

# BMJ Open A bibliometric analysis of cancer research in South Africa: study protocol

Jennifer Moodley,<sup>1,2</sup> Vedantha Singh,<sup>1</sup> Benjamin M Kagina,<sup>3,4</sup> Leila Abdullahi,<sup>3,4</sup> Gregory D Hussey<sup>3,4</sup>

**To cite:** Moodley J, Singh V, Kagina BM, *et al.* A bibliometric analysis of cancer research in South Africa: study protocol. *BMJ Open* 2015;**5**:e006913. doi:10.1136/bmjopen-2014-006913

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2014-006913>).

Received 14 October 2014  
Revised 22 January 2015  
Accepted 26 January 2015



CrossMark

<sup>1</sup>Cancer Research Initiative, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

<sup>2</sup>Women's Health Research Unit, Faculty of Health Sciences, School of Public Health & Family Medicine, University of Cape Town, Cape Town, South Africa

<sup>3</sup>Vaccines for Africa Initiative, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa

<sup>4</sup>Division of Medical Microbiology, Department of Clinical Laboratory Sciences, University of Cape Town, Cape Town, South Africa

## Correspondence to

Professor Jennifer Moodley; [jennifer.moodley@uct.ac.za](mailto:jennifer.moodley@uct.ac.za)

## ABSTRACT

**Introduction:** Cancer is an important and growing public health burden in South Africa (SA). Over the past few decades, there has been considerable scientific activity in cancer in SA. However, there has been limited analysis of cancer scientific publications. In this paper, we present a protocol for bibliometric analysis of cancer research conducted in SA.

**Methods and analysis:** A comprehensive search of the journal databases PubMed, SCOPUS, Web of Science and EBSCO will be conducted to identify and retrieve data from primary peer-reviewed cancer research articles using a set of consensus search words. Articles that involve cancer research conducted in SA or using biological or clinical data from South African participants and published between 2004 and 2014 will be included in the study. Two independent researchers will screen the articles for eligibility. Bibliometric indicators and study characteristics will be extracted, entered into a database and analysed. The cancer disease site will be recorded and research will be classified using the Common Scientific Outline system. Data obtained will be analysed to determine SA's publication productivity index in cancer research. Annual trends in bibliometric indicators and the type of cancer research will be determined. The degree of collaboration in research conducted in SA will be analysed using co-authorship matrix software. A publication to disease type ratio will be used to assess scientific production relative to disease burden.

**Ethics and dissemination:** As this analysis will draw on publicly available data and does not directly involve human participants, ethical review is not required. We anticipate that the bibliometric analysis will identify the trends in cancer research productivity and the extent to which cancer research is aligned to the local burden of disease. Results will be published in peer-reviewed journals and presented in a user-friendly format to relevant policymakers and funders.

## INTRODUCTION

Cancer is a complex disease that places significant burden on individuals, families, health services and society. In 2012, an estimated 14.1 million new cancer cases and 8.2 million cancer deaths occurred worldwide.<sup>1</sup> The latter represents 15% of all deaths. More

## Strengths and limitations of this study

- The study will identify trends in cancer research publications and the extent to which cancer research is aligned to the local burden of disease.
- This study will provide information of the extent and pattern of collaboration with researchers in Africa and the rest of the world.
- A limitation of bibliometric analysis is that it does not measure the quality of research outputs.
- A limitation of this study will be language bias due to the selection of English-only articles.
- If the author affiliations of all members of collaborating groups or consortia are not reported either in the text or in an appendix, research studies with contributions from South Africa may inadvertently be excluded at the screening stage.

than half of all cancers (56.8%) and almost two-thirds of all cancer deaths (64.9%) occurred in less developed regions of the world. Of concern is the fact that the global cancer burden is projected to increase to 21.4 million new cancer cases and 13.2 million cancer deaths by 2030, the majority of this burden will be in developing countries.

Similar to other developing countries, in South Africa (SA) the overall burden of disease attributable to cancer is rising. Although communicable diseases continue to be a significant cause of morbidity and mortality in SA, cancer is emerging as a critical public health problem. The proportion of all deaths attributed to cancer has increased steadily over the past few years: 7.3% in 2011, 7.8% in 2012 and 8.3% in 2013.<sup>2</sup> According to the most recent South African National Cancer Registry statistics, there were 55 385 new cancer cases in 2008, with an age standardised incidence rate of 113.2 and 101.1 per 100 000 in males and females, respectively.<sup>3</sup> Leading incident cancers include lung, breast, cervix, colorectal tract and prostate cancer.<sup>3</sup> The number of new cancer cases in SA is projected to increase by 46% by 2030.<sup>1</sup> Life expectancy in

SA has increased from 54 years in 2005 to 60 years in 2011.<sup>4</sup> It is imperative that the gains from the stabilising communicable disease rates and increasing life expectancy are not offset by the rising burden of cancer. One way to mitigate the rising cancer burