

# Open access to the medical literature: How much content is available in published journals?

To understand how much of the published medical journal literature is currently openly available from publishers, we studied 174 journal titles in research and clinical medicine from Thomson Scientific's ISI Citation Databases. Nearly one third of the journals studied have some or all of their recent content freely available through the journal's web site. Forty-six journals (26%) offered their most recent content immediately available without charge; an additional 12 journals (7%) make all content available within a specified period after initial publication. We also considered content back to 1992, a file depth approximately twice the cited half-life of the journals in aggregate. Although 93% of the journals have their current issue and one or more back years available electronically, we estimated that less than 60% of article content since 1992 is electronically available and 21% is open access (OA). Considering other paths to OA (partial open access or e-print archives) would increase this percentage.



**MARIE E McVEIGH**

Product Development Manager



**JAMES K PRINGLE**

Vice President, Development

Thomson Scientific, Philadelphia, USA

Access to medical research has provoked vigorous debate in recent months, especially because of developments at the United States National Institutes of Health (NIH) regarding federally supported research. An NIH proposal to encourage or even to require NIH grant recipients to deposit articles in the National Library of Medicine's PubMed Central repository shortly after publication in a scholarly journal remains under review at the time of this writing.<sup>1</sup>

Implicit in this proposal is the assumption that access to medical literature is not 'open'. This article seeks to answer the question of how much of the medical journal literature, as provided by journal publishers, is openly available today.

Open access (OA) is generally defined as the availability of a published item without charge to the user for the right to read, download, copy, print, or distribute the item. While the definition is simple to state, open access in practice exists as a complex mixture of distribution, time, and economic criteria that determine whether and when a given article can be accessed by a potential reader.

In 2003, Willinsky described nine 'flavors' of open access publishing,<sup>2</sup> offering a useful model by which we can consider how much of the relevant literature is, in fact, open.

In this study, we consider three types of journal-level open access:

- *Unqualified open access*: unrestricted, immediate and complete access to a journal's contents upon publication.
- *Delayed open access*: complete issue contents made freely available after a specified interval of subscription-only access.
- *Partial open access*: only selected content in the journal is freely available, either upon publication or after a specified interval of subscription-only access.

While the number of journals offering unqualified open access has increased in the past year, open access journals still make up a relatively small component of the published literature overall. Across Thomson Scientific's multidisciplinary ISI Citation Databases, OA journals account for less

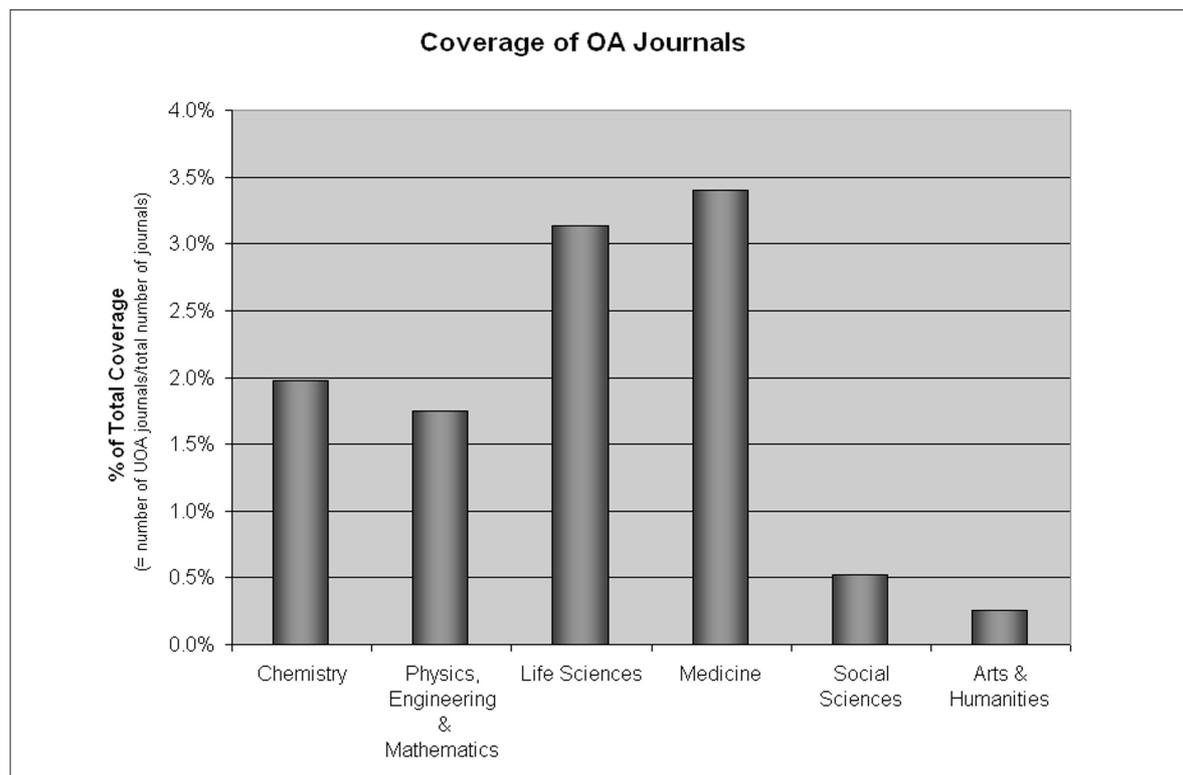


Figure 1. Percentage of journal coverage that offers unqualified open access.

than 3% of the covered titles, although this percentage differs among broadly defined subject areas (see Figure 1).<sup>3</sup> Focusing only on unqualified open access content masks an important dimension of open access to the literature – the role of archival content. In the current study, we also examine the contribution of delayed open access to the volume of content available to users.

In order to perform a journal-by-journal analysis with a significant level of detail, we restricted our study to a subset of journals that focus on a wide range of subject areas in research and clinical medicine. These journals represent the type of content that would most likely be subject to the NIH public access policies. They exert significant influence in the medical literature and also carry prestige among non-expert audiences. They thus help us understand how policies such as that proposed by the NIH might alter the current environment.

## Methods

The journals we selected for study represent those currently covered in the *Web of Science* and indexed in the 'Medicine', General & Internal' category, or

the 'Medicine, Research & Experimental' category. Each of the titles was researched individually on the Internet to find the first year of full text electronic content (if any); the first year of OA content (if any); the most recent year of OA content (if any); and whether all or only selected content is available through OA. The start year of each journal was derived from *Ulrich's Periodicals Directory* at [ulrichsweb.com](http://ulrichsweb.com).<sup>4</sup> The data given reflect the access model and content of the journals to the best of our ability, and at the time of our analysis (ending 31 December 2004). Journals were considered to be 'electronically available' only if the full text of all content was available via the Internet. We were careful to distinguish content that was available electronically to subscribers from content that was available to all users.

To estimate the total article output of each title over the past 13 years, we used *Web of Science*<sup>5</sup> coverage for the years 1999 through the final update of the product in 2004, and limited the file to two document types: articles and reviews. Searching for each source title, and using the 'Analyse Results' feature, we isolated the number of items in the database for each publication year. The average number of articles per year was determined based on these data. Multiplying the

average articles per year by the number of volume-years the journal published since 1992, we obtained an estimate of the total number of articles and reviews published in the most recent 13 years. Similar methods were used to estimate the number of electronically available articles, and the number of open access articles produced in each journal since 1992. These calculations considered the data we had acquired on the year of the first available electronic issues and the oldest and most recently available OA issues.

Although journals that were partial open access were noted, we did not attempt to estimate the number of articles that were made available under this model as there was no way to determine an annual average of such content. We did not include in the count of partial open access journals those titles that offer a free sample issue, only those publications with a formal policy for providing free access to selected content from each issue.

