American Journal of Evaluation OnlineFirst, published on January 12, 2010 as doi:10.1177/1098214009354774

American Journal of Evaluation
000(00) 1-18

DOI: 10.1177/1098214009354774
http://aje.sagepub.com

Bibliometrics as a Performance Measurement Tool for Research Evaluation: The Case of Research Funded by the National Cancer Institute of Canada

David Campbell,[1] Michelle Picard-Aitken,[1] Grégoire Côté,[1] Julie Caruso,[1] Rodolfo Valentim,[2] Stuart Edmonds,[3] Gregory Thomas Williams,[2] Benoît Macaluso,[4] Jean-Pierre Robitaille,[4] Nicolas Bastien,[4] Marie-Claude Laframboise,[4] Louis-Michel Lebeau,[4] Philippe Mirabel,[4] Vincent Larivière,[4] and Éric Archambault[1]

Abstract

As bibliometric indicators are objective, reliable, and cost-effective measures of peer-reviewed research outputs, they are expected to play an increasingly important role in research assessment/management. Recently, a bibliometric approach was developed and integrated within the evaluation framework of research funded by the National Cancer Institute of Canada (NCIC). This approach helped address the following questions that were difficult to answer objectively using alternative methods such as program documentation review and key informant interviews: (a) Has the NCIC peer-review process selected outstanding Canadian scientists in cancer research? (b) Have the NCIC grants contributed to increasing the scientific performance of supported researchers? (c) How do the NCIC-supported researchers compare to their neighbors supported by the U.S. National Cancer Institute? Using the NCIC evaluation as a case study, this article demonstrates the usefulness of bibliometrics to address key evaluation questions and discusses its integration, along complementary indicators (e.g., peer ratings), in a practice-driven research evaluation continuum.

Keywords

bibliometrics, citation analysis, grant program, performance measurement, research evaluation

[1]Science-Metrix, Montréal, Québec, Canada
[2]Canadian Cancer Society, Toronto, Ontario, Canada
[3]Canadian Partnership Against Cancer, Toronto, Ontario, Canada
[4]Observatoire des sciences et des technologies (OST), Centre interuniversitaire de recherche sur la science et la technologie (CIRST), Université du Québec à Montréal (UQAM), Montréal, Québec, Canada

Corresponding Author:
David Campbell, Science-Metrix, 1335A Avenue du Mont-Royal Est, Montréal, Québec, Canada H2J 1Y6.
Email: david.campbell@science-metrix.com

Research evaluation consists of monitoring, through the use of quantitative and/or qualitative methods, ongoing research programs/initiatives to assess the efficiency and effectiveness with which they are being implemented, to determine the extent to which they are achieving their targeted objectives, and, when appropriate, to recommend adjustments. Ultimately, research evaluation helps determine the overall success of research programs/initiatives through the measurement of their impacts (e.g., effect on research performance) and outcomes (e.g., long-term outcomes, such as socioeconomic benefits). Research evaluation also deals with the assessment of the scientific performance of applicants, which usually relies on peer review during the adjudication process of grant competitions (Luukkonen, 2002). Generally, research evaluation makes use of a variety of indicators to draw as complete a picture as possible of the complex aspects that account for the performance of research organizations (King, 1987).

The main objective of this article is to demonstrate the usefulness of bibliometrics as a performance measurement tool that can be used to address key research evaluation questions, taking the National Cancer Institute of Canada's (NCIC) research grants program as a case in point. The article describes the development and application of a bibliometric approach that was used to answer evaluation questions about the NCIC's research grants program and discusses how bibliometrics could be integrated, along complementary indicators (e.g., peer ratings, statistics on patents, and academic appointments/promotions), in a practice-driven research evaluation continuum.

Evaluation Context—The NCIC

The NCIC was established in 1947, as a joint initiative of the Department of National Health and Welfare (now Health Canada) and the Canadian Cancer Society.[1] The mission of the NCIC was "to undertake and support cancer research and related programs in Canada that will lead to the reduction of the incidence, morbidity, and mortality from cancer" (National Cancer Institute of Canada [NCIC], 2009). The NCIC awarded over Can$60 million to cancer research (Canadian Cancer Research Alliance [CCRA], 2007) in 2005 (NCIC, 2005). The NCIC's revenue was contributed by the Canadian Cancer Society and the Terry Fox Foundation, its funder partners.

In 2005, the NCIC launched a new strategic plan that clearly positioned its research programs as drivers to improve cancer control. Cancer control aims to prevent cancer, cure cancer, and increase survival and quality of life for those who develop cancer, by converting the knowledge gained through research, surveillance, and outcome evaluation into strategies and actions (Luciani & Berman, 2000). The new strategic plan also committed the NCIC to systematically incorporate comprehensive evaluations of its research programs.

The NCIC programs fund a broad spectrum of cancer research, the majority of which has been historically allocated through peer-reviewed research grants. Evaluating the achievement of the NCIC's targeted impacts—reducing morbidity and mortality from cancer—through its research grants presents both theoretical and practical challenges. There is a long series of events—and typically a lengthy time period—between the funding of research and subsequent changes in cancer-related morbidity and mortality. To begin evaluating the impact of its research grants, the NCIC constructed a simple logic model that postulates a linear movement from funding to impact on its mission (Figure 1).

Using this model, an evaluation of the NCIC's research grants program was designed in three phases (Figure 1). In Phase I, a qualitative review of the research grants application and peer-review process was conducted using self-administered surveys of grant applicants and review panel members. Results of Phase I provided lessons about the efficiency and effectiveness of the research grants application and peer-review process of the NCIC's research grants program and recommendations as to how to improve it. Phase II, and the subject of this article, used quantitative methods to

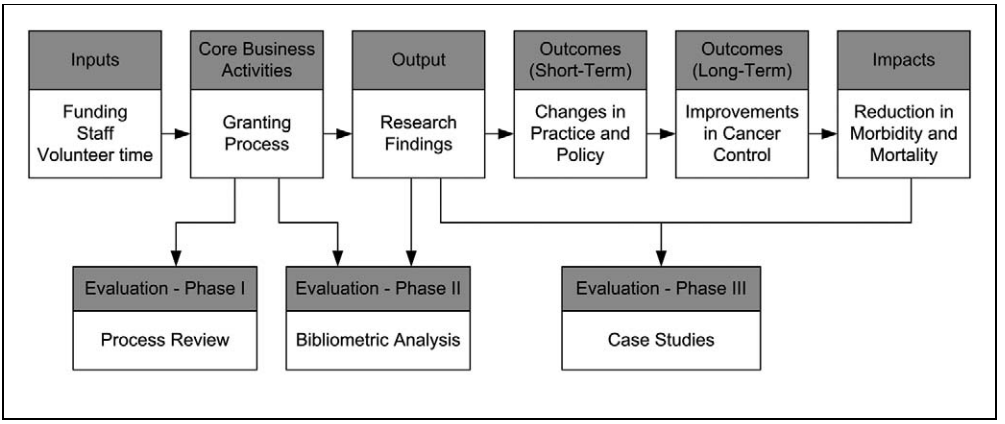

Figure 1. Logic model of the National Cancer Institute of Canada (NCIC) research grants and corresponding evaluation phases.

assess the peer-review process as well as the output of research funding (i.e., research findings). Phase III will apply case study methodology to examine how research funding leads to changes in practice and/or policy that affect the lives of people with cancer or with a risk of cancer.

