# Four decades of health economics through a bibliometric lens

Adam Wagstaff [a,*], Anthony J. Culyer [b,c]

[a] Development Research Group, The World Bank, Washington, DC, USA
[b] Department of Health Policy, Management and Evaluation, University of Toronto, 155 College Street, Toronto, Ontario M5T 3M6, Canada
[c] Department of Economics and Related Studies, University of York, Heslington, York YO1 5DD, UK

## ABSTRACT

In this paper, we take a bibliometric tour of the last forty years of health economics using bibliographic "metadata" from EconLit supplemented by citation data from Google Scholar and our own topical classifications. We report the growth of health economics (we find 33,000 publications since 1969—12,000 more than in the economics of education) and list the 300 most-cited publications broken down by topic. We report the changing topical and geographic focus of health economics (the topics 'Determinants of health and ill-health' and 'Health statistics and econometrics' both show an upward trend, and the field has expanded appreciably into the developing world). We also compare authors, countries, institutions and journals in terms of the volume of publications and their influence as measured through various citation-based indices (Grossman, the US, Harvard and the JHE emerge close to or at the top on a variety of measures).

© 2012 World Bank. Published by Elsevier B.V. All rights reserved.

## 1. Introduction

According to Google's Books Ngram Viewer, the terms "health economics" and "Health Economics" started appearing in books only in the 1940s, four decades after the terms "agricultural economics" and "international economics" surfaced. The first two decades of health economics were slow, but in the mid-1960s use of the terms "health economics" and "Health Economics" increased sharply, and the growth has mostly continued since then.

In this paper, we take a bibliometric tour of the last forty years of health economics. Our bibliographic "metadata" are drawn from EconLit which dates from 1969 although it includes many earlier classic papers that were reprinted as chapters in collections published after 1969, including for example Arrow (1963). Unlike other databases such as the Social Science Citation Index, SCOPUS, and Medline, EconLit allows for a relatively clean definition of health economics based on EconLit's inclusion criterion ("a substantial economics content") and a publication's JEL codes. We run risks of omission and commission:there are many non-economics and multi-disciplinary journals that have published many papers by health economists, and some authors may not have chosen a health

JEL code despite their paper having a substantial health content; conversely, our net will catch some items that are devoid of economics (and not written by economists), as well as items where health is not the main focus of the publication. In any event, for reasons to be explained below, we had little choice. And our pragmatic definition of the sub-discipline enables us to avoid many esoteric considerations regarding, for example, the emphasis to be placed in characterizing health economics as the topics its practitioners study (e.g. financial aspects of health services) rather than the discipline (viz. economics) characteristically applied to understanding and explaining phenomena in health and health care (Culyer, 1981; Williams, 1979).

Our EconLit metadata include author, title, journal, the year of publication, author's institution, country of focus, keyword, and three-digit JEL code. EconLit does not include any citation data; for 78% of records in our dataset we were able to obtain citation data from Google Scholar. In addition to getting at influence through citations, we also want to say something about the topics that health economists work on. The JEL sub-categories are not especially illuminating[1] and the keywords in the EconLit metadata are freely chosen by authors and unsurprisingly are highly

---

* Corresponding author at: The World Bank, 1818 H Street NW, Washington, DC 20433, USA. Tel.: +1 202 473 0566; fax: +1 202 522 1153.
E-mail address: awagstaff@worldbank.org (A. Wagstaff).

[1] They comprise: (a) Health: General, (b) Analysis of Health Care Markets, (c) Health Production, (d) Health: Government Policy; Regulation; Public Health, and (e) Health: Other. Recently two categories have been added: (f) Health and Inequality, and (g) Health and Economic Development.

heterogeneous. We therefore used our own topical classification system, and assigned a topic to a subset of our records manually. We aggregate our metadata to the author, institution and journal levels, showing which of each have published the most, and which of each have had the greatest impact in terms of citations. We show publication counts for both "country of origin" and "country studied". We present frequency distributions of the words used in titles and abstracts and keywords. We also show how the subtopic focus has changed over time and varies across the top institutions.

Two previous exercises are worth mentioning, although neither comes close in terms of comprehensiveness and ambition. Rubin and Chang (2003) undertook a bibliometric analysis of health economics articles over the period 1991–2000 using EconLit metadata. They also report author rankings (using number of articles and number of pages) but in most other respects their analysis is much more limited, covering a shorter time period, excluding citations, using the limited JEL topic classification system, and omitting many of the cuts of the data that we present. Also of note is the exercise Culyer undertook as background for his four-volume collection of 78 articles on health economics (Culyer, 2006). This was based on a survey of teaching members of the International Health Economics Association (iHEA) in 2004 that asked for their student reading lists, the objective being to ensure representation in the 2006 collection of those articles most frequently cited for student reference by people teaching the subject. His poll of iHEA yielded usable responses from 112 members and 164 articles that were included on at least two reading lists. We use this source as basis for making some intertemporal comparisons between health economics then (2004) and now (2011), and also for comparing the frequency of citations in general (in 2011) with literature recommended by specialist health economics teachers (in 2004).

