THE EFFECT OF COLLABORATIVE NETWORKS ON HEALTHCARE RESEARCH PERFORMANCE

Imperial College London

Vanash Mahendra Patel

Department of Surgery and Cancer
Imperial College London

This dissertation is submitted for the degree of Doctor of Philosophy
October 2015

I lovingly dedicate this thesis to my family, who supported me each step of the way.

Declaration

This dissertation is the result of my own work and includes nothing, which is the outcome of work done in collaboration except where specifically indicated in the text. It has not been previously submitted, in part or whole, to any university or institution for any degree, diploma, or other qualification.

Signed:

Date: 30/10/2015

Vanash Mahendra Patel

MBBS MRCS(Eng) MSc

ABSTRACT

We can all use assessment and appraisal to help us improve our performance in any area of life. Healthcare researchers are no exception. For healthcare researchers a system is required to measure research performance according to an accepted global benchmark.

While there are existing systems that have been created to measure research performance in general, and healthcare research performance has been appraised with several bibliometric indicators, there is a lack of evidence to prove their validity and a deficiency of indicators that embrace social behaviours such as collaboration.

In this thesis we endeavoured to enhance knowledge on healthcare research performance assessment, which has the potential to be integrated into systems that specifically appraise healthcare research performance. Ultimately, these systems may promote a performance-based culture that better reflects the quality and impact of healthcare research.

Acknowledgements

I would like to express the deepest gratitude to my supervisor Professor Thanos Athanasiou, who has shown the assertiveness and the substance of a genius: he continually and influentially conveyed a spirit of adventure in regard to research and an enthusiasm in regard to teaching. Without his supervision and constant support this thesis would not have been possible.

I would like to express my special appreciation and thanks to my supervisor Professor Ara Darzi, who has been an inspirational role model and a tremendous mentor for me. His guidance on both research as well as on my career have been invaluable.

As well as my efforts, the achievement of any research project depends to a great extent on the encouragement, advice and guidance of many others. I would like to sincerely thank my friend and colleague Mr. Hutan Ashrafian for the continuous support, engagement, useful comments and remarks throughout the learning process of this thesis. In chapter 2, Mr. Hutan Ashrafian assisted with selecting studies for the systematic review. I take this opportunity to express my gratitude to Dr. Lutz Bornmann and Dr. Rüdiger Mutz who have educated and directed me in the field of scientometrics. In chapter 4, Dr. Lutz Bornmann advised on study design and Dr. Rüdiger Mutz guided me with the statistical analysis. I would also like to express my appreciation to Dr. Pietro Panzarasa and Dr. Tim Evans for enlightening and instructing me in the field of social network analysis. In chapter 5, Dr. Pietro Panzarasa guided me with the statistical analysis and Dr. Tim Evans aided me with the network illustrations. I have written this thesis in the first person plural to reflect the invaluable contributions of the aforementioned collaborators.

I would like to thank Sage Publications for granting permission to include Chapter 2 and Chapter 4 of my thesis, which were originally published in the Journal of the Royal Society of Medicine (2011; 104(6): 251-61 and 2013; 106(1): 19-29 respectively). I would like thank S. Karger AG for granting permission to include Chapter 3 of my thesis, which was originally published in Medical Principles and Practice (2013; 22(2): 178-83).

Last, but not least, I would like to thank my family. Thank you to my devoted parents for instilling in me the love of education and the continuous aspiration for more knowledge. To my two adorable children, Eashan and Amelie, thank you for the joy you bring into my life and for reminding me about the essential things in life. Above all, I thank my wife, Rohina, for her continuous love, patience, support, and for believing in me.

CONTENTS

LIST OF TABLES

List of Figures

List of Abbreviations and Acronyms

AHSC	Academic Health Science Centre
ANOVA	Univariate analysis of variance
AWCR	Age Weighted Citation Ratio
B	Unstandardized B coefficient
CI-L	Confidence interval - lower
CI-U	Confidence interval -upper
CR	Contingent reward
GDP	Gross Domestic Product
ISI	Thomson Scientific's Institute for Scientific Information database
MANOVA	Multivariate analysis of variance
MeSH	Medical Subject Headings
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
SE	Standard error
SRM	Segmented regression model
UK	United Kingdom

List of Appendices

1 INTRODUCTION

Healthcare research is the basic, applied or translational research conducted to assist and support the enrichment of knowledge in the discipline of medicine. It encompasses clinical and non-clinical research concerned with safeguarding and promoting public health; and may be undertaken by government organisations, industries, charities, research councils and universities within the health and social care structures.(*1*)

Healthcare research has unearthed many cutting-edge discoveries, which have been translated into tremendous advances in patient care.(*2*) It has led to huge benefits in life-saving intervention and to major advances in the quality of life of patients with chronic disease.(*2*) Unsurprisingly, these developments and innovation in healthcare have increased global life expectancy by 6 years since 1990.(*3*) Developing a breakthrough of a new drug or treatment requires massive amounts of healthcare research and engages many healthcare professionals.(*4*) The pursuit of new knowledge through healthcare research requires significant resources and investment, which aims to generate benefit to society.(*4*) Whether healthcare research is privately, publicly or charitably funded, it is essential that the stakeholder can appreciate the quality and value of the research to justify the economic costs that are incurred. However, the structures, processes and outcomes that drive quality and value of healthcare research are poorly understood, and assessment of healthcare research performance has been largely dismissed as a field of genuine scientific study.(*5*)

The lack of research performance assessment may have supported inequality between clinical and non-clinical research, whilst discouraging translation, technology and innovation, education, and social and global responsibility. (*6-8*) To confront this shortfall, it is imperative that academic healthcare researchers are evaluated to a recognized global benchmark.(*5*) This may drive innovation and excellence in research leading to novel breakthroughs, which can be translated into better-quality healthcare. Presently, there are attempts to appraise overall research produced by higher education institutions through systems such as the Research Excellence Framework (REF) in the UK and Excellence in Research for Australia (ERA).(*1, 9*) It can be difficult to apply these broad systems to healthcare research because of the multifaceted and multidimensional magnitude of the specialty. Therefore it is essential to gain a deeper understanding of the approaches to measuring healthcare research performance before establishing systems that specifically appraise healthcare research performance.

Conventionally, the contribution of healthcare research to knowledge has been gauged using bibliometric indicators such as publication number and citation count in the peer reviewed literature.(*10*) In 2005 Hirsch introduced the h-index to objectively quantify an individual's scientific research output by combining the number of publications and citations in a single number.(*10*) The h index is defined as the number of h papers published by an author that have received at least h citations.(*10*) Hirsch maintains that the h index provides a broad assessment and that it can predict future academic productivity.(*11*) However there are potential limitations of the h-index and its role needs to be determined and validated in academic healthcare setting.(*12*)

Though bibliometric indicators have a role in research performance evaluation, they measure a single facet of performance and maybe limited in the breadth of their applicability. There is an emergent body of research exploring the effects of social capital on performance, where social capital is defined as the network of social connections that exist between individuals in a particular society, enabling that society to function effectively.(*13*) For example, there is a strong conviction among policy-makers and seemingly most scientists that research collaboration has positive outcomes on research productivity.(*14*)

An academic healthcare unit has to deliver high-quality evidence-based clinical care, provide sound medical education at both undergraduate and postgraduate level and execute first class research.(*15, 16*) These challenges, which are not only the responsibility for the Departmental Head but everyone working within the institution, require individuals that have the ability to understand an evolving environment, and to capitalize on changing trends.(*17, 18*) Researchers should recognize that a dense network of interpersonal relationships and internal communications could strengthen knowledge.(*19*)

Although there are prevailing systems in place to measure research performance in general, and a variety of bibliometric indicators have been used to evaluate healthcare research performance, there is a paucity of evidence to demonstrate their validity and a lack of indicators that encompass social behaviours such as collaboration.

In this thesis, we systematically reviewed the literature to identify indicators that have been used to assess academic performance of healthcare research. We evaluated the utility of these indicators in terms of individual, specialty, institutional and global assessment, and determined their feasibility, validity, reliability and acceptability.

There are several citation databases that bibliometric performance measures can be extracted from, so we investigated whether there were any differences in calculations of these performance measures from the mainstream databases, which included Google Scholar, Web of Science and Scopus.

We explored the validity of the Hirsch index (a bibliometric tool which is increasingly used to assess and appraise an individual's research performance) and whether it could be improved to better measure the academic performance and citation profile for individual healthcare researchers.

Intellectual collaboration and authorship attribution play a critical role in successful scientific research. However their impact upon research performance remains poorly quantified, and is seldom appraised in formal academic assessment processes. We applied social network analysis to a co-authorship network within a large healthcare research faculty to examine whether there was a causal relationship between collaborative and authorship patterns of individual researchers and their biomedical research depending on their career stage.

2 How Has Healthcare Research Performance Been Assessed?

2.1 INTRODUCTION

Academic healthcare is the synergy between studying disease mechanisms, identifying new treatments, improving patient care and training healthcare professionals.(*7, 8, 20*) Although the contribution of research to healthcare over the last century has been remarkable, academic healthcare often endures the inequality and lack of transition between basic and clinical research, fails to drive technology and innovation in clinical practice, underrates the role of education, and disregards social and global accountability.(*6-8*)

To tackle these deficits, a system is required for academic healthcare researchers to measure research performance according to an accepted global benchmark, so that innovation and quality of research can be improved and new discoveries can be translated into medical advances.(*5*) Currently, systems such as The Research Assessment Exercise

(UK) and Institutional Assessment Framework (Australia) has attempted to appraise academic research in general.(*21, 22*) The Research Excellence Framework and Excellence in Research for Australia are new developments to assess the quality of research in UK and Australian higher education institutions,(*1, 9*) although currently there are no validated systems to accurately measure performance in healthcare research. This has been difficult to implement due to the operational complexity of the discipline (**Figure 1**).(*5*) To design a system that can successfully measure healthcare research performance, it is imperative to determine which indicators can measure this more accurately.

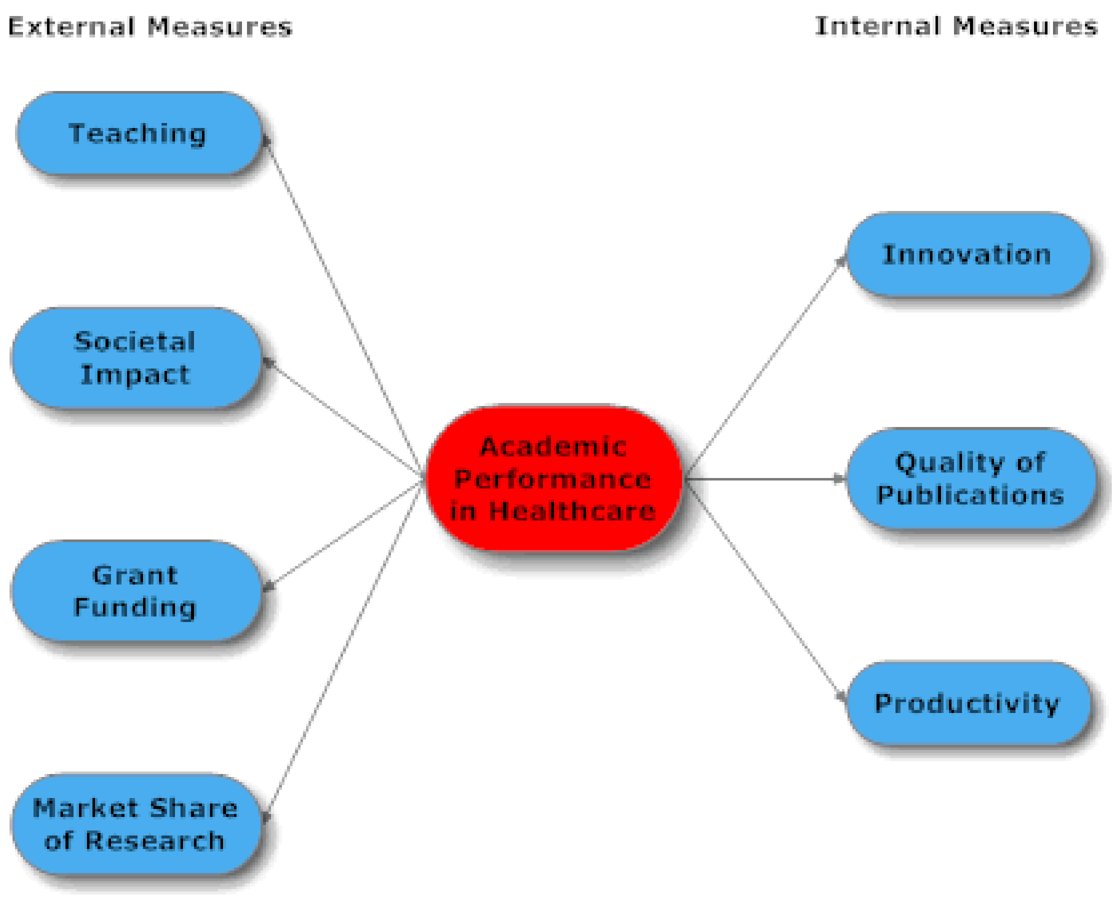

Figure 1: Elements of academic healthcare performance.(*5*)

The objectives of this chapter are to: (i) identify existing indicators which specifically assess healthcare research performance; (ii) assess each indicator to determine its feasibility, validity, reliability and acceptability; and (iii) evaluate the utility of each indicator in terms of individual, specialty, institutional and global perspective.

2.2 METHODS

This study was performed following guidelines from the preferred reporting items for systematic reviews and meta-analyses (PRISMA).(*23*)

2.2.1 Data sources and searches

Studies to be included in the review were identified by searching the following databases: (i) EMBASE (1980 - September 2010), (ii) PsycINFO (1967 – September 2010), (iii) Ovid MEDLINE (1950 – September 2010), and (iv) Cochrane Library.

All databases were searched using the following free text search: "academic OR university OR education OR scientific OR institution" AND "performance OR competence OR quality OR productivity" AND "assessment OR evaluation OR indicator OR peer review" AND "index OR bibliometric OR impact factor OR citation OR benchmark" AND "health care OR medicine OR surgery OR physician OR biomedical OR hospital OR scientist". The search was expanded by using all possible suffix variations of the keywords. Additional studies were identified by searching the bibliographies of the studies that had been identified through the electronic search. A keyword search was chosen rather than Medical Subject Headings (MeSH), because there was a lack of established MeSH terms in this area of research.

2.2.2 Study selection

We included all original studies that evaluated research performance indicators, which measured performance across individuals, specialties, institutions and countries in healthcare. For this study, healthcare was defined as the prevention, treatment, and management of illness and the preservation of mental and physical well-being through the services offered by the medical and allied health professions.(*24*) There were no language restrictions. We excluded all studies that did not have data relevant to healthcare.

Another researcher and I independently reviewed the titles and abstracts of the retrieved articles, and selected publications to be included in this review. We reviewed the full texts of these publications, and selected the relevant articles for inclusion in the review. When there was disagreement, a third researcher was consulted and a decision was made by agreement of all researchers.

2.2.3 Data extraction and quality assessment

Another researcher and I independently extracted data from the full text, which included source of article, study design, study period, type of performance indicator, data source, study population and their sample size, type of statistical analysis, outcomes and methodological limitations. Disagreements in data extraction were resolved by discussion and consensus between all authors. Study quality was assessed using the Oxford Centre for Evidence-based Medicine Levels of Evidence classification.(*25*)

2.2.4 Data synthesis and analysis

The methodology of the included studies was heterogeneous, therefore it was not possible to pool data and statistically analyse the results. The indicators that were identified were analysed in terms of their: (i) utility (the usefulness of indictors at individual, specialty, institutional and global levels); (ii) feasibility (measure of whether the indicator is capable of being used); (iii) validity (measure of the relevance of the indicator: content, convergent and discriminant validity); (iv) reliability (measure of the reproducibility or consistency of an indicator); and (v) acceptability (the extent to which the indicator is accepted by researchers).(*26, 27*)

2.3 RESULTS

2.3.1 Study selection

We retrieved 6705 potentially relevant articles, of which 1185 duplicate articles were identified and excluded. Of the remaining articles, 5385 were excluded after title and abstract review. Review of the full text and bibliography of the remaining 135 articles identified 50 studies for inclusion in the review (**Figure 2**) (**Table 15 - Appendix 1**). The agreement for inclusion of the studies between the authors was satisfactory ($\kappa = 0.86$, $p < .001$).

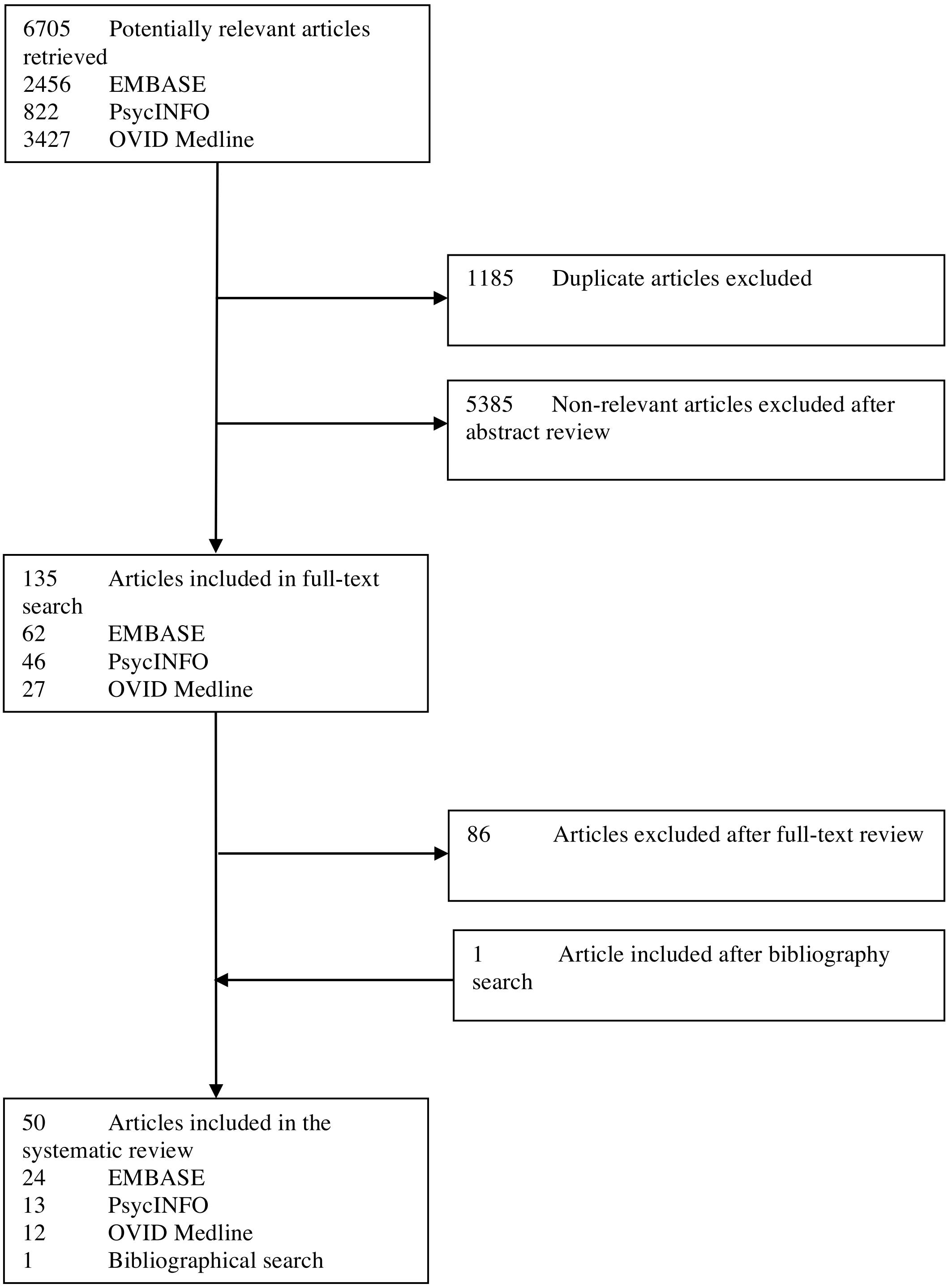

Figure 2: Selection of articles for the systematic review.

2.3.2 Study characteristics

All evidence was level 4 according to the Oxford Centre for Evidence-based Medicine.(*25*)

The plurality of studies were performed in North America (n = 20)(*28-47*) and Western Europe (n = 19).(*48-66*) Fewer studies were performed in Eastern Europe (n =5),(*67-71*) South America (n = 3),(*72-74*) Asia (n = 2)(*75, 76*) and Australia (n = 1).(*77*) The studies were published from 1973 until 2009, but the majority of the studies were published after the millennium (n = 34).(*31, 33, 38, 41, 42, 45-50, 52-54, 56-61, 63-65, 67-77*) The design of each study was retrospective and observational.

Forty-two studies used Thomson Scientific's Institute for Scientific Information database (ISI).(*28-32, 34-44, 46-48, 50, 51, 53-60, 63, 65, 67-70, 72-77*) Out of these, ten studies used one additional database: Scopus (n = 2),(*54, 77*) MEDLINE (n = 5),(*38, 42, 58, 66, 76*) PsycINFO (n = 1),(*40*) National Institutes of Health (NIH) (n = 1)(*41*) and institutional (n = 1)(*72*); three studies used two additional databases: MEDLINE and PsycINFO (n = 1)(*68*), EMBASE AND MEDLINE (n = 1)(*65*) and PsycINFO and NIH (n = 1).(*31*) Out of the studies that did not use ISI, four studies used one database: MEDLINE (n = 2)(*49, 61*) and institutional (n = 2)(*63, 71*); four studies used two databases: institutional and MEDLINE (n = 1),(*52*) Scopus and Spanish Office of Patents and Trademarks (n = 1),(*64*) and NIH and MEDLINE (n = 1),(*33*) Scopus and Google (n = 1).(*45*)

Only seven studies assessed research performance over a lifetime(*45, 54, 66, 68, 72-74*) in comparison to 24 studies assessing research performance over a 1 – 5 year period.(*29, 30, 33-37, 41-44, 49, 50, 57, 59, 61-65, 69, 71, 75, 76*)

The main methodological limitation was the use of a single bibliometric database as the only information source in 32 studies.(*28-30, 32, 34-37, 39, 41, 43, 44, 46-51, 53, 55-57, 59-61, 63, 65-67, 69, 70, 73-75*)

2.3.3 Type of indicators

The types of indicator that were used to measure research performance in each study included number of publications (n = 38),(*28, 30-44, 46-54, 56-58, 61, 62, 65-71, 74*) number of citations (n = 27),(*28-31, 34-37, 40-42, 44, 46-48, 50, 51, 53, 55-57, 59, 65, 66, 69, 71, 77*) Impact Factor (n = 15),(*34, 38, 42, 49, 52, 56, 58, 60, 62, 63, 66, 68, 70, 75, 76*) research funding (n = 10),(*31-33, 41, 43, 44, 46, 47, 49, 70*) degree of co-authorship (n = 9),(*34, 45, 51, 52, 55, 63, 66, 70, 71*) population size (n = 6),(*38, 47, 54, 58, 63, 77*) gross domestic product (n = 5),(*38, 47, 54, 58, 63*) h index (n = 5),(*41, 45, 72-74*) peer review (n = 6),(*46, 48, 49, 57, 65, 66*) g index (n =1),(*45*) Age Weighted Citation Ratio (AWCR) (n = 1),(*45*) number of conference presentations (n = 1),(*42*) number of patents (n = 1),(*64*) number of doctoral students (n = 1),(*31*) number of editorial responsibilities (n = 1)(*31*) and gender (n = 2).(*42, 66*) Twelve studies evaluated one indicator only,(*29, 39, 55, 59-61, 64, 67, 72, 73, 75, 76*) whereas 16 studies evaluated 2 indicators,(*28, 32, 33, 37, 43, 50, 53, 62, 68, 69, 74, 77*) 7 studies evaluated 3 indicators,(*44, 48, 51, 52, 54, 57, 71*) 9 studies evaluated 4 indicators(*34, 41, 45, 46, 48, 49, 56, 58, 70*) and 4 studies evaluated 5 indicators.(*31, 42, 47, 63*)

2.3.3.1 Number of publications

The simplest measure of research productivity in healthcare is the number of published articles a researcher or group of researchers produce within a time span.(*28, 30-44, 46-58, 61-63, 65-71, 74, 77*) This indicator can be presented by document type so that letters, editorials, reviews, and conference papers can be excluded.(*61*) It is relatively easy to calculate using bibliometric databases such as ISI, MEDLINE and Scopus, but these databases will ignore non-journal publications. It can be difficult to retrieve all the publications for certain researchers because of the commonality of names.(*32*) The number of publications does not take into account the size of the research group, the type of research or the quality of the publication. To address this problem, publications per author, population size or publications in top ranked journals can be considered.(*61, 63*) Although the number of publications is commonly used to measure research performance in individuals, specialties, institutions and countries, often as a benchmark, there are no studies formally validating this indicator in healthcare. Despite the lack of validity, a few studies have shown significant correlation between the number of publications and other measures of research performance, such as citations, peer review and research funding.(*33, 44, 46, 48, 49*)

2.3.3.2 Number of citations

The impact of healthcare research can be measured by counting the number of citations received by a researcher or group of researchers from published articles within a time span.(*28-31, 34-37, 40-42, 44, 46-48, 50, 51, 53, 55-57, 59, 65, 66, 69, 71, 77*) Bibliometric databases such as ISI, Scopus, and Google Scholar are required to extract citation counts, which are subject to error because the databases are affected by commonality of names, typographical errors, variation of literature sources and geographical bias.(*59, 77*) Citation analysis assumes that there is a positive association between the citing and referenced article, which does not account for articles that can be cited for negative impact. Citation counts are typically higher in older articles, falsely elevated by self-citations, and can vary between document type and speciality.(*50, 59*) In order to make comparisons across specialties relative citation factor can be used to normalise citation counts.(*50, 59*) As well as specialties, the number of citations has been used to measure research performance in individuals, institutions and countries but there are no studies formally validating this indicator in healthcare. One study, with a small sample size, has demonstrated a low correlation between number of publications and citations.(*36*) However, the majority of studies have shown significant correlation between the number of citations and other measures of research performance, such as publications, co-authorship, peer review and research funding.

2.3.3.3 Impact Factor (IF)

The Journal IF is calculated by dividing the number of current year citations to the source items published in that journal during the previous two years.(*67*) It is an evaluation tool provided by ISI Thomson Reuters Journal Citation Reports® which is used to measure the scientific impact of journals.(*68*) Evaluating research performance using IF can have a marked affect on performance rankings.(*34, 38, 49, 52, 56, 60, 62, 68, 70, 75, 76*) However, the IF is influenced by publication language, document type, citation patterns, open access journals, fast track publications and co-authorship, as well as disregarding publications from zero impact journals.(*75*) More importantly, there is large IF variation between healthcare specialties. For this reason, IF may not reflect quality of research performance, but instead the different publication and citation patterns within specialties.(*75*) Normalizing the IF can provide a more realistic assessment of research quality, which has been demonstrated at an institutional level.(*76*)

2.3.3.4 H index

The h index of a researcher is the number of 'h' publications with at least 'h' citations each during a time span.(*10*) Initially the h index was introduced in physics to address the limitations of publication number, which does not account for research quality, and citation number, which can be disproportionately influenced by a small number of highly cited papers.(*10*) The h index simultaneously evaluates the quality and sustainability of research productivity,(*10*) and can be calculated without difficulty by bibliometric databases such as ISI, Scopus and Google Scholar. In healthcare, the h index has been shown to be a useful statistic to evaluate a researcher's contribution within a given specialty and may even be helpful as a promotional tool.(*41, 45, 72-74*) General drawbacks of bibliometrics, such as commonality of names and publication language are shared by the h index, which is also positively biased to senior researchers with older publications.(*45, 72, 73*)

Indicators such as the g index and AWCR have been proposed to address these limitations, but there is strong correlation between both of these measures and the h index.(*45*) In addition, the h index has been shown to overcome the disadvantages of multiple authorship and self citation.(*45*) There is consensus that the h index cannot be used to measure research performance between different specialties because of diverse publication and citation practices.(*41, 45, 72-74*)

2.3.3.5 Research funding

Research funding is a term covering any financial support for scientific research. This indicator poses an analytical problem, because it is an example of circular cause and effect. Based on bibliometrics, it is difficult to differentiate whether more research funding improves a researcher's performance or if superior performing researchers receive more research funding. Regardless, most of studies show significant correlation between research funding and research performance at an individual and institutional level.(*31, 33, 41, 44, 46, 49, 70*) Developed countries with higher research spend also have higher research productivity.(*32, 47*)

2.3.3.6 Degree of co-authorship

Co-authorship determines the extent a researcher or research group collaborates with others to publish articles. Authors can collaborate at an international, institutional, departmental or individual level. In healthcare several studies have demonstrated that research performance is improved with international collaboration.(*56, 63, 70*) The role of co-authorship at an organizational level has been shown to have a positive impact on performance and has been considered as a novel evaluation tool.(*51, 52*) However, the role of co-authorship at an individual level is uncertain, but indicators such as the h index overcome this potential limitation.(*34, 45, 71*)

2.3.3.7 Gross Domestic Product (GDP) and population size

GDP is a measure of a country's overall economic output and population size is the number of individuals in a region. Adjusting research performance indicators for GDP and population size allows fairer comparison of global performance.(*38, 47, 58*) However, GDP and population size may also be markers of performance in their own right.(*54, 63, 77*)

2.3.3.8 Uncommon indicators

It is difficult to quantify the value of indicators such as peer review, number of conference presentations, number of patents, number of doctoral students, number of editorial responsibilities, and gender because of limited research in these areas.(*31, 42, 46, 48, 49, 57, 64-66*)

2.3.4 Feasibility, validity, reliability and acceptability

Feasibility of using publications, grants, doctoral students and editorial responsibilities to measure research performance was assessed by a survey in one study.(*31*) The respondents generally agreed with the use of these four indicators. Seven studies measured convergent validity by correlating number of publications with number of citations.(*30, 35, 37, 40, 44, 51, 66*) One study demonstrated significant reliability of textbook citations to measure research performance ($p < .001$).(*40*) No other studies assessed research performance indicators in terms of feasibility, validity, reliability and acceptability.

2.3.5 Utility

Twenty-one studies compared research performance between individuals(*28, 33-36, 39, 42, 44-46, 48, 49, 52, 60, 65-69, 71, 74*) and 14 between specialties.(*29, 30, 34, 35, 43, 45, 50, 54-56, 59, 60, 73, 75*) All individuals were researchers in a range of healthcare specialties, and the most common specialties were medicine in general (14 studies)(*29, 32-34, 41, 49, 50, 52-54, 60, 61, 70, 72*) and psychology (11 studies).(*30, 35-37, 39, 40, 51, 55, 57, 68*) Eleven studies compared research performance between institutions, which included universities, national academies and hospitals in the USA, UK, Canada, Australia, New Zealand, France, Germany, Italy, Switzerland, Finland, Serbia, Croatia, Romania, Brazil and Iran.(*37, 39, 41, 51, 56, 62, 67, 68, 70, 73, 76*) Thirteen studies compared research performance between countries, of which 9 studies assessed performance globally and 3 studies assessed performance of the USA with the UK, Europe and Brazil.(*29, 32, 38, 47, 53, 54, 57, 58, 61, 62, 64, 73, 77*)

2.4 DISCUSSION

This is the first systematic review that identifies indicators for assessment of research performance in healthcare. The most widely used indicators include bibliometrics such as number of publications, number of citations and IF, h index, g index and AWCR. Less commonly used indicators include degree of co-authorship, number of conference presentations, number of patents, research funding, number of doctoral students, number of editorial responsibilities, peer review, gender, GDP and population size. The utility of these indicators in assessing research performance in individuals, specialties, institutions and countries has been well described, but their feasibility, validity, reliability and acceptability have not been formally evaluated.

Measuring the number of publications and their citations are simple ways to signify influence. Although they are the most commonly used methods, it is hard to compare them amongst specialties or career stages. However, this shortcoming can be overcome by normalising these indicators to scientific disciplines and experience at both individual and institutional levels.

The h-index considers both the research productivity and its impact, although its use is limited by variations in individuals' age and their discipline. Several other variants of the h-index have been developed to address these drawbacks, for instance the *g*-index, which provides higher scores for increased numbers of citations.

The journal IF should be cautiously used, preferably as an adjunct to other methods, this is because it only considers the impact of journals and does not assess the performance of individual researchers or the impact of their publications.(*78*)

Research funding and degree of co-authorship can be used in addition to the above-mentioned indicators to measure individual, speciality and institutional performance. When measuring performance at global level, GDP and population size should be added to the performance assessment metrics.

Bibliometric research outputs are readily accessible from databases such as ISI, Scopus and Google Scholar. The methods of extracting these outputs should be transparent in all databases so that researchers are able to make an informed decision on the sources of their performance statistics (**Table 1**).

	Web of Science	Scopus	Google Scholar
Date of inauguration	Since early 1960s, but accessible via internet in 2004	11/2004	11/2004
Number of Journals	10969	16500 (> 1200 open access journals)	Not revealed (theoretically all electronic resources)
Language	English (plus 45 other languages)	English (plus more than 30 other languages)	English (plus any language)
Subject coverage	Science, social science and arts and humanities	Science and social science	Not revealed
Period covered	1900-present	1966- present	Not revealed
Updating	Weekly	1-2/week	Monthly
Developer	Thompson Scientific (US)	Elsevier (Netherlands)	Google Inc. (US)
Fee-based	Yes	Yes	No
H index calculation	Yes	Yes	Only using Harzing's Publish or Perish software

Table 1: Characteristics and differences between the citation indexing databases. (*79-84*)

A universally accepted framework needs agreement by the decision-makers in academia to standardise research outputs, so that the economic and societal impact of research can be measured. A recent example includes the STAR METRICS working group in the United States (Science and Technology in America's Reinvestment – Measuring the EffecT of Research on Innovation, Competitiveness and Science) who are developing a common empirical infrastructure.(*85*)

There are several limitations at a study and review level. Studies will be biased when authors evaluate their own performance or the performance of their affiliated specialties, institutions or countries. There is different coverage of peer-reviewed publications between bibliometric databases, so a source level bias will exist in studies that use a single data source. This systematic review was limited by the poor quality of the studies. In addition, meta-analysis could not be performed because of the diversity of the studies, which did not have homogeneous methods or results.

This study has several implications: (i) Further studies are needed to determine the feasibility, validity, reliability and acceptability of current and future research performance indicators; (ii) Specifically, it is important to assess the value of the h index because it measures the importance, broad impact consistency and sustainability of a scientist's research; (iii) Co-authorship networks and changes in collaboration patterns over time should be analysed to establish whether they are important tools to assess and develop research performance; (iv) The use of the IF to evaluate a researcher's performance needs to be investigated, since the IF has only been designed to measure journal performance; (v) Researchers and policy makers can then debate what role the indicators should play, both in terms of the weighting and the level they should be incorporated into the decision making process; (vi) The balanced scorecard is a performance measurement framework that adds strategic non-financial performance measures to traditional financial metrics (**Figure 3**).(*86*) Although designed for business and industry, the balanced scorecard can be modified for non profit and non manufacturing research institutions.(*87*) This approach needs to be adapted by institutions to present a more unbiased view of research performance. This multifaceted method of research performance evaluation will require a multidimensional model of analysis utilising a broad range of robust analytical techniques(*21*); (vii) Enhanced healthcare research indices should be translated into improved healthcare outcomes because the principal aim of healthcare research is to improve patient wellbeing. It is now imperative to consider healthcare outcomes as opposed to research outputs. The use of healthcare outcomes can then determine important factors such as the societal and economic impact of healthcare research.(*88-92*)

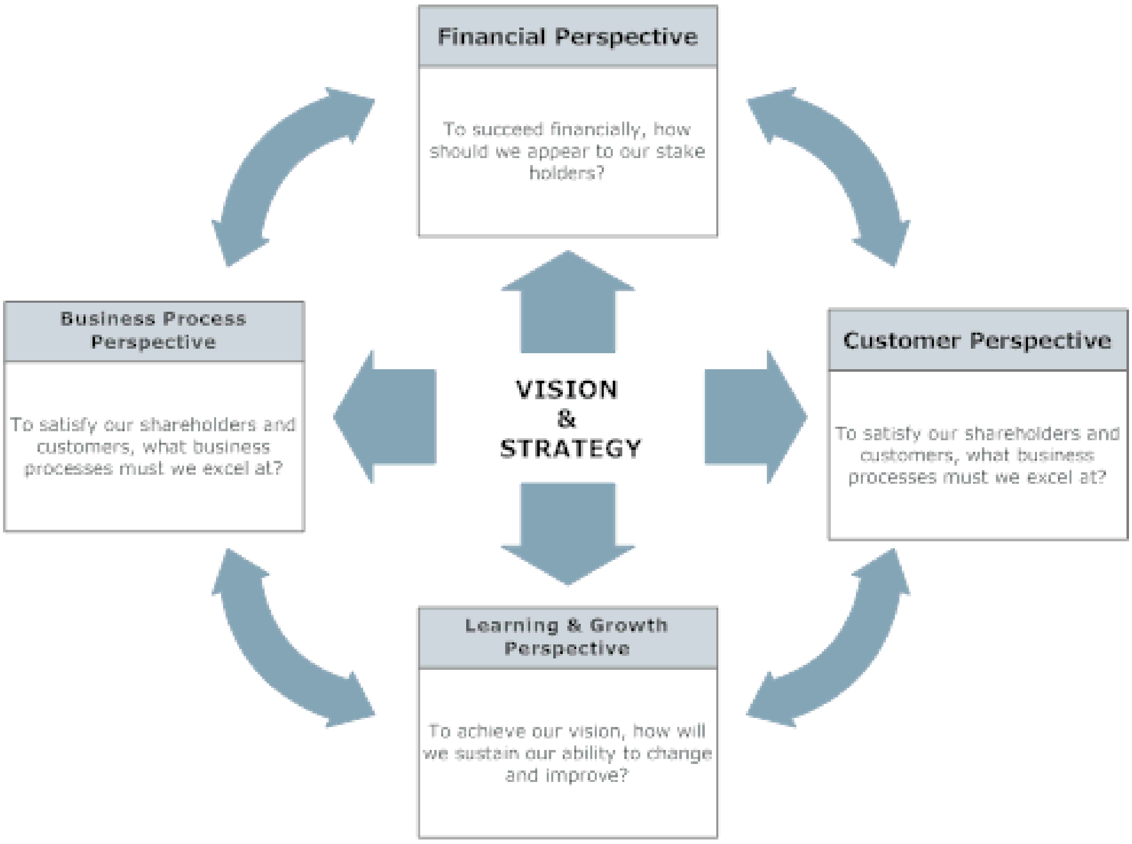

Figure 3: Balanced scorecard showing performance areas of an organization.(*86*)

2.5 CONCLUSION

Recently, there has been greater awareness of the importance of research performance indicators in healthcare. As a result the prevalence and usage of metrics such as number of publications, number of citations, IF and h index has increased. However, the assessment of feasibility, validity, reliability and acceptability of these indicators has been poorly investigated. Future studies are required to improve the current standards and accuracy of performance evaluation. It is imperative to have a balanced approach when measuring research performance in healthcare, which should consider quality and innovation. There is an increased need to consider the role of healthcare research outcomes in achieving societal and economic impact. The ultimate aim is to accurately quantify the research performance of healthcare individuals and institutions to cultivate an environment that can support translational medicine to improve the quality of patient care.

2.6 CHAPTER SUMMARY

The objectives of this chapter were firstly to identify indicators that have been used to assess healthcare research performance; secondly to evaluate the utility of these indicators in terms of individual, specialty, institutional and global assessment, and finally to determine their feasibility, validity, reliability and acceptability.

We systematically reviewed relevant studies published in the English language, which were identified by searching EMBASE, PsycINFO, Ovid MEDLINE and Cochrane Database of Systematic Reviews until April 2010. The initial search identified 6705 potentially relevant articles. We included 50 studies that evaluated academic performance indicators, which measured performance across individuals, specialties, institutions and countries in healthcare.

We extracted source of article, study design, study period, type of performance indicator, data source, study population and their sample size, type of statistical analysis, outcomes and methodological limitations.

Most studies were performed in North America (n = 20) and Western Europe (n = 19). The majority of the studies used Thomson Scientific's Institute for Scientific Information database (n = 42). The most common academic performance indicators were: number of publications (n = 38), number of citations (n = 27), impact factor (n = 15), research funding (n = 10), and degree of co-authorship (n = 9). Furthermore, the h index was acknowledged as an emerging bibliometric indicator. The utility of these indicators was adequately described but there was limited investigation of feasibility, validity, reliability and acceptability.

Therefore in the next chapter we focused on investigating the reliability of bibliometric indicators with particular emphasis on the h index.

3 MEASURING HEALTHCARE RESEARCH PERFORMANCE WITH THE H INDEX: WHICH SEARCH TOOL SHOULD BE USED?

3.1 INTRODUCTION

Research performance has traditionally been evaluated through bibliometrics that included the total number of publications and the total number of citations.(*93*) Total paper counts do not reflect research quality and citation numbers do not provide an accurate account of research breadth. As a result, there is an accepted need to utilise improved markers of research performance to quantify research excellence with increased precision and objectivity.(*5*) In 2005, Hirsch proposed the 'h index', which measures the importance, significance, and broad impact of a scientist's cumulative research contributions. A scientist with an index of h has published h number of papers, each of which has been cited by others at least h times.(*10*) The use of the h index to measure research performance is rapidly increasing, although the use of the h index has not been previously validated for medical researchers. The h index has been criticized for the technical shortcomings associated with self-citations, field dependency and multiple authorship (these are common to the majority of current bibliometrics).(*94, 95*) Specifically, an h index value taken as an isolated measure has been criticized for yielding an unreliable interpretation of the research performance of an individual and may not be fully reflective of those publications with the greatest impact.(*94, 95*)

The strongest indication of its acceptability is that its calculation has been integrated into the citation databases of Web of Science, Scopus and Google Scholar.(*80, 81, 83, 84, 96*) There are differences in the scope of these databases (**Table 1**)(*82*), which use disparate systems to count citations.(*93*) The databases have been shown to produce quantitatively and qualitatively different citation counts for general medical research.(*93*)

A number of differences in h index calculation between databases have been reported when evaluating researchers' performance in the fields of computing science and mathematics.(*97, 98*) It is unclear whether the same h index calculation discrepancy exists when assessing medical researchers. Differences in h index when calculated across different databases may have implications if used in assessing a medical researcher's performance or for decisions on academic promotion.

The aim of this study was to compare the results of the most popular bibliometric databases for a universally accepted cohort of medical scientists in terms of their h index scores.

3.2 METHODS

The cohort of Nobel Laureates in Physiology and Medicine (who represent healthcare research excellence and quality) from 1901 to 2009 was chosen to investigate whether or not the h index of medical scientists varies between Web of Science, Scopus and Google Scholar. The same cohort was used to determine whether a scientist's age, country of institutional affiliation at the time of the award, and whether the Laureate was a physician or not (physician status) had an influence on the calculation of the h index by each database.

A list of all the Nobel Laureates in Physiology and Medicine (1901 to 2009) was obtained by searching the official website of the Nobel Prize (www.nobelprize.org). These individuals were selected because not only are they high scientific achievers in medical research, but also they are representative of global medical science in terms of age, country of institutional affiliation at the time of the award, and physician status. The full names, date of birth, date of bereavement, country of institutional affiliation at the time of the award, and physician status for each Laureate were extracted by reading their biography from the official website. These specific parameters were chosen to evaluate any influence such demographics may have upon the determination of bibliometric indices. For example, it is known that a disproportionately higher number of Laureates originate from the United States and Europe; however it is not known whether this factor predicts higher bibliometric scores. The same rationale was applied to the other demographics.

Another researcher and I calculated the year of first publication, year of last publication, total number of publications, total number of citations and h index for each Laureate from

Google Scholar, Web of Science, and Scopus. Publications after 31st December 2009 were excluded, and all data were collected within 7 days from all 3 citation databases. Another researcher independently crosschecked the data to confirm accuracy of data collection. For Laureates that had deceased before 31st December 2009, publications after the date of bereavement were excluded. The specific methodology for bibliometric extraction from each database is outlined below.

3.2.1 Google Scholar

Harzing's Publish or Perish software, which analyses raw citations from Google Scholar, was used to calculate a series of bibliometrics.(*81*) The quoted initial and surname were inputted in the Author Impact Analysis field to generate a list of publications authored by a specific Laureate. The search was not restricted to any specialty. Publications from the list that were not authored by the specific Laureate were deselected. The bibliometrics were extracted from the results field (**Figure 13 - Appendix 2**).

3.2.2 Web of Science

The surname and initial were inputted to search for publications authored by a specific Laureate. The Distinct Author Set feature uses citation data to create sets of articles likely written by the same person. This feature was used as a tool to focus the search to compile a list of all the publications for each Laureate. The bibliometrics were extracted by creating a citation report from this list (**Figure 13 -Appendix 2**).

3.2.3 Scopus

The surname and initials or first name were inputted to search for publications authored by a specific Laureate. The Scopus Author Identifier uses an algorithm that matches author names based on their affiliation, address, subject area and source title, dates of publication, citations and co-authors. This identifier was used as a tool to focus the search to compile a list of all the publications for each Laureate. The bibliometrics were extracted by viewing the citation overview of this list (**Figure 13 - Appendix 2**).

3.3 STATISTICAL ANALYSIS

One-way analysis of variance (ANOVA) was used to assess differences of bibliometrics between the databases. Cronbach's alpha statistic was used to measure the reliability of bibliometrics between the databases. Multivariate linear regression analysis was used to explore whether the h index from each database was influenced by a scientist's age, country of institutional affiliation at the time of the award, and physician status. Data were analysed with the use of SPSS for Windows (Rel. 18.0.0. 2009. Chicago: SPSS Inc.).

3.4 RESULTS

There were 101 Nobel Prizes in Physiology and Medicine awarded to 195 Laureates between 1901 and 2009, with a median age of 56 years (32 - 87) at the time of the award; 10 Laureates were female and 125 Laureates were deceased. Of the 101 Nobel Prizes, 37(36.6%) were given to one Laureate only, 31(30.7%) were shared by two Laureates and 32(31.7%) were shared between three Laureates. At the time of the award 92(47.2%) Laureates were affiliated to North American countries, 92(47.2%) Laureates were affiliated to European countries, and 11(5.6%) Laureates were affiliated to other countries (**Table 2**). Ninety-nine (99, 50.8%) Laureates were non-physicians and 96(49.2%) were physicians. Three Laureates (prize winning years 1948, 1986 and 2007) were excluded from the analysis because accurate data could not be retrieved due to commonality of names. Data were not available for 68 Laureates in Web of Science (prize winning years ranging from 1901- 1976) and 29 Laureates in Scopus (prize winning years ranging from 1901- 1985) (**Table 3**).

	Number of Laureates (N = 195)
North America	
USA	90
Canada	2
Europe	
UK	31
Germany	16
France	10
Sweden	8
Switzerland	6
Austria	5
Denmark	5
Belgium	4
Italy	2
Netherlands	2
Hungary	1
Portugal	1
Spain	1
Others	
Australia	6
Russia	2
Argentina	1
Japan	1
South Africa	1

Table 2: Country of affiliation at time of award.

	James Watson	Francis Crick	Maurice Wilkins
Nobel Prize year	1962	1962	1962
Birth year	1928	1916	1916
Country of affiliation at time of award	USA	UK	UK
Physician status	Non-physician	Non-physician	Non-physician
Google Scholar			
Year of first publication	1948	1945	1934
Year of last publication	2009	2004	2003
Total number of publications	984	180	91
Total number of citations	23551	10180	2247
H index	57	32	23
Web of Science			
Year of first publication	1972	1970	1970
Year of last publication	2008	2007	1995
Total number of publications	161	25	12
Total number of citations	5567	3336	1091
H index	36	15	8
Scopus			
Year of first publication	1969	1950	1948
Year of last publication	2008	2007	1995
Total number of publications	21	87	38
Total number of citations	994	6562	244
H index	6	33	8

Table 3: Example of bibliometric outcomes of Nobel Laureates in Physiology and Medicine.

The median year of first publication was significantly different between the 3 databases with the lowest year in Google Scholar (1933; 1843-1989) compared to Scopus (1957; 1880-1996) and Web of Science (1970; 1968-1991) ($p < 0.0001$ for all comparisons). The median year of last publication was not significantly different in Google Scholar (1997; 1903-2009) and Scopus (2000; 1880-2009) ($p = 0.408$), but was significantly higher in Web of Science (2008; 1970-2009) ($p < 0.0001$ for both comparisons). The median total number of publications was significantly different between the 3 databases with the highest in Google Scholar (236; 3-1000) compared to Web of Science (109; 1-1255) and Scopus (67; 1-992) ($p < 0.0001$ for all comparisons). The median total number of citations was significantly different between the 3 databases with the highest in Web of Science

(10579; 2-95460) compared to Google Scholar (6521; 4-97988) and Scopus (991; 0-41470) ($p = 0.029$ Web of Science versus Google Scholar, $p < 0.0001$ for other comparisons). The median h index was not significantly different between Google Scholar (35; 1-166) and Web of Science (43; 1-161) ($p = 0.066$), but was significantly lower in Scopus (13; 0-111) ($p < 0.0001$ for both comparisons).

The h index was the most reliably calculated bibliometric across the 3 databases (Cronbach's Alpha = 0.90) (**Table 4**). This reliability was greater between Web of Science and Scopus (Cronbach's Alpha = 0.91), than Google Scholar and Scopus (Cronbach's Alpha = 0.85) or Google Scholar and Web of Science (Cronbach's Alpha = 0.82).

	Google Scholar	**Web of Science**	**Scopus**	**Cronbach's Alpha**
Total Nobel Laureates n = 195				
Data available	192	127	166	
Data not retrievable/available	3	68	29	
Bibliometrics, median (range)				
Year of first publication	1933 (1843-1989)	1970 (1968-1991)	1957 (1880-1996)	0.670
Year of last publication	1997 (1903-2009)	2008 (1970-2009)	2000 (1880-2009)	0.832
Total number of publications	236 (3-1260)	109 (1-1255)	67 (1-992)	0.759
Total number of citations	6521 (4-97988)	10579 (2-95460)	991 (0-41470)	0.789
H index	35 (1-161)	43 (1-166)	13 (0-111)	0.900

Table 4: Reliability of bibliometrics from Google Scholar, Web of Science and Scopus.

Univariate regression analysis demonstrated that younger Laureates were more likely to have significantly higher h indexes returned by each of the 3 databases ($p < 0.0001$) (**Table 5**). The h index from Google Scholar was higher in Laureates from North America in comparison to Laureates from European and other countries respectively ($p < 0.0001$). Physicians had a lower h index than non-physicians in Scopus ($p = 0.018$).

Multivariate regression model (**Table 5**) showed that the impact of birth year on h index remained significant. There was no significant influence of country of institutional affiliation at time of award on h index in any database. The non-physician bias of Scopus became non-significant, but physicians had higher h index in Google Scholar and Web of Science ($p = 0.025$ and $p = 0.029$ respectively).

Characteristic	Univariate analysis[a]	*p* value	Multivariate analysis[a,b]	*p* value
Birth year				
Google Scholar	0.616 (0.500-0.731)	< 0.0001	0.561 (0.449-0.672)	< 0.0001
Web of Science	0.273 (0.209-0.336)	< 0.0001	0.280 (0.218-0.343)	< 0.0001
Scopus	0.643 (0.542-0.744)	< 0.0001	0.594 (0.497-0.690)	< 0.0001
Country of affiliation at time of award				
Google Scholar	-0.006 (-0.009--0.003)	< 0.0001	-0.004 (-0.007-0.000)	0.067
Web of Science	-0.001 (-0.004-0.001)	0.352	-0.002 (-0.06-0.001)	0.159
Scopus	-0.003 (-0.007-0.000)	0.053	-0.002 (-0.007-0.003)	0.467
Physician status				
Google Scholar	-0.002 (-0.004-0.001)	0.223	0.003 (0.000-0.007)	0.025
Web of Science	0.001 (-0.001-0.003)	0.521	0.003 (0.000-0.006)	0.029
Scopus	-0.003 (-0.006--0.001)	0.018	0.003 (0.000-0.007)	0.079

[a]Values are expressed as unstandardized B coefficients (95% confidence intervals)
[b]Variables included birth year, country of affiliation at time of award and physician status

Table 5: Regression analyses of Laureate characteristics and h index values from Google Scholar, Web of Science and Scopus.

3.5 DISCUSSION

The h index was more consistently calculated between the databases than other bibliometrics. However, in general, it is important to consider which database to use for computing the h index for a variety of reasons. As demonstrated by our study, h index scores are influenced by age and physician status. Resultantly, individual scores may be discordant between databases as each covers a different time span and list of journals. The scope of each database must be transparent including algorithms used such that researchers are able to make informed decisions on which source to base their performance metrics on. This study showed that physician's status could constitute an important component of objective research assessment in healthcare.

There appeared to be a paucity of Laureates with h index scores of greater than 60 before circa 1900 as demonstrated in **Figure 4**. This could be due to several factors: firstly, there could have been fewer publications in circulation during this period, hence probability of being citation. Secondly, with the advent of electronic publishing of manuscripts in the contemporary era, manuscripts receive wider audiences and thus are more likely to be cited. Hence finally, what was represented is the degree to which journals indexed their historic manuscripts.

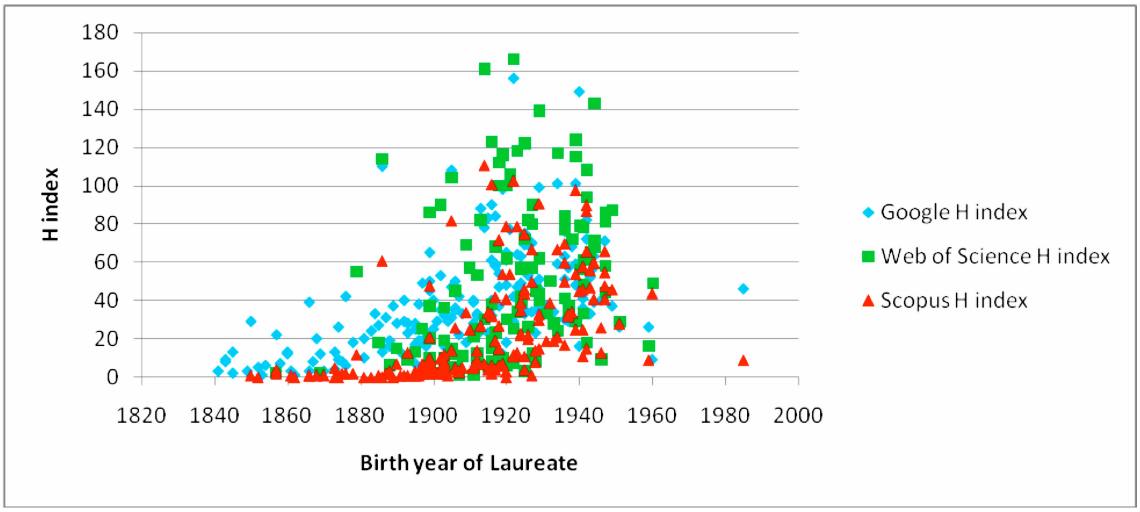

Figure 4: Scatter plot of birth year of Laureate versus their h index.

The use of the h index in Nobel Laureates reflects the application of this bibliometric tool as a valuable measure of research excellence and quality. The future use of the h index could include its application in assessment, promotion and support of healthcare scientists.

The strengths of this study included using a large cohort of researchers who had been selected by peer review for receipt of the Nobel Prize for their achievements in medical research. The broad range of birth year of the Laureates allowed the impact of the time period covered by each database on the h index to be investigated. Most of the Laureates were equally affiliated at the time of the award to either North American or European countries, which strengthened the evaluation of geographical bias of the databases in relation to the h index. There were similar numbers of physician and non-physician Laureates, so the physician bias of the databases in relation to the h index could be assessed. The methodology of data collection was robust given the agreement of two independent authors with the retrieval of the results and furthermore the short information retrieval period occurring within one week.

