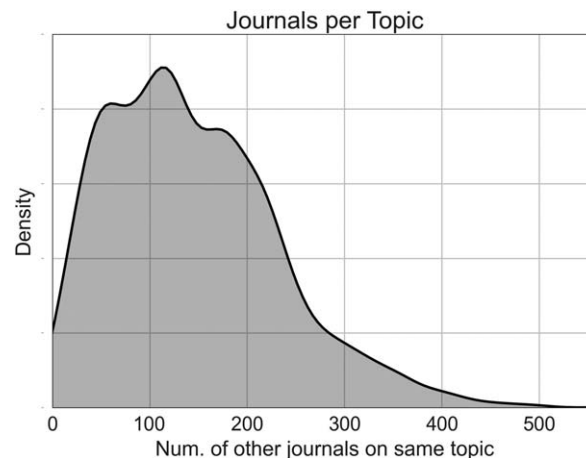

FIG. A2. Distribution of the topical neighborhood sizes of journals. On average journals had 144.8 other journals that addressed the same topic(s).

guish the candidate articles that are at risk for reference from those that do not belong on Wikipedia. Yet, to model referencing decisions with *all* articles—including the dozens of millions of articles never referenced on Wikipedia—is likely to result in a model that predicts that no article will ever be referenced. Consequently most studies have voluntarily hobbled themselves by simply modeling only on the subset of referenced articles.

Here we propose a compromise strategy based on "demonstrated demand." The idea is simple: Articles are at risk for reference if *other* articles on the same topic are referenced. Topical reference indicates that there is demand from Wikipedia editors for literature on the topic

TABLE A1. Most common sources referenced using the *cite-journal* template that are not indexed by Scopus.

Journal name	Times referenced
BILLBOARD	1,630
NATIONAL PARK SERVICE	539
ROYAL GAZETTE	523
BULL AMER MATH SOC	506
FLIGHT INTERNATIONAL	455
BAH NEWS	385
NEW YORK TIMES	369
ROLLING STONE	360
ENTERTAINMENT WEEKLY	342
J BOMBAY NAT HIST SOC	314
WHOS WHO	312
BIZJOURNALSCOM	288
THE GUARDIAN	287
THE LANCET	281
VARIETY MAGAZINE	270
INSIDE SOAP	239
TIME MAGAZINE	238
BIOCHIMICA ET BIOPHYSICA ACTA	201
EDMONTON JOURNAL	201
SPORTS ILLUSTRATED	197

TABLE A2. Percent of journal data that are not used in estimates language-specific models (demonstrated demand = 0). These percentages indicate the portion of research areas for which there is no demonstrable demand from (language-specific) Wikipedia editors.

Wikipedia language	Percent excluded (weight = 0)
Chinese	1.28
Russian	1.28
Japanese	1.36
Arabic	1.47
Vietnamese	2.01
Portuguese	2.16
German	2.30
Spanish	2.62
Indonesian	2.95
Hindi	3.62
Hungarian	3.73
Ukrainian	4.17
Slovenian	4.25
Persian	4.40
Serbian	4.92
Greek	4.99
Turkish	5.63
Serbo-Croatian	6.35
Malay	6.85
Bulgarian	7.12
Dutch	7.31
Catalan	7.88
Danish	8.36
Swedish	9.78
Romanian	10.08
Korean	10.08
Estonian	11.73
Galician	12.13
Norwegian(Bokmal)	12.36
Czech	12.84
French	13.68
Croatian	14.52
Hebrew	20.93
Armenian	22.62
Waray-Waray	25.32
Esperanto	26.37
Finnish	26.81
Italian	29.16
Lithuanian	34.00
Norwegian(Nynorsk)	40.01
Latin	43.67
Slovak	49.87
Polish	62.69
Uzbek	78.91
Minangkabau	80.88
Kazakh	81.44
Basque	86.30
Cebuano	86.95
Volapuk	99.98

and that an article's characteristics (e.g., accessibility, status) may determine which of the candidate articles an editor finds and references. Conversely, if articles on a given topic are never referenced, it is likely that Wikipedia editors do not "demand" literature on this topic, no matter the accessibility or status of the supply. Demonstrated demand exists at the level of topics and, like accessibility and status, we identify an article's topic at the journal level. Demonstrable demand is also a language-dependent metric: Some Wikipedias may lack editors with expertise or interest in, for example, dentistry, thereby consigning all dentistry journals to irrelevance with regards to referencing decisions (but not irrelevant for analysis of coverage, of course). To calculate demonstrated demand we identify for each journal its topical "neighbors" and assign demand of 0 if none of these neighbors are ever referenced in a particular Wikipedia.

We calculate demonstrated demand for a journal through its topical neighbors, which are defined as other journals that share at least one narrow (ASJC) subject code. Only one journal, *Prevenzione & Assistenza Dentale*, had no neighbors while the mean neighborhood size was 144.8. Figure 4 displays the distribution (kde) of neighborhood size.

Table A2 contains the percentages of journals that were excluded from estimating models in each language. This percentage varies widely. For example, only 0.17% of journals were not used for the English Wikipedia model, 49.87% for Slovak, and a 100% for Volapuk. These numbers correspond directly to demonstrable demand for various research literatures by the editors of each Wikipedia. While the English Wikipedia references ~32,000 articles from top journals, the Slovak Wikipedia references only 108 and Volapuk references 0.

