1-1-2019


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*

Let us know how access to this document benefits you

Follow this and additional works at: [https://jdc.jefferson.edu/surgeryfp](https://jdc.jefferson.edu/surgeryfp)

Part of the [Surgery Commons](https://jdc.jefferson.edu/surgeryfp)

**Recommended Citation**

Till, Brian M.; Rudolfson, Niclas; Saluja, Saurabh; Gnanaraj, Jesudian; Samad, Lubna; Ljungman, David; and Shrime, Mark, "Who is pirating medical literature? A bibliometric review of 28 million Sci-Hub downloads." (2019). *Department of Surgery Faculty Papers*. Paper 165.  
[https://jdc.jefferson.edu/surgeryfp/165](https://jdc.jefferson.edu/surgeryfp/165)

---

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University’s [Center for Teaching and Learning (CTL)](https://ctl.jefferson.edu). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in Department of Surgery Faculty Papers by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: [JeffersonDigitalCommons@jefferson.edu](mailto:JeffersonDigitalCommons@jefferson.edu).
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, 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.

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.

The use of illegal online “shadow libraries” such as Sci-Hub has also emerged as a means of accessing scientific literature. 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. 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. 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
Correspondence

The highest download densities in countries. There are several probable factors underlying this trend. First, institutions in most upper-middle-income countries do not qualify for HINARI.\textsuperscript{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\textsuperscript{8} and rapidly growing medical industries,\textsuperscript{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,\textsuperscript{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.

Copyright © 2018 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY-NC-ND 4.0 license.

*Brian M Till†, Niclas Rudolfson†, Saurabh Saluja, Jesudian Gnanaraj, Lubna Samad, David Ljungman, Mark Shrimne

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, Sinuvari Main Road, Karunya Nagar, Coimbatore, Tamil Nadu, India (SG); Indus Health Research Center, Korangi Crossing, Karachi, Pakistan (LS); Department of Surgery, Institute of Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Goteborg, Sweden (DL); Massachusetts Eye and Ear Infirmary, Boston, MA, USA (MS)

†Equal contributors.


See Online for appendix