Implementing Phase II of the NCIC Research Grants Evaluation

Through a review of best practices on evaluating research performance and in consultation with its Advisory Committee on Research, the NCIC defined a set of specific evaluation questions that were to be addressed during Phase II of the evaluation using quantitative methods. The key evaluation questions were as follows:

1. Has the NCIC peer-review process been successful in selecting outstanding Canadian cancer researchers?
2. Has the NCIC funding contributed to increasing the scientific performance, in terms of scientific impact, of the papers authored by the researchers it supported?
3. How do NCIC-supported researchers compare, in terms of scientific output, to their U.S. counterparts who received support from the National Cancer Institute (NCI)?

In the natural sciences and engineering (NSE) that include the health sciences, the results of scientific research are mainly disseminated through the publication of peer-reviewed papers in scholarly journals (Larivière, Archambault, Gingras, & Vignola-Gagné, 2006; Moed, 2005). Consequently, the scientific performance of researchers funded by the NCIC, a central aspect of the above three questions, can be efficiently assessed through the use of bibliometrics: a set of methods and procedures used in the quantification of bibliographic records and for which the basic units of measurement are bibliographic data derived from peer-reviewed publications.

As such, a bibliometric approach was devised and applied to answer the key questions raised in Phase II of the NCIC research grants evaluation. To answer Question 1, the scientific impact—as measured using citation analysis—of NCIC-supported researchers was compared with that of Canadian cancer researchers overall (excluding papers by NCIC-supported researchers). To answer Question 2, the scientific impact of papers authored by researchers while they were receiving funds from the NCIC (i.e., *supported papers*) was then compared to that of their papers authored while they were not funded by the NCIC (i.e., *nonsupported papers*) to establish whether the support provided by the NCIC might have had a positive effect on the scientific performance of the supported researchers. To answer Question 3, the NCIC was compared to the U.S. National Institute of Health's National Cancer Institute (NIH-NCI), looking at the size and impact of their respective scientific production. Analyzing the performance of comparables of the assessed entity allows for conclusions to be drawn about its relative strengths and weaknesses.

Because bibliometric indicators are based on a set of internationally recognized standards, they are considered to be the most objective and reliable measures of academic research output, in addition to being cost-effective in their usage (Archambault & Côté, 2008). Recently, Archambault, Campbell, Gingras, and Larivère (2009) analyzed the robustness of bibliometric indicators as measures of scientific production by comparing statistics obtained from two major bibliographic databases of peer-reviewed literature that differ in terms of scope, volume of data, and coverage policies, namely the Web of Science ([WOS] Thomson Reuters) and Scopus (Reed Elsevier). They showed that indicators based on counts of refereed articles and on counts of citations to these papers are stable at the country level in various fields of science, regardless of the database used.

Although bibliometrics was initially developed by information scientists, it is now being applied more frequently, in combination with peer review, to the field of research evaluation. One of the reasons of this is that governments are increasingly recognizing the importance of accountability for public spending in research, which has increased the need to balance the more traditional assessment by peer review with objective methods such as bibliometrics (Hicks, Tomizawa, Saitoh, & Kobayashi, 2004; King, 1987; Martin, 1996; Trochim, Marcus, Mâsse, Moser, & Weld, 2008).

Bibliometric Approach

This section describes the construction of the data sets used in producing reliable indicators of scientific production for NCIC-funded applicants (core data set) and NCI-funded applicants (benchmark data set). It also presents the taxonomy used in determining the domains of activity of NCIC and NCI researchers, the bibliometric indicators used to quantify scientific outputs, and the statistical analysis performed on bibliometric indicators.

Databases

In this study, Thomson Reuters' WoS, which includes three databases (the Science Citation Index Expanded™ [SCI Expanded], the Social Sciences Citation Index™, and the Arts & Humanities Citation Index™) covering nearly all fields of science, was used to produce statistics on the scientific production of the NCIC researchers and their comparables. Although the vast majority of papers relevant to the area of cancer research are indexed in SCI Expanded, using all of WoS allowed for the retrieval of additional papers in complementary fields (e.g., socioeconomics of cancer). All WoS-indexed papers published by the NCIC researchers and their comparables are included in the analysis, irrespective of the nature of their results—negative or positive. However, for obvious reasons of coverage, unpublished papers (presenting either negative or positive results) are not included in the analysis.

The WoS data were received from Thomson Reuters as flat text files and were then transformed and imported into a relational structured query language (SQL) Server database. Although the WoS lists several types of documents, only articles, research notes, and review articles were retained in producing the bibliometric indicators, as these are considered the main types of documents through which new knowledge is disseminated in the NSE. In addition, all of these documents have been subject to peer review prior to being accepted for publication, ensuring that the research is of good quality and constitutes an original and robust contribution to scientific knowledge. These three types of publications make up nearly 70% of the source documents indexed in the WoS.

Constitution of the Core Data Set: NCIC-Funded Applicants

A bibliometric data set for an institution is usually built by retrieving papers in which the name of the institution is found in the authors' addresses. Because the NCIC supports research, as opposed to being an institute that performs intramural research, its name is not expected to be found in the authors' addresses on published papers. As a result, to build a data set of NCIC-supported papers, a publication portfolio had to be reconstituted for each researcher who received funding from the NCIC.

The construction of the data set followed a two-part process. First, the NCIC provided a list of principal investigators (PIs) who were awarded Operating and Program Project Grants between 1994 and 2006 ($n = 685$ PIs). The names of these researchers were then used to retrieve their publications indexed in the WoS. To avoid overestimates created by homograph problems, each researcher's portfolio was manually cleaned to remove papers belonging to another researcher with the same surname and initials (i.e., false positives).

Bulk researchers' name query. Before executing the automatic retrieval of papers, NCIC researchers' names were transformed to match the format of author names in the WoS, which include only their initials rather than the full first name of the authors. For example, "John W. Smith" is transformed into "Smith-JW" and also into "Smith-J". The latter form ensures the retrieval of publications wherein the middle name (or its initial) is omitted. Subsequently, the formatted names are queried against the database to retrieve, for each researcher, all of the papers bearing his or her name as an author between 1994 and 2006. The search is limited to papers bearing a Canadian address to minimize the occurrence of false positives resulting from homographs in researchers' names. This can lower the recall rate (i.e., does not retrieve all papers by a researcher) in cases where the researcher is newly established in Canada or has been a visiting scientist outside Canada, although in this case scientists normally sign their papers specifying both affiliations (i.e., the Canadian and visiting addresses). Although imperfect, this approach favors precision over exhaustivity, which is appropriate when working with large samples such as the one used in this study ($n = 685$ PIs).

Due to the prevalence of homograph problems, the automatic query overestimated the number of publications in many paper portfolios, especially for researchers with a common surname (e.g., Smith). Moreover, the occurrence of these problems is increased by two limitations of the WoS database.

1. It includes only the initials of the first name—John Smith, James Smith, and Joan Smith are all identified as "Smith-J".
2. It does not contain any information on the relation between the names of the authors and their institutional addresses. For example, in retrieving papers by "Smith-J" who is affiliated with McGill University, a paper coauthored by "Smith-J" of McMaster University and "Anderson WC" of McGill University would be selected. This is due to the fact that, in the absence of links between author names and their addresses, "Smith-J" could be from either McGill or McMaster University. Hence, the presence of homographs is not limited to researchers located in the same institution.

Because there is no a priori regarding which researchers will be overestimated and which will not, the papers retrieved automatically were validated manually for each researcher (Figure 2).

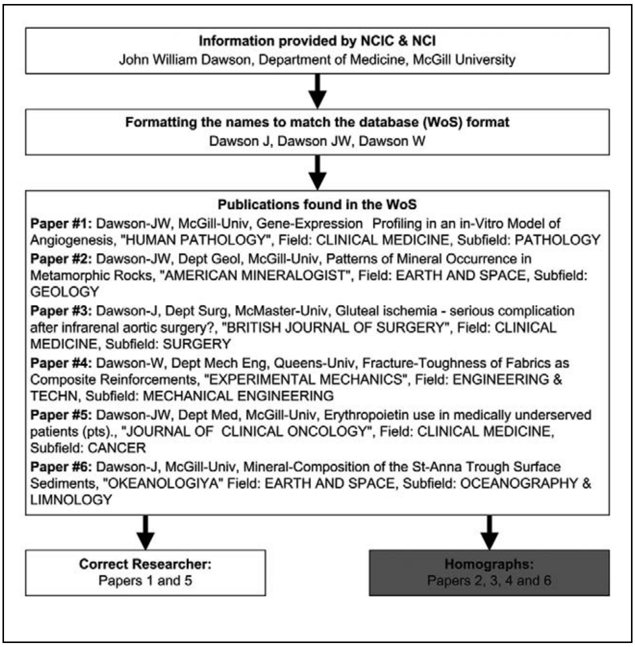

Figure 2. Steps taken in building a researchers' publication portfolio.

Portfolio cleaning. Cleaning the publication portfolios consisted of manually removing papers authored by homographs that were erroneously assigned to a researcher by the automatic query. In so doing, careful attention was paid to the disciplines and specific topics of papers belonging to a publication portfolio. Several questions arise when analyzing whether a set of papers belong to a given researcher (e.g., are the papers consistent with respect to the discipline of the researcher as revealed by his or her departmental affiliation? Is the scope of the papers broader than the products of only one individual researcher?). For example, the attribution of an engineering paper to a biologist or a physics paper to an historian would be seriously questioned. However, given the commonness of multidisciplinarity in science, it is not sufficient to rely mechanically on departmental affiliations of researchers to validate the publications of their portfolio. For example, a philosopher may publish articles dealing with medical ethics in clinical medicine journals, and an engineer may collaborate on papers dealing with environmental problems published in biology or earth sciences journals. The institutional addresses may provide additional clues because they often include the authors' departments. However, the WoS does not contain any information on the relation between the authors' names and their institutional addresses.

In cases where the previous actions failed to determine whether a paper should or should not be considered a part of a researcher's portfolio, the publication was downloaded when it was electronically available through libraries or open access. The article's signatures on the paper itself often provide a link between each author's name and her or his institutional address (including departmental affiliation), which normally allows one to unambiguously identify false positives (Figure 2).

Besides false positives, another issue relates to false negatives; papers authored by a researcher that were not retrieved by the automatic query. These "absent papers" reflect the fact that the WoS only covers a fraction of all of the work published worldwide by researchers. For example, journals of local interest, books, and various publications that are generally referred to as "gray literature" (such as in-house research reports), in addition to most conference and symposium proceedings, are not indexed in Thomson Reuters' scientific databases. Therefore, the publications in the WoS do not encompass the entire curriculum vitae (CV) of researchers funded by the NCIC. Of the scientific output (mostly papers) compiled in the NCIC's annual reports, 22% were not published in journals indexed in the WoS (data not shown). Nevertheless, the WoS indexes the portion of their publications that is the most visible and the most frequently cited by the scientific community. Another factor that can contribute to false negatives is that the name of a researcher as it appears in the list of PIs provided by the NCIC may not be identical to the name under which he or she published articles. This can, for example, result from inaccuracies in the names in the input file and from changes in the name of female researchers due to marriage/divorce.

As the cleaning of publication portfolios involves judgment on the part of individuals performing the task, errors inevitably occur. In this respect, a validation of this procedure was previously performed, which demonstrated that when working with aggregated portfolios (i.e., a number of researchers associated with a given organization), the error rate is negligible (<1%), enabling the production of reliable indicators of scientific production (unpublished data). Altogether, manual cleaning of publication portfolios is a time- and resource-consuming process requiring careful attention. Yet, it is the only way to guarantee that results are sufficiently robust to evaluate important questions, such as the impact of funding on specific groups of researchers.

In building the data set for NCIC-funded applicants, the cleaning process removed half of the initial number of papers automatically assigned to researchers, leaving 24,208 papers authored by NCIC-supported researchers at any given time between 1994 and 2006. Papers were considered *NCIC-supported* if they were published between the year after the start of the grant and the year after the end of the grant. For example, if researchers were supported by the NCIC from 1997 to 2000, their papers published between 1998 and 2001 were counted as *NCIC-supported papers*. As the study period begins in 1994, the inclusion of papers begins with those published from 1995 onward to allow for a 1-year lag following funding distributed in 1994. Between 1995 and 2006, a total of 22,793 papers were authored by NCIC-supported researchers, of which 54% (12,244) were *NCIC-supported papers*. The balance (22,793 − 12,244 = 10,549) is referred to as *nonsupported papers* (these are represented by all other papers by the NCIC researchers, those written before receiving the funding or more than 1 year after the end of the funding period). It should be noted that researchers who received the NCIC funding could also have received grants from other funders; papers considered as *NCIC-supported* may thus have been also partly supported by other sources of funding. Therefore, the authors of this paper do not assume that the performance of these papers, in terms of scientific impact, can be entirely attributed to the NCIC. The approach used to tag papers as supported or nonsupported is discussed at length in the Discussion section.

Constitution of the Benchmark Data Set: NCI-Funded Applicants

The benchmark data set consists of a set of papers from an organization comparable to the NCIC, namely the U.S. NCI. As was the case for the NCIC, inclusion of the NCI in the address field of papers published by the researchers it funds is not expected (i.e., for the extramural researchers considered here). Therefore, a publication portfolio had to be reconstituted for each researcher in a sample of PIs supported by the NCI using the method described above for NCIC-supported researchers. The list of funded researchers was made available to the NCIC by the NCI for the period 2000–2006.

To compare the scientific output of NCIC- and NCI-supported researchers, a random sample of NCI-supported researchers was paired to the population of NCIC-supported researchers to obtain an equal number of researchers on both sides of the comparison. Data on researchers who received NCI grants were only made available from 2000 onward. Because the list of NCI-funded researchers did not provide information on the type and amount of grants awarded to them, researchers having received funding from 2000 to 2006 from different types of NCI grants (i.e., different types of individual research and team grants are included whereas center grants are excluded) were selected by randomly sampling NCI-supported researchers. As such, to obtain the comparable population of NCIC-supported researchers, Canadian researchers were selected so long as they obtained the NCIC funding from 2000 to 2006, regardless of the type of grants they received (NCIC-Program Project grant or NCIC-Operating grant). A sample of 523 NCIC-supported researchers was thus obtained and compared with a random sample of 523 NCI-supported researchers. All papers published by both groups of researchers from 2000 to 2006 were counted, regardless of whether these researchers were funded over the whole period.

Disciplinary Classification

The categories and methods used to delineate the various domains of activity of the NCIC and NCI researchers are, by and large, those used by the U.S. National Science Foundation (NSF) in the Science and Engineering Indicators series (see for example, National Science Foundation [NSF], 2006). The taxonomy is a journal-based classification that has been in use since the 1970s. Because the NSF classification does not entirely satisfy the needs in the SSH, this taxonomy was modified using a customized classification of journals for the social sciences.

The resulting taxonomy has one important advantage over other classifications (such as that used by Thomson Reuters)—it is mutually exclusive, which means that each paper is attributed to a single field or subfield based on the journal in which it is published. One limitation of this classification is that, for example, a paper on the environment that is published in a journal specialized in chemical engineering would be classified as belonging to the field of chemistry and the subfield of chemical engineering, even though its subject is the environment. These anomalies have little effect when large numbers are considered; however, their impact is greater when the number of papers considered is small (e.g., below 30). Some of the subfields are categorized as general (e.g., general biomedical research), and this reflects the fact that in many fields there are some journals that address a broader readership.

Bibliometric Indicators

Using researcher portfolios built using the aforementioned methods as well as papers computed at the country level (for Canada and the United States), the following indicators were calculated:

Number of papers. Whole counting of scientific papers (i.e., when a paper is coauthored, the paper counts once for each coauthor instead of being fractionated) written by authors associated with a funding organization (i.e., NCIC or NCI) based on author names or with a country based on author addresses. At the country level, each paper is counted only once per country even if there are two or more authors from the same country.

Average of relative citations (ARC). ARC is an indicator of the *actual* scientific impact of papers produced by an entity (i.e., a country, a researcher; Moed, de Bruin, & van Leeuwen, 1995; Van Raan, 2004). Citations for each paper were counted over a 3-year window (i.e., for papers published in 1995, citations were counted during the 1995–1997 period; for papers published in 1996, citations were counted during the 1996–1998 period; and so on). The exceptions are papers published in 2005, which have a citation window of 2 years (2005 and 2006), and in 2006, which have a citation window of 1 year, because there were no citation data for subsequent years. The use of a fixed citation window is preferred over total citations counts as it corrects for differences in the age of publications (older papers have accumulated citations over a longer period than recent ones such that they tend to have higher citation counts). To account for different citation patterns across subfields of science (e.g., there are more citations in biomedical research than mathematics), the citation count of a paper in a given subfield (see previous subsection for information on the classification of papers by subfield) is divided by the average count of all papers published the same year in its subfield within the WoS, to obtain a relative citation count (RC). The ARC of a given entity is the average of the RC of papers belonging to it. When the ARC is above 1, an entity scores better than the world; when it is below 1, an entity publishes papers that are cited less often than the world average. Self-citations are excluded.

Although indicators of scientific impact based on journal impact factors (e.g., the average of relative impact factors [ARIF]) are the most widespread, the indicator used in this study (i.e., the ARC which is sometimes referred to as a "normalized citation impact indicator") is generally recognized as the most appropriate bibliometric measure of scientific impact for the purpose of establishing the performance of a research group (Moed, 2009). This is because the ARC is a *direct* measure of the scientific impact of papers by a research group based on their citation counts whereas the ARIF is an *indirect* measure based on the impact factors of the journals in which the papers of the group were published. In the former case, the citation counts of the papers by the group are specific to the authors in the group, whereas in the latter case the impact factors of the papers by the group are not specific to the authors in the group; many authors from many institutions in various countries that are outside the group being measured contribute to the impact factors of the journals in which their papers were published. In fact, journal impact factors correlate poorly with actual citations of individual articles such that the ARC and ARIF indicators are not strongly correlated at microaggregation levels (e.g., research groups; Moed, 2009; Seglen, 1997). The strength of the association between them increases as we move toward higher aggregation levels (i.e., from research groups to countries).

Statistical Analyses

To establish whether there were significant differences between the scientific impact of various entities, a series of statistical tests were performed in statistical package for the social sciences (SPSS). For each statistical test, the difference in scientific impact was considered significant at $p < .05$, very significant at $p < .01$, and highly significant at $p < .001$. Because data on scientific impact are not normally distributed, nonparametric tests were used. The Mann-Whitney U test was used to compare the ARC of two independent samples, whereas the Z test was used to compare the proportion of papers of two independent samples that are in the 5% of papers with the highest citation counts.

Limitations of Bibliometrics

The use of citation analysis to infer the scientific impact of scientific publications has been criticized on several grounds. King (1987) reviewed a number of objections to its use for performance assessment in scientific research.

1. Citation analysis assumes that referenced articles disclosed knowledge that was essential to the work depicted in the citing article. Although it is true of most citations, some of them will be "perfunctory."
2. Incorrect work may be highly cited. However, it has been argued that cited work, even if incorrect, has made an impact or has stimulated further research efforts whereas poor quality work will usually be ignored.
3. Researchers may artificially increase their scientific impact through citing their own work. The indicator used in this study correct for this by removing self-citations.
4. Errors in counting the number of citations of an entity (e.g., institution, country) could occur in the WoS (as it could in any other database) due to indexing errors arising from different ways of citing the name of an author and/or institution (e.g., Can For Serv, CFS, Canadian Forest

Service, and Can Forest Service) or to homograph problems (i.e., authors with the same name). To limit these types of errors, Observatoire des sciences et des technologies (OST) and Science Metrix analysts spent an appreciable amount of time harmonizing the name of researchers, institutions, and countries in the database.

5. Another factor affecting citation counts is the difference in citation practices between disciplinary subfields. For instance, it is well known that mathematicians cite less than biomedical researchers. Hence, one should not directly compare publications and citation counts between subfields. The indicator used in this study corrects for this by normalizing citation counts by subfields.
6. The WoS coverage of scientific literature has a bias in favor of countries that publish in English-language journals. Thus, for countries whose researchers would tend to publish more in other languages, their scientific production is underestimated in the context of international comparisons. In the NCIC-versus-NCI comparison, the impact is limited, given that researchers of both groups are from Anglo-Saxon countries, namely Canada and the United States.

A common misconception about indicators based on citation counts is the idea that they inform on the "quality" of the underlying work. In fact, these indicators provide a partial measure of the impacts published articles have had on the scientific community. Complementary measures of scientific performance such as publications counts, peer ratings, and distinctive awards/prizes have their own set of limitations; the authors refer readers to King (1987) for a thorough discussion of these limitations in the context of research evaluation. As such, citation analysis should be used as one of many partial indicators of scientific performance in the toolbox of research evaluators rather than as a stand-alone tool. When all of the indicators converge, the conclusions drawn from the evaluation are regarded as being more reliable than those based on a single indicator (King, 1987).

Summary of Results

There were 685 researchers supported through the NCIC during 1994–2006, of which 679 published at least one paper during the period. Overall, just over half of scientific papers by these researchers at any time over the 1994–2006 period were published with financial support from the NCIC: 12,244 out of 22,793 papers by NCIC-funded researchers were published after their first year of the NCIC funding up until the year after the end of the grant. The proportion of NCIC-supported papers has been shrinking since 1994, as the number of nonsupported papers (i.e., all other papers by the NCIC researchers) has increased steadily, while the number of NCIC-supported papers published per year has been more stable.

Scientific Impact of NCIC-Supported Researchers' Total Scientific Production

An important indicator of scientific performance is the *actual* impact that the published papers have had on the scientific community. The ARC is an indicator of the number of times an entity's papers (e.g., a group of researchers, a country) are actually cited relative to the world average and can therefore be used as a *direct* measure of scientific impact. All entities considered in this section score above the world average (i.e., ARC above 1).

Researchers who were supported by the NCIC at any one time between 1994 and 2006 have a substantially higher scientific impact than nonsupported Canadian researchers in all fields (highly significant for the three sets of papers by NCIC-funded researchers [all papers by NCIC researchers, NCIC-supported papers, and nonsupported papers], $p < .001$; Figure 3). In addition, the impact of

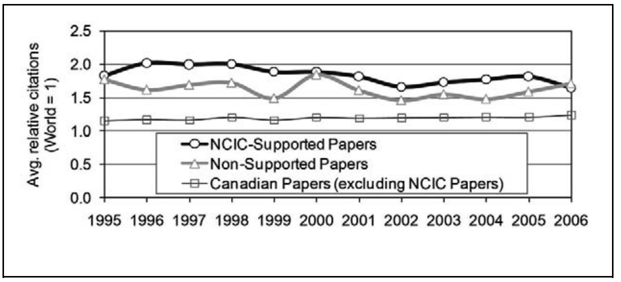

Figure 3. Average of relative citations of papers by National Cancer Institute of Canada (NCIC)–supported and nonsupported Canadian researchers.

papers by NCIC-funded scientists was greater when they were authored with NCIC support than when authored without this source of financing (highly significant, $p < .001$). However, the difference between the scientific impact of NCIC-supported papers and those without NCIC support vanished in 2006. This could be due to the relatively short time that this recent year affords for citations. In general, papers reach their citation peak (i.e., year in which they have received the most citations) about 2–3 years after publication. Therefore, articles published in 2006 will not have reached their citation peak, potentially leading to misleading citation scores by subfield for 2006; the average of citation counts for articles in a subfield in 2006 might be underestimated leading to biases in the normalized citation counts of papers. As a result, the ARC indicator is most reliable over the 1996–2005 period.

Scientific Impact of NCIC-Supported Researchers' Production in Cancer-Related Journals

Researchers who received support from the NCIC at any given time from 1994 to 2006 contributed to nearly half of the Canadian papers published in the subfield of cancer research (delineated using specialist oncology journals based on the NSF classification, see Methods) from 1995 to 2006, and 26% of Canadian papers in cancer research were produced by these authors while they were receiving funds from the NCIC (data not shown). Within this subfield, the scientific impact of papers by researchers who were supported by the NCIC (NCIC-supported papers together with nonsupported papers) was higher than the scientific impact of Canadian papers, and the difference was highly significant ($p < .001$) for the three following sets of papers: all papers by the NCIC researchers, NCIC-supported papers, and nonsupported papers (Figure 4). As was the case for the total output of NCIC researchers (all subfields included), the impact of their papers in cancer research when they were supported by the NCIC was generally (for 10 out of 13 years) higher than when they were not supported, and the difference was significant ($p < .05$; Figure 4).

NCIC-Supported Researchers Compared to NCI-Supported Researchers

For the 2001–2007 period, the average level of funding awarded per researcher by the NCI was 4.5 times that awarded by the NCIC (data not shown). On average, NCI-supported researchers received about U.S.$600,000 at purchasing power parity (PPP) in 2006 compared to about U.S.$150,000 PPP for NCIC-supported researchers. It should be noted, however, that NCI grants also include PI salaries, whereas NCIC grants do not provide for personal salary support of the PIs and/or coapplicants. Furthermore, based on financial data provided by both institutes, grants awarded to U.S. researchers by the NCI covered, on average, 3.85 years of research out of 7, compared to 3.64 years out of 7 for Canadian researchers who received the NCIC grants. Thus, although the amount of funding is disproportionately different, the length of the funding period is highly comparable.

The scientific production of the two groups was similar throughout the 2000–2006 period, although the output of NCI-supported researchers had a slightly stronger increase than the output

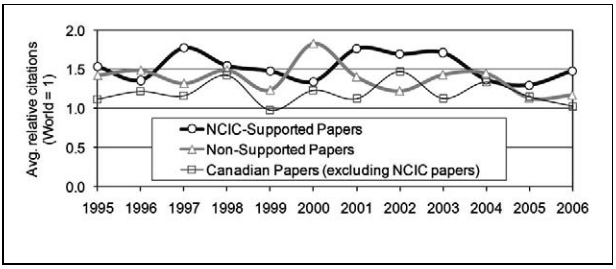

Figure 4. Average of relative citations of papers by National Cancer Institute of Canada (NCIC)–supported and nonsupported Canadian researchers in cancer research.

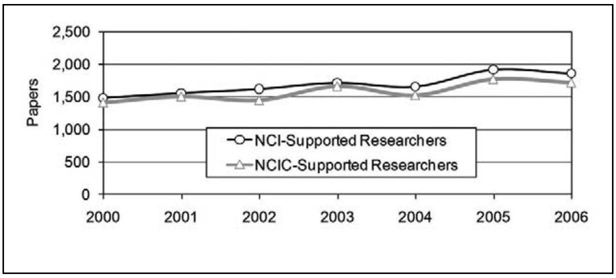

Figure 5. Scientific production by National Cancer Institute of Canada (NCIC) –supported ($n = 523$) versus National Cancer Institute (NCI) −supported ($n = 523$) researchers.

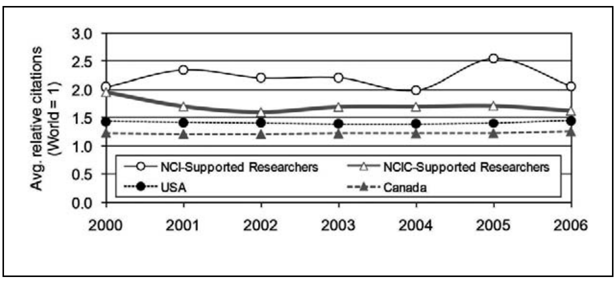

Figure 6. Average of relative citations of papers by National Cancer Institute of Canada (NCIC)–supported ($n = 523$) versus National Cancer Institute (NCI)–supported ($n = 523$) researchers.

of NCIC-supported researchers (net increase of 25% vs. 21%; Figure 5). In total, NCIC-supported researchers published nearly as many papers as did NCI-supported researchers (11,019 vs. 11,794), despite a level of funding that is nearly five times smaller (considering only financial support from the NCI and the NCIC).

However, the papers produced by NCI-supported researchers have had stronger scientific impact than the papers published by NCIC-supported scientists (highly significant, $p < .001$), with the exception of the year 2000, during which both groups had similar ARC values (Figure 6). For the 2000–2006 period, papers by NCI-supported researchers received, on average, about 120% more citations than the average world paper, while those of NCIC researchers received about 70% more citations. Both groups produced papers that achieved greater scientific impact than their respective nation's average paper (highly significant, $p < .001$, Figure 6).

Discussion

The bibliometric approach employed by the NCIC as part of its research grants evaluation allowed the NCIC to answer questions related to the performance of its research funding that could not be addressed objectively using alternative methods such as program documentation review, surveys, and key informant interviews. The bibliometric analysis built on the findings from Phase I (process review) and helped validate some of its results—especially, the peer-review process is successful in selecting outstanding researchers for the attribution of NCIC research grants. Most importantly, it provided NCIC with key insight into the value for money of its investment, which will aid in strategic planning and priority setting going forward.

The bibliometric analysis of NCIC's research grants addressed the following questions: (a) Has the NCIC peer-review process been successful in selecting outstanding Canadian cancer researchers? (2) Has the NCIC funding contributed to increasing the scientific performance, in terms of scientific impact, of the papers authored by the researchers it supported? (c) How do NCIC-supported researchers compare, in terms of scientific output, to their U.S. counterparts who received support from the NCI?

The identification of papers that are published with and without the financial support of a specific organization is not only a key feature of any bibliometric approach designed to answer questions of this type, but it also represents an important methodological challenge for research evaluation. In this study, papers published between the year after the start of an NCIC grant and the year after the end of an NCIC grant were tagged as *NCIC-supported papers*, and the balance were tagged as *nonsupported papers*.

The choice of the time window for tagging papers as being supported or nonsupported is arbitrary and may lead to the inclusion of a number of nonsupported papers in the set of NCIC-supported papers, and vice versa. However, this approach, when used with a large number of researchers (e.g., in this project, $n = 685$ PIs) over a long time period (12 years), is believed to provide a representative sample of papers published with the financial support of the NCIC.

Other studies have used funding acknowledgments when comparing sets of supported and nonsupported papers (Cronin & Shaw, 1999; Rangnekar, 2005). Although this approach can more precisely identify supported papers, its recall rate can be as low as 50%, as authors of scientific papers do not systematically acknowledge sources of funding (Lewison & Carding, 2003). This can, in turn, lead to the inclusion of supported papers in the set of nonsupported papers. Because the papers' acknowledgments were not indexed in the database used to perform this study and because it was not feasible, given the large data sample used (>20,000 papers), to manually download each paper to examine the acknowledgments section, the former approach was chosen.

As neither of these two approaches is perfect, care must be taken when interpreting differences between supported and nonsupported papers. To strengthen the analysis, future work could examine how supported and nonsupported papers classified based on grant periods compare to supported and nonsupported papers classified based on funding acknowledgments. To perform this comparison, the data set used in this study could be reduced to a manageable size when capturing data on funding acknowledgments (e.g., using the 3 most recent years). In addition, further refinements of the approach could be achieved by investigating the time lag from reception of funding to the first and last publication associated with it. Such investigations should consider the time lag variation between scientific fields.

Researchers receiving the NCIC funding also receive grants from other funders, such that papers considered as NCIC-supported may have also been partially supported by other sources of funding. In fact, the NCIC awarded about 25% (over Can$60 million in 2005) of all funds for cancer research in Canada, distributed by members of the Canadian Cancer Research Alliance (CCRA), which includes federal and provincial governments as well as voluntary funding organizations like the NCIC (CCRA, 2007). As such, we do not assume that the scientific impact of NCIC-supported papers can be entirely attributed to the NCIC. Nevertheless, because NCIC-funded researchers also receive funding from other sources when they *do not* receive support from the NCIC, everything being equal, the protocol suggested here takes into account the variation in financing that should accrue to the NCIC.