Potential limitations of this study include difficulty in distinguishing between articles belonging to authors with similar names. Web of Science and Scopus use identifiers to group an author's publications together, and Scopus claims to have achieved 99% certainty for 95% of its records.(*83, 84, 99*) These individual author sets were used to focus our search for a Laureate's publications to ensure that data collection was comprehensive. Harzing's Publish or Perish software does not have an author identifier feature, which made data extraction from Google Scholar more difficult and time consuming. The reliability of Google Scholar is unknown because the coverage and methods of the database are not transparent.(*100*) This study did not take into account the age at which the Nobel Prize was awarded, and the follow-up citations after the Nobel Prize award may also have affected final citation and h index results. When considering the significant influence of birth year of the Laureate on h index, it was not possible to determine whether publications of older Laureates had a longer time to get cited or whether publications of younger Laureates were more likely to get cited because a greater proportion of their work was more accessible. The language bias of native English speakers in relation to the h index was not tested. The results of this study cannot be translated to other scientific disciplines because the value of the h index is discipline dependent.(*94*) The overlap of publications and citations between the databases was not considered, so it was not possible to determine the degree to which the h index increases by combining the unique publications and citations from each database.(*79, 97, 101*)

3.6 CONCLUSION

Google Scholar and Web of Science returned significantly higher h index scores for Laureates in comparison to Scopus. However, h index was the most reliably calculated bibliometric across the 3 databases. Researcher-specific characteristics should be considered when calculating h index from citation databases.

3.7 CHAPTER SUMMARY

The objectives of this chapter were to compare bibliometric indicators for healthcare researchers returned by: Google Scholar, Web of Science and Scopus databases and to assess whether a researcher's age, country of institutional affiliation and physician status influences the calculations.

One hundred and ninety five (195) Nobel Laureates in Physiology and Medicine from 1901-2009 were considered. Year of first and last publications, total publications and citation counts, and h index for each Laureate were calculated from each database. Cronbach's alpha statistics was used to measure reliability of h index scores between the databases. Laureate characteristic influence on the h index was analysed using linear regression.

There was no concordance between the databases when considering the number of publications and citation count per Laureate. The h index was the most reliably calculated bibliometric across the 3 databases (Cronbach's Alpha=0.900). All databases returned significantly higher h index scores for younger Laureates ($p<0.0001$). Google Scholar and Web of Science returned significantly higher h index for physician Laureates ($p=0.025$ and $p=0.029$ respectively). Country of institutional affiliation did not influence h index in any database.

The h index appeared to be the most consistently calculated bibliometric between the databases for Nobel Laureates in Physiology and Medicine. Researcher-specific characteristics constituted an important component of objective research assessment.

Based on the findings of this study, we decided to use the h index to explore the validity and feasibility of bibliometric indicators in the next chapter. Because we demonstrated physician bias in Google Scholar and Web of Science, we decided to use Scopus to calculate bibliometric data in the subsequent studies.

4 ENHANCING THE H INDEX FOR THE OBJECTIVE ASSESSMENT OF HEALTHCARE RESEARCHER PERFORMANCE

4.1 INTRODUCTION

Academic healthcare institutions employ various systems to quantify research performance. This may enhance research quality and innovation so that the latest discoveries can be translated into medical advances.(*5*, *82*) The simplest measure of research performance is the number of publications a researcher produces; however this bibliometric fails to recognise the quality of a researcher's performance.(*82*) The number of citations received by a researcher's publications may determine the scientific impact of his or her research, however this may be distorted by self-citations and negatively associated citations.(*102, 103*)

The h index is an alternative bibliometric designed to measure the importance, significance, and broad impact of a scientist's cumulative research contributions.(*10*) A researcher has index h if h of his or her publications (Np) have at least h citations each and the other (Np – h) publications have ≤h citations each.*(10)* For example, an academic with an h index of 20 must have 20 publications with at least 20 citations each. Evaluating research performance using the h index is rapidly increasing. The strongest indication of its acceptability is that its calculation has been integrated into the contemporary citation databases of Web of Science and Scopus.(*83, 84, 104*)

The h index is frequently criticised for the problems of self-citations, field dependency and multiple authorship, which are common to the majority of current bibliometrics.(*105*) If researchers are evaluated only on the basis of the h index, they may be unfairly treated mainly due to three reasons. Firstly, the h index does not fully consider highly cited papers, so that the impact of a researcher with a low h index and a few very highly cited papers will be undervalued. Secondly, the h index does not fully consider productivity of a researcher, so that publications with citations less than the h index are not included. Thirdly, the h index draws a sharp distinction between publications that have higher impact (publications with citations equal to the h index or more) and publications with less importance (for example, the h index assumes that in a researcher with an h index of 20 a publication with 19 citations is not important). In addition to the h index a more reasonable criterion is necessary which separate papers with visible impact from publications with less or no impact.(*105*)

Recently, a small number of indicators were developed in order to solve these three limitations. The first two limitations are addressed by specifying the percentage of citations that fall into the high, core and low visibility areas of a researcher's publications.

The third limitation is addressed by a special form of regression which determines the threshold of publications with visible impact from publications with less or no impact.(*105*)

At present the assessment and appraisal of academic healthcare researchers relies on a peer review process to judge an individual's research performance, clinical commitment and educational contribution. These domains are further analysed to identify an individual's impact and external visibility.(*15, 106*) The h index has been considered as an objective tool that can assess research performance through citation analysis, although it has not been formally validated in healthcare researchers.(*82*)

In this study we aimed to: (i) quantify the high, core and low visibility publications of individual researchers within the Faculty of Medicine at Imperial College London; (ii) statistically estimate the number of high visibility publications of each researcher within our group by using a segmented regression model (sRM); (iii) demonstrate the construct validity of the h index and our proposed adjuncts by comparing them with a qualitative peer-reviewed process of academic promotion in healthcare research; and (iv) determine the convergent validity of the h index and our proposed adjuncts by correlating them with conventional bibliometric indicators.

4.2 METHODS

4.2.1 Sample population

The Faculty of Medicine Imperial College London was established in 1997, and is one of Europe's largest medical institutions in terms of its staff and student population and its research income.(*107*) It incorporates eight campuses in North and West London, and is divided into six divisions: Clinical Sciences, Kennedy Institute of Rheumatology, Medicine, National Heart and Lung Institute, School of Public Health, and Surgery and Cancer.(*107*) We generated a list of all Faculty of Medicine Imperial College academics in employment on the 31st December 2009 from the intranet database. We included academics that were ranked in the hierarchical system in the following grades: lecturer, senior lecturer, reader and professor. Academics can be promoted from lecturer to senior lecturer to reader and then to professor. Each promotion has to be approved by an academic promotions committee, which takes into account contributions to the following: education, research, leadership, management, professional and clinical practice.(*107*) We excluded academics that were research fellows, associates, officers and assistants. We extracted the first name, surname, academic rank, physician status (whether the academic was a physician or not) and department for each of the academics included in the study.

4.2.2 Generating an academic's publication list

We used SciVerse Scopus to evaluate research output because it offers sophisticated tools to search, analyse, and visualize research, as well as export to citation management software (**Table 6**).(*84, 108*)

Launched November 2004
More than 18,000 titles: 17,000 peer-reviewed journals (including 1,200 Open Access journals) 600 trade publications 350 book series 3.7 million conference papers from proceedings and journals
42.5 million records: 22 million records with references back to 1996 20.5 million records pre-1996 which go back as far as 1823
359 million scientific web pages indexed via Scirus (science-specific internet search engine)
24 million patent records from five patent offices
"Articles-in-Press" from over 3,500 journals

Table 6: Content coverage of SciVerse Scopus.(*84, 104*)

The Scopus Author Identifier uses an algorithm that creates a publication set for an author based on their affiliation, address, subject area and source title, dates of publication, citations and co-authors. We used the surname and initials to search for the publication set authored by each academic. When more than one publication set was identified for the academic's name, we viewed the titles and abstracts of each publication set to determine which sets should be combined. After we selected all the relevant publication sets, we used the 'Show documents' button to generate a list of the academic's publications, with their corresponding citation counts. We selected publications from this list that were published between 1st January 2000 until 31st December 2009, and we exported the list to Microsoft Excel. We sorted the publication list in order of descending citations, and calculated the total number of publications, total number of citations and h index.

4.2.3 Calculating h^2 upper, h^2 centre, and h^2 lower

We differentiated the geometrical area of the citation count into the high visibility area (h^2 upper), the core visibility area (h^2 centre), and the low visibility area (h^2 lower). **Figure 5a** shows a schematic curve of number of citations versus publications ranked in order of decreasing citations. The h index refers to the area h×h (h^2 centre), which does not take into consideration the areas starting at h citations (h^2 upper) or starting at h publications (h^2 lower).

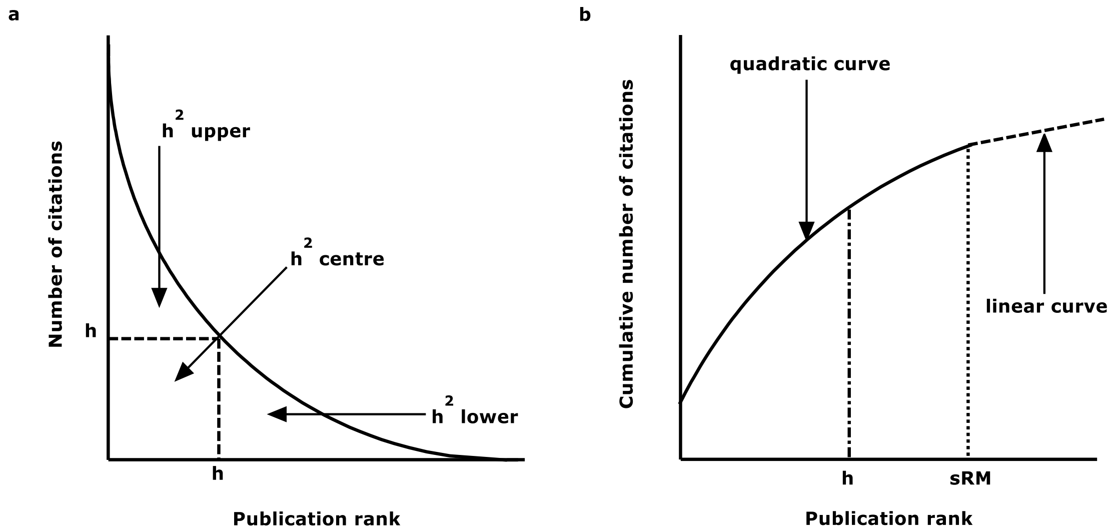

Figure 5: (a) Schematic curve of number of citations versus publication rank showing h^2 upper, h^2 centre, and h^2 lower, and (b) schematic curve of cumulative number of citations versus publication rank showing sRM value.(*105*)

We calculated h^2 upper, h^2 centre, and h^2 lower as a percentage of total citation counts using the following equations:

$$h^2 \text{ upper} = \frac{\sum_{j=1}^{h} \left(cit_j - h\right)}{\sum_{j=1}^{n} \left(cit_j\right)} \times 100$$

$$h^2 \text{ centre} = \frac{h \times h}{\sum_{j=1}^{n} \left(cit_j\right)} \times 100$$

$$h^2 \text{ lower} = \frac{\sum_{j=h+1}^{n} cit_j}{\sum_{j=1}^{n} cit_j} \times 100$$

The equations demonstrate that h^2 upper was calculated by subtracting h×h from the sum of citations of publications with citation counts greater than an academic's h index value. The sum of h^2 upper and h^2 centre as a percentage of total citation counts was subtracted from 100 to calculate h^2 lower.

4.2.4 Calculating sRM value

We defined the statistical estimate of the number of high visibility publications as the sRM value. **Figure 5b** shows a schematic curve of cumulative citations versus publications ranked in order of decreasing citations. The quadratic curve in the first part depicts those publications with high citations that contribute most to the total citation count of a researcher ('core' publications with high visibility). The linear curve in the second part illustrates those publications that have little contribution to the total citation count (publications with low visibility). The h index underestimates the point of intersection of the two curves (**Figure 14 - Appendix 3**), so we used segmented regression to statistically model this point to obtain the sRM value. The following model for the cumulative citation counts y_j was assumed, whereby x_j is the rank of the publication j and z_0 was the point of intersection of the two curves: if $x_j < z_0$

$$y_j = b_0 + b_1 x_j + b_2 x^2{}_j + e_j\,, \quad e_j \sim N(0, \sigma^2{}_e)$$

otherwise,

$$y_j = b_0 + b_1 z_0 + b_2 z^2{}_0 + b_3\left(x_j - z_0\right) + e_j\,, \quad e_j \sim N(0, \sigma^2{}_e)$$

The x values for j ranged from 1 to k. The z_0 value was an unknown parameter that was defined as the maximum of the quadratic function:

$$z_0 = \frac{-b_1}{2b_2}$$

The size of the residual variance ($\sigma^2{}_e$), or the proportion of explained variance to total variance (R^2), can quantify the model fit. The sRM value was calculated for an academic if their publication set met the following criteria: (i) the citation count could be clearly distinguished into two parts, (ii) the algorithm converged, (iii) R^2 was greater than 0.90, (iv) the point of intersection of the curves lay within the range of publications, and (v) there were at least 15–20 publications in the set.

4.3 Statistical analysis

We statistically analyzed for construct validity, as well as convergent validity: The bibliometric indicators were construct valid, if they showed significantly mean differences according to the hierarchy of academic rank (lecturer, senior lecturer, reader and professor). The bibliometric indicators were convergent valid, if they showed at least low to moderate correlations amongst the set of indicators, as well as low to moderate correlations with the h index. Bibliometric data as a rule and the calculated bibliometric indicators are not normally distributed, so we used Box-Cox power transformation to normalize the data.(*109*) Bibliometric data often have extreme values or outliers (e.g., highly cited publications), which might distort the variance. To adjust for extreme values we weighted the data: extreme values were assigned a smaller weight than the other values (M-estimator). Weights were estimated using Robust Regression.(*110, 111*) We used two-factorial multivariate analysis of variance (MANOVA) on unweighted power transformed data followed by a two-factorial ANOVA based on weighted power transformed data to analyze the differences between bibliometric indicators and academic rank or physician status.(*112*) We also applied post hoc Scheffe's test to explore significant differences between the groups. As the numbers of observations in each group were unequal, we employed Type III sums of squares. We adjusted for the year of first publication of an academic within the time frame of this study by using covariance adjustment. We also used Pearson-correlation coefficient to calculate correlations on the power transformed data. We used SAS® 9.2 for all analyses. (SAS Institute Inc., SAS® 9.2 Enhanced Logging Facilities, Cary, NC: SAS Institute Inc., 2008.)

4.4 RESULTS

The bibliometric data for 501 academics in the Faculty of Medicine were analyzed. The cohort was unbalanced because the academic ranks were heterogeneous in terms of the number of academics and physician status ($\chi^2(3)=48.1$ $p<0.05$, Cramer`s V=0.31); professors and senior lecturers were more likely to be physicians than lecturers and readers (**Table 7**). Out of the 501 academics, 398 (79.4%) had at least one publication in the year 2001. The criteria for the calculation of sRM value were fulfilled in 438 academics. The h^2 upper values were not power transformed because they were approximately normally distributed.

	Physician status				
	Non-physician		Physician		Total n (%)
Academic Rank	Observed n (%)	Expected n (%)	Observed n (%)	Expected n (%)	
Lecturer	84 (31.8)	57 (21.6)	25 (10.6)	52 (21.9)	109 (21.8)
Senior Lecturer	44 (16.7)	68 (25.8)	84 (35.4)	60 (25.3)	128 (25.5)
Reader	40 (15.1)	33 (12.5)	23 (9.70)	30 (12.7)	63 (12.6)
Professor	96 (36.4)	106 (40.2)	105 (44.3)	95 (40.1)	201 (40.1)
Total n (%)	264 (100.0)	264 (100.0)	237 (100.0)	237 (100.0)	501(100.0)
Row percent	52.7	52.7	47.3	47.3	100

Table 7: Composition of the academics in the Faculty of Medicine. The expected value demonstrates the relationship between academic rank and physician status in terms of over and under representation.

In the MANOVA, construct validity was demonstrated for total number of publications, total number of citations, h index, h^2 upper, h^2 centre, h^2 lower, and sRM value, because there were statistically significant differences between the different academic ranks (Wilks` Lambda=0.62, $F(18, 1377.9)=14.11$, $p<0.05$) (**Table 8**). There were statistically significant differences between all the bibliometrics and the groups with different physician status (physician or non-physician) (Wilks` Lambda=0.90, $F(6,487)=8.55$, p<0.05) and a statistically significant interaction effect between academic rank and physician status (Wilks` Lambda=0.94, F(18, 1377.9)=1.68, p<0.05).

	$N_p$tot↑	$N_c$tot↑	h index↑	h^2 upper↓	h^2 centre↑	h^2 lower↑	sRM value↑
<u>Non-physician</u>							
Lecturer	20.2	491.7	10.0	65.5	23.8	9.5	12.8
Senior lecturer	32.7	766.3	13.1	57.4	28.0	14.6	20.9
Reader	37.6	835.6	14.4	53.4	29.3	17.3	24.4
Professor	69.9	2045.8	21.4	53.4	27.6	19.0	35.7
<u>Physician</u>							
Lecturer	16.7	256.6	6.4	65.2	20.4	10.4	11.9
Senior lecturer	31.4	562.3	10.2	56.0	28.0	13.5	18.6
Reader	59.8	964.0	15.6	49.9	28.2	21.9	28.5
Professor	82.5	2388.8	21.9	53.0	26.7	20.2	36.6

$N_p$tot = total number of publications, $N_c$tot = total number of citations, ↑= increases when academic rank increases, ↓= decreases when academic rank increases

Table 8: This table demonstrates the significant increase or decrease (p < 0.05) in the mean values of each bibliometric indicator with academic rank.

In the ANOVA, construct validity was demonstrated for total number of publications, total number of citations, h index, h^2 upper, h^2 centre, h^2 lower, and sRM value, because there were statistically significant differences between the different academic ranks (p<0.05) (**Figure 6**) (**Table 9**). Groups with different physician status showed statistically significant differences regarding total number of citations, h index, h^2 upper and sRM value (p<0.05). Academic rank and physician status showed a statistically significant interaction effect with respect to total number of publications and h index (p<.0.05). Professors (physician or non-physician) had more publications than the other academic ranks (p<0.05). Physician professors and physician readers had more publications than their non-physician counterparts (p<0.05). The more recent the year of first publication of an academic the lower were his or her total number of citations (regression coefficient=-0.12, p<0.05) and h^2 upper (regression coefficient=-0.011, p<0.05). Lecturers had the smallest h index and total number of citations, but they showed the highest h^2 upper (about 65%) of all groups.

Effect		SSQ	df	MSQ	F	R^2
Total number of publications^p						
	Year of first publication	1.22	1	1.22	1.09	
	Academic rank	245.78	3	81.92	72.69*	
	Physician status	0.26	1	0.26	0.23	
	Academic rank x physician status	10.35	3	3.45	3.06*	
	Error	554.55	492	1.13		
	Total	858.43	500			0.35
Total number of citations^p						
	Year of first publication	25.30	1	25.30	5.34*	
	Academic rank	889.83	3	296.61	62.60*	
	Physician status	18.60	1	18.60	3.93*	
	Academic rank x physician status	37.18	3	12.39	2.62	
	Error	2312.13	488	4.74		
	Total	3373.65	496			0.31
h index^p						
	Year of first publication	2.69	1	2.69	2.75	
	Academic rank	196.42	3	65.47	66.81*	
	Physician status	5.73	1	5.73	5.85*	
	Academic rank x physician status	8.23	3	2.74	2.80*	
	Error	478.46	488	0.98		
	Total	716.46	496			0.33
h^2 upper^{np}						
	Year of first publication	0.25	1	0.25	11.10*	
	Academic rank	1.00	3	0.33	14.98*	
	Physician status	0.001	1	0.001	0.06	
	Academic rank x physician status	0.03	3	0.01	0.43	
	Error	10.96	492	0.02		
	Total	12.30	500			0.11
h^2 centre						
	Year of first publication	0.04	1	0.04	3.03	
	Academic rank	0.21	3	0.07	5.69*	
	Physician status	0.04	1	0.04	3.02	
	Academic rank x physician status	0.01	3	0.004	0.36	
	Error	5.90	485	0.01		
	Total	6.17	493			0.05
h^2 lower^p						
	Year of first publication	0.18	1	0.18	2.22	
	Academic rank	5.92	3	1.97	23.80*	
	Physician status	0.05	1	0.05	0.55	
	Academic rank x physician status	0.28	3	0.09	1.11	
	Error	40.78	492	0.08		
	Total	48.44	500			0.16
sRM value^p						
	Year of first publication	0.02	1	0.02	0.69	
	Academic rank	3.62	3	1.21	34.87*	
	Physician status	0.15	1	0.15	4.18*	
	Academic rank x physician status	0.10	3	0.03	0.97	
	Error	14.38	416	0.04		
	Total	18.53	424			0.22

p = power transformed, np = not power transformed, SSQ = sums of squares, df = degree of freedom, MSQ = mean square, F = F-test, R^2 = coefficient of determination; Type III sums of squares. * = $p < 0.05$

Table 9: Results of seven two-factor analyses of variance (academic rank, physician status, interaction) with covariate-adjustment (year of publication).

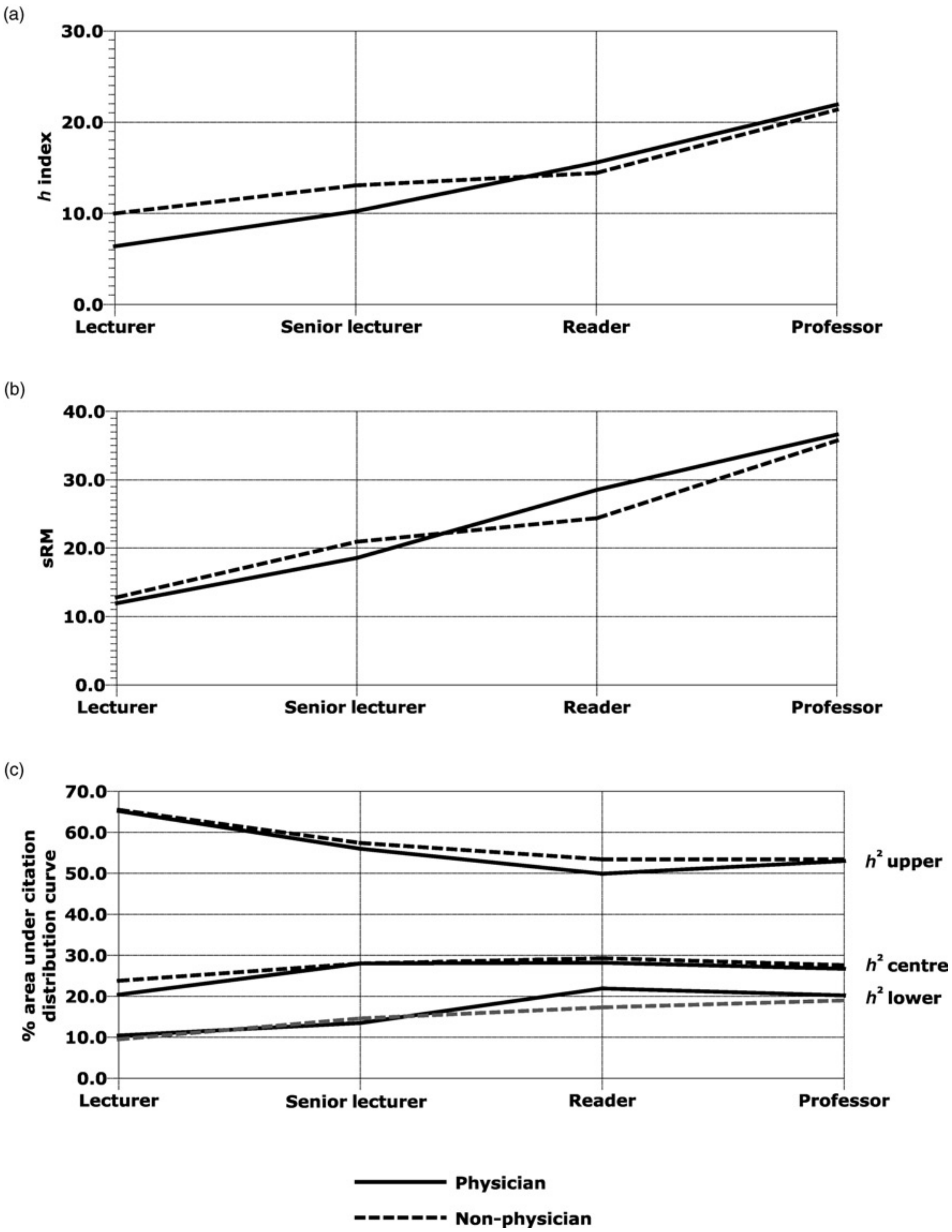

Figure 6: Differences between academic rank, physician status and (a) h index, (b) sRM value, (c) h^2 upper, h^2 centre and h^2 lower.

Convergent validity of the h index and sRM value was demonstrated by significant correlations with total number of publications (r=0.89 and 0.86 respectively, p<0.05) and total number of citations (r=0.96 and r=0.65 respectively, p<0.05) (**Table 10**).

	Nptot	Nctot	h index	h^2 upper	h^2 centre	h^2 lower	sRM
Nptot	1						
Nctot	.84*	1					
h index	.89*	.96*	1				
h^2 upper	-.39*	0.06	-.17*	1			
h^2 centre	.21*	0.04	.29*	-.64*	1		
h^2 lower	-.74*	-.41*	-.53*	.66*	-.42*	1	
sRM	.86*	.65*	.84*	-.76*	.61*	-.76*	1

Nptot = total number of publications, Ncto t= total number of citations, * = p < 0.05

Table 10: Matrix of inter-correlations (Pearson) between different bibliometrics (power transformed variables, N=438).

4.5 DISCUSSION

We have demonstrated how to calculate h^2 upper, h^2 centre, h^2 lower and sRM value in order to enhance bibliometric assessment and appraisal of research performance of an academic healthcare researcher. The sRM value is the first indicator that is based on statistical and not numerical concepts. Our study supports construct and convergent validity of the h index and sRM value.

Criticisms of the h index have led to the development of numerous variants of the *h* index.(*12, 105*) However, many of these variants are highly correlated with the h index and offer limited value in research performance assessment.(*12, 105, 113*)

Our study presents h^2 upper, h^2 centre, h^2 lower, and the sRM value, which complements and provides additional information to the h index. For example, in our study the lecturers had few publications and had a low h index, but their small publication set was well cited. A multi-dimensional bibliometric approach to research performance assessment may identify patterns of productivity, such as prolific researchers (high h^2 centre and upper, low h^2 lower), mass producers (high h^2 lower, low h^2 centre and upper) and perfectionists (high h^2 upper, low h^2 centre and lower).(*105, 114*) This can provide practical information that can be conveyed in various forms such as graphical images (**Figure 7**) to support academics and supervisors in planning and managing careers, as well as assisting in the decisions for academic promotion and funding.(*115*)

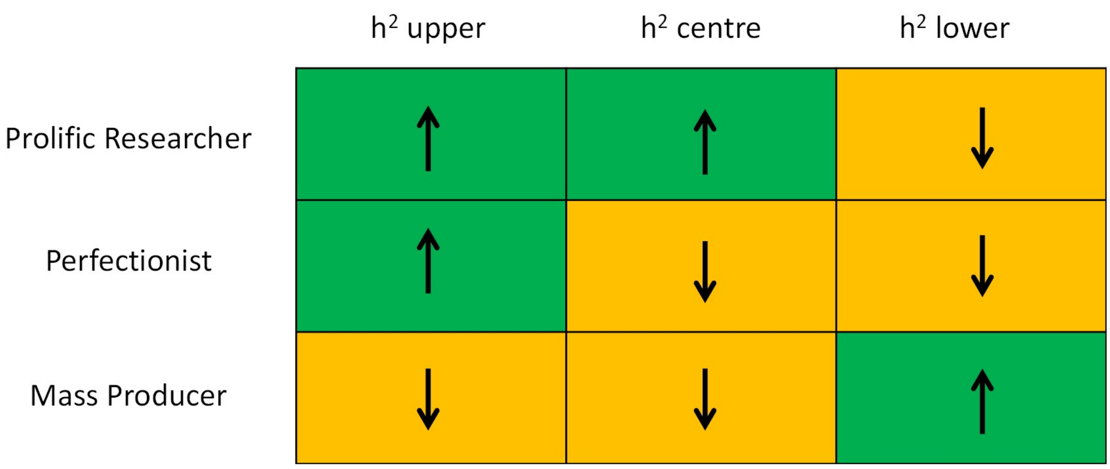

Figure 7: Representative Grid categorizing researchers as prolific, perfectionist or mass producers according to h^2 upper, h^2 centre and h^2 lower characteristics.