### Defining open access to medical journals

Among broad subject areas covered in the ISI Citation Databases, medical journals generally have the highest percentages of unqualified open access titles. To understand the availability of this literature in more detail, we focused on journals indexed in the categories 'Medicine, General & Internal', and 'Medicine, Research & Experimental.' These two categories include journals with coverage of subjects in novel medical research and general clinical practice, from pre-clinical trials to case reports. Among the prominent titles in these categories, and which are representative of the subject areas covered, are the *New England Journal of Medicine*, *JAMA – Journal of the American Medical Association*, *The Lancet*, *Nature Medicine*, the *Journal of Experimental Medicine*, and the *Journal of Clinical Investigation*. These journals present critical new developments across a broad range of medical fields and are of interest not only to highly specialized researchers, but also to researchers in related fields and to a knowledgeable but non-expert population. The list of current coverage comprises 174 titles from 110 unique publishers, exclusive of nine book or monograph series with irregular publication schedules.

Electronic access is a prerequisite for all current models of OA, and the degree to which electronic access has shaped the environment of scholarly

publication over the past several years is evident in the fact that 93% (161) of the journals had their most recent issues available in electronic form. Surprisingly, 40% of these journals offer some or all of their most recent issues as open access under unqualified, delayed and partial open access models. Table 1 shows the distribution of the population of 174 medical journals according to the type of access.

Type of access	Number of journals	Percentage of total
Unqualified open access	46	27%
Delayed open access	12	7%
Partial open access	13	7%
Subscription access	90	52%
<b>No electronic content</b>	<b>13</b>	<b>7%</b>

Table 1. Distribution of medical journals by type of access

### How deep is open access?

Although these journals show a surprisingly high percentage of open access in their most recent issues, this figure has different meanings for different types of reader. There are two primary types of interaction with the published literature: current awareness (browsing) and retrospective, in-depth review (searching), and these are affected differently by a time-lag between publication and access. For current awareness, to stay abreast of the emerging literature in a particular subject, a delay in availability can materially affect the course of ongoing research. The 26% of journals in this study with unqualified open access define the minimum amount of current awareness available openly from publishers. However, at least equal in its importance is retrospective searching, mining a larger collection of literature to collect a body of work on a given subject. In this case, the depth of file to which one has access is of great importance. To understand the effectiveness of the open access literature for searching, we need to examine the availability of the prior years' content of the medical journals in the sample.

We found that, while many of the journals have many years of continuous publication, the depth of content that is available electronically is quite limited (see Figure 2). Over 80% of the 174 titles studied were launched prior to 1970, with 21 titles having published continuously for over 100 years. However, the vast majority of the journals have an electronic file containing no more than five years of content.