The rest of the paper is organized as follows. Section 2 presents our methods, including the bibliometric methods used to analyze citations and the classification of health economics subtopics. Section 3 introduces our data. Section 4 presents our findings, and Section 5 our conclusions.

## 2. Counting influence through bibliometrics

Publication counts capture the volume of output of an author, institution, journal or country. They do not necessarily capture the *influence* of that author, institution or journal on the research community. Citations are a commonly used measure of scholarly influence, the assumption being that a heavily cited paper is more influential than a less heavily cited paper, and that an individual who has authored several highly cited papers has been more influential among his or her peers than one who has authored a similar number of less cited papers.

The *h*-index (Hirsch, 2005), which we refer to below as $I^1$, is the most widely used citation-based summary measure of scholarly influence, reflecting both the number of publications and the number of times they have been cited. An *h*-index of 20 means that an author (or institution or journal) has 20 publications to its name each of which has been cited at least 20 times. Fig. 1 illustrates. It shows the citation curves for two publication portfolios (e.g. for two individuals). The horizontal axis plots the cumulative number of articles up to the total *N*, ranked in descending order of citations. The *y*-axis plots the number of times the *n*th article has been cited. The *h*-index is found by drawing a 45° line on the chart, and reading off the x-axis or the *y*-axis. Hirsch argues that this index is a robust and relevant measure of ". . .the importance, significance and broad impact of a scientist's cumulative research contributions" and the index is widely used. It is reported, for example, in citation software such as Publish or Perish (Harzing, 2010), SCOPUS and the Web of
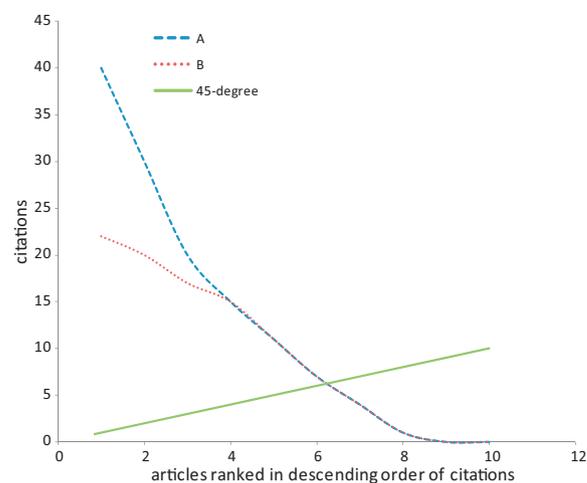


Note: The chart show the hypothetical citation curves of two individuals A and B. The portfolios of each are ranked in descending order of citations. Both have an *h*-index of 6 meaning that A and B have 6 publications to their name each of which has been cited at least 6 times. The *h*-index is found by reading off either axis at the point where the citation curve intersects the 45-degree line.

**Fig. 1.** Measuring influence through citations (i).
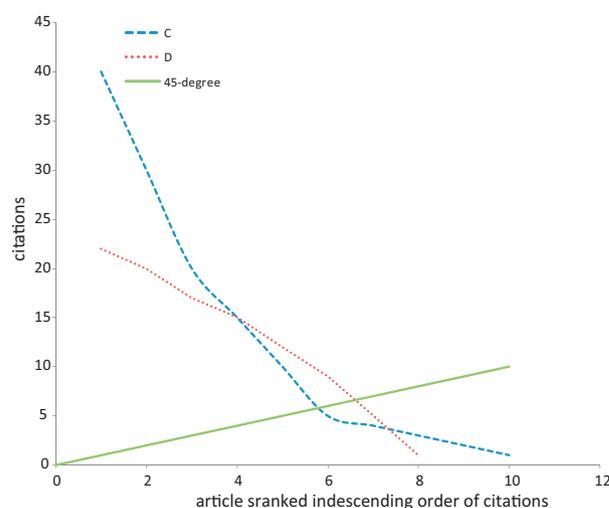


Note: The chart show the hypothetical citation curves of two individuals C and D. The portfolios of each are ranked in descending order of citations. D has an *h*-index of 7 meaning that he or she has 7 publications to his or her name each of which has been cited at least 7 times, while C has an *h*-index of 6. The *h*-index is found by reading off either axis at the point where the citation curve intersects the 45-degree line.

**Fig. 2.** Measuring influence through citations (ii).

Science (which implemented the *h*-index within two years of the publication of Hirsch's paper).[2]

The *h*-index has shortcomings. Both A and B in Fig. 1 have the same *h*-index even though A's first few articles are cited far more often. In Fig. 2, D has a higher *h*-index than C though C's citation curve might reasonably be argued to be the more impressive one on balance. Alternatives to the *h*-index have been proposed, but Ravallion and Wagstaff (2011) have argued they have unattractive properties. They proposed instead an axiomatic approach to measuring scholarly influence based on an influence function where influence is strictly increasing in the number of citations. If a publication portfolio has a citation curve that lies everywhere above that of another publication portfolio, the influence function approach will say the former has greater influence than the latter. Where citation curves intersect, Ravallion and Wagstaff impose the assumption of concavity—influence increases in citations but at

---