During the 1995–2006 period, just over half of the total scientific production of researchers who received funding from the NCIC at any time between 1994 and 2006 was produced with financial support from the institute. These researchers were authors on nearly half of the Canadian papers published in cancer-related journals from 1995 to 2006. Moreover, the results of this study suggest that a quarter of Canadian papers in these journals were published with the NCIC support. Thus, it seems that the NCIC's financial contribution—25% of Canada's total funding for cancer research (CCRA, 2007)—resulted in a proportionately sized scientific output—25% of Canada's production in cancer journals—by the researchers it supported.

The NCIC strives to advance cancer research by supporting outstanding researchers selected through peer review. A number of studies found a positive correlation between peer ratings and citation scores of researchers in grant competitions (Lawani & Bayer, 1983; McAllister, Anderson, & Narin, 1980). When both measures of scientific excellence are used for the purposes of researcher assessment, the selection of awardees is generally regarded as being more reliable than when only peer ratings are considered (King, 1987). Because peers are making increased use of bibliometric indicators in rating researchers, as appears to be the case among members of the NCIC's review panel (according to NCIC representatives), positive correlations between peer ratings and citation scores will appear even more frequently in the future, especially if reviewers make formal use of bibliometric indicators (using measures provided by statistical offices, companies, or specialized university departments), rather than solely using their tacit knowledge about how well a researcher performs with respect to these indicators.

If the NCIC's peer-review process has been efficient in selecting leading researchers, papers authored by successful applicants while they were not financially supported by the NCIC are expected to have, on average, greater scientific impact than other Canadian papers. In the present case, the results, using the ARC, support this hypothesis when considering papers in all fields, as well as when limiting the analysis to papers in cancer journals. Considering that the scientific impact of NCIC researchers is significantly higher than that of other Canadian researchers who were not supported by NCIC, who themselves score above the world average, it is quite clear that researchers funded by the NCIC are of high caliber.

The analysis of the *actual* scientific impact also supports the hypothesis that the funds provided by the NCIC contributed to increasing the scientific performance of successful applicants: the ARC of NCIC-supported papers is significantly higher than that of nonsupported papers. While some citations in scientific papers may be nonessential, they generally substantiate a finding or assumption, and in cases where the cited work is incorrect or outdated, the act of citing still reflects a stimulatory effect of the cited work (King, 1987). Thus, the fact that the ARC of NCIC-supported papers is higher than that of nonsupported papers suggests that the NCIC contributed to increasing the usefulness, for the scientific community, of the knowledge produced by NCIC awardees.

The results of this bibliometric study provide evidence of the successful selection of outstanding Canadian researchers by the NCIC peer-review panel and of the NCIC's positive effect on the scientific impact of the researchers it supported. However, a more detailed analysis of the research outputs and of funding at the level of individual NCIC researchers is required to provide a deeper understanding of these results.

For instance, the current study compared the scientific impact of NCIC-supported researchers with that of other Canadian researchers using scientific articles published in cancer journals. It was recently shown that only 50% of cancer-related research is being published in specialist oncology journals and that oncologists publish their most influential work in more general journals or in journals covering other specialities (López-Illescas, Moya-Anegón, & Moed, 2008). Thus, it is safe to assume that about 50% of cancer research was sampled in this study, such that the scientific impact of NCIC-supported researchers would likely remain above that of other Canadian researchers if the field was expanded to include cancer-related research not published in specialist oncology journals. Nevertheless, it could be interesting to compare the results of the current study with those obtained with an expanded data set.

Another issue relates to the type of research supported by the NCIC or the type of research generally performed by the researchers it supported. For instance, if the NCIC supports a specific type of cancer research, such as clinical investigations, then efforts should be made to examine the citation habits of researchers performing basic research as opposed to clinical research. Indeed, it might be found that clinical research is generally more cited than basic research, in which case the difference in the scientific impact of NCIC-supported researchers and other Canadian researchers would relate to citation patterns rather than to the scientific excellence of the research performed. Finally, to more precisely assess whether NCIC's peer-review process has been efficient in selecting leading researchers, future comparisons should also compare the scientific impact of successful applicants to that of unsuccessful applicants.

Compared to a random sample of researchers funded by the NCI of the same size as the population of NCIC-funded researchers, NCIC-supported researchers produced, on average, nearly as many papers per year from 2000 to 2006, despite receiving about five times less funding per researcher (taking into account only financial support from the NCI and the NCIC). Considering that the NCI provides approximately 70% of the total public funding for cancer research in the United States (Jönsson & Wilking, 2007), compared to the NCIC's contribution of about 25% in Canada (CCRA, 2007), the average NCIC researcher likely has access to less public funds than the average NCI researcher even if one takes account of the fact that NCI funding includes salaries while NCIC funding does not. If further research examining the relative amount of public funding available to the two groups supports this hypothesis, then NCIC researchers would have produced more papers per public dollar investment than NCI researchers. Such an analysis was performed in the field of biomedical research at the level of countries and showed that Canada ranked 1st, slightly ahead of the United States, for its number of articles produced per dollar investment on research and development (R&D; Soteriades, Rosmarakis, Paraschakis, & Falagas, 2006).

As investigations into the causes of cancer and its potential cures require the use of increasingly sophisticated and expensive technologies, the costs of cancer research have also increased significantly, requiring ever greater investment on the part of funding bodies. As such, access to greater financial resources may not directly translate into more scientific publications but might instead result in projects that could not have been accomplished otherwise (e.g., large scale, long term, or innovative projects, such as the Human Genome Project). The higher scientific impact of the papers produced by NCI researchers (as compared to those published by NCIC researchers) might then be explained, at least in part, by NCI researchers performing more research projects of this type, which are likely to attract more attention from the scientific community. Here again, more data are required on the type of research performed and on the distribution of available resources (e.g., infrastructure, financial and human resources) among researchers to provide a detailed interpretation of the results.

As a final issue, to provide further "evaluative" value in the final recommendations for NCIC's research grants, it would be interesting to conduct a more thorough analysis comparing the scientific output of both successful and *unsuccessful* applicants prior to and after their application to the program. Indeed, the unsuccessful applicants are more directly comparable to successful applicants than are all Canadian researchers not funded by the NCIC.

In spite of the limitations discussed above, this study demonstrates the usefulness of bibliometrics as a performance measurement tool: bibliometrics can be used to address important evaluation questions related to the performance of researchers supported by funding programs, including questions that are difficult to answer objectively through other methods such as program documentation review, key informant interviews, and surveys. More specifically, the use of bibliometrics, especially citation analysis, was shown to provide valuable information on the effect of funding, through a comparison of scientific papers produced with and without financial support from a funding organization. Citation analysis also provided a means to answer a key evaluation question pertaining to NCIC's peer-review process, which demonstrates the usefulness of bibliometrics as a formal and uniform approach to assessing the general performance of this process. As such, the authors believe that the use of bibliometrics in ex-post analyses aimed at assessing the peer-review process and the effect of funding on research performance should be integrated within a practice-driven research evaluation continuum.