Previous studies have shown construct validity of the h index in healthcare specialties such as anaesthesia, neurosurgery, radiology and urology.(*45, 116-119*) Our study is the largest construct validation study of the h index in academic healthcare researchers within one of the largest Faculties of Medicine in Europe. It also takes into account the length of academic career and physician status. Academic physicians (clinician scientists) will spend approximately half of their time in clinical care, and many physicians believe that bibliometric analyses favour basic scientific research over clinical research.(*88, 120, 121*)

Our study confirms that differences do exist between physician and non-physician academic healthcare researchers. This should be considered when measuring the research performance of academic physicians with bibliometrics such as the h index.

Our study has limitations associated with the disadvantages of the h index, which does not consider degree of co-authorship, gratuitous authorship, age of the researcher, context of citations, self-citations and scientific discipline.(*12, 105, 117*) We did not consider the impact of document type on research performance. We did not account for the bias that may exist of researchers preferring to cite work from their own country. We have extracted bibliometric data for academics from a single centre specialising in medicine, so the findings of this study may not be generalisable to other academic institutions or specialties. The h index does give an indication of the absolute number of citations (h×h), but this information is deficient in the sRM value. We collected data over a defined timeframe, so our results cannot be used as an absolute benchmark for comparing academic healthcare researchers. The sRM value is not easily calculable because it relies on segmented regression modelling using statistical software.

There are differences in the publication sources for Web of Science, Scopus and Google Scholar, which produce quantitatively and qualitatively different citation counts for healthcare researchers.(*93*) We only used Scopus to extract the bibliometric data of the academic healthcare researchers, and it is unknown whether physician-bias exists for any of the citation databases. Miscalculation of bibliometric data is more likely if an author has a common name, if they have published using a combination of names, if they have worked in a variety of institutions or if they have extensive research interests.(*117*)

Future research may include studies that explore the relationship between our measures of high visibility (h^2 upper and sRM) and the impact and external visibility of an individual's research. This may objectively measure the benefit of an individual's research to the wider economy and society. It may also offer impartial evidence for selection and promotion committees to guide and support healthcare researchers in achieving enhanced academic goals.

4.6 CONCLUSION

Our study supports construct validity of the h index as a measure of academic rank for healthcare researchers. However, the h index describes only a small proportion of information about the academic output of a researcher. As a result, it may not accurately reflect an individual's research performance. The use of h^2 upper, h^2 centre, h^2 lower and sRM value in combination with the h index may provide additional objective evidence to assess and appraise the impact of an academic healthcare researcher. This may identify patterns of academic activity that can support research quality and productivity of innovative researchers. It may also offer guidance for less creative researchers to optimise their academic profiles.

4.7 CHAPTER SUMMARY

The objectives of this chapter were to investigate whether the h index could be improved to better measure the academic performance and citation profile for individual healthcare researchers.

This was a Cohort study set in the Faculty of Medicine, Imperial College London, United Kingdom. Publication lists were extracted from 1st January 2000 until 31st December 2009 for 501 academic healthcare researchers from the Faculty of Medicine. The *h* index for each researcher was calculated over a 9-year period. The citation count for each researcher was differentiated into high (h^2 upper), core (h^2 centre), and low (h^2 lower) visibility areas. Segmented regression model was used to statistically estimate number of high visibility publications (sRM value). Validity of the *h* index and other proposed adjuncts were analyzed against academic rank and conventional bibliometric indicators.

Construct validity was demonstrated for *h* index, h^2 upper, h^2 centre, h^2 lower, and sRM value (all $p<0.05$). Convergent validity of the *h* index and sRM value was shown by significant correlations with total number of publications ($r=0.89$ and 0.86 respectively, $p<0.05$) and total number of citations ($r=0.96$ and $r=0.65$ respectively, $p<0.05$). Significant differences in *h* index and sRM value existed between non-physician and physician researchers ($p<0.05$).

This study supports the construct validity of the h index as a measure of healthcare researcher academic rank. It also identifies the assessment value of our developed indices of h^2 upper, h^2 centre, h^2 lower and sRM. These can be applied in combination with the *h* index to provide additional objective evidence to appraise the performance and impact of an academic healthcare researcher.

Our research so far has demonstrated feasibility, validity and reliability of bibliometric indicators, such as the h index. The systematic review highlighted that collaboration should be considered as an indicator to assess and develop research performance. In the next chapter we investigated the effects of collaborative networks and authorship attribution on healthcare research performance.

5 THE EFFECTS OF COLLABORATIVE ACADEMIC HEALTHCARE NETWORKS AND AUTHORSHIP ATTRIBUTION ON SCIENTIFIC SUCCESS IN BIOMEDICAL RESEARCH

5.1 INTRODUCTION

Healthcare researchers require assessment and appraisal to maintain or improve their research performance.(*5, 82*) Objective bibliometric indicators of research performance, such as number of publications or citations, are commonly used for quantitative and qualitative evaluation. It is widely recognised that collaboration plays a fundamental role in sustaining the performance of healthcare research, and, for this reason, should be encouraged by institutions and academic centres.(*122*) However, current academic evaluation systems do not rely on a systematic use of indicators that explicitly take into account the network of collaborative relationships in which researchers are embedded.(*122, 123*)

Collaborative research typically results in multiple authors listed on a single publication. For biomedical publications, the sequence of co-authors' names is often determined by contribution- and supervision-related credit, so that the first author is the researcher that has contributed the most, and the last author has the most supervisory role.(*124*) Occasionally, researchers who appear in the middle of the sequence of co-authors are likely to be those who have failed to demonstrate adequate contribution, and for this reason have been offered authorship gratuitously.(*125, 126*) Conversely, co-authors of a publication that have contributed equally to the scientific work are typically listed in alphabetical order in the sequence of names.

In this study, drawing on a unique co-authorship network in healthcare research, we embraced a network-based perspective (*122, 123*) to investigate: (i) the effects of collaborative patterns on research performance; and (ii) the role that various criteria for authorship attribution have in sustaining or hindering research performance.

5.2 METHODS

5.2.1 Sample population

The Faculty of Medicine Imperial College London was established in 1997, and is one of Europe's largest medical institutions.(*107, 108*) It is organised into the Institute of Clinical Sciences, Kennedy Institute of Rheumatology, Department of Medicine, National Heart and

Lung Institute, School of Public Health, and Department of Surgery and Cancer.(*107, 108*) We used the database from the university intranet to create a list of all academics from the Faculty of Medicine Imperial College that were in employment on the 31st December 2009. We included academics that were ranked in the hierarchical system in the following grades: lecturer, senior lecturer, reader and professor. We excluded academics that were research fellows, associates, officers, assistants or staff with honorary academic status. For each of the academics included in the study, we extracted the first name, surname, gender, academic rank, and physician status (i.e., whether the academic was a physician or not).(*108*) We constructed indicator (i.e., binary) variables for gender and physician status. We used the academic rank of lecturer as the reference category against which we controlled for rank-related differences in performance between academics.

For each academic, we recorded the Institute, School or Department with which they were affiliated. Among the six institutional units, we used the Department of Surgery and Cancer as the reference category, against which we controlled for institution-related differences in performance between academics.

5.2.2 Generating an academic's publication list

We used SciVerse Scopus Author Identifier to generate the publication list authored by each academic.(*84, 108*) If the search tool identified more than one publication list for an academic's name, then we combined the appropriate publication lists. For each academic, we examined the publication list and excluded any publications that were not attributable to the individual academic. Finally, all publication lists were divided into three time periods (January 1st 2001 to December 31st 2003; January 1st 2004 to December 31st 2006; and January 1st 2007 to December 31st 2009). This enabled us to carry out a longitudinal analysis for detecting the structural determinants of research performance over time. Measures of research performance included publication and citation count. For each academic and time period, we used total citation counts only for the publications in that time period, so that the effects of collaborative patterns on the academic's research performance could be unambiguously assessed. We extracted citation data in January 2013 to allow sufficient time for recent publications to be cited.

5.2.3 Controlling for productivity and solo versus multiple authorship

For each academic and time period, we extracted the number of publications in the preceding time period. Using this variable, we tested whether an academic's productivity in a time period had an impact upon the number of citations the academic received in connection with the publications in the subsequent time period.

For each academic and time period, we recorded the number of publications in which they appeared as the solo author. For each academic, we also calculated the median number of co-authors per publication. To assess the effects of multi-authorship on performance, for each academic, and across all their multi-authored publications in each time period, we extracted the minimum number of co-authors per publication.

5.2.4 Creating the co-authorship network

For each time period, we combined the publication lists of all academics into a single list, which was stored as a comma-separated value (.csv) file. Each file was loaded into Network Workbench Software, which is a software for the analysis, modelling and visualisation of large-scale networks.(*127*) The software filtered out the duplicate publications resulting from the combination of the publication lists of academics that co-authored publications. The co-authorship network was then extracted from the remaining publications and was stored as a network (.net) file. The nodes of the network were the authors, and links were assumed to exist between two nodes when the corresponding authors had co-authored one or more scientific publications. The network so constructed was therefore undirected and unweighted. We used the software to calculate six network measures: (i) degree centrality, (ii) closeness centrality, (iii) betweenness centrality, (iv) eigenvector centrality, (v) local clustering coefficient and (vi) constraint.

5.2.5 Measuring authors' network-based centrality

The normalised degree centrality of a node is the number of links incident upon the node, divided by its maximum possible value (i.e., the number of nodes in the network minus one) **(Figure 8a)**. Formally, for an undirected network of *n* nodes and no self-edges, the degree centrality of node *i* can be expressed in terms of the adjacency matrix *A* as:

$$C_D(i) = k_i = \sum_j A_{ij} ,$$

where

$$A_{ij} = \begin{cases} 1 \text{ if there is a link between nodes } i \text{ and } j \ (i \neq \text{j}); \\ 0 \text{ otherewise.} \end{cases}.$$

To obtain the normalized degree centrality of node *i*, $C_D^{'}(i)$, we simply divide $C_D(i)$ by its maximum value, i.e., *n – 1*:

$$C_D^{'}(i) = \frac{C_D(i)}{n-1} .$$

A large body of literature has suggested that highly connected nodes have a greater chance of receiving information and having more influence or prestige than poorly connected ones.(*128, 129*) We tested the hypothesis that academics with more collaborators (i.e., with a higher normalized degree centrality) were more likely to be exposed to a larger amount of information and opportunities and could therefore achieve a better performance than academics with fewer collaborators (i.e., with a low normalized degree centrality).

Degree centrality is a local measure of centrality, and as such does not depend on the global structure of the network. Although a node may be highly connected, it may not be suitably located so as to reach others and receive or send information quickly within the network. For this reason, we also tested the effects of global measures of centrality on academics' performance.

Eigenvector centrality measures the importance of a node in a network as a function of the connections the node has to other nodes that are themselves important **(Figure 8a).**(*130*) Instead of awarding a node only one score for each of its neighbours, eigenvector centrality awards the node a score that is proportional to the sum of the scores of its neighbours. Formally, we have:

$$C_E(i) = \kappa_1^{-1} \sum_j A_{ij} C_D(j),$$

where κ_1 is the largest eigenvalue of the adjacency matrix *A*. The measure is therefore premised on the idea that the centrality of a node is high to the extent that the node's neighbourhood includes many nodes or nodes that also have a high centrality, or both. We tested the hypothesis that academics with a higher value of eigenvector centrality could achieve a better performance than academics with a lower value.

Betweenness centrality measures the extent to which a node lies on the shortest paths between other nodes in the network **(Figure 8a)**.(*131*) Betweenness centrality captures the ability of a node to control interactions and information flows between pairs of other nodes in the network, and thus to act as the gatekeeper or broker between others. Formally, we have:

$$C_B(i) = \sum_{j,l} \frac{g_{j,l}(i)}{g_{j,l}},$$

where $g_{j,l}$ is the number of geodesics (i.e., shortest paths) linking nodes *j* and *l*, for *i* distinct from *j* and *l*; $g_{j,l}(i)$ is the number of such geodesics that contain node *i*; and $\frac{g_{j,l}(i)}{g_{j,l}} = 0$ if both $g_{j,l}(i)$ and $g_{j,l}$ are zero. Nodes with higher betweenness centrality are expected to have a higher status, power or influence on others than nodes with lower betweenness. We therefore tested the hypothesis that academics with higher betweenness could achieve better research performance than academics with lower betweenness.

Finally, a node's closeness centrality is defined as the inverse of the sum of the shortest distances separating the node from all other nodes, and thus measures how close the node is to all other nodes in the network **(Figure 8a).**(*123*) Formally, we have:

$$C_C(i) = \left[\sum_j d_{i,j} \right]^{-1},$$

where $d_{i,j}$ is the length of the geodesic path from node *i* to node *j*, i.e., the number of links along the path. Nodes with higher closeness centrality are expected to obtain information more promptly, and exert more influence on others, than nodes with lower closeness. We therefore tested the hypothesis that academics with higher closeness could obtain a higher performance than academics with lower closeness.

5.2.6 Measuring authors' social capital: closed versus open network structures

We investigated the role of social capital in facilitating academics' performance, by testing the effects that closed and open network structures had on academics' citation counts. We relied on two network measures: the local clustering coefficient and network constraint.

Local clustering coefficient is defined as the ratio between the number of links connecting pairs of a node's neighbours and the total number of pairs of the node's neighbours **(Figure 8b).**(*132*) Formally, we have:

$$Clust(i) = \begin{Bmatrix} \dfrac{K[N_i]}{k_i(k_i-1)/2} & \text{for } k_i \geq 2 \\ 0 & \text{for } k_i = 0,1 \end{Bmatrix},$$

where $K[N_i]$ is the number of links connecting pairs of the neighbours N_i of node *i*, and k_i is the degree (i.e., the number of neighbours) of node *i*. The clustering coefficient has traditionally been used to operationalize conceptions of social capital predicated on the mechanism of social cohesion and network closure.(*13*) From this perspective, clustering captures the extent to which a node can derive benefits from being socially embedded within cohesive social structures, rich in third-party relationships. Among these closure-based sources of social capital are normative control, deviance avoidance, the enhancement of one's sense of belonging and trust, the creation of a common culture, and the facilitation of cooperation and of the exchange of fine-grained, complex, tacit and proprietary information. We tested the hypothesis that academics whose local network was more socially cohesive (i.e., with a higher clustering coefficient) could obtain a better performance than academics in a less cohesive network (i.e., with lower clustering).

Innovative healthcare research often necessitates strong support from colleagues who are experts in similar areas, but also requires access to the diverse sources of knowledge in other specialties. While closed networks facilitate social support and knowledge flows, individuals can also benefit from participating in open structures that are rich in cleavages and opportunities of brokerage. This is the idea underpinning an alternative conception of social capital: by gaining exposure to a greater variance and novelty of information, individuals embedded in brokered structures will be creative and successful in their endeavours.(*133-135*) Structural holes are opportunities for individuals to broker between otherwise disconnected individuals. Individuals closely linked with one another are likely to possess similar ideas: the more an individual's contacts are connected with each other, the less likely they are to take the individual closer to valuable sources of knowledge and resources that the individual is not already able to access. Highly cohesive networks may thus create isolation and resistance to information and knowledge flowing from outside the network. By contrast, when an individual's contacts are disconnected from each other, the presence of structural holes may provide the individual with opportunities for gaining access to new and non-redundant social circles in which other individuals are likely to have different ideas and resources.

Network constraint measures the extent to which a node is connected to other nodes that are already connected with each other **(Figure 8c)**. Formally, the constraint of node *i* has been defined by Burt as (*133*):

$$Constr(i) = \left(\sum_{j} p_{i,j} + \sum_{q} p_{i,q} p_{q,j} \right)^2, \ q \neq i, j,$$

where $p_{i,j}$ is the entry of the transition matrix *P* and measures the proportion of node *i*'s network time and energy invested in the relationship with node node *j*, and (for undirected networks) is defined as

$$p_{i,j} = \frac{w_{i,j}}{\sum_m w_{i,m}},\ i \neq j\ ,$$

where $w_{i,j}$ is the weight of the link connected nodes *i* and *j* (*135*). Constraint thus captures the lack of structural holes in a network. A low value of network constraint means that a node can broker between otherwise disconnected others, and can therefore benefit from discontinuities in the social structure. On the other hand, a large value of network constraint implies paucity of connections to non-redundant others, and is therefore associated with network closure. We tested the hypothesis that academics with a lower value of network constraint could achieve better performance than academics with a higher value.

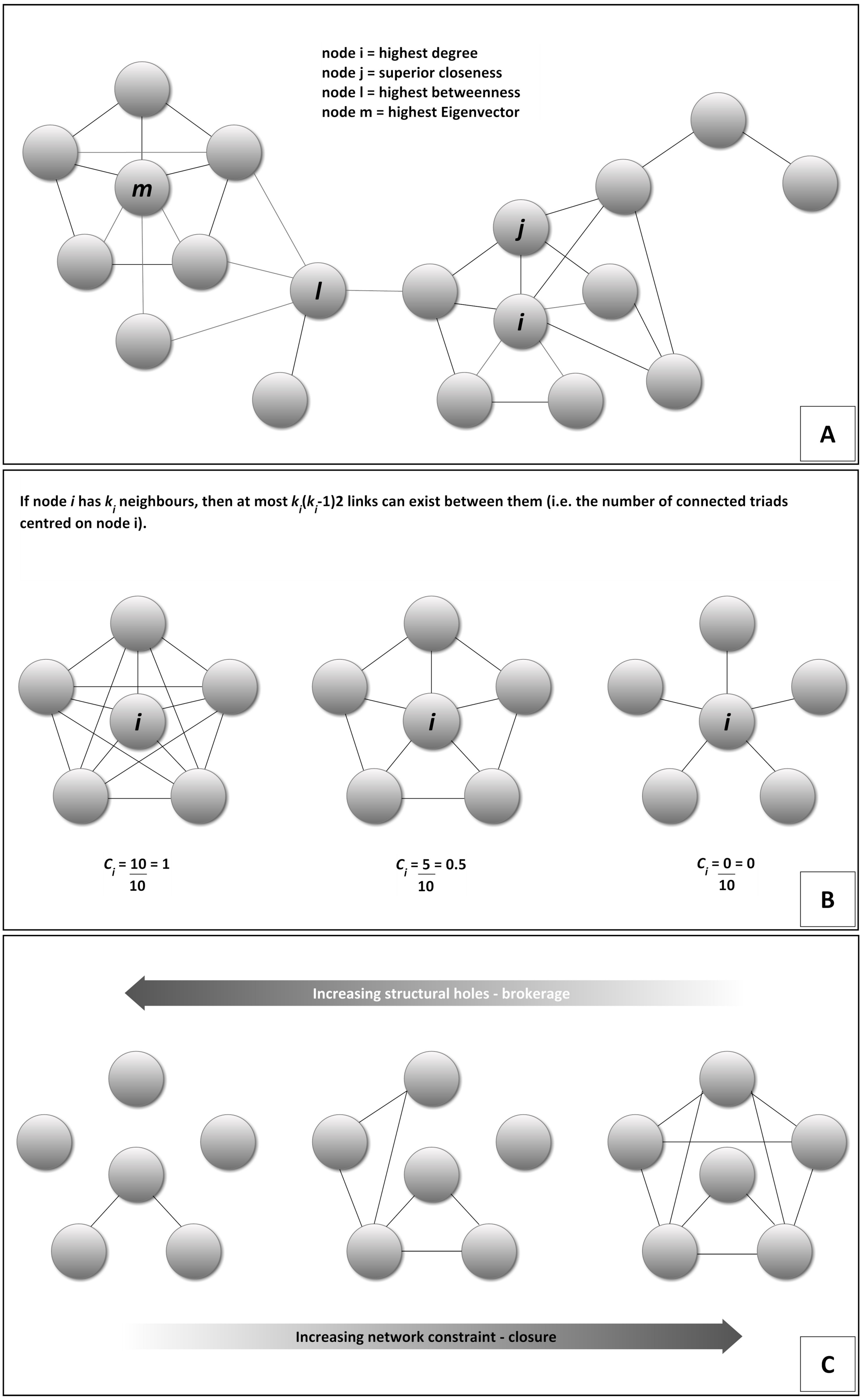

Figure 8: Examples of network measures of centrality and social capital.

5.2.7 Extracting authors' positions in sequences of co-authors

The position of an author in the list of co-authors of a publication is often used to indicate the importance of the contribution of the author to the scientific work (citation).(*136-141*) Our data retained this important information on authorship credit, which we used to assess the association between an author's position and their research performance. For each multi-authored publication, and distinguishing between publications in which co-authors were listed in alphabetical and non-alphabetical order, we extracted the four most important positions, classified as follows:

- The first-listed author on any multi-authored publication was always recorded as "first author".
- The last-listed author was recorded as "last author" provided the publication had two or more co-authors.
- The second-listed author was recorded as "second author" provided the publication had at least three co-authors.
- The penultimate author was recorded as "penultimate author" provided there were at least four co-authors on the publication.
- Any remaining position was recorded as "other".

For each academic and time period, we extracted the number of multi-authored publications in which co-authors were listed alphabetically and non-alphabetically. For each of these two groups of publications, we recorded the number of publications in which the academic appeared as listed in each of the above five positions. In total, for each academic and time period, we therefore constructed ten position-related variables, and then tested the effect of these variables on the academic's performance.

5.3 STATISTICAL ANALYSIS

For all variables, **Table 11** shows the means prior to centering, the standard deviations, and the minimum and maximum values. **Table 12** shows zero-order correlations between variables.

	Variable	Mean	Standard Deviation	Minimum	Maximum
1	Number of citations	414.419	655.472	.000	7883.000
2	Number of past publications	29.270	58.252	.000	1013.000
3	Gender	.705	.456	.000	1.000
4	Physician status	.480	.500	.000	1.000
5	Institute of Clinical Sciences	.027	.161	.000	1.000
6	Kennedy Institute of Rheumatology	.046	.209	.000	1.000
7	Department of Medicine	.409	.492	.000	1.000
8	National Heart Lung Institute	.197	.398	.000	1.000
9	School of Public Health	.109	.311	.000	1.000
10	Senior Lecturer	.257	.437	.000	1.000
11	Reader	.126	.332	.000	1.000
12	Professor	.410	.492	.000	1.000
13	(Median number of co-authors/publication)^2	6.600	3.313	2.000	40.000
14	(Minimum number of co-authors/publication)^2	2.900	1.620	2.000	26.000
15	Degree	.013	.021	.000	.302
16	Betweenness	.080	.150	.000	2.180
17	Closeness	.010	.000	.000	.010
18	Eigenvector	.010	.030	.000	.410
19	Clustering coefficient	.320	.250	.000	1.000
20	Constraint	.160	.328	.000	10.583
21	Solo authored publications	1.078	2.986	.000	48.000
Publications in non-alphabetical author sequence					
22	First author	1.833	2.540	.000	35.000
23	Last author	4.380	6.900	.000	89.615
24	Second author	1.816	2.418	.000	23.000
25	Penultimate author	2.317	4.327	.000	87.000
26	Other author	4.419	7.089	.000	104.000
Publications in alphabetical author sequence					
27	First author	.400	.928	.000	7.597
28	Last author	.611	1.398	.000	14.000
29	Second author	.104	.429	.000	6.000
30	Penultimate author	.025	.168	.000	2.000
31	Other author	.014	.133	.000	2.000

Table 11: Means prior to centering, standard deviations, and minimum and maximum values for all variables:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1	1.00																														
2	0.56	1.00																													
3	0.12	0.12	1.00																												
4	0.05	0.08	0.15	1.00																											
5	-0.02	-0.03	0.03	0.01	1.00																										
6	0.01	0.01	-0.02	-0.03	-0.04	1.00																									
7	-0.09	-0.01	-0.01	0.07	-0.14	-0.18	1.00																								
8	0.09	0.09	-0.01	-0.02	-0.08	-0.11	-0.41	1.00																							
9	0.11	-0.03	-0.06	-0.09	-0.06	-0.08	-0.29	-0.17	1.00																						
10	-0.18	-0.17	-0.08	0.21	0.01	0.02	0.01	-0.05	0.02	1.00																					
11	-0.07	-0.06	-0.03	-0.09	0.01	-0.06	0.07	0.00	0.02	-0.22	1.00																				
12	0.39	0.35	0.18	0.08	-0.07	-0.02	-0.02	0.10	0.01	-0.49	-0.32	1.00																			
13	0.14	0.01	0.00	-0.02	0.04	-0.03	-0.01	-0.02	0.08	-0.03	-0.03	0.04	1.00																		
14	-0.18	-0.15	-0.06	-0.14	0.03	-0.03	0.04	-0.07	-0.06	0.02	-0.03	-0.22	0.36	1.00																	
15	0.55	0.40	0.08	0.06	0.00	-0.02	-0.09	0.07	0.10	-0.11	-0.02	0.26	0.23	-0.11	1.00																
16	0.48	0.28	0.11	0.09	-0.02	0.01	-0.08	0.05	0.07	-0.13	-0.03	0.29	0.15	-0.14	0.84	1.00															
17	0.39	0.38	0.07	0.05	0.01	0.00	-0.03	0.07	0.00	-0.06	0.02	0.19	0.03	-0.04	0.49	0.27	1.00														
18	0.23	0.21	0.05	0.02	0.00	0.01	-0.04	0.08	-0.01	-0.03	0.02	0.05	0.02	-0.03	0.73	0.52	0.37	1.00													
19	-0.30	-0.24	-0.10	-0.06	0.01	0.00	0.04	-0.03	-0.06	0.12	-0.02	-0.36	-0.02	0.29	-0.29	-0.36	-0.13	-0.05	1.00												
20	-0.14	-0.11	-0.09	-0.04	-0.02	0.00	0.00	-0.03	0.05	0.10	-0.02	-0.16	-0.13	0.04	-0.17	-0.16	-0.16	-0.08	0.30	1.00											
21	0.27	0.37	0.09	0.14	-0.04	0.00	0.01	0.12	-0.02	-0.10	-0.04	0.24	-0.05	-0.16	0.12	0.15	0.07	0.01	-0.15	-0.07	1.00										
22	0.11	0.04	0.00	0.08	-0.03	-0.02	-0.08	0.02	0.03	0.07	0.06	-0.11	-0.08	-0.18	0.06	0.07	0.08	0.01	-0.10	-0.01	0.18	1.00									
23	0.56	0.49	0.14	0.11	-0.03	-0.02	-0.08	0.10	0.00	-0.23	-0.06	0.47	-0.03	-0.25	0.32	0.43	0.14	0.07	-0.34	-0.13	0.33	0.04	1.00								
24	0.36	0.23	0.05	0.04	-0.01	-0.04	-0.11	0.03	0.08	-0.09	0.03	0.15	0.00	-0.20	0.25	0.29	0.12	0.09	-0.24	-0.09	0.14	0.35	0.38	1.00							
25	0.46	0.42	0.14	0.06	-0.01	-0.01	-0.08	0.07	0.03	-0.20	-0.05	0.38	0.06	-0.15	0.31	0.44	0.10	0.07	-0.29	-0.11	0.26	0.03	0.68	0.38	1.00						
26	0.50	0.25	0.07	0.06	0.03	-0.01	-0.10	0.04	0.16	-0.14	-0.02	0.28	0.53	-0.13	0.47	0.44	0.10	0.08	-0.28	-0.14	0.15	0.13	0.39	0.41	0.47	1.00					
27	0.14	0.12	0.04	0.06	-0.01	0.01	-0.07	0.05	0.04	-0.01	0.02	0.05	-0.14	-0.22	0.07	0.10	0.05	0.01	-0.12	-0.04	0.20	0.12	0.21	0.18	0.15	0.06	1.00				
28	0.21	0.21	0.08	0.09	-0.06	0.08	-0.02	0.08	-0.05	-0.11	-0.08	0.27	-0.11	-0.23	0.12	0.19	0.04	0.00	-0.19	-0.07	0.27	0.18	0.34	0.15	0.27	0.13	-0.07	1.00			
29	0.10	0.08	0.04	-0.03	0.00	-0.02	-0.05	0.01	0.09	-0.05	-0.02	0.11	-0.07	-0.09	0.06	0.10	0.03	0.01	-0.12	-0.05	0.06	0.08	0.15	0.17	0.13	0.07	0.06	0.06	1.00		
30	0.09	0.04	0.07	0.06	-0.03	-0.03	-0.02	0.06	0.00	-0.05	-0.02	0.09	-0.02	-0.05	0.08	0.11	0.05	0.09	-0.07	-0.03	0.04	0.11	0.13	0.09	0.09	0.06	0.01	0.11	0.12	1.00	
31	0.07	0.10	0.07	0.04	0.01	-0.02	-0.03	0.08	-0.04	-0.04	-0.01	0.05	0.00	-0.01	0.06	0.04	0.03	0.06	-0.03	-0.02	0.04	-0.01	0.08	0.01	0.07	0.05	0.01	0.09	0.00	0.04	1.00

Table 12: Zero-order correlations between variables.

To investigate the non-linear effects of number of co-authors per publication on performance, we squared the median number of co-authors, and jointly tested the main and quadratic terms in the model. To avoid problems of multi-collinearity arising from high correlation between variables, before squaring the median, we centred it around its mean (i.e., we subtracted the mean from the variable).(*142*)

Models 1 through to 3 in **Table 14** include various centrality measures. These measures could not be included simultaneously in one single model because the high degree of correlation between them would have created problems of multi-collinearity.

We estimated interaction effects between an author's academic rank and two position-related variables: the number of publications in which the author appeared as last-listed in non-alphabetical order, and the number of publications in which the author was first-listed in alphabetical order. We centred the two position-related variables, and multiplied each of them by each of the three rank-related indicator variables. We thus obtained six interaction terms.

Finally, we estimated interaction effects of authors' academic rank, brokerage opportunities, and position in by-line. To construct the interaction terms, we multiplied each of the three rank-related indicator variables by the product between centred constraint and centred number of publications in which authors appeared as last-listed in non-alphabetical order. We thus obtained three additional interaction terms.

The statistical analysis and modelling strategy is described in detail in **Appendix 4**.

5.4 RESULTS

We analysed 525 academics in the Faculty of Medicine at Imperial College London, one of Europe's largest medical institutions.(*107, 108*) The composition of the academics in terms of academic rank, gender and physician status is shown in **Table 13.**

Academic Rank	Physician status n (%)		Gender n (%)		Total n (%)
	Non-physician	Physician	Male	Female	
Lecturer	85 (77)	25 (23)	66 (60)	44 (40)	110 (21)
Senior Lecturer	45 (34)	89 (66)	87 (65)	47 (35)	134 (25)
Reader	42 (64)	24 (34)	44 (67)	22 (33)	66 (13)
Professor	101 (47)	114 (53)	173 (80)	42 (20)	215 (41)
Total n (%)	273 (52)	252 (48)	370 (70)	155 (30)	525 (100)

Table 13: Composition of the academics in the Faculty of Medicine.

We estimated maximum likelihood negative binomial panel regression models with beta-distributed random effects and bootstrapped standard errors. **Table 14** reports the estimates for the coefficients of control and theoretical variables. Models 1, 2 and 3 show the main effects of all variables on authors' performance. Model 4 also includes the interactions between independent variables.

	Model 1		Model 2		Model 3		Model 4	
	EC	SE	EC	SE	EC	SE	EC	SE
Control variables								
Number of past publications	.002**	.001	.002**	.001	.000	.001	.002**	.000
Gender	.074	.066	.070	.062	.067	.060	.074	.065
Physician status	-.203**	.058	-.195**	.057	-.222**	.050	-.204**	.057
Senior lecturer	.055	.074	.041	.078	.028	.086	.546	.293
Reader	.110	.107	.093	.102	.116	.093	.670*	.291
Professor	.420**	.095	.411**	.081	.407**	.093	.940**	.293
Solo versus multi-authorship								
Solo authored papers	.010	.015	.009	.015	.013	.014	.007	.016
Median number of co-authors per publication	.034**	.013	.032*	.014	.069**	.012	.030*	.013
(Median number of co-authors per publication)^2	-.003	.001	-.003**	.001	-.003**	.001	-.003**	.001
Minimum number of co-authors per publication	-.033	.017	-.032*	.016	-.053**	.017	-.037*	.017
Network-based measures of centrality								
Degree	6.519**	1.572					6.881**	1.209
Eigenvector			2.196*	1.080				
Betweenness			.360*	.172				
Closeness					204.775**	9.843		
Network-based measures of social capital								
Clustering coefficient					-.788**	.132		
Constraint	-.974**	.342	-1.074*	.451			-1.143**	.254
Author's position in publication								
Position in non-alphabetical sequence:								
First author	.044**	.011	.044**	.011	.041**	.009	.046**	.010
Last author	.019*	.009	.019*	.009	.021**	.004	-.140*	.067
Second author	.012	.018	.012	.016	.022	.016	.019	.019
Penultimate author	-.010	.007	-.011	.007	-.003	.008	-.009	.009
Other author	.020**	.007	.023**	.005	.021**	.006	.022**	.007
Position in alphabetical sequence:								
First author	.090**	.024	.088**	.032	.074**	.030	.084*	.041
Last author	.014	.021	.014	.027	.026	.023	.020	.021
Second author	.037	.051	.037	.048	.053	.048	.031	.043
Penultimate author	-.064	.100	-.117	.142	-.070	.103	-.067	.098
Other author	.119	.153	.133	.143	.068	.166	.098	.123
Interactions: academic rank and position								
Senior lecturer X non-alphabetical last position							.128	.069
Reader X non-alphabetical last position							.102	.072
Professor X non-alphabetical last position							.187**	.067
Senior lecturer X alphabetical first position							.087	.069
Reader X alphabetical first position							-.076	.085
Professor X alphabetical first position							.004	.052
Interactions: academic rank, position, and brokerage								
Senior lecturer X non-alphabetical last position X constraint							-.116	.126
Reader X non-alphabetical last position X constraint							.047	.178
Professor X non-alphabetical last position X constraint							.245**	.065

EC = estimated coefficient; SE = standard error

* $p \leq 0.05$, ** $p \leq 0.01$

Table 14: Results of random-effects negative binomial panel regressions with main and interaction effects.

We found that an author's past productivity had positive effects on the author's success: a good record of publications in a given period of time was likely to be associated with publications of high impact in the subsequent time period. Based on the estimated coefficients from Model 4, for every additional publication in a three-year period, an author's expected citation count on publications in the subsequent three-year period increased by a factor of 1.002 (i.e., by 0.20%), holding all other variables constant.

We demonstrated that there were no statistically significant gender-related effects on performance.

We found, on average, that non-physician academics obtained higher research performance than physician academics. Being non-physician increased the expected citation count by a factor of 0.81 (i.e., by 18.44%), holding all other variables constant.

We further found evidence in favour of academic rank-based statistical discrimination: estimates from Model 4 indicate that being a professor increased the expected number of citations with respect to the citations received by a lecturer by a factor of 2.561 (i.e., by 156%), holding the interacted (centred) variables at their means (i.e., zero) and all other variables constant.

We examined the effects of solo versus multiple authorship on research performance (*143*). While solo-authored publications did not have any statistically significant effect on performance, we found a non-linear effect of number of co-authors per publication on citations. This was clearly indicated by the statistically significant positive and negative effects of, respectively, the linear and quadratic terms of median number of co-authors per publication. Based on the estimated coefficients from Model 4, and holding all other variables constant at their means, **Figure 9** shows the inverse U-shaped relationship between the (centred) median number of co-authors per publication and the expected number of citations received by a female professor at the Institute of Clinical Sciences.

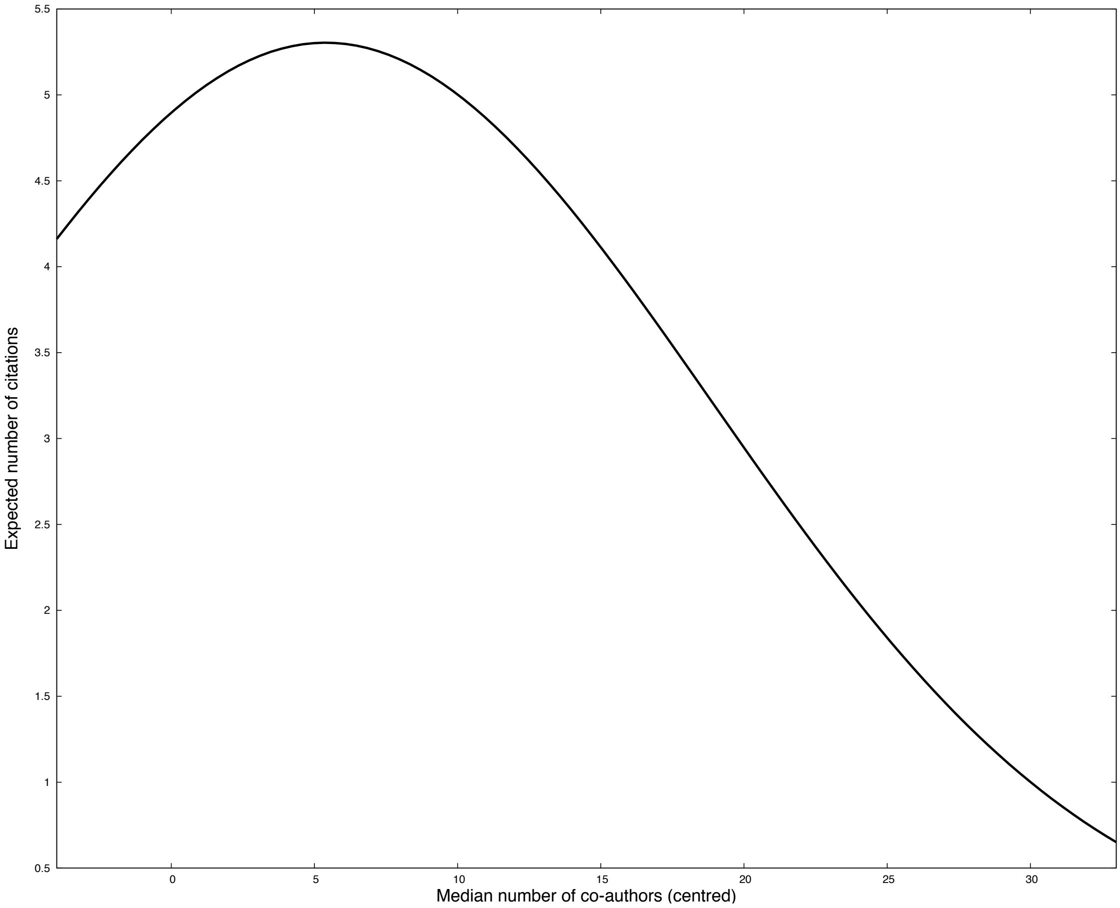

Figure 9: Impact of the (centred) median number of co-authors per publication on the expected citation count of a female professor at the Institute of Clinical Sciences, when all other independent variables are held constant at their means.

More generally, an author's citations increased as their median number of co-authors per publication began to increase. However, these effects of co-authorship on performance remained positive only up to a certain threshold, namely in correspondence of a median centred number of five (or, equivalently, a non-centred number of 12) co-authors per publication. Beyond this threshold, a further increase in the median number of co-authors per publication degraded the author's performance. For example, for an author with a centred median number of two co-authors per publication (i.e., one standard deviation below five, corresponding to a non-centred number of nine co-authors), a one-unit increase in the median led to an increase in the expected citation count by a factor of 1.019 (i.e., by 1.88%). However, for an author with a centred median number of eight (or a non-centred number of 15) co-authors per publication (i.e., one standard deviation above five), a one-unit increase in the median led to a decrease in the expected citation count by a factor of 0.986 (i.e., by 1.43%). The statistically significant negative effects of the minimum number of co-authors per publication on authors' performance further corroborate these results. If the minimum number of co-authors with whom an author has ever published increased by one, the author would experience an expected decrease in citation count by a factor of 0.964 (i.e., by 3.59%), holding all other variables constant.

We further examined the effects of collaborative patterns on performance. A graphical representation of the co-authorship network is shown in **Figure 10 and 11**.

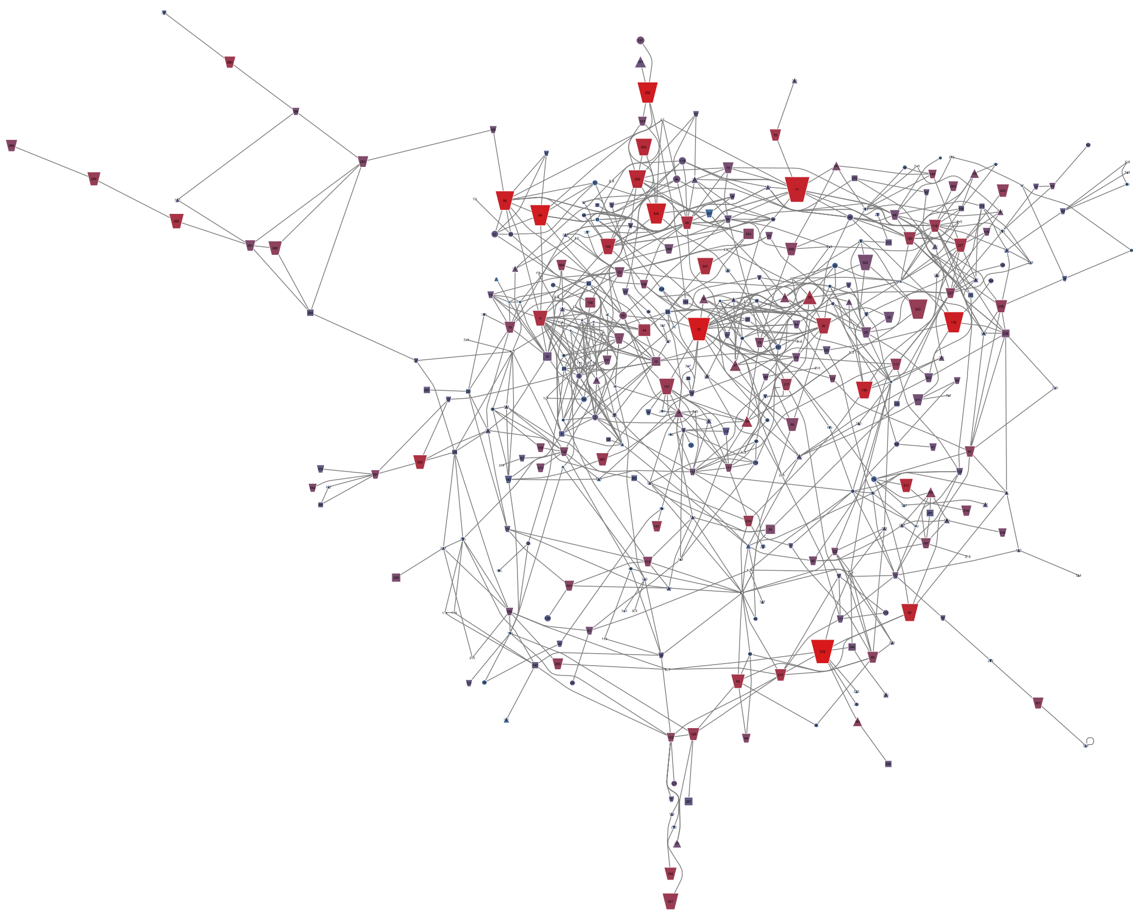

Figure 10 The largest connected component of academics from Faculty of Imperial College Medicine extracted from the entire co-authorship network. Academic rank sets the z coordinate (professor nodes on top and lecturer nodes on bottom) and shape indicates rank (circle = lecturer, triangle = senior lecturer, square = reader, trapezius = professor). Increasing node area indicates increasing citation count. Node colour is determined by h-index (red = max, blue = min).

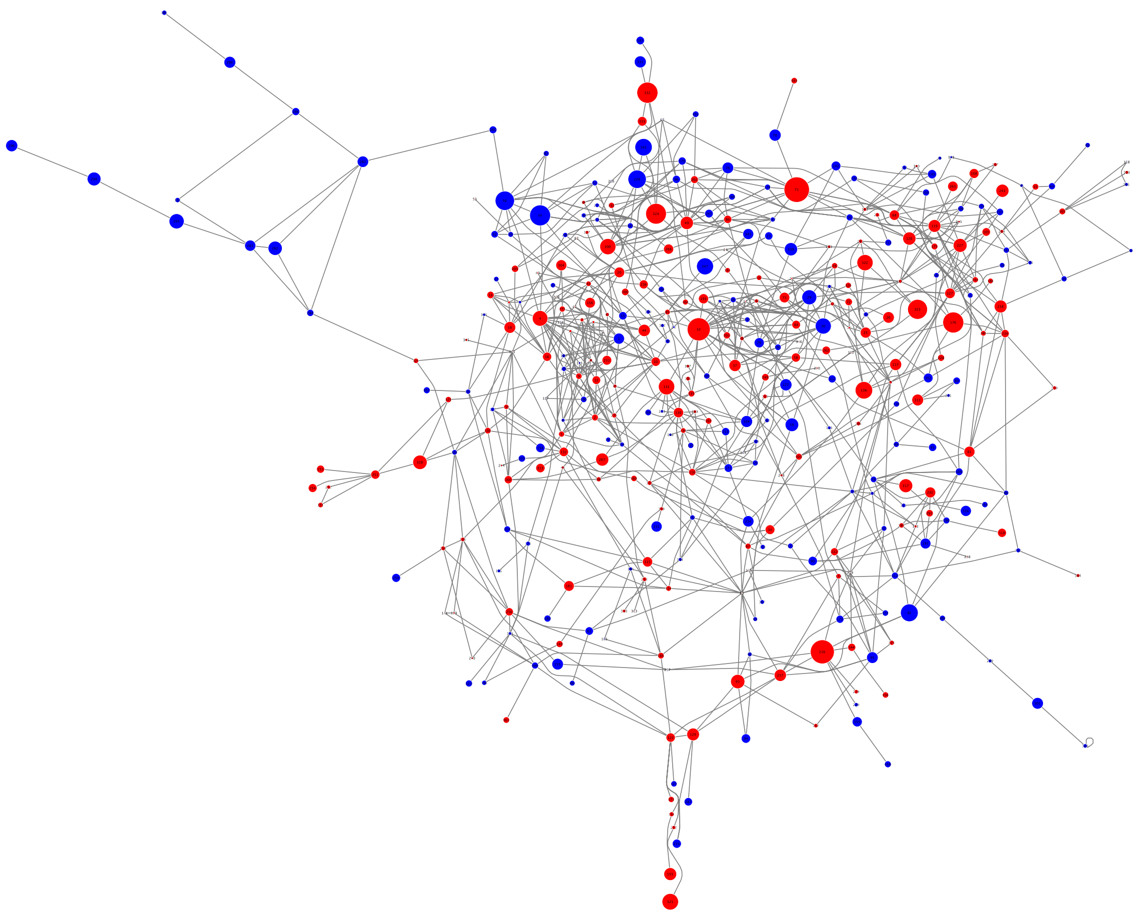

Figure 11: Increasing node area indicates increasing citation count. Node colour is determined by physician status (red = physician, blue = non-physician)

Models 1 through to 3 in **Table 14** include various centrality measures that have a statistically significant positive effect on performance. Having many collaborators (i.e., high degree), being connected to many others who are, in turn, highly central themselves (i.e., high eigenvector), lying on the shortest paths between many pairs of others (i.e., high betweenness), and being able to reach many others in very few steps (i.e., high closeness) contributed towards an increase of an author's performance.

Our findings provide evidence in favour of the positive effects of open network structures and brokerage on performance. As indicated by Model 3, clustering had a statistically significant negative effect on citation counts, while, consistently, Models 1, 2 and 4 suggest that lack of network constraint (i.e., availability of brokerage opportunities) had positive effects on performance. Collaborating with others who were not collaborating themselves thus enabled authors to improve their performance.

We then estimated the effects of an author's different positions in the sequences of co-authors on the author's total number of citations received. We distinguished between non-alphabetised and alphabetised multi-authorship. In the case of publications with authors sequenced non-alphabetically, and when only the main effects of position-related variables were tested (i.e., in Models 1 through to 3), the first and last positions had a statistically significant positive effect on total number of citations. The middle position in the by-line also had a positive effect on performance. For publications with authors sequenced alphabetically, only the first position had a positive effect on total number of citations.

Our results changed when interaction effects between positions in by-line and academic ranks were added (Model 4). In the case of publications with co-authors sequenced non-alphabetically, the first position in the list of co-authors maintained its positive effect on total number of citations. Authors with one additional publication in which they were first-listed in non-alphabetical order experienced an increase in their expected number of citations by a factor of 1.047 (i.e., by 4.71%), holding all other variables constant. However, the effect of being last-listed on performance changed sign from positive to negative for lecturers, but remained positive for higher academic ranks, especially for professors. For lecturers with one additional publication in which they were last-listed in non-alphabetical order, the expected number of citations decreased by a factor of 0.869 (i.e., by 13.04%), holding all other variables constant. By contrast, a professor with an average value of (centred) constraint (i.e., zero) experienced an increase by a factor of 1.048 (i.e., by 4.85%), holding all other variables constant. Finally, the first position on publications with co-authors sequenced alphabetically retained its positive effects on performance only for lecturers. Being first-listed alphabetically in one additional publication enabled a lecturer to obtain an increase in the expected citation count by a factor of 1.087 (i.e., by 8.78%), holding all other variables constant. Thus, even when all co-authors had equally contributed to the joint publication, the first-listed author with a junior academic position could secure extra credit or attention by the community.

Structural holes interacted with academic rank and last position in the non-alphabetised sequence of co-authors. An increase in brokerage opportunities was found to be less beneficial to professors that appeared as the last-listed co-authors in a large number of publications than to professors that were last-listed in fewer publications. Alternatively, the positive effects that an increase in the number of publications in which a professor was last-listed had on the professor's research performance were mitigated by an increase in the professor's brokerage opportunities. Unlike other academics in junior positions, authors in senior academic roles acting as the coordinators of many groups of researchers, as reflected by their last position in a large number of publications, could therefore gain advantages from socially cohesive network structures. Based on the estimated coefficients from Model 4, **Figure 12** shows the combined effects of structural holes and last position in the by-line on the expected performance of a female professor at the Institute of Clinical Sciences, when all other variables are held constant at their means.

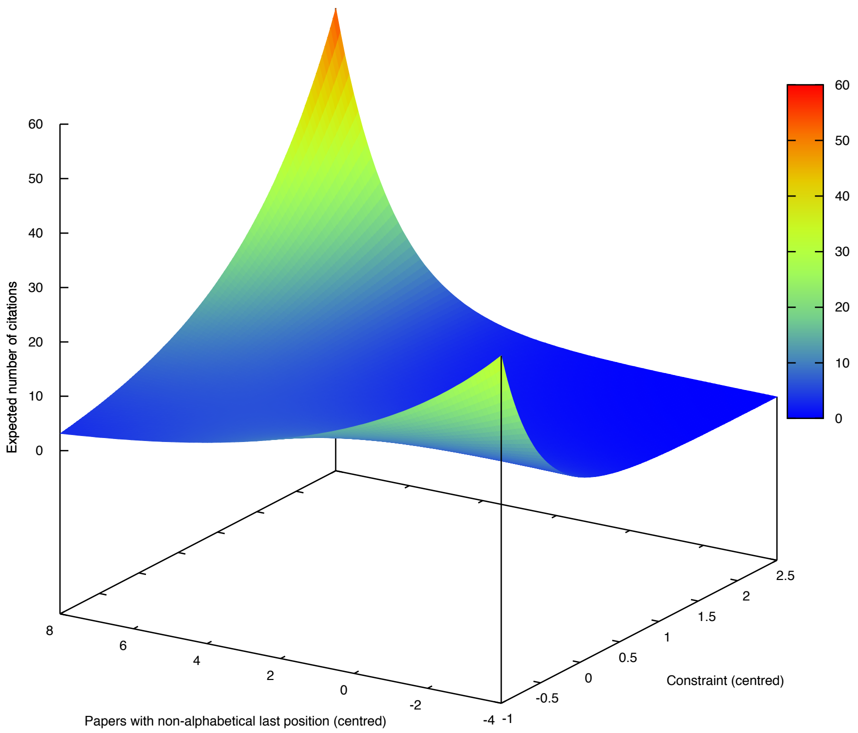

Figure 12: Combined effects of structural holes and last non-alphabetical position on the expected citation count of a female professor at the Institute of Clinical Sciences, when all other independent variables are held constant at their means.

5.5 DISCUSSION

The principal findings of our study suggest that social network analysis can be suitably applied to the investigation of academic career status, progression and the extent of collaboration in healthcare research. It confirms that collaboration in healthcare research has non-trivial effects on healthcare research performance throughout all academic ranks. It also offers an additional approach to the evaluation of academic success, which can augment current appraisal strategies in connection with the performance of academic research in healthcare.

Our results have also implications for the role of gender in modern academic healthcare research progression. In the past women were very likely to face discrimination in academic medicine, which hindered peer-review publications and delayed advancement in their careers.(*144*) The "gender gap" may still exist, but our study corroborates other findings suggesting that gender equality is becoming more widespread in academic medicine.(*145*)

Healthcare research depends heavily on non-clinical scientific research, and consequently non-clinical scientific research is cited three to five times more than clinical research.(*78*) Additionally, academic physicians tend to split their time between clinical and research activities, which may contribute to a lower research output.(*108*) Not surprisingly, our study has shown that academic physicians had lower citation counts than academic non-physicians. Current systems for assessing academic performance often do not consider these differences, though decisions on academic promotion are largely based on scholarly productivity.(*146, 147*) Academic medicine requires excellent physicians, who greatly facilitate translation of laboratory research into clinical practice.(*121*) Unless assessment and promotion systems recognise the dual role of academic physicians, then there is a risk that they will become increasingly isolated from academic medicine, which would ultimately lead to a decline in research quality.(*121*) Furthermore, the discrepancy in citations between academic physicians and academic non-physicians highlights the underlying need for greater collaboration between these two groups of researchers and for higher inter-group social cohesion that may facilitate translational research.

Inappropriate authorship and the subsequent lack of transparency and accountability threaten the integrity of scientific research.(*126, 147*) Honorary authorship, where individuals are named as authors but have not contributed significantly to the work, has driven a large increase in the number of co-authors per publication over the last fifty years.(*125*) Our study identified non-linear effects of number of co-authors per publication on citations, and suggested that publications with more than five co-authors were disadvantageous to a healthcare researcher's performance. Our study also suggests that in publications where all co-authors provided an equal contribution as indicated through an alphabetical listing of their names, perhaps other researchers unconscientiously gave the first author extra credit or attention. This supports the notion that scientific journals should reinforce authorship policies and ensure that each author has made a genuine contribution to the published work.(*126*) Conversely, we did not find support in favour of the hypothesis that solo-authored publications were associated with higher citation counts, which emphasises the importance of collaboration and multi-authorship in scientific research. These findings are in qualitative agreement with recent studies that have suggested a paradigm shift towards teamwork in scientific research, based on the evidence that teams tend to produce more frequently cited research than individuals.(*148*) Combined, our results therefore suggest that researchers should strike a balance between solo authorship and excessive co-authorship so as to optimise the size of their collaborative teams and, ultimately, sustain their citations and research quality.

Over the years teamwork has witnessed an increase in the number of co-authors per publication. As a result, it has become imperative that a proper assessment of research performance through multi-authored publications be based on the identification of the most important co-authors that played fundamental roles in the scientific work.(*124*) Typically, the first-listed co-author is the researcher who has carried out most of the work, and is usually a student, research fellow or junior lecturer.(*124*) The last-listed co-author often contributes substantially to research concept, experimental design and guidance, and is conventionally the mentor or supervisor of the first-listed co-author.(*124*) Our study has suggested that to progressively improve research performance, junior researchers should concentrate on producing publications in which they appear as the first-listed authors. Subsequently, when they are promoted to more senior academic roles, they should aim for publications in which they are the last-listed authors, thus emphasising their mentoring and supervisory roles. If promotion committees use bibliometric indicators to measure research performance, then to produce accurate and fair assessments they should consider the number of co-authors and authorship position on each publication.

Over the years contemporary medicine has witnessed increasingly blurring boundaries between distinct specialties.(*14, 149*) The lone researcher struggles to answer prominent questions without collaborating with other scientists, often from other disciplines or institutions.(*149*) Evidence suggests that healthcare studies are moving towards a more collaborative and team-based approach to research.(*14, 149*) However, fostering a collaborative culture in academic medicine is challenging because promotion committees and tenure systems often discourage collaboration by focusing on a researcher's independent contributions.(*149*) Furthermore, the conventional mind-set of funding agencies and scientific journals supports the independent researcher over the collaborator.(*149*) Our study has shown that, overall, collaboration has positive effects on research performance and, more importantly, that researchers should embrace different collaborative patterns as their academic career progresses. For example, junior researchers, in addition to securing the first position in the sequence of co-authors of their publications, should aim for low network constraint by forging ties that span a large number of structural holes. They should build diverse and innovative research teams that will provide them with opportunities of brokerage between otherwise disconnected collaborators from whom they will be able to acquire new non-overlapping knowledge, ideas, and insights. The converse is true when researchers are promoted to more senior academic roles and stages of career, at which some degree of network closure is likely to facilitate the development of an established scientific vision and research strategy. Previous studies have shown that networks rich in structural holes may thrive in generating innovative ideas through novel recombination of heterogeneous pools of knowledge.(*133*) Our results have extended these studies by suggesting that the benefits of structural holes may be contingent on the researcher's stage of career. Unlike a junior researcher, a scientist in a senior position, acting as the coordinators of a large group of researchers, can leverage on a cohesive network structure rich in third-party relationships to promote the transfer and sharing of complex knowledge, consolidate the group's identity, and minimise the risk of missing important opportunities of cross-fertilisation of ideas within the group. (*150*)

Our study has four fundamental implications for healthcare research: (i) there are advantages of intellectual cooperation over solo authorship across all academic ranks; (ii) while researchers may gain from expanding their network of collaborators, they should consider redistributing collaborators across multiple publications, so as to refrain from publishing papers with an excessively large number of co-authors; (iii) there are differential benefits of authorship attribution and collaborative strategies depending on the stage of a researcher's career; (iv) academic institutions and funding agencies should encourage and reward a team-based approach to research by engaging in a systematic and transparent measurement of collaboration with a view to promoting the best research practice that would translate into the highest-quality patient care.

Our research has several limitations. It seems reasonable to argue that co-authorship represents one of the main forms of scientific collaboration.(*123*) However, there are certainly other peripheral or indirect forms of collaboration that are not reflected in formal co-authorship, and yet represent genuine instances of intellectual co-operation. Researchers may motivate, inspire and contribute to each other's research without always being listed as co-authors, for example by mentoring and supervising junior colleagues, or by providing commentary at conferences, workshops, and professional meetings.(*122*) Although the results from this study were derived from a large-scale collaboration network, the data were obtained from a single institution and a single research speciality. This may affect the degree of generalisability of our findings to other contexts. We used a single citation database to extract bibliometric data, but Scopus, Web of Science, and Google Scholar may produce different citation counts for healthcare researchers.(*104*) Even though special care has been taken to disambiguate authors' names, multiple databases could be used to further validate the bibliometric data concerning authors who have a common name, have worked in numerous institutions, or have diverse research interests.

5.6 CONCLUSION

In this study we have demonstrated that it is possible to gauge scientific collaboration by studying the structure of co-authorship networks. Our results indicate a robust association between healthcare research collaboration and superior research performance. Whilst junior researchers (lecturers) and senior researchers (professors) may benefit from distinctive network strategies, there are advantages of intellectual cooperation and joint publications over solo research and authorship, across all academic ranks. As a consequence we suggest that academic institutions should encourage research collaboration. Most importantly, they should also aim to engage in systematic measurements of scientific collaboration with a view to helping researchers to improve their academic productivity and quality. Healthcare promotion committees and tenure systems should not ignore physician status to support research output, but rather encourage collaborative activity between academic physicians and non-physician academics in order to optimise cross-boundary research and improve performance. Scientific journals and funding agencies should also reward teamwork to foster collaboration. Ultimately, the incentives for scientific collaboration will be greater if transparency is improved in healthcare research. The salience of collaborative networks for healthcare research can inspire the development of effective measures for encouraging new academic associations, sustaining the dissemination of novel research concepts, and addressing complex multifaceted healthcare demands. This, in turn, may help promote the best research practice that translates into the highest-quality patient care.

5.7 CHAPTER SUMMARY

The objective of this chapter was to understand the impact of collaborative patterns and norms for credit assignment upon research performance in biomedical science remains. This chapter aimed to investigate the association between individual biomedical researchers' scientific success and positions in collaboration networks and in bylines of publications.

We examined the collaborative and authorship practices of 525 biomedical researchers within a large healthcare research institution in the United Kingdom over a nine-year period. We constructed the co-authorship network in which nodes are the researchers, and links are established between any two researchers if they co-authored one or more articles. For each researcher, we recorded the positions held in the bylines of all articles published in each three-year interval, and calculated the number of citations these articles subsequently accrued. We estimated maximum likelihood negative binomial panel regression models.

Our analysis suggests that collaboration sustained success, yet excessive co-authorship did not. Researchers could benefit from different authorship and collaborative practices depending on their career stage. Last positions in non-alphabetized bylines were beneficial for higher academic ranks, but not for junior ones. A professor could increase the expected citation count by 4.85% if last-listed non-alphabetically in one additional publication; yet, a lecturer suffered from a reduction of 13.04%. First positions in alphabetized bylines sustained performance, but only for junior academics. A lecturer could increase the expected citation count by 8.76% if first-listed alphabetically in one additional publication. Moreover, whilst junior researchers amplified success by brokering among otherwise disconnected collaborators, senior researchers prospered from socially cohesive networks, rich in third-party relationships.

These results may help biomedical scientists shape successful careers and research institutions develop effective assessment and recruitment policies that will ultimately sustain the quality of healthcare research and patient care.

In the next chapter we will summarise the results of this thesis, discuss the limitations, offer direction on further work, and finally comment on potential impact of this research.

6 CONCLUSION

We have demonstrated that the most common research performance indicators that are currently being used are number of publications, number of citations, Impact Factor, research funding, degree of co-authorship, and h index. However, there was limited investigation of feasibility, validity, reliability and acceptability, although the utility of these indicators was adequately described.

We have identified that differences exist between the citation databases Google Scholar, ISI and Scopus. The Hirsch index emerged as the most consistently calculated bibliometric between the databases in healthcare researchers, but this was influenced by researcher specific characteristics such as age and physician status.

Our findings support the construct validity of the h index as a tool to assist academic promotion committees. It also enhances the assessment value of the h index through the development of the indices h^2 upper, h^2 centre, h^2 lower and sRM. These can be applied in combination with the *h* index to provide additional objective evidence to appraise the performance and impact of a healthcare researcher.

We have proposed a number of suitable network metrics for studying the structure of a co-authorship network in healthcare research, and have suggested objective measures of social capital. Although researchers at different stages of their career may benefit from a spectrum of network approaches to improve research performance, overall, academic collaboration benefits research performance throughout the academic hierarchy.

There are several limitations to this thesis, which should be considered when interpreting the results. In this thesis we have principally used bibliometric indicators, such as publication number, citation count and h index, to quantify research output. Quantitative assessment of publication and citation data has grown globally, particularly in university and government organisations, as well as by policymakers, research companies, and researchers themselves.(*151*) Despite the large-scale use of these bibliometric indicators to measure research performance, they are often too simple to comprehend the intricacy and multi-dimensional nature of research output. For instance, commonly used bibliometrics have recognised flaws, which have previously been described in this thesis. More importantly, using bibliometrics may alter a scientist's behaviour, and this phenomenon has already been described in the field of economics as Goodhart's law: "when a measure becomes a target, it ceases to become a good measure".(*152*) A purely bibliometric based system to measure performance may force supervisors to capitalise on these indicators, who may coerce researchers into popular research topics. Subsequently this may inhibit research driven by freedom, inquisitiveness and innovation. Crucially, fundamental attributes in university researchers such as teaching and mentoring of students are difficult to assess with bibliometrics.(*153*) Although bibliometrics are disadvantaged, these quantitative measures have been preferred by assessment organisations such as the REF, because they are easier to acquire. However, the goals of assessment systems are usually quality driven, and there is no convincing evidence that quantity predictably produces quality.(*154*)

In the principal studies of this thesis, which evaluated research performance and collaboration, the sample population was derived from a single institution (Imperial College London) and specialty (Faculty of Medicine). Although a large sample population was investigated and the results from these studies may have practical implications for research performance assessment at Imperial College London, the differences between institutions and specialties makes it difficult to generalize the findings. Furthermore, the studies focused on individual performance, but did not consider performance assessment at departmental, institutional or global level. Our objective measures of collaboration were based on the assumption that co-authorship represents scientific collaboration.(*123*) But, we know that not all forms of intellectual co-operation are echoed through co-authorship. Researchers may stimulate, encourage and support another researcher's work without being a co-author. Examples include teaching, mentoring and coaching junior researchers.(*122*)

The outcomes of this thesis suggest that further studies are essential to validate objective measures of research performance so that they can be used more reliably in assessment and appraisal of healthcare researchers. Future studies should consider research performance assessment across specialties, as well as investigating assessment tools at departmental, institutional and global levels. Furthermore, studies are required to investigate the true causal relationship between grant/funding allocation and enhanced research performance. Eventually, future work may include the development of a tool, such as a standardised balanced scorecard, to measure research performance across specialties and institutions. Contemporary measures of research performance such as the h index have a role in academic promotion decisions, but for fair and accurate evaluation, promotion committees may require multiple sources to quantify these measures and use a balanced approach integrating peer-review. Academic healthcare organizations should promote intellectual collaboration at all levels to drive research performance, as well as fostering collaboration between academic physicians and non-physicians, which may narrow the divide between basic science and translational medicine. The ultimate aim is to support the translation of research to encourage healthcare innovation with greater societal and economic impact.

7 REFERENCES

1. REF, Research Excellence Framework. http://www.hefce.ac.uk/Research/ref/. (2010).
2. C. Lenfant, Shattuck lecture--clinical research to clinical practice--lost in translation? *The New England journal of medicine* **349**, 868 (Aug 28, 2003).
3. C. Global Burden of Disease Study, Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, (Jun 7, 2015).
4. J. Avorn, The $2.6 billion pill--methodologic and policy considerations. *The New England journal of medicine* **372**, 1877 (May 14, 2015).
5. H. Ashrafian, C. Rao, A. Darzi, T. Athanasiou, Benchmarking in surgical research. *Lancet* **374**, 1045 (Sep 26, 2009).
6. S. Awasthi *et al.* (International Campaign to Revitalise Academic Medicine 2005).
7. J. P. Ioannidis, Academic medicine: the evidence base. *BMJ* **329**, 789 (Oct 2, 2004).
8. D. Wilkinson, ICRAM (the International Campaign to Revitalise Academic Medicine): agenda setting. *BMJ* **329**, 787 (Oct 2, 2004).
9. ARC, Excellence in Research for Australia: http://www.arc.gov.au/excellence-research-australia. (2010).
10. J. E. Hirsch, An index to quantify an individual's scientific research output. *Proc Natl Acad Sci U S A* **102**, 16569 (Nov 15, 2005).
11. J. E. Hirsch, Does the H index have predictive power? *Proc Natl Acad Sci U S A* **104**, 19193 (Dec 4, 2007).
12. L. Bornmann, H. D. Daniel, The state of h index research. Is the h index the ideal way to measure research performance? *EMBO Rep* **10**, 2 (Jan, 2009).
13. J. S. Coleman, Social capital in the creation of human capital. *American Journal of Sociology* **94**, S95 (1988).
14. S. Lee, B. Bozeman, The Impact of Research Collaboration on Scientific Productivity. *Social Studies of Science* **35**, 673 (2005).
15. V. M. Patel *et al.*, What does leadership in surgery entail? *ANZ J Surg* **80**, 876 (Dec, 2010).
16. W. W. Souba, The new leader: new demands in a changing, turbulent environment. *J Am Coll Surg* **197**, 79 (Jul, 2003).

17. G. P. Latham, L. Saari, The application of social learning theory to training supervisors through behavioral modeling. *Journal of Applied Psychology* **64**, 239 (1979).
18. G. A. Yukl, *Leadership in Organizations*. (Prentice Hall, Englewood Cliffs, NJ, ed. 4th, 1998).
19. R. M. Kanter, Collaborative Advantage: The Art of Alliances. *Harvard Business Review* **72**, 96 (1994).
20. S. Kanter, What is academic medicine? *Academic Medicine* **8**, 205 (2008).
21. F. M. E. Uzoka, A fuzzy-enhanced multicriteria decision analysis model for evaluating university Academics' research output. *Information Knowledge Systems Management* **7** 273 (2008).
22. IAF, Institutional Assessment Framework . http://www.deewr.gov.au/HigherEducation/Resources/Pages/InstPerfPortfolios.aspx. (2009).
23. A. Liberati *et al.*, The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* **339**, b2700 (2009).
24. AHMD, The American Heritage® Medical Dictionary. (2010).
25. EBM, Oxford Centre for Evidence-based Medicine. http://www.cebm.net/levels_of_evidence. (2009).
26. K. Ahmed, H. Ashrafian, G. B. Hanna, A. Darzi, T. Athanasiou, Assessment of specialists in cardiovascular practice. *Nat Rev Cardiol* **6**, 659 (Oct, 2009).
27. V. Wass, C. Van der Vleuten, J. Shatzer, R. Jones, Assessment of clinical competence. *Lancet* **357**, 945 (Mar 24, 2001).
28. P. W. Armstrong, M. M. Caverson, L. Adams, M. Taylor, P. M. Olley, Evaluation of the Heart and Stroke Foundation of Canada Research Scholarship Program: research productivity and impact. *Can J Cardiol* **13**, 507 (May, 1997).
29. C. G. Brown, H. R. Coward, R. C. Stowe, Assessing the scientific strength of Chile. *Arch Biol Med Exp (Santiago)* **24**, 37 (1991).
30. A. R. Buss, Evaluation of Canadian psychology departments based upon citation and publication counts. *Canadian Psychological Review/Psychologie canadienne*.**17**, 143 (1976).
31. J. P. Byrnes, C. C. McNamara, Evaluating doctoral programs in the developmental sciences. *Developmental Review*.**21**, 326 (2001).
32. J. Divic, Survey of Croatian science from Science Citation Index between 1985 and 1992 scientometrics - one possible approach. *Periodicum Biologorum* **96**, 187 (1994).
33. B. G. Druss, S. C. Marcus, Tracking publication outcomes of National Institutes of Health grants. *American Journal of Medicine* **118**, 658 (2005).
34. L. B. Ellwein, M. Khachab, R. H. Waldman, Assessing research productivity: evaluating journal publication across academic departments. *Acad Med* **64**, 319 (Jun, 1989).
35. N. S. Endler, Research productivity and scholarly impact of Canadian psychology departments. *Canadian Psychological Review/Psychologie canadienne*.**18**, 152 (1977).
36. N. S. Endler, Where the "stars" are: The 25 most cited psychologists in Canada (1972-1976). *Canadian Psychological Review/Psychologie canadienne*.**20**, 12 (1979).
37. N. S. Endler, J. P. Rushton, H. L. Roediger, Productivity and scholarly impact (citations) of British, Canadian, and U.S. departments of psychology (1975). *American Psychologist*.**33**, 1064 (1978).
38. M. E. Falagas, P. A. Papastamataki, I. A. Bliziotis, A bibliometric analysis of research productivity in parasitology by different world regions during a 9-year period (1995-2003). *BMC Infectious Diseases* **2334**, 56 (2006).

39. R. A. Gordon, Research productivity in master's-level psychology programs. *Professional Psychology: Research and Practice*.**21**, 33 (1990).
40. R. A. Gordon, P. J. Vicari, Eminence in social psychology: A comparison of textbook citation, Social Sciences Citation Index, and research productivity rankings. Feb 1992. *Personality and Social Psychology Bulletin*.**18**, 26 (1992).
41. D. Hendrix, An analysis of bibliometric indicators, National Institutes of Health funding, and faculty size at Association of American Medical Colleges medical schools, 1997-2007. *Journal of the Medical Library Association* **96**, 324 (2008).
42. N. Housri, M. C. Cheung, L. G. Koniaris, T. A. Zimmers, Scientific Impact of Women in Academic Surgery. *Journal of Surgical Research* **148**, 13 (2008).
43. P. Kaplan, W. J. Mysiw, W. S. Pease, Academic productivity in physical medicine and rehabilitation. *American Journal of Physical Medicine and Rehabilitation* **71**, 81 (1992).
44. G. Koren *et al.*, Patient-based research in a tertiary pediatric centre: A pilot study of markers of scientific activity and productivity. *Clinical and Investigative Medicine* **20**, 354 (1997).
45. J. Lee, K. L. Kraus, W. T. Couldwell, Use of the h index in neurosurgery. Clinical article. *J Neurosurg* **111**, 387 (Aug, 2009).
46. M. A. Lichtman, D. Oakes, The productivity and impact of The Leukemia & Lymphoma Society scholar program: The apparent positive effect of peer review. *Blood Cells, Molecules, and Diseases* **27**, 1020 (2001).
47. E. S. Soteriades, M. E. Falagas, Comparison of amount of biomedical research originating from the European Union and the United States. *British Medical Journal* **331**, 192 (2005).
48. L. Bornmann, G. Wallon, A. Ledin, Does the committee peer review select the best applicants for funding? An investigation of the selection process for two European molecular biology organization programmes. *PLoS ONE [Electronic Resource]* **3**, e3480 (2008).
49. P. A. Bovier, H. Guillain, T. V. Perneger, Productivity of medical research in Switzerland. *Journal of Investigative Medicine* **49**, 77 (2001).
50. C. Castellano, F. Radicchi, On the fairness of using relative indicators for comparing citation performance in different disciplines. *Archivum Immunologiae et Therapiae Experimentalis* **57**, 85 (2009).
51. A. M. Colman, S. P. Grant, E. Henderson, Performance of British university psychology departments as measured by number of publications in BPS journals. Win, 1992-1993. *Current Psychology: Research & Reviews*.**11**, 362 (1992).
52. P. Devos, H. Lefranc, E. Dufresne, R. Beuscart, From bibliometric analysis to research policy: the use of SIGAPS in Lille University Hospital. *Stud Health Technol Inform* **124**, 543 (2006).
53. G. A. Fava, J. Guidi, N. Sonino, How Citation Analysis Can Monitor the Progress of Research in Clinical Medicine. *Psychotherapy and Psychosomatics*.**73**, 331 (2004).
54. B. Groneberg-Kloft *et al.*, Institutional operating figures in basic and applied sciences: Scientometric analysis of quantitative output benchmarking. *Health Research Policy and Systems* **6**, (2008).
55. J. M. Innes, The utility of a citation index as a measure of research ability in psychology. Jul 1973. *Bulletin of the British Psychological Society*.**26**, 227 (1973).
56. J. Koskinen *et al.*, How to use bibliometric methods in evaluation of scientific research? An example from Finnish schizophrenia research. *Nordic Journal of Psychiatry* **62**, 136 (2008).
57. G. Lewison, G. Thornicroft, G. Szmukler, M. Tansella, Fair assessment of the merits of psychiatric research. *British Journal of Psychiatry* **190**, 314 (2007).
58. A. Michalopoulos, M. E. Falagas, A bibliometric analysis of global research production in respiratory medicine. *Chest* **128**, 3993 (Dec, 2005).

59. F. Radicchi, S. Fortunato, C. Castellano, Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 17268 (2008).
60. A. Rostami-Hodjegan, G. T. Tucker, Journal impact factors: a 'bioequivalence' issue? *British Journal of Clinical Pharmacology* **51**, 111 (2001).
61. O. Tutarel, Geographical distribution of publications in the field of medical education. *BMC Medical Education* **2**, 1472 (2002).
62. D. Ugolini, A. Bogliolo, S. Parodi, C. Casilli, L. Santi, Assessing research productivity in an oncology research institute: The role of the documentation center. *Bulletin of the Medical Library Association* **85**, 33 (1997).
63. D. Ugolini, C. Casilli, G. S. Mela, Assessing oncological productivity: Is one method sufficient? *European Journal of Cancer* **38**, 1121 (2002).
64. E. Campos Jimenez, A. Campos Ferrer, Technological analysis of immunology in Spain, 2004-2005. *Inmunologia* **26**, 55 (2007).
65. L. Bornmann, H. Daniel, Selecting scientific excellence through committee peer review – A citation analysis of publications previously published to approval or rejection of post-doctoral research fellowship applicants. *Scientometrics* **68**, 427 (2006).
66. C. Wenneras, A. Wold, Nepotism and sexism in peer-review. *Nature* **387**, 341 (May 22, 1997).
67. S. Brkic, The impact of biomedical literature published in the province of Vojvodina on researchers in the world and in Yugoslavia. *Medicinski Pregled* **54**, 21 (2001).
68. D. David, M. Moore, A. Domuta, Romanian psychology on the international psychological scene: A preliminary critical and empirical appraisal. *European Psychologist***.7**, 153 (2002).
69. M. Jokic, Scientometric evaluation of the projects in biology funded by the Ministry of Science and Technology, Republic of Croatia, in the 1991-1996 period. *Periodicum Biologorum* **102**, 129 (2000).
70. L. Puljak, K. Vukojevic, S. L. Kojundzic, D. Sapunar, Assessing clinical and life sciences performance of research institutions in Split, Croatia, 2000-2006. *Croatian Medical Journal* **49**, 164 (2008).
71. L. Vuckovic-Dekic, B. Ribaric, B. Vracar, Implementation of various criteria for evaluating the scientific output of professional scientists and clinicians-scientists. *Archive of Oncology* **9**, 103 (2001).
72. A. W. Kellner, L. C. Ponciano, H-index in the Brazilian Academy of Sciences: comments and concerns. *An Acad Bras Cienc* **80**, 771 (Dec, 2008).
73. R. Mugnaini, A. L. Packer, R. Meneghini, Comparison of scientists of the Brazilian Academy of Sciences and of the National Academy of Sciences of the USA on the basis of the h-index. *Braz J Med Biol Res* **41**, 258 (Apr, 2008).
74. N. Torro-Alves, R. D. Herculano, C. A. S. Tercariol, O. K. Filho, C. F. O. Graeff, Hirsch's index: A case study conducted at the Faculdade de Filosofia, Ciencias e Letras de Ribeirao Preto, Universidade de Sao Paulo. *Brazilian Journal of Medical and Biological Research* **40**, 1529 (2007).
75. R. J. Epstein, Journal impact factors do not equitably reflect academic staff performance in different medical subspecialties. *Journal of Investigative Medicine* **52**, 531 (2004).
76. N. Rezaei-Ghaleh, F. Azizi, The impact factor-based quality assessment of biomedical research institutes in Iran: Effect of impact factor normalization by subject. *Archives of Iranian Medicine* **10**, 182 (2007).
77. I. B. Hickie, H. Christensen, T. A. Davenport, G. M. Luscombe, Can we track the impact of Australian mental health research? *Australian and New Zealand Journal of Psychiatry* **39**, 591 (2005).
78. P. O. Seglen, Why the impact factor of journals should not be used for evaluating research. *BMJ* **314**, 498 (Feb 15, 1997).

79. M. E. Falagas, E. I. Pitsouni, G. A. Malietzis, G. Pappas, Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J* **22**, 338 (Feb, 2008).
80. Google. (2004).
81. A. W. Harzing, Publish or Perish: http://www.harzing.com/pop.htm. . (2010).
82. V. M. Patel *et al.*, How has healthcare research performance been assessed? – a systematic review. *J R Soc Med* **104**, 251 (2011).
83. Thomson-Reuters, ISI Web of Knowledge Website: http://www.isiknowledge.com/. (2004).
84. Elsevier, Scopus Website: http://www.scopus.com. (2004).
85. C. Macilwain, Science economics: What science is really worth. *Nature* **465**, 682 (Jun 10).
86. R. S. Kaplan, D. P. Norton, The balanced scorecard--measures that drive performance. *Harv Bus Rev* **70**, 71 (Jan-Feb, 1992).
87. W. N. Zelman, D. Blazer, J. M. Gower, P. O. Bumgarner, L. M. Cancilla, Issues for academic health centers to consider before implementing a balanced-scorecard effort. *Acad Med* **74**, 1269 (Dec, 1999).
88. H. Ashrafian, V. M. Patel, P. Skapinakis, T. Athanasiou, Nobel Prizes in medicine: are clinicians out of fashion? *J R Soc Med* **104**, 387 (Sep, 2011).
89. M. Buxton, S. Hanney, How can payback from health services research be assessed? *Journal of health services research & policy* **1**, 35 (Jan, 1996).
90. R. Smith, Measuring the social impact of research. *BMJ* **323**, 528 (Sep 8, 2001).
91. A. P. Weiss, Measuring the impact of medical research: moving from outputs to outcomes. *The American journal of psychiatry* **164**, 206 (Feb, 2007).
92. J. Spaapen, H. Dijstelbloem, F. Wamelink, *Evaluating Research in Context: A method for comprehensive assessment*. (Consultative Committee of Sector Councils for Research and Development (COS), The Hague, ed. 2nd, 2007).
93. A. V. Kulkarni, B. Aziz, I. Shams, J. W. Busse, Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. *JAMA* **302**, 1092 (Sep 9, 2009).
94. L. Bornmann, H. Daniel, What Do We Know About the h Index? *Journal of the American Society for Information Science and Technology* **58**, 1381 (2007).
95. L. Bornmann, R. Mutz, H. Daniel, The h index research output measurement: Two approaches to enhance its accuracy. *Journal of Informetrics* **4**, 407 (2010).
96. J. Kulasegarah, J. E. Fenton, Comparison of the h index with standard bibliometric indicators to rank influential otolaryngologists in Europe and North America. *Eur Arch Otorhinolaryngol* **267**, 455 (Mar).
97. J. Bar-Ilan, Which h-index? — A comparison of WoS, Scopus and Google Scholar *Scientometrics* **74**, 257 (2008).
98. L. I. Meho, Y. Rogers, Citation counting, citation ranking, and h-index of human-computer interaction researchers: A comparison of Scopus and Web of Science. *Journal of the American Society for Information Science and Technology* **59**, 1711 (2008).
99. J. Qiu, Scientific publishing: identity crisis. *Nature* **451**, 766 (Feb 14, 2008).
100. J. Mingers, E. Lipitakis, Counting the citations: a comparison of Web of Science and Google Scholar in the field of business and management. *Scientometrics* **85**, 613 (2010).
101. N. Bakkalbasi, K. Bauer, J. Glover, L. Wang, Three options for citation tracking: Google Scholar, Scopus and Web of Science. *Biomed Digit Libr* **3**, 7 (2006).
102. L. Bornmann, H.-D. Daniel, What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation* **64**, 45 (2008).
103. F. Radicchi, S. Fortunato, C. Castellano, Universality of citation distributions: toward an objective measure of scientific impact. *Proc Natl Acad Sci U S A* **105**, 17268 (Nov 11, 2008).

104. V. M. Patel *et al.*, Measuring academic performance for healthcare researchers with the H index: which search tool should be used? *Med Princ Pract* **22**, 178 (2013).
105. L. Bornmann, R. Mutz, H. Daniel, The h index research output measurement: Two approaches to enhance its accuracy. *Journal of Informetrics* **4**, 407 (2010).
106. C. R. Lucey, D. Sedmak, M. Notestine, W. Souba, Rock stars in academic medicine. *Acad Med* **85**, 1269 (Aug, 2010).
107. Imperial College London: http://www1.imperial.ac.uk/medicine/about/. (2011).
108. V. M. Patel *et al.*, Enhancing the h index for the objective assessment of healthcare researcher performance and impact. *Journal of the Royal Society of Medicine* **106**, 19 (2013).
109. G. E. P. Box, D. R. Cox, An analysis of transformations. , Series B, 26, 211–252. . *Journal of the Royal Statistical Society, Series B* **26**, 211 (1964).
110. P. J. Huber, *Robust Statistics*. (John Wiley & Sons, Inc. , New York, 1981).
111. F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*. (John Wiley & Sons, Inc., New York, 1986).
112. J. P. Stevens, *Applied multivariate statistics for the social sciences*. (Taylor & Francis, New York, 2009).
113. L. Bornmann, R. Mutz, S. E. Hug, H.-D. Daniel, A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants. *Journal of Informetrics* **5**, 346 (2011).
114. S. Cole, J. R. Cole, Scientific output and recognition: a study in the operation of the reward system in science. *Am Sociol Rev* **32**, 377 (Jun, 1967).
115. J. P. Ioannidis, More time for research: fund people not projects. *Nature* **477**, 529 (Sep 29, 2011).
116. B. M. Benway, P. Kalidas, J. M. Cabello, S. B. Bhayani, Does citation analysis reveal association between h-index and academic rank in urology? *Urology* **74**, 30 (Jul, 2009).
117. M. D. Bould *et al.*, h-indices in a university department of anaesthesia: an evaluation of their feasibility, reliability, and validity as an assessment of academic performance. *Br J Anaesth* **106**, 325 (Mar, 2011).
118. A. E. Rad, W. Brinjikji, H. J. Cloft, D. F. Kallmes, The H-index in academic radiology. *Acad Radiol* **17**, 817 (Jul, 2010).
119. C. M. Spearman, M. J. Quigley, M. R. Quigley, J. E. Wilberger, Survey of the h index for all of academic neurosurgery: another power-law phenomenon? *J Neurosurg* **113**, 929 (Nov, 2010).
120. S. C. Durso, C. Christmas, S. J. Kravet, G. Parsons, S. M. Wright, Implications of academic medicine's failure to recognize clinical excellence. *Clin Med Res* **7**, 127 (Dec, 2009).
121. D. J. Sheridan, Reversing the decline of academic medicine in Europe. *Lancet* **367**, 1698 (May 20, 2006).
122. R. Yousefi-Nooraie, M. Akbari-Kamrani, R. A. Hanneman, A. Etemadi, Association between co-authorship network and scientific productivity and impact indicators in academic medical research centers: a case study in Iran. *Health research policy and systems / BioMed Central* **6**, 9 (2008).
123. M. E. Newman, The structure of scientific collaboration networks. *Proc Natl Acad Sci U S A* **98**, 404 (Jan 16, 2001).
124. P. Mattsson, C. J. Sundberg, P. Laget, Is correspondence reflected in the author position? A bibliometric study of the relation between corresponding author and byline position. *Scientometrics* **87**, 99 (2010).
125. W. Al-Herz, H. Haider, M. Al-Bahhar, A. Sadeq, Honorary authorship in biomedical journals: how common is it and why does it exist? *Journal of medical ethics*, (Aug 17, 2013).

126. J. S. Wislar, A. Flanagin, P. B. Fontanarosa, C. D. Deangelis, Honorary and ghost authorship in high impact biomedical journals: a cross sectional survey. *BMJ* **343**, d6128 (2011).
127. NWB Team. (2006). Network Workbench Tool. Indiana University, Northeastern University, and University of Michigan. .
128. G. Ahuja, Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly* **45**, 425 (2000).
129. J. E. Perry-Smith, C. E. Shalley, The social side of creativity: a static and dynamic social network perspective. Academy of Management Review 28 (1), 89-106. *Academy of Management Review 28 (1), 89-106.* **28**, 89 (2003).
130. P. F. Bonacich, Power and centrality: A family of measures. *American Journal of Sociology* **92**, 1170 (1987).
131. L. C. Freeman, A set of measures of centrality based upon betweennessSociometry, 40, 35-41 (1977). *Sociometry* **40**, 35 (1977).
132. D. J. Watts, S. H. Strogatz, Collective dynamics of 'small-world' networks. *Nature* **393**, 440 (1988).
133. R. S. Burt, in *Harvard University Press*. (Cambridge MA, 1992).
134. R. S. Burt, Structural holes and good ideas. *American Journal of Sociology* **110**, 349 (2004).
135. M. Granovetter, The strength of weak ties. *American Journal of Sociology* **78**, 1360 (1973).
136. T. Tscharntke, M. E. Hochberg, T. A. Rand, V. H. Resh, J. Krauss, Author sequence and credit for contributions in multiauthored publications. *PLoS biology* **5**, e18 (Jan, 2007).
137. D. N. Laband, R. D. Tollison, Intellectual collaboration. *Journal of Political Economy* **108**, 632 (2000).
138. A. Dance, Authorship: Who's on first? *Nature* **489**, 591 (Sep 27, 2012).
139. M. O. Baerlocher, M. Newton, T. Gautam, G. Tomlinson, A. S. Detsky, The meaning of author order in medical research. *Journal of investigative medicine : the official publication of the American Federation for Clinical Research* **55**, 174 (May, 2007).
140. V. Yank, D. Rennie, Disclosure of researcher contributions: a study of original research articles in The Lancet. *Annals of internal medicine* **130**, 661 (Apr 20, 1999).
141. J. P. Drenth, Multiple authorship: the contribution of senior authors. *JAMA* **280**, 219 (Jul 15, 1998).
142. J. Jaccard, R. Turrisi, C. Wan, *Interaction effects in multiple regression*. (Sage , Newbury Park, CA, 1990).
143. S. Wuchty, B. F. Jones, B. Uzzi, The increasing dominance of teams in production of knowledge. *Science* **316**, 1036 (May 18, 2007).
144. N. A. Press, in *National Academy of Sciences; National Academy of Engineering; Institute of Medicine, Committee on Maximizing the Potential of Women in Academic Science and Engineering*. (Washington, DC, 2006).
145. R. Jagsi *et al.*, The "gender gap" in authorship of academic medical literature--a 35-year perspective. *The New England journal of medicine* **355**, 281 (Jul 20, 2006).
146. A. A. Atasoylu *et al.*, Promotion criteria for clinician-educators. *Journal of general internal medicine* **18**, 711 (Sep, 2003).
147. R. Wootton, A simple, generalizable method for measuring individual research productivity and its use in the long-term analysis of departmental performance, including between-country comparisons. *Health research policy and systems / BioMed Central* **11**, 2 (2013).
148. S. J. B. F. U. B. Wuchty, The increasing dominance of teams in production of knowledge. *Science* **316**, 1036 (1036).
149. J. J. Cohen, E. K. Siegel, Academic medical centers and medical research: the challenges ahead. *JAMA* **294**, 1367 (Sep 21, 2005).

150. G. Guerrero-Medina *et al.*, Supporting diversity in science through social networking. *PLoS biology* **11**, e1001740 (Dec, 2013).
151. Whitepaper using bibliometrics: A guide to evaluating research performance with citation data. http://ip-science.thomsonreuters.com/m/pdfs/325133_thomson.pdf. *Thomson Reuters*, (2008).
152. C. A. E. Goodhart, Problems of Monetary Management: The U.K. Experience. *Papers in Monetary Economics* (1975).
153. C. Macilwain, Halt the avalanche of performance metrics. *Nature* **500**, 255 (Aug 15, 2013).
154. C. F. Louis, G. Reed, The Current State and Recommendations for Meaningful Academic Research Metrics Among American Research Universities. A Report of the Research Metrics Working Group. *US Research Universities Futures Consortium*, 9 (2013).
155. A. C. Cameron, P. K. Trivedi, Regression Analysis of Count Data. *Cambridge: Cambridge University Press*, (1998).
156. J. Heckman, Ed., *Heterogeneity and State Dependence*, vol. University of Chicago Press (Chicago, 1981), vol. University of Chicago Press, pp. 91-139.
157. C. Hsiao, Analysis of Panel Data. *Cambridge: Cambridge University Press*, (2003).

8 APPENDICES

APPENDIX 1

Source	Study design	Study period	Indicator	Data source	Feasibility, validity, reliability, acceptability	Utility	Study population and setting	Sample sizes	Type of statistical analysis	Outcomes	Methodological limitations
Innes et al, 1973 (UK)(*55*)	RO	1948-1952	Citations	ISI	ND	Specialties	Speciality publications from the Journal of Experimental Psychology	73	Chi square test	Papers with 6 or more citations refer to same citation more often than other papers ($p < 0.01$)	Single data source
Buss et al, 1976 (Canada)(*30*)	RO	1973-1975	Publications, citations	ISI	Convergent validity	Specialties	Psychology departments in Canada	32	Correlation	Publication and citation measures highly correlated (range 0.69 - 0.92)	Single data source
Endler et al, 1977 (Canada)(*35*)	RO	1975	Publications, citations	ISI	Convergent validity	Individuals, specialties	Psychology departments in Canada	35	Correlation	Correlations between publications and citations within departments is 0.76 and within individuals correlation reduced to 0.41	Single data source
Endler et al, 1978 (Canada)(*37*)	RO	1975	Publications, citations	ISI	Convergent validity	Institutions	Psychology departments in UK, Canada and USA	180	Pearson's correlation	Institutional ranking changes when considering citation rates; correlation between all indicators; citations strongly correlated with reputational rank ($p < 0.001$)	Single data source
Endler et al, 1979 (Canada)(*36*)	RO	1972-1976	Publications, citations	ISI	ND	Individuals	Professors of psychology Ontario Universities, McGill and Memorial Hospital of New Foundland, Canada	25	Pearson's correlation	Low correlation between publications and citations	Single data source
Ellwein et al, 1989 (USA)(*34*)	RO	1985-1987	Publications, citations, co-authorship, IF	ISI	ND	Individuals, specialties	Research from medical school departments at University of Nebraska College of Medicine, USA	362 faculty, 12 departments	Descriptive	Departmental rankings significantly changed when Impact Factor considered	Single data source
Gordon et al, 1990 (USA)(*39*)	RO	1980-1987	Publications	ISI	ND	Institutions	Master's level programs in Psychology in USA universities	77	Correlation	Strong correlation between publications and American Psychological Association productivity score (log transformation of average standardised score on total publications and APA publications per faculty)	Single data source
Carole Ganz Brown et al, 1991 (USA)(*29*)	RO	1984	Citations	ISI	ND	Specialties, global	Scientific research in Chile vs. the World and strength in specialties	15 countries, 16 specialties	Descriptive	Chile internationally strong in biomedicine and clinical medicine but has relatively low citation numbers in these specialties in comparison to other countries	Single data source
Kaplan et al, 1992 (USA)(*43*)	RO	1988-1990	Publications, grant funding	ISI	ND	Specialties	Sports medicine departments in USA universities	20	Descriptive	Many departments successful without grant support; 8 departments produced 1 article per faculty member in 3 years; most productive departments based in Northeastern and Midwest USA	Single data source
Gordon et al, 1992 (USA)(*40*)	RO	1980-1990	Publications, citations	ISI, PsycINFO	Reliability, Convergent validity	Individuals	Psychologists	104	Correlation	Textbook citations and psycINFO correlated with ISI citation measures ($r = 0.36$, $p < 0.001$ and, $r = 0.28$, $p < 0.001$ respectively)	Textbooks not peer reviewed
Colman et al, 1993 (UK)(*51*)	RO	1983-1989	Publications, citations, co-authorship	ISI	Convergent validity	Institutions	Psychology departments in UK Universities	42	Pearson's correlation	Strong correlation between co-authorship performance and publication/citation ranking (r 0.43 - 0.62, $p < 0.01$)	Single data source

Source	Study design	Study period	Indicator	Data source	Feasibility, validity, reliability, acceptability	Utility	Study population and setting	Sample sizes	Type of statistical analysis	Outcomes	Methodological limitations
Divic et al, 1994 (USA)(*32*)	RO	1985-1992	Publications, grant funding	ISI	ND	Global	Scientific research in Croatia vs. the World	11 countries	Descriptive	Croatia < 0.5 papers per researcher (low compared to European countries but average compared to communist countries); main contributors are Institute Ruder Boskovic (40%), schools of medicine, hospitals (44%)	Single data source
Wennarås et al 1997 (Sweden)(*66*)	RO	Earliest publication - 1994	Publications, citations, co-authorship, gender, IF, peer review	MEDLINE, ISI	Convergent validity	Individuals	Scientists applying for postdoctoral fellowships at the Swedish Medical Research Council	114	Multiple regression analysis	Peer reviewers scored female applicants lower than males even though they had similar scientific productivity measured with bibliometrics ($p < 0.0001$)	Single data source
Ugolini et al, 1997 (Italy)(*62*)	RO	1991-1993	Publications, IF	Institutional	ND	Institutions	Oncology research in National Institute for Research on Cancer, Italy	1 (Publications n = 564)	Descriptive	66% of publications had a normalised IF score between 8-10; formula for fund allocation developed	Assumes IF is a quality measure for research
Koren et al, 1997 (Canada)(*44*)	RO	1993-1995	Publications, citations, grant funding	ISI	Convergent validity	Individuals	Clinical programs at The Hospital for Sick Children, Canada	44	Regression analysis	Strong correlation between funding and publications ($p < 0.001$, $p = 0.61$) and citation impact ($p < 0.001$, $p = 0.59$), and publications and citation impact ($p < 0.001$, $\rho = 0.88$)	Single data source
Armstrong et al, 1997 (Canada)(*28*)	RO	1980-1990	Publications, citations	ISI	ND	Individuals	Scholars and unsuccessful scholars from Heart and Stroke Foundation Canada	66 awarded scholarship, 126 not awarded scholarship	Students t test, ANOVA, Chi square test	Awarded Scholars had significantly more publications ($p = 0.0001$) and citations ($p = 0.026$); no significant difference in IF	Single data source, does not determine whether it is selection process that increases productivity.
Jokic et al, 2000 (Croatia)(*69*)	RO	1991-1996	Publications, citations	ISI	ND	Individuals	Biology projects funded by The Ministry of Science and Technology, Croatia	91 projects, 90 principal investigators , 494 researchers	Descriptive	21 projects , 31 PIs and 233 researchers with no publications indexed in ISI	Single data source
Rostami-Hodjegan et al, 2000 (UK)(*60*)	RO	1994-2000	IF	ISI	ND	Individuals	Professors of medicine at University of Sheffield Faculty of Medicine, UK	8	None	Professors 1-3 ranked on weighted impact are ranked 5-8 when using IF	Single data source
Lichtman et al, 2001 (USA)(*46*)	RO	1981-1990	Publications, citations, grant funding, peer review	ISI	ND	Individuals	Scholars and unsuccessful applicants of The Leukemia & Lymphoma Society America	124 scholars, 123 non-funded applicants	T test and Wilcoxon test	Funded scholars produced 3859 (60%) more publications with significantly greater citations/subject ($p < 0.0001$); when compared to control both scholar (30% more) and non-funded applicants (10% more) had more citations	Single data source, does not assess impact of the research on productivity
Bovier et al, 2001 (Switzerland)(*49*)	RO	1990-1994	Publications, IF, grant funding, peer review	MEDLINE	ND	Individuals	Medical research projects from Swiss National Science Foundation	399	Analysis of variance and linear regression, non parametric regression (LOWNESS), multiple linear regression, logistic regression	Grant amount and peer review scores strongly related to productivity	Assumes IF is a quality measure for research
Byrnes et al, 2001 (USA)(*31*)	RO	1992-1998	Publications, citations, grant funding, doctoral students, editorial responsibilities	ISI, PsycINFO, NIH	Feasibility	Individuals	Doctoral programs in developmental sciences in USA and Canadian universities	97 (faculty n = 802)	Correlation	Highly productive individuals skew results; when citations used to rank top programs only 73% of programs included from previous ranking; programs receiving most funding not necessarily the most productive; editorial responsibilities not correlated with productivity	Poor quality feasibility study
Vuckovic-Dekic et al, 2001 (Serbia)(*71*)	RO	1996-1999	Publications, citations, co-authorship	Institutional	ND	Individuals	Clinicians and basic scientists at Institute for Oncology and Radiology of Serbia	6 Clinicians, 4 basic scientists	Descriptive	Ranking of clinicians and basic scientists differed greatly when adjusted for publications, first authorship, co-authorship and partial authorship	Small sample size

Source	Study design	Study period	Indicator	Data source	Feasibility, validity, reliability, acceptability	Utility	Study population and setting	Sample sizes	Type of statistical analysis	Outcomes	Methodological limitations
Brkic et al, 2001 (Serbia)(*67*)	RO	1986-1997	Publications	ISI	ND	Individuals, institutions	Biomedical literature in Serbia	6979 publications	Descriptive	Highest production in 1995 and 1987; 16421 authors; mean authors/paper = 2.35	Single data source
Ugolini et al, 2002 (Italy)(*63*)	RO	1995	Publications, co-authorship, IF, GDP, population	ISI	ND	Global	Oncology research in European Union countries	16	Descriptive	Highest output ranked by publications: UK, Italy, Germany, France, The Netherlands; top five countries remained the same when adjusted for population; UK dropped to fourth when adjusted for GDP	Single data source
Tutarel et al, 2002 (Germany)(*61*)	RO	1995-2000	Publications	MEDLINE	ND	Global	Geographical distribution of medical education research in Academic Medicine and Medical Education Journals	2207 Academic Medicine publications, 746 Medical Education publications	Descriptive	Academic Medicine journal 25 different countries (USA 87.49%) and Medical Education 50 different countries (UK 42.63%, Australia 11.66%, USA 10.46%)	Single data source, MEDLINE only has the address of the 1st author, medical education articles not limited to the two journals
David et al, 2002 (Romania)(*68*)	RO	Earliest publication - 2002	Publications, IF	ISI, MEDLINE, PsycINFO	ND	Individuals, institutions	Academic psychologists in academic psychology programs in Romanian universities	7 famous psychologists, 6 psychology academic programs	Descriptive	Individuals: publications range 3 – 54; Institutions: publications range 0 – 84 and IF: 0 – 8	Does not include educational and administrative criteria; does not include all Romanian institutions
Fava et al, 2004 (Italy)(*53*)	RO	1981-2000	Publications, citations	ISI	ND	Global	Top ranking countries in clinical medicine research in the World	30	Descriptive	Publication rank = USA, UK Japan, Germany, France; citation rank = USA, Switzerland, Poland, Saudi Arabia, Spain; citation/paper rank = Netherlands, USA, Canada, Finland, Denmark	Single data source
Epstein et al, 2004 (Hong Kong)(*75*)	RO	2003	IF	ISI	ND	Specialties	IF of sample journals from four medical specialties: medicine, oncology, genetics, public and occupational health	Medicine journals = 100, oncology journals = 100, genetics journals = 100, public and occupational health journals = 89	Non-parametric Kruskal-Wallis test	Significant differences in IFs between specialties ($p < 0.006$)	Single data source
Michalopoulos et al, 2005 (Greece)(*58*)	RO	1995-2003	Publications, IF, GDP, population	ISI, MEDLINE	ND	Global	Worldwide respiratory medicine research	Western Europe, US, Japan, Canada, Asia, Oceania, Eastern Europe, Central and Latin America, Africa	Descriptive	Publication ranking: Western Europe, US, Japan, Canada, Asia, Oceania, Eastern Europe, Central and Latin America, Africa; adjusted for IF/population/GNIPC : Canada, Oceania, US, Western Europe, Japan, Eastern Europe, Africa, Asia, Latin America and the Caribbean	Only includes journals from JCR and MEDLINE, , country of publication only identified by address of 1st author
Soteriades et al, 2005 (USA)(*47*)	RO	1994-2004	Publications, citations, grant funding, GDP, population	ISI	ND	Global	Biomedical research in USA and European Union countries	29 European Union countries (Publications EU n = 1485749, USA n = 1356805)	Descriptive	Productivity of 15 EU member states 76% of USA; including 10 newest EU members declines to 66%; productivity higher for all EU vs. USA when adjusted for funding (797 vs. 563 papers per $bn R&D)	Single data source

Source	Study design	Study period	Indicator	Data source	Feasibility, validity, reliability, acceptability	Utility	Study population and setting	Sample sizes	Type of statistical analysis	Outcomes	Methodological limitations
Druss et al, 2005 (USA)(*33*)	RO	1996	Publications, grant funding	MEDLINE, NIH	ND	Individuals	R01 grants from the National Institutes of Health	18211	Multivariate analysis	Mean publications per grant = 7.58; 6.4% of R01 grants not associated with any publication; new grants produced less publications than competing renewals (6.53 vs. 7.43, p < 0.001); grants reviewed by basic science committees produced more publications than those reviewed by clinical study sections (8.39 vs. 5.82, p < 0.001); high academic prestige associated with greater number of manuscripts (7.33 vs. 6.88, p = 0.02)	Only NIH grants taken into account, only MEDLINE database used, methods rely on authors to identify sources of funding
Hickie et al, 2005 (Austrailia)(*77*)	RO	1993-2003	Citations, population	ISI, Scopus	ND	Global	Impact of Mental Health Researchers in Institutions in Australia and New Zealand vs. the world	15 Countries, 16 institutions, 30 (ESI) and 53 (WOS) individuals	Descriptive	Psychiatry/psychology: Australia ranks higher than NZ for citations, but lower when adjusted per paper and per capita; Neuroscience/behaviour: Australia ranks higher than NZ for citations, but lower when adjusted per paper	Lack of capacity to distinguish between common names in scientometric databases,
Bornman et al 2006 (Switzerland)(*65*)	RO	1990-1995	Publications, citations, peer review	EMBASE, ISI, MEDLINE	ND	Individuals	Biomedical researchers from Boehringer Ingelheim Fonds (BIF) foundation	64 approved applicants and 333 rejected applicants	Multiple regression analysis	Peer review process successful in selecting scientists who perform on a higher level	Single data source
Falagas et al, 2006 (USA)(*38*)	RO	1995-2003	Publications, IF, GDP, population	ISI, MEDLINE	ND	Global	Worldwide parasitology research	Western Europe, US, Japan, Canada, Asia, Oceania, Eastern Europe, Latin America & Caribbean, Africa (Publications n = 18377)	Descriptive	Publication ranking: Western Europe, US, Latin America & the Caribbean; IF ranking: USA highest; adjusted for GNIPC ranking: Oceania highest; Eastern Europe tripled productivity from 1995 to 2003; Latin America & the Caribbean and Asia doubled productivity from 1995 to 2003; Africa productivity remains low	Only JCR-cited journals included, only parasitology category included, MEDLINE only has the address of the 1st author
Devos et al, 2006 (France)(*52*)	RO	2001-2004	Publications, co-authorship, IF	MEDLINE, institutional	ND	Individuals	Doctors in Lille University Hospital	700	Descriptive	Publications n = 2814; mean IF 2.26.	Single centre study
Housri et al, 2007 (USA)(*42*)	RO	2002-2004	Publications, citations, IF, gender, conference presentations	ISI, MEDLINE	ND	Individuals	Principal investigators from abstracts presented at Association for Academic Surgery (AAS) and Society of University Surgeons (SUS)	37 women, 300 men (649 AAS abstracts, 337 SUS abstracts)	Fisher's exact and Student's t-test	AAS: Significantly less female authors (p < 0.0001); publication rate higher in females (p = 0.0132); citation rate not significantly different; SUS: Significantly less female authors (p < 0.0001); publication and citation rate not significantly different; female publications in higher impact journals (p = 0.0082)	Assumed final author on the abstract was PI, does not account for rejected abstracts.
Torro-Alves et al, 2007 (Brazil)(*74*)	RO	Earliest publication - 2006	Publications, h index	ISI	ND	Individuals	Professors linked to undergraduate and postgraduate scientific programs at University of Sao Paulo, Brazil	142 undergraduate professors and 134 postgraduate professors	ANOVA, Posteriori post hoc comparison, Pearson correlation	Undergraduate: chemistry and biology better performance than physics/mathematics and psychology and education for all indicators than (p ≤ 0.001); Graduate: psychology had lower h index and few publications in ISI than other programs (p ≤ 0.001); Significant correlation between all indicators p = 0.05	Single data source
Rezaei-Ghaleh et al, 2007 (Iran)(*76*)	RO	2004	IF	ISI, MEDLINE	ND	Institutions	Biomedical research in Iranian Universities	20 (876 publications with IF)	Pearson's and Spearman's correlation	Mean rank normalised IF = 0.471±0.252, mean Journal to field impact score = 0.670±0.446, standardised IF = -0.108±0.771 (Iranian articles published in slightly weaker journals); Strong correlation between rnIF, JFIS and SIF	Does not account for journals without IF

Source	Study design	Study period	Indicator	Data source	Feasibility, validity, reliability, acceptability	Utility	Study population and setting	Sample sizes	Type of statistical analysis	Outcomes	Methodological limitations
Lewison et al, 2007 (UK)(*57*)	RO	1996-1998	Publications, citations, peer review	ISI	ND	Global	Psychiatric research in the UK and USA	2	Descriptive	World scale value: USA superior performance in psychiatric genetics (155) and mental health services research (144); UK slightly better than average in PG (106) but below average in MHS (81); World scale value calculated using relative esteem values: USA = 98 and UK = 116	Single data source
Campos Jimènez et al , 2007 (Spain)(*64*)	RO	2004-2005	Patents	Scopus, OEPMPAT	ND	Global	Immunological patents in Spain and the World	44 Countries, 2518 worldwide patents	Descriptive	Spaniards accounted for 1.99% of patents worldwide 2004 and 1.75% in 2005; Spaniards were 1st inventors/applicants in 10% of their patents; most patentees were private enterprises	Scopus database used to search for world patents and OEPMPAT used to search for Spanish patents
Radicchi et al , 2008 (Italy)(*59*)	RO	1990, 1999, 2004	Citations	ISI	ND	Specialties	Scientific specialties	14 (publications range = 266-9761)	Descriptive	Citation rates vary between specialties; performance between specialties can be measured using relative citation rate; can be used to generalise the h index	Single data source
Hendrix et al, 2008 (USA)(*41*)	RO	1997-2000	Publications, citations, h index, grant funding	ISI, NIH	ND	Institutions	USA medical schools which were members of the Association of American Medical Colleges	31	Principal components analysis	Gross research productivity variables – publications, citations, average number of faculty; research impact variables – citations/article, impact index, % articles with 0 citations; individual research productivity - average number of articles, average citations and average funding per faculty member	Single data source
Bornmann et al, 2008 (Germany)(*48*)	RO	1993-2006	Publications, citations, peer review	ISI	ND	Individuals	Scientists from Long-Term Fellowship (LTF) and Young Investigator (YI) programmes of the European Molecular Biology Organisation (EMBO)	668 LTF applicants, 297 YI applicants	Multiple regression analysis	Peer review process successful in selecting scientists who perform on a higher level	Single data source, single centre study, does not account for funding
Groneberg-Kloft et al, 2008 (Germany)(*54*)	RO	1961-2007	Publications, GDP, population	ISI, Scopus	ND	Specialties, global	Medical specialties and worldwide countries	22 specialties, 32 countries (5527558 publications)	Descriptive	Publications – USA ranked 1; adjusted for GDP – USA ranked 14; adjusted for population – USA ranked 10;	Specialties classified differently in databases
Koskinen et al, 2008 (Finland)(*56*)	RO	1996-2005	Publications, citations, co-authorship, IF	ISI	ND	Specialties, institutions	Schizophrenia research in Finnish institutions	6 institutions (265 publications from 1st author with a background Finnish institution)	Descriptive	Number of publications: Helsinki 42, Kuopio 22, Oulu 33, Tampere 16, Turku 45, National Public Health Institute 33, other Finnish 19, non Finnish 55; clinical-epidemiological (49%), biological (38%) and pharmacological (13%). Median citation = 9; international collaboration 43% of articles	Single data source
Pulijak et al, 2008 (Croatia)(*70*)	RO	2000-2006	Publications, co-authorship, IF, grant funding	ISI	ND	Institutions	Clinical and Life Sciences performance of research institutions in Split, Croatia	4 (350 publications)	Descriptive	Publications tripled from 2000 (30) to 2006 (76); mean IF increased from 2000 (2.03) to 2006 (2.89); most common specialties is medicine (general and internal) followed by oncology and paediatrics; University of Split School of Medicine - 89% of all articles, 78 research articles published by 12 PI's receiving total grant of 2.3 million Euros; average cost of article was 29,210 Euros	Single data source
Mugnaini et al, 2008 (Brazil)(*73*)	RO	Earliest publication - 2005	h index	ISI	ND	Specialties, institutions, global	Scientific specialties from Brazilian Academy of Sciences (BAS) and National Academy of Sciences, USA (NAS-USA)	10 (389 BAS members, 410 NAS-USA members)	Descriptive	Median h index: Health sciences BAS = 20, NAS-USA = 82.5; Biomedical sciences BAS = 22, NAS-USA = 66; other scientific specialties = range 3 – 18	Single data source, existence of sub-specialties, some scientists members of both BAS and NAS-USA;

Source	Study design	Study period	Indicator	Data source	Feasibility, validity, reliability, acceptability	Utility	Study population and setting	Sample sizes	Type of statistical analysis	Outcomes	Methodological limitations
Kellner et al, 2008 (Brazil)(*72*)	RO	Earliest publication - 2008	h index	ISI, institutional	ND	Specialties	Scientific specialties from Brazilian Academy of Sciences (BAS)	10 (357 BAS members)	Descriptive	Mean h index: Health and Biomedicine (n = 107) = 22; other scientific specialties (n = 250) = 12	Some journals, books, meetings not indexed in ISI and institutional databases
Castellano et al, 2009 (Italy)(*50*)	RO	1990, 1999, 2004	Publications, citations	ISI	ND	Specialties	Scientific specialties	14 (publications range = 266-9761)	Descriptive	Determines a function that gives a minimum value of relative citation factor for a paper to be in the top q%	Single data source
Lee et al, 2009 (USA)(*45*)	RO	Earliest publication - 2008	h index, co-authorship, g index, AWCR	Google, Scopus	ND	Individuals, specialties	Neurosurgeons from neurosurgery training programs	120	Correlation	H index significantly increased with academic rank ($p < 0.0001$), co-authorship does not affect h index, strong correlation between g index and AWCR ($p < 0.0001$), variation of mean h index for different medical specialties	Only evaluates neurosurgery

Table 15: Description of studies included in the systematic review.