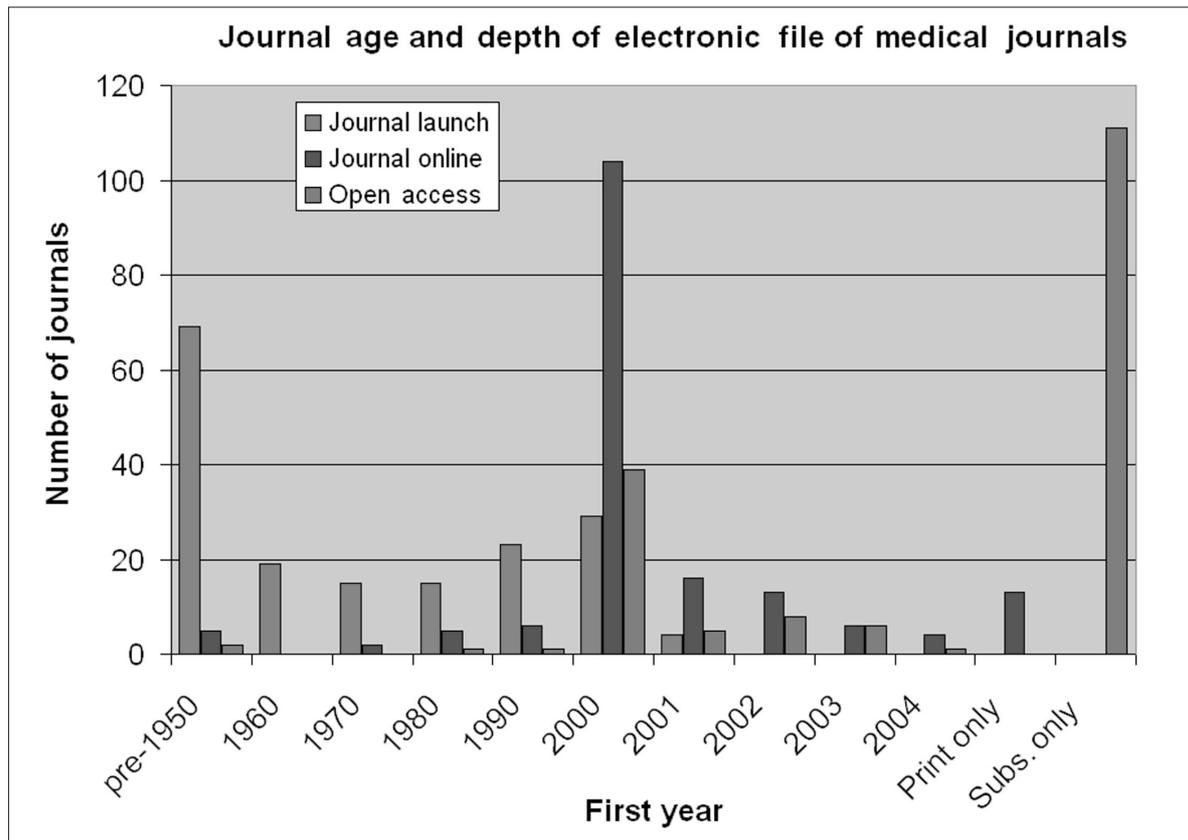


Figure 2. Age of journal contents compared to the depth of electronic file.

Journal age was determined by the launch year as recorded in *ulrichsweb.com*; depth of online file was established from surveying each journal's web site.

### Is depth Important? Citation patterns

Is this older literature important? Citation patterns show that it does play a role in continuing medical research, although its relevance to the general public is less clear. We collected all references to the journals in the 'Medicine, General & Internal', and 'Medicine, Research & Experimental' categories from the 2003 *Journal Citation Reports* to define the age of the literature in these subjects that is being actively cited in recent research articles. The cited half-life of this body of literature was 6.6 years; thus half of all citations in 2003 were to articles published prior to 1999. Putting this fact together with the analysis of archival accessibility, we can say that fewer than half of the 940,000 citations to the journals in this study were referencing content that was accessible electronically. This suggests that, despite the ready availability of current materials online, and the subset of OA material, the majority of references are derived from literature accessed through traditional pathways.

### Print, electronic and open access to active content at the article level

With a cited half-life of 6.6 years, we undertook to analyze the distribution of articles among the various types of access across a period twice the cited half-life, or 13 years. Because this file depth accounts for a large majority of the citations to these journals in the most recently published work, we considered this to represent the active life-span of most articles.

Since any given journal will have articles available through two or three access models across this year-span, we attempted to estimate the number of research articles and reviews published per journal in each of these models. Although the news, summaries, correspondence and editorial content of journals are an important part of scientific communication, we considered the research-oriented content the critical element for this analysis, as it represents the materials that will fall under the NIH mandate for public access. We estimated that, of approximately 157,000 articles

published since 1992, 60% are available electronically, but only 21% of the total are open access. Our estimates did not include, however, the contribution of partial open access journals. Thirteen titles in our study offer selected content as open access, either immediately or after a specified delay. Often, the content that is made freely available through partial open access includes articles that are judged to be of the greatest importance either scientifically or in consideration of public health, effectively increasing the amount of OA content available.

## Conclusion

Considering a subset of the medical and biomedical research journals covered in the Thomson ISI Citation Databases, we examined how three models of open access are applied at the journal level. We found that the journals on the subject of general and research medicine were marked by a particularly high proportion of OA content, primarily through immediate and unrestricted access to content as soon as it is published. Of the 174 journals studied, 93% have their most recent issues available electronically, but very few have more than five years of full articles available in electronic form, and even fewer have more than five years of OA content. We found that, for the subject studied, 26% of the journals made their most recent issues open access, and 21% of articles since 1992 were available as open access. If articles available under partial open access model are included, the total percentage of content openly accessible is likely to remain in the 20–30% range. Citation data from the year 2003 *Journal Citation Reports*<sup>6</sup> demonstrate that more than half the articles in these journals are still being cited more than six years after their initial publication. This suggests that a significant portion of the use of journal articles is still dependent on traditional forms of access, rather than electronic access.

Other ways of increasing access are also being offered by publishers, including country-wide or site-wide subscriptions and special initiatives like the Health InterNetwork Access to Research Initiative (HINARI)<sup>7</sup> that make journals available at low or no cost in developing countries. In addition, many publishers now permit author or institutional archiving of articles, further increasing the percentage of published literature that is available free of charge to the user.

Once a matter of concern largely to librarians trying to stretch a shrinking budget around an expanding corpus of scholarly journals, 'access' has become an issue for the research community, for funding bodies and government agencies, and for the general public whose tax monies contribute to the support of the scientific enterprise. The question of who has and who needs access to the literature describing scientific research and results has been a subject of passionate discussion as the number of stakeholders increased. If more readers have access to the literature, will more of them read it? Will it benefit a doctor's care of his/her patient? Will it improve a patient's understanding and management of their healthcare needs? Will it expand the possibility of advanced medical research and patient care in developing nations? While we can measure with citation the effect older literature has on continuing research, and so measure the benefit to the scholarly literature of an openly accessible or a subscription-accessible article, we cannot measure these other benefits.

In much of the discussion of open access, public good is assumed to arise directly, and solely, from increased access, and to require availability within a short period after initial publication. While this study cannot address these broader questions, we have demonstrated that a significant amount – between 20% and 30% – of current, relevant and valued medical research is available now to all interested readers directly from publisher sites. The benefit from such access can only be seen over time.

## References

1. National Institutes of Health; 'Notice: Enhanced Public Access to NIH Research Information.' Published 3 September 2004. <http://grants1.nih.gov/grants/guide/notice-files/NOT-OD-04-064.html> (accessed 10 January 2005).
2. Willinsky, J., 'The Nine Flavors of Open Access Scholarly Publishing', *Journal of Post-Graduate Medicine* 49 (2003): 263–7.
3. McVeigh, M.E., 'Open Access Journals in the ISI Citation Databases: Analysis of Impact Factors and Citation Patterns.' October 2004; <http://www.isinet.com/media/presentrep/essayspdf/openaccesscitations.pdf> (accessed 10 January 2005).
4. <http://www.ulrichsweb.com> (accessed December 2004).

5. *Web of Science*®. Published by the Thomson Corporation (accessed December 2004).
6. *Journal Citation Reports*® on the Web. Science Edition – 2003. Published by the Thomson Corporation (accessed December 2004).
7. HINARI:  
<http://www.healthinternetwork.org/scipub.php>

Article © Marie E McVeigh and James K Pringle

---

■ **Marie E McVeigh**  
Product Development Manager  
Thomson Scientific  
3501 Market Street, Philadelphia, PA 19104, USA  
Tel: +1-215-386-0100 Ext. 1105  
E-mail: [marie.mcveigh@thomson.com](mailto:marie.mcveigh@thomson.com)

■ **James K Pringle**  
Vice President, Development  
Thomson Scientific  
3501 Market Street, Philadelphia, PA 19104, USA  
Tel: +1-215-386-0100 Ext. 1485  
E-mail: [james.pringle@thomson.com](mailto:james.pringle@thomson.com)

URL: <http://www.thomson.com/scientific>

---

To view the original copy of this article, published in *Serials*, the journal of the UKSG, click here:

<http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=18&issue=1&spage=45>

For a link to the full table of contents for this article, please click here:

<http://serials.uksg.org/openurl.asp?genre=issue&issn=0953-0460&volume=18&issue=1>