Results

From each Wikipedia's model, two parameters are of focal interest: the odds ratio (of probability of referencing) when OA is True, and odds ratio when (log of) impact factor increases by 1 unit. Table A3 shows these odds ratios and associated *p*-values for each Wikipedia. Ratios significant at the 0.05 level are bolded.

While earlier results showed that both accessibility and status increase the odds that a journal will be referenced in the English Wikipedia, the relative strength of these effects varies across languages. Some Wikipedias, like the Turkish, prioritize a journal's academic status over accessibility; the odds of referencing high status journals are nearly 200% higher than lower status journals. Other Wikipedias, like the Serbian, prioritize accessibility over status; the odds of referencing an OA journals are ~275% higher than a paywall journal.

Intuition and previous work suggest that reliance on OA literature is associated negatively with a country's level of economic development (Evans & Reimer, 2009), yet this pattern is not apparent in Table A3. For example, India and Ukraine, relatively poor countries naturally associated with the Hindi and Ukrainian Wikipedias, actually exhibit a small preference against OA literature, while a relatively wealthy country like Sweden has a Wikipedia that exhibits a huge preference for OA literature. The unexpected patterns may be because of the influence of bots (Steiner, 2014). For example, about a third of all articles on the Swedish Wikipedia were

TABLE A3. Odds ratios and associated *p*-values for *open access* and (log) *impact factor* for 50 Wikipedias. Statistically significant odds ratios are bolded.

Wikipedia Language	open access odds ratio	open access *p*-value	Ln (impact factor) odds ratio	Ln (impact factor) *p*-value
Arabic	0.923	0.258	**2.189**	**0.000**
Armenian	1.052	0.802	**2.669**	**0.000**
Basque	0.000	1.000	1.059	0.890
Bulgarian	0.936	0.623	**2.256**	**0.000**
Catalan	**1.452**	**0.000**	**1.963**	**0.000**
Cebuano	0.000	1.000	1.989	0.093
Chinese	**1.337**	**0.000**	**2.257**	**0.000**
Croatian	**0.651**	**0.009**	**2.230**	**0.000**
Czech	1.247	0.230	**2.258**	**0.000**
Danish	0.722	0.120	**2.190**	**0.000**
Dutch	**1.743**	**0.000**	**2.238**	**0.000**
English	**1.471**	**0.000**	**1.878**	**0.000**
Esperanto	1.114	0.591	**1.245**	**0.001**
Estonian	1.221	0.151	**2.705**	**0.000**
Finnish	0.666	0.300	**1.576**	**0.000**
French	0.850	0.550	**2.030**	**0.000**
Galician	**1.464**	**0.000**	**2.176**	**0.000**
German	**1.755**	**0.000**	**2.264**	**0.000**
Greek	0.798	0.078	**2.008**	**0.000**
Hebrew	1.191	0.531	**1.906**	**0.000**
Hindi	**0.757**	**0.029**	**2.113**	**0.000**
Hungarian	**0.749**	**0.005**	**1.804**	**0.000**
Indonesian	0.886	0.242	**2.467**	**0.000**
Italian	0.638	0.299	**2.072**	**0.000**
Japanese	**2.577**	**0.000**	**1.865**	**0.000**
Kazakh	0.000	1.000	1.759	0.114
Korean	1.246	0.055	**1.944**	**0.000**
Latin	0.812	0.637	**1.975**	**0.000**
Lithuanian	1.035	0.923	**2.345**	**0.000**
Malay	1.211	0.157	**2.214**	**0.000**
Minangkabau	2.927	0.314	1.661	0.224
Norwegian(bokmal)	0.866	0.437	**2.328**	**0.000**
Norwegian(nynorsk)	0.588	0.109	**1.510**	**0.000**
Persian	0.941	0.678	**2.210**	**0.000**
Polish	0.588	0.480	**2.330**	**0.000**
Portuguese	**1.527**	**0.000**	**2.076**	**0.000**
Romanian	0.903	0.545	**2.178**	**0.000**
Russian	**1.419**	**0.000**	**2.086**	**0.000**
Serbian	**3.824**	**0.000**	**1.516**	**0.000**
Serbo-Croatian	**3.761**	**0.000**	**1.518**	**0.000**
Slovak	1.943	0.157	**3.249**	**0.000**
Slovenian	0.926	0.487	**2.389**	**0.000**
Spanish	**1.913**	**0.000**	**1.698**	**0.000**
Swedish	**3.745**	**0.000**	**2.094**	**0.000**
Turkish	**1.262**	**0.021**	**2.956**	**0.000**
Ukrainian	**0.818**	**0.030**	**2.566**	**0.000**
Uzbek	0.000	1.000	**2.963**	**0.012**
Vietnamese	0.966	0.682	**2.143**	**0.000**
Volapuk	0.588	0.480	**2.330**	**0.000**
Waray-Waray	**2.104**	**0.017**	**2.172**	**0.000**

created by a bot (Jervell, 2014). Idiosyncrasies of the small number of human and nonhuman entities that edit science in non-English Wikipedias may thus play a larger role than gross cross-national patterns.

It bears emphasis that this analysis of references in non-English Wikipedias is exploratory. Further work should extract references in a way that is sensitive to each Wikipedia's language and conventions. Such analysis may reveal differences in how scientific content found in Wikipedia across languages is differentially embedded in or husbanded by local scientific communities.