APPENDIX 2

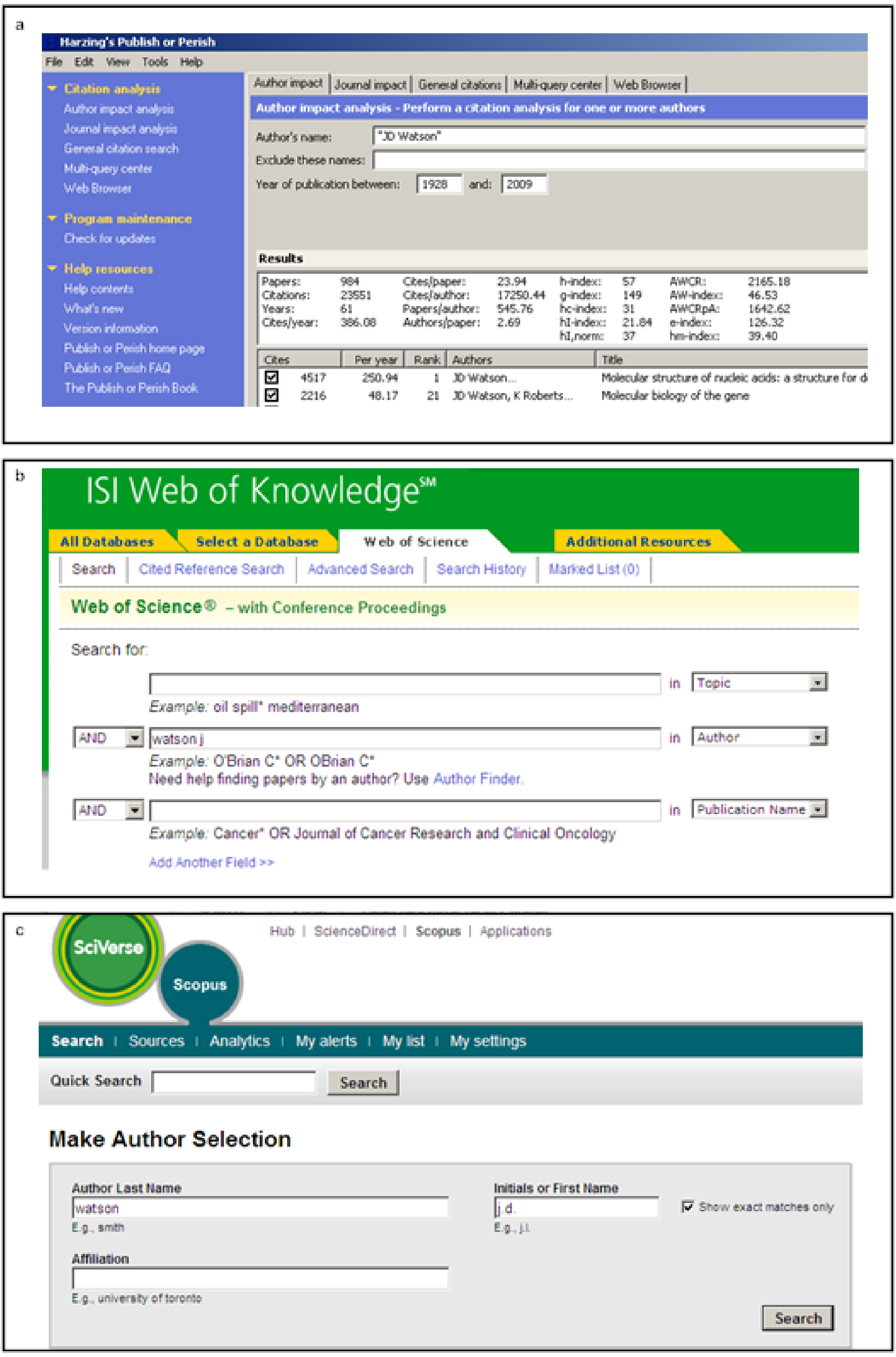

Figure 13: Screenshots from the citation databases.

APPENDIX 3

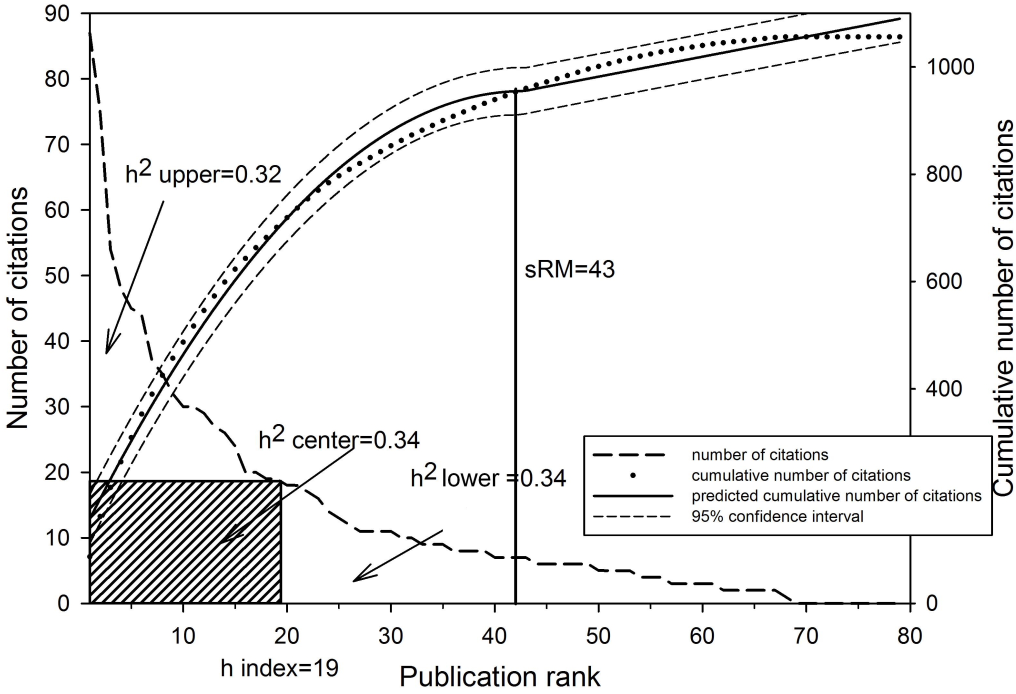

Figure 14: Calculation of *h*^{2} upper, *h*^{2} centre, *h*^{2} lower and sRM value for one of the professors of clinical medicine (R^2_{sRM}=.99).

APPENDIX 4

8.1 Statistical analysis

As shown by **Table 11**, the dependent variable (citations) is considerably overdispersed, since the sample variance of 655.47 is about 1.58 as large as the sample mean of 414.42. The default standard errors for both cross-section and panel estimator would therefore understate the true standard errors.

Both the dependent variable and the regressors can potentially vary over time (within variation) and with individuals (between variation). As indicated by **Table 16**, time-invariant regressors (e.g., gender) have zero within variation. For most of the other regressors as well as the dependent variable, there is more variation across individuals than over time. The coefficients of regressors with relatively little within variation estimated with a fixed-effects model would be imprecise (and not identified when there is no within variation at all). Within estimation may therefore lead to considerable efficiency loss.

		Standard Deviation		
		Overall	**Between**	**Within**
Control Variables	Number of citations	655.472	543.952	366.264
	Number of past publications	58.252	39.503	42.836
	Gender	.456	.457	.000
	Physician status	.500	.500	.000
	Senior lecturer	.437	.437	.021
	Reader	.332	.332	.000
	Professor	.492	.492	.000
	Institute of Clinical Sciences	.161	.161	.000
	Kennedy Institute of Rheumatology	.209	.209	.000
	Department of Medicine	.492	.492	.021
	National Heart Lung Institute	.398	.397	.021
	School of Public Health	.311	.311	.000
Solo versus multi-authorship	Solo authored papers	2.986	2.701	1.277
	Median number of co-authors per publication	3.313	2.568	2.078
	(Median number of co-authors per publication)^2	62.842	45.325	42.707
	Minimum number of co-authors per publication	1.620	1.216	1.087
Network-based measures of centrality	Degree	.021	.015	.014
Network-based measures of social capital	Constraint	.328	.202	.258
Author's position in publication Position in non-alphabetical sequence	First author	2.540	1.972	1.603
	Last author	6.900	6.321	2.775
	Second author	2.418	1.907	1.488
	Penultimate author	4.327	3.659	2.312
	Other author	7.089	6.156	3.523
Author's position in publication Position in alphabetical sequence	First author	.928	.702	.608
	Last author	1.398	1.189	.736
	Second author	.429	.291	.316
	Penultimate author	.168	.112	.126
	Other author	.133	.081	.105
Interactions: academic rank and position	Senior lecturer X non-alphabetical last position	1.810	1.559	.920
	Reader X non-alphabetical last position	1.566	1.088	1.127
	Professor X non-alphabetical last position	5.963	5.487	2.341
	Senior lecturer X alphabetical first position	.437	.326	.292
	Reader X alphabetical first position	.331	.215	.252
	Professor X alphabetical first position	.635	.498	.394
Interactions: academic rank, position, and brokerage	Senior lecturer X non-alphabetical last position X constraint	1.049	.630	.839
	Reader X non-alphabetical last position X constraint	.213	.133	.165
	Professor X non-alphabetical last position X constraint	.687	.587	.322

Table 16: Within and between variation of the dependent variable and the regressors.

To examine the variation of the dependent variable in more detail, we calculated the transition probabilities from one period to the next, after aggregating citations into appropriate categories. **Table 17** shows that there was considerable persistence in performance from one period to another one. Over 65% of the authors with no citations in one time period also did not receive any citation in the subsequent period, while over 38% of the authors with more than 1,000 citations in one period received more than 1,000 citations in the subsequent period. **Table 18** shows the first-order autocorrelations at all lags. Clearly, citations were correlated over time, and autocorrelations varied little with lag length.

Citations	0-49	50-99	100-199	200-299	300-399	400-499	500-799	800-999	1000-	Total
0-49	65.03	11.19	11.19	4.90	1.40	0.70	4.90	0.70	0.00	100.00
50-99	45.78	22.89	19.28	6.02	1.20	2.41	1.20	1.20	0.00	100.00
100-199	27.78	27.78	27.22	6.67	6.11	1.11	3.33	0.00	0.00	100.00
200-299	17.78	20.74	34.81	9.63	5.19	6.67	2.22	2.22	0.74	100.00
300-399	12.94	20.00	30.59	15.29	12.94	2.35	4.71	1.18	0.00	100.00
400-499	7.69	16.92	26.15	18.46	13.85	4.62	6.15	4.62	1.54	100.00
500-799	3.88	9.30	18.60	24.81	12.40	10.08	13.18	1.55	6.20	100.00
800-999	2.33	6.98	13.95	25.58	2.33	11.63	20.93	4.65	11.63	100.00
1000-	1.34	0.00	4.03	7.38	6.71	12.75	22.15	7.38	38.26	100.00
Total	22.63	15.42	20.45	11.46	6.72	5.53	8.30	2.37	7.11	100.00

Table 17: Transition probabilities.

Time period	1	2	3
1	1.0000		
2	0.7905	1.0000	
3	0.6889	0.6870	1.0000

Table 18: First-order autocorrelation of citations from one period to another.

8.2 Modelling strategy

Because the dependent variable is a count variable, the use of linear models would have resulted in inefficient, inconsistent, and biased estimates. We estimated maximum-likelihood negative binomial panel models instead of Poisson panel estimators owing to the overdispersion of the dependent variable. Indeed the negative binomial estimator can explicitly handle overdispersion, and thus it accounts for unobserved heterogeneity among observations. By contrast, in case of overdispersion, the Poisson estimator would produce consistent, but inefficient estimates and standard errors that are biased downward.

Table 19 shows the estimated coefficients for institutional affiliation that were not reported in the four random-effects negative binomial panel models of **Table 14**. Findings indicate statistically significant differences in performance between academic departments (e.g., between the School of Public Health and the Department of Surgery and Cancer).

	Model 1		Model 2		Model 3		Model 4	
Control variables	EC	SE	EC	SE	EC	SE	EC	SE
Institute of Clinical Sciences	.24*	.11	.25*	.11	.21	.15	.23	.14
Kennedy Institute of Rheumatology	.15	.10	.12	.11	.20	.12	.24*	.12
Department of Medicine	.02	.06	.02	.06	.07	.07	.04	.07
National Heart Lung Institute	.16	.10	.17*	.08	.14	.08	.16	.08
School of Public Health	.24*	.10	.26**	.10	.26**	.10	.21*	.09

EC = estimated coefficient; SE = standard error
* $p \leq 0.05$, ** $p \leq 0.01$

Table 19: Results of negative binomial random-effects panel regressions with main effects of institutional affiliation.

Table 20 - 22 summarizes estimated coefficients and standard errors for different estimators.

Model 5 refers to the pooled quasi-maximum likelihood Poisson estimator with cluster-robust standard errors that control for both overdispersion and serial correlation over time for a given individual. Note that the default standar errors (not reported here) that impose the restriction of mean-variance equality are smaller. Correcting for overdispersion using the sandwich variance matrix estimate would increase the standard errors estimates (not reported here). This points to the importance of controlling for overdispersion (see below). Moreover, controlling for serial correlation over time for a given individual produces even larger cluster-robust standard error estimates (Model 5). Similarly, Model 6 refers to the pooled quasi-maximum likelihood negative binomial estimator with cluster-robust standard errors that control for both overdispersion and serial correlation. Efficiency gains can be obtained if estimation is based on a specified model for the dependence over time for a given individual. To this end, we also estimated generalized estimating equations (GEE) or population-averaged models with different correlation structures. Model 7 refers to the GEE Poisson model with unstructured error correlation, i.e., placing no restriction on the correlation of errors over time aside from their equality across individuals. Model 8 refers to the GEE negative binomial model with unstructured error correlation. Model 9 is the GEE Poisson estimator with equicorrelated errors. Model 10 is the GEE negative binomial estimator with equicorrelated errors. Model 11 is the Poisson maximum likelihood panel estimator with gamma-distribued random effects and cluster-robust bootstrapped standard errors (with 100 replications). Model 12 refers to the conditional maximum likelihood fixed-effects negative binomial panel estimator with both individual- and time-specific effects and cluster-robust boostrapped standard errors (with 100 replications). The computation of all models was implemented using Stata 64/MP 10.1.

The coefficient estimates of the pooled models are quite similar to those from the corresponding population-averaged models. Compared with the parameter estimates from the population-averaged models, the random-effects estimated coefficients differ roughly by 20-30%. The random-effects Poisson and negative binomial estimates and their standard errors are similar. Notice that the negative binomial fixed-effects estimator (Model 12) is unusual in that, unlike other fixed-effects linear estimators, it provides estimates for time-invariant regressors in addition to time-varying ones (*155*). However, notice that fixed-effects estimates of most time-invariant regressors are not statistically significant.

For all panel models, likelihood-ratio (LR) tests of model specification indicates that they are more appropriate than the corresponding pooled models, i.e., $p\left(\chi^2(1) > LR\right) < 0.001$.

	Model 5		Model 6		Model 7		Model 8	
	EC	SE	EC	SE	EC	SE	EC	SE
Control variables								
Number of past publications	.002**	.000	.003**	.001	.002**	.000	.002**	.001
Gender	.169*	.076	.145*	.065	.172*	.084	.142*	.063
Physician status	-.132	.073	-.220**	.062	-.073	.073	-.203**	.061
Senior lecturer	.461*	.200	.259	.236	.464	.356	.304	.240
Reader	.516**	.201	.452	.236	.556	.351	.490*	.249
Professor	1.026**	.190	.755**	.230	1.080**	.345	.794**	.234
Solo versus multi-authorship								
Solo authored papers	.006	.008	.008	.014	.013	.008	.012	.010
Median number of co-authors per publication	.029	.015	.049**	.015	.024	.019	.049**	.013
(Median number of co-authors per publication)^2	-.003**	.001	-.003**	.001	-.002**	.001	-.002**	.001
Minimum number of co-authors per publication	-.079**	.027	-.072**	.019	-.039	.034	-.070**	.024
Network-based measures of centrality								
Degree	6.810**	1.251	14.039**	3.329	5.247**	1.103	9.259**	1.392
Network-based measures of social capital								
Constraint	-1.032**	.266	-.797**	.269	-.967**	.344	-.873**	.251
Author's position in publication								
Position in non-alphabetical sequence:								
First author	.034**	.011	.065**	.016	.024*	.012	.061**	.013
Last author	-.120*	.047	-.104	.055	-.128	.090	-.104	.060
Second author	.030*	.014	.043**	.014	.033**	.013	.037	.014
Penultimate author	-.008	.009	.003	.012	-.004	.009	.003	.010
Other author	.028**	.007	.032**	.010	.029**	.010	.040**	.007
Position in alphabetical sequence:								
First author	-.002	.036	.003	.054	.078*	.033	.039	.051
Last author	.012	.020	.030	.024	.016	.020	.031	.023
Second author	.017	.040	.084	.056	.067	.045	.113*	.049
Penultimate author	.020	.135	-.103	.129	-.186	.159	-.219	.172
Other author	.047	.147	.039	.174	.041	.135	.036	.163
Interactions: academic rank and position								
Senior lecturer X non-alphabetical last position	.102*	.050	.091	.058	.124	.092	.106	.061
Reader X non-alphabetical last position	.066	.052	.085	.066	.103	.092	.097	.063
Professor X non-alphabetical last position	.160**	.047	.157**	.055	.166	.090	.155**	.060
Senior lecturer X alphabetical first position	.186	.105	.231*	.114	.182	.135	.195	.123
Reader X alphabetical first position	-.045	.086	-.016	.091	-.074	.071	-.033	.081
Professor X alphabetical first position	.050	.049	.008	.064	-.010	.047	-.008	.061
Interactions: academic rank, position, and brokerage								
Senior lecturer X non-alphabetical last position X constraint	.005	.214	-.202*	.081	-.153	.232	-.230**	.076
Reader X non-alphabetical last position X constraint	.084	.151	.347	.333	.230	.130	.400**	.134
Professor X non-alphabetical last position X constraint	.201**	.049	.333**	.082	.220**	.061	.300**	.055

EC = estimated coefficient; SE = standard error

* p ≤ 0.05, ** p ≤ 0.01

Table 20: Summary of estimated coefficients and standard errors for different estimators.

	Model 9		Model 10		Model 11		Model 12	
	EC	SE	EC	SE	EC	SE	EC	SE
Control variables								
Number of past publications	.002**	.000	.003**	.000	.002**	.001	.001**	.000
Gender	.176*	.077	.142*	.065	.099	.080	.032	.175
Physician status	-.131	.073	-.227**	.061	-.238*	.102	-.420**	.136
Senior lecturer	.448*	.207	.276	.214	.436	.310	.231	.323
Reader	.531**	.204	.485*	.222	.781**	.286	.448	.350
Professor	1.046**	.193	.796**	.208	1.351**	.298	.391	.315
Solo versus multi-authorship								
Solo authored papers	.007	.008	.009	.010	.067**	.021	.015	.009
Median number of co-authors per publication	.028	.016	.039**	.013	-.017	.024	.032*	.014
(Median number of co-authors per publication)^2	-.003**	.001	-.002**	.001	.000	.001	-.002**	.000
Minimum number of co-authors per publication	-.076**	.026	-.065**	.019	-.035	.025	-.053**	.019
Network-based measures of centrality								
Degree	6.837**	1.254	14.151**	1.318	7.780**	2.117	2.104*	.829
Network-based measures of social capital								
Constraint	-1.000**	.267	-.785**	.217	-.582	.397	-1.440**	.273
Author's position in publication								
Position in non-alphabetical sequence:								
First author	.035**	.011	.070**	.013	.079**	.015	.035**	.011
Last author	-.123**	.048	-.109*	.050	-.147*	.070	.019	.066
Second author	.029*	.014	.041**	.013	.030	.016	.011	.016
Penultimate author	-.008	.009	.004	.009	-.009	.010	.004	.005
Other author	.028**	.008	.032**	.008	.003	.011	.030**	.005
Position in alphabetical sequence:								
First author	.002	.035	.019	.050	.082	.084	.206*	.084
Last author	.011	.020	.028	.023	.002	.045	.032	.023
Second author	.020	.040	.095	.049	.062	.068	.102*	.050
Penultimate author	.016	.133	-.128	.131	-.175	.190	-.059	.097
Other author	.051	.142	.059	.156	.329*	.152	.012	.173
Interactions: academic rank and position								
Senior lecturer X non-alphabetical last position	.194	.104	.091	.052	.079	.074	-.012	.071
Reader X non-alphabetical last position	-.047	.086	.087	.053	.136	.074	-.018	.071
Professor X non-alphabetical last position	.046	.048	.164**	.050	.198**	.069	.012	.066
Senior lecturer X alphabetical first position	-.007	.218	.224*	.108	.106	.137	-.107	.111
Reader X alphabetical first position	.082	.150	-.027	.083	-.070	.122	-.199*	.093
Professor X alphabetical first position	.200**	.049	-.010	.058	-.047	.098	-.157	.088
Interactions: academic rank, position, and brokerage								
Senior lecturer X non-alphabetical last position X constraint			-.199**	.067	-.106	.202	-.263	.157
Reader X non-alphabetical last position X constraint			.291*	.137	-.002	.230	.025	.155
Professor X non-alphabetical last position X constraint			.331**	.051	.263*	.126	.222**	.057
Time period								
Time period 1							1.104**	.060
Time period 2							.877**	.041

EC = estimated coefficient; SE = standard error
* $p \leq 0.05$, ** $p \leq 0.01$

Table 21: Summary of estimated coefficients and standard errors for different estimators.

Model	5	6	7	8	9	10	11	12
Number of observations	1285	1285	1285	1285	1285	1285	1285	1257
Number of groups	490	490	490	490	490	490	490	463
Wald chi square	1851.380	850.880	1079.680	1283.600	1868.340	1453.110	586.790	2149.850
Prob > ChiSq	.000	.000	.000	.000	.000	.000	.000	.000
Log pseudolikelihood	-157048.170	-8484.234					-70132.897	-4574.788
Pseudo R2	.606							

Table 22: Summary of estimated coefficients and standard errors for different estimators.

8.3 Specification tests

8.3.1 Testing for overdispersion

There are several methods of testing for overdisperion. A regression-based overdispersion test statistic can be computed by estimating the random-effects panel Poisson model, running the auxiliary OLS regression (without constant) of the generated dependent variable, $\left\{\left(y-\hat{\mu}\right)^2-y\right\}/\hat{\mu}$, on $\hat{\mu}$, and conducting a *t* test of whether the coefficient of $\hat{\mu}$ is zero, where $\mu=\exp\left(x'\beta\right)$ (Cameron and Trivedi, 2005). We obtained a *t* value of 31.25 (p<0.001), which is an indication of significant overdispersion.

Because the log-likelihood functions of both the panel Poisson model and panel negative binomial model can be easily obtained, the LR test statistic can also be used to test for overdispersion. We conducted a LR test that compares the estimates from the random-effects negative binomial panel model with those from the random-effects panel Poisson model, where standar errors were estaimated through 100 bootstrap replications. The null hypothesis is $H_0:\alpha=0$, where the scalar parameter α specifies the conditional variance $Var\left(y_i\left|x_i\right.\right)=\mu_i+\alpha\mu_i^2$. Thus, the null hypothesis is that there is no overdispersion, and implies that the negative binomial model reduces the Poisson one. The LR test statistic is: LR = -2(LLF_r – LLF_u), where LLF_r is the maximised value of the restricted log-likelihood function of the Poisson model, and LLF_u is the maximised value of the unrestricted log-likelihood function of the negative binomial model. Asymptotically LR follows the χ^2 distribution. Since there is only one constraint, the degree of freedom is one. We obtained: LR = -2(-7,0132.897 + 8,460.0336) = 123,345.7268. Thus, $p\left(\chi^2(1) > 123{,}345.7268\right) < 0.001$, which provides further support in favour of the hypothesis of overdispersion. That is, because α is significantly different from zero, the LR test suggests that the Poisson distribution is not appropriate.

8.3.2 Unobserved heterogeneity: fixed versus random effects

The analysis of panel data is often affected by the problem of unobserved time-invariant effects known as "unobserved heterogeneity" (*156*). This is particularly relevant to our study since a prior history of successful publications may affect the future likelihood of further successful publications. We dealt with this possibility of "state dependence" (i.e., the likelihood of an event being a function of the state of the unit) by including in our models the number of past publications among the covariates. However, we did not include a lagged dependent variable.

In addition to state dependence, there is another potential problem that, if not properly accounted for, could lead to spurious results. Authors may differ in their ability to produce papers of high impact because of unobserved factors. These factors could arise from permanent differences among the authors, such as intellecutal skills, not captured by the independent variables. If this noise were systematic for the same authors over time, it could lead to a serial correlation among the error terms for those authors, and would produce consistent but inefficient estimated coefficients. Moreover, past productivity may seem to promote future scientific performance simply because it is a proxy for time-invariant unobservable factors that facilitate or hinder the publication of articles of high impact. Failure to address this "spurious state dependence" (Heckman, 1981) can also induce biases in the estimates.

The problem of unobserved heterogeneity is directly related to model specification. If the model does not suffer from a problem of omitted variable, no such problem would occur. However, most statistical models are not fully specified. One possible solution would be to refine the sample. In our study we included all faculty members in the set of scientists at risk of publishing an article. However, it may be the case that some of these academics were in fact not at risk of publishing high-impact articlesor even publishing any article in some or all observation periods, while others had a higher propensity to publish. This may suggest the possibility of misspecification of the sample population, and may justify attempts to clean up the risk set by eliminating observations unlikely to experience the event. However, differences in propensity to publish high-impact articles were likley to originate from unobservable individual effects. Simply filtering out a subset of individuals from the sample would therefore have been inappropriate and would have biased the sample itself.

Two models traditionally used to address problems of unobserved heterogeneity are the fixed-effects and random-effects models. Fixed-effects models treat the unobserved individual-specific effect as invariant over time and compute it for each panel (author). This method would thus estimate a constant term for each distinct author. By contrast, random-effects models treat the indicidual-specific effects as randomly drawn from some underlying probability distribution. There is a vast body of literature concerned with the strenghts and shortcomings of fixed-effects models versus random-effects ones in the linear case(*157*), and the same comparative assessement extends to the case of non-linear models (*155*).

To address concerns of heterogeneity, in this study we employed a random-effects panel negative binomial model, which introduces two additional parameters to account for both overdispersion and within correlation. Our choice of the random-effects estimator was motivated as follows. First, unlike random-effects models, fixed-effects ones would produce biased estimates when panels extend over relatively short periods (*156, 157*). Because all authors in our sample were present for only three periods of time, the random-effects model was clearly the favored estimator. Second, fixed-effects models (but with the exception of the negative binomial estimator; see (*155*)) cannot include time-independent regressors because they would be absorbed into the individual-specific effects and would not be identified. In our case, this limitation would have implied the exclusion of a number of covariates, such as gender, physician status and academic rank, and as a result the analysis would have been severely limited. For instance, we could not have estimated the interaction effects between academic rank and position in by-line. Third, we used a random-effects panel regression model so that we could obtain an unconditional inference. Consequently, the results are not restricted to the particular individuals sampled, but can be generalised to the population of medical scientists from which the sample was drawn. Finally, to test whether individual-specific unobservables are uncorrelated with individual-specific observables, we conducted a Hausman test to compare the estimated coefficients of the two-way fixed-effects negative binomial estimator with both individual- and time-specific effects with the estimated coefficients of the random-effects negative binomial estimator with time dummies. The null hypothesis is that there is no statistically significant difference between the estimates of the models, and thus that there is no need for fixed-effects estimation. The test produced: $H = 38.40 < \chi^2_{.05}(38) \approx 55.758$ or, alternatively, $p\left(\chi^2(38) > 55.758\right) = 0.4514 > 0.05$. Thus, the test does not reject the null hypothesis that the individual-specific effects are uncorrelated with the regressors and that the random-effects estimator produces consistent (and efficient) estimates.

We also addressed concerns of heterogeneity by replicating the analysis using two groups of increasingly restrictive definitions of the risk set. To restrict the analysis to authors of comparable scientific productivity, the first three sets included all authors who, across all three periods of time, had a history of, respectively, at least one publication (n=479), five publications (n=449), and 10 publications (n=369). To restrict the analysis only to authors with comparable propensity to enagage in collaborative teams, we produced three additional risk sets: one including only authors who published at least five multiple-authored articles (n=450); another including only authors who never published any solo-authored article (n=397); and another including only authors who published at least five multiple-authored publication and never published any solo-authored article (n=326). We finally created two more risk sets: one including only authors who, across all periods of time, had a history of at least one publication and published at least five multiple-authored articles (n=424); the other including only authors with a history of at least one publication and who never published any solo-authored article (n=383). The results obtained with different subsets were qualitatively similar, and we reported only those based on the complete sample.