In the current case, the results pertaining to assessing NCIC peer-review process support the hypothesis that it has been effective at selecting outstanding Canadian researchers pursuing cancer research. However, this conclusion is based on an analysis of the scientific impact of funded researchers in an aggregated manner, thereby providing information on the *average* performance of the peer-review process. As noted by Moed (2009), the positive correlations observed in some studies between peer ratings and citation scores of researchers in grant competitions are not perfect such that when looking at researchers individually, disagreement between the two measures do occur. Thus, the authors believe that to reinforce the grant adjudication process, bibliometrics could be integrated as early as in the assessment of research proposal, instead of being used only in performing ex-post analyses on the performance of the peer-review process, to reduce the risk of "bad decisions" (i.e., cases in which disagreement between peer and bibliometric ratings are unjustified) and for clear reasons of transparency. In so doing, a funding agency could provide peer-review panels with initial rankings of the applicants based on a formal bibliometric analysis and let the peers justify why their final decisions differ from the initial rankings (Moed, 2009).

Evaluators, program managers, peer reviewers, and bibliometric practitioners have much to gain by working collectively on developing integrated assessment models that incorporate and maximize the usefulness of both qualitative (i.e., peer-review) and quantitative (i.e., bibliometrics) assessment methods. An integrated approach is the key to overcoming the imperfections inherent in all of these tools (Archambault & Côté, 2008; Moed, 2009).

Acknowledgments

The authors wish to thank the National Cancer Institute of Canada (NCIC) and the U.S. National Institute of Health's National Cancer Institute (NIH-NCI) for having provided listings of the principal investigators they have funded and Dr. Henk F. Moed of CWTS for his constructive comments on this study.

Declaration of Conflicting Interest

The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding

This work was supported by the Canadian Cancer Society.

Note

1. On February 1, 2009, the Canadian Cancer Society integrated the operations of the NCIC, creating the Canadian Cancer Society Research Institute. Note that because the period being examined in this study precedes the integration, the text will reference the NCIC.

References

Archambault, É., Campbell, D., Gingras, Y., & Larivière, V. (2009). Comparing bibliometric statistics obtained from the Web of Science and Scopus. *Journal of the American Society for Information Science and Technology*, *60*, 1320-1326.

Archambault, É., & Côté, G. (2008). Better bibliometrics needed for greater accountability of research expenditures. *Research Money*, *22*, 8.

Canadian Cancer Research Alliance. (2007). *Cancer research investment in Canada, 2005*. Retrieved January 16, 2009, from http://www.ccra-acrc.ca/PDF%20Files/CCRA_07_EN_Full%20Report.pdf

Cronin, B., & Shaw, D. (1999). Citation, funding acknowledgment and author nationality relationships in four information science journals. *Journal of Documentation*, *55*, 402-408.

Hicks, D., Tomizawa, H., Saitoh, Y., & Kobayashi, S. (2004). Bibliometric techniques in the evaluation of federally funded research in the United States. *Research Evaluation*, *13*, 78-86.

Jönsson, B., & Wilking, N. (2007). Cancer research and development and the drug development process. *Annals of Oncology*, *18*, iii49-iii54.

King, J. (1987). A review of bibliometric and other science indicators and their role in research evaluation. *Journal of Information Science*, *13*, 261-271.

Larivière, V., Archambault, É., Gingras, Y., & Vignola-Gagné, É. (2006). The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology*, *57*, 997-1004.

Lawini, S. M., & Bayer, A. E. (1983). Validity of citation criteria for assessing the influence of scientific publications: New evidence with peer assessment. *Journal of the American Society for Information Science*, *34*, 59-74.

Lewison, G., & Carding, P. (2003). Evaluating UK research in speech and language therapy. *International Journal of Language & Communication Disorders*, *38*, 65-84.

López-Illescas, C., de Moya-Anegón, F., & Moed, H. F. (2008). The actual citation impact of European oncological research. *European Journal of Cancer*, *44*, 228-236.

Luciani, S., & Berman, N. J. (2000). Canadian strategy for cancer control. *Chronic Diseases in Canada*, *21*, 23-25.

Luukkonen, T. (2002). Research evaluation in Europe: State of the art. *Research Evaluation*, *11*, 81-84.

Martin, B. R. (1996). The use of multiple indicators in the assessment of basic research. *Scientometrics*, *36*, 343-362.

McAllister, P., Anderson, R. C., & Narin, F. (1980). Comparison of peer and citation assessment of the influence of scientific journals. *Journal of the American Society for Information Science*, *31*, 147-152.

Moed, H. F., de Bruin, R. E., & van Leeuwen, T. N. (1995). New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications. *Scientometrics*, *33*, 381-442.

Moed, H. F. (2005). *Citation analysis in research evaluation.* Dordrecht: Springer.

Moed, H. F. (2009). New developments in the use of citation analysis in research evaluation. *Archivum Immunologiae Et Therapiae Experimentalis*, *57*, 13-18.

National Cancer Institute of Canada. (2005). *NCIC 2015: Driving excellent cancer research to improve cancer control*. Retrieved January 16, 2009, from http://www.ncic.cancer.ca/~/media/NCIC/Files%20list/English-20files/PDF/Strategic%20plan%20-%20EN%20-%20PDF%20-%202005_530236530.ashx

National Cancer Institute of Canada. (2009). *Canadian Cancer Society Research Institute*. Retrieved January 16, 2009, from http://www.ncic.cancer.ca/About%20us%20and%20news/History%20and%20mission.aspx

National Science Foundation. (2006). *Science and engineering indicators*. Retrieved January 16, 2009, from http://www.nsf.gov/statistics/seind06/pdfstart.htm

Rangnekar, D. (2005). Acknowledged: Analysing the bibliometric presence of the Multiple Sclerosis Society. *Aslib Proceedings*, *57*, 247-260.

Seglen, P. O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, *314*, 497.

Soteriades, E. S., Rosmarakis, E. S., Paraschakis, K., & Falagas, M. E. (2006). Research contribution of different world regions in the top 50 biomedical journals (1995–2002). *The Faseb Journal*, *20*, 29-34.

Trochim, W. M., Marcus, S. E., Mâsse, L. C., Moser, R. P., & Weld, P. C. (2008). The evaluation of large research initiatives: A participatory integrative mixed-methods approach. *American Journal of Evaluation*, *29*, 8-28.

Van Raan, A. F. J. (2004). Measuring science. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems.* Dordrecht: Kluwer Academic Publishers.