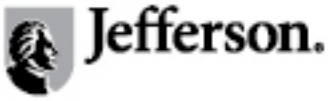

1-1-2019

Who is pirating medical literature? A bibliometric review of 28 million Sci-Hub downloads.

Brian M. Till
Harvard Medical School; Thomas Jefferson University, Brian.Till@jefferson.edu

Niclas Rudolfson
Harvard Medical School; Lund University

Saurabh Saluja
Harvard Medical School; New York Presbyterian Hospital

Jesudian Gnanaraj
SEESHA Hospital

Lubna Samad
Indus Health Research Center

See next page for additional authors

Authors
Brian M. Till, Niclas Rudolfson, Saurabh Saluja, Jesudian Gnanaraj, Lubna Samad, David Ljungman, and Mark Shrime

Who is pirating medical literature? A bibliometric review of 28 million Sci-Hub downloads

Access to the medical literature is essential for both the practice of evidence-based medicine and meaningful contribution to medical sciences. Nonetheless, only 12% of newly published papers are freely accessible online,[1] and, as of 2014, only 3 million of the 26·3 million articles indexed on PubMed were available on the site's repository of free materials, PubMed Central.[2] Access to paywall-protected literature remains primarily through institutional subscriptions. Such subscriptions are costly and many struggle to afford access. The result is a disparity in access to the medical literature, particularly for those in low-income and middle-income countries (LMICs).

In the early 2000s, the rapid expansion of online publication was recognised as an opportunity to iron out these inequities. The WHO-led Health InterNetwork Access to Research Initiative (HINARI) was developed to offer free access to medical journals for not-for-profit medical facilities and research institutions in some LMICs. Yet knowledge of the programme's existence remains poor, the platform is cumbersome, and there is evidence that access to the highest-impact-factor journals has been restricted.[3]

The use of illegal online "shadow libraries" such as Sci-Hub has also emerged as a means of accessing scientific literature.[4] An analysis of requests to the site logged from September 2015 to January 2016 revealed that Sci-Hub had successfully satisfied 99·3% of queries.[5] Due to copyright protections, such sharing remains illegal, and Elsevier, the largest academic publisher, has taken legal action against Sci-Hub, winning a $15 million settlement in US courts.

We aimed to define the proportion of downloads on Sci-Hub that are medical in nature and to consider these data at the national level, evaluating the relation between density of medical literature downloads and scientific output, national income classifications, and indicators of internet penetrance.

We did a bibliometric review of previously compiled Sci-Hub download requests logged between September, 2015, and February, 2016.[6] Data points included date, time, country of request, and digital object identifier (DOI). For each DOI, we obtained associated metadata using the CrossRef application programming interface. We obtained statistics on per-country scientific publications from Scimago, and relied on World Bank Development Indicators for income classifications and internet penetrance.

To determine whether articles were published by medical journals, we relied on Elsevier's Scopus classifications. Scopus uses four major categories (health sciences, life sciences, physical sciences, and social sciences) and 27 major subject headings. We deemed journals with at least 50% of classification terms in health sciences (excluding veterinary medicine) to be medical journals.

We sought to control for differences in country size and level of engagement with medical sciences (number of universities, number of scientists per university) using scientific productivity. We divided the number of Sci-Hub queries from each country by the corresponding volume of publications (in the same Scopus categories) attributable to the country during the period.

To analyse the effect of income group, we did linear regressions of

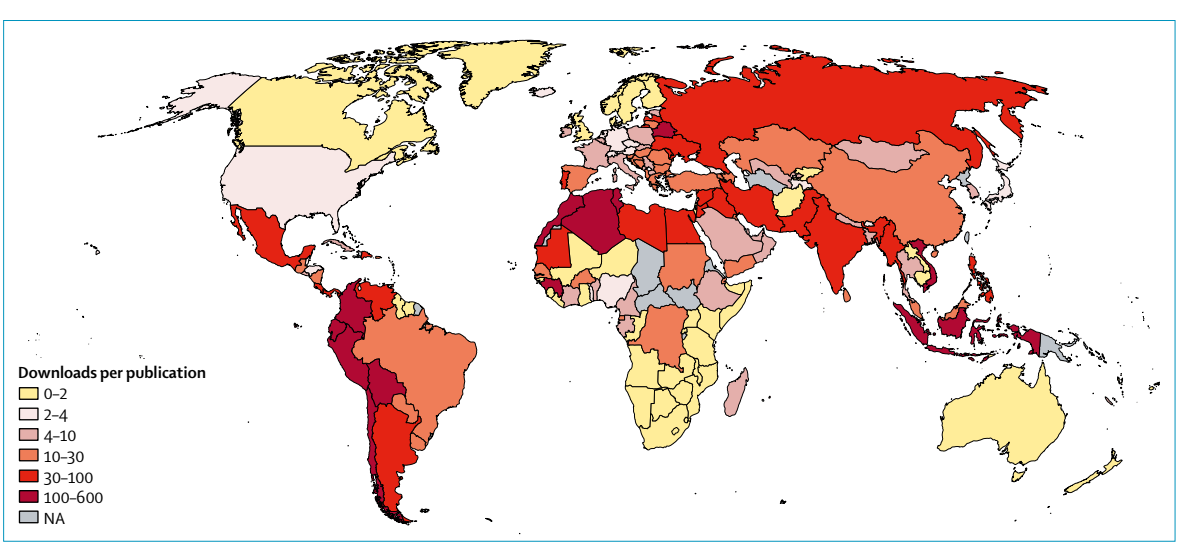

Figure: Sci-Hub downloads in health sciences per scientific publication from the same country

logarithmically transformed data with downloads per country publication as the dependent variable. In a second regression, we controlled for internet penetrance. Data and code are available upon request.

Of 27·8 million download queries, 23·2 million were requests for journal articles. We categorised 94% of requests using Scopus terms. 4·7 million requests (22%) were for medical journal articles.

Most queries for medical literature originated in LMICs (3·3 million, 69%). Almost half (2·2 million, 47%) originated in upper-middle income countries (appendix). In absolute numbers, the most frequent countries of origin were India, China, the USA, Brazil, and Iran. When controlling for scientific output, the highest densities came from Peru, Algeria, Ecuador, Morocco and Indonesia (figure).

See Online for appendix

Compared with those from high-income countries (HICs), downloads per country publication were significantly lower in low-income countries (LICs) (β=–1·20, 95% CI –2·08 to –0·32; p=0·008) and significantly higher in upper-middle-income countries (β=1·06, 95% CI 0·38 to 1·73; p=0·002). After adjusting for internet penetrance, there was no longer a significant difference between HICs and LICs (β=–0·20, 95% CI –1·93 to 1·52; p=0·819), while both lower-middle-income and upper-middle-income countries had significantly more downloads per publication than HICs (β=1·33, 95% CI 0·06 to 2·60; p=0·041 and β=1·48, 95% CI 0·56 to 2·40; p=0·002, respectively; appendix).

Nearly 1 million articles published by medical journals are downloaded on Sci-Hub each month. Although demand for pirated medical literature bridges national income classifications, the density of these downloads differs significantly between HICs and LMICs, with higher rates observed in LMICs.

The highest download densities in this sample are from middle-income countries. There are several probable factors underlying this trend. First, institutions in most upper-middle-income countries do not qualify for HINARI.[7] Those in lower-middle-income countries may qualify for a fee-based version of the system depending on a range of factors. This exclusion, when coupled with increasing levels of educational attainment[8] and rapidly growing medical industries,[9] has probably led to increased demand for medical literature while legal avenues for access remain limited. With efforts undertaken by WHO focused on expanding access in the poorest countries, and with academic publishers focused on the sale of bundled journal packages to library consortia concentrated in high-income states,[10] there appears to be a trough in access for middle-income countries.

There are two notable limitations to this research. First, the available metadata limits our classifications to journal of publication rather than the specific content of a downloaded paper. Given the size of our sample, analysis at the level of the individual article is impractical. Second, we have only analysed publications with Scopus classifications. It is possible that this has led to the exclusion of journal articles in lower-profile publications.

Our analysis illuminates the large volume of medical literature being downloaded illegally in nearly all countries of the world. A significant relation exists between the nation of origin and the density of these requests. This continued inequity in legal access to the medical literature demands the attention of both the publishing industry and policymakers.

We declare no competing interests.

Brian M Till†, Niclas Rudolfson†, Saurabh Saluja, Jesudian Gnanaraj, Lubna Samad, David Ljungman, Mark Shrime
brian.till@jefferson.edu

Harvard Medical School Program in Global Surgery and Social Change, Boston, MA 02115, USA (BMT, NR, SS, DL, MS); Thomas Jefferson University, Department of Surgery Philadelphia, PA, USA (BMT); WHO Collaborating Center for Surgery and Public Health, Department of Clinical Sciences, Lund University, Lund University Hospital, Lund, Sweden (NR); Weill Cornell Medicine, Department of Surgery, New York Presbyterian Hospital, New York, NY, USA (SS); SEESHA Hospital, Siruvani Main Road, Karunya Nagar, Coimbatore, Tamil Nadu, India (JG); Indus Health Research Center, Korangi Crossing, Karachi, Pakistan (LS); Department of Surgery, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Göteborg, Sweden (DL); Massachusetts Eye and Ear Infirmary, Boston, MA, USA (MS)

†Equal contributors.

1 Laakso M, Björk B-C. Anatomy of open access publishing: a study of longitudinal development and internal structure. *BMC Med* 2012; **10:** 1–9.
2 Dunn AG, Coiera E, Mandl KD. Is Biblioleaks inevitable? *J Med Internet Res* 2014; **16:** e112.
3 Villafuerte-Gálvez J, Curioso WH, Gayoso O. Biomedical journals and global poverty: is HINARI a step backwards? *PLoS Med* 2007; **4:** 220.
4 Mejia CR, Valladares-Garrido MJ, Miñan-Tapia A, et al. Use, knowledge, and perception of the scientific contribution of Sci-Hub in medical students: study in six countries in Latin America. *PLoS One* 2017; **12:** e0185673.
5 Himmelstein DS, Romero AR, McLaughlin SR. Sci-Hub provides access to nearly all scholarly literature. *eLife* 2018; **7:** e32822.
6 Bohannon J. Who's downloading pirated papers? Everyone. *Science* 2016; **352:** 508–12.
7 WHO. Hinari eligibility 2017. http://www.who.int/hinari/eligibility/en/ (accessed Aug 1, 2017).
8 Barro RJ, Lee J. A new data set of educational attainment in the world, 1950–2010. *J Dev Econ* 2013; **104:** 184–98.
9 Dieleman J, Murray, CJ, Kahn Case M, et al. Financing global health 2016: development assistance, public and private health spending for the pursuit of universal health coverage. Seattle: Institute for Health Metrics and Evaluation, 2017: 70–71.
10 Ware M, Mabe M. The STM report: an overview of scientific and journal publishing. March, 2015. http://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf (accessed Aug 3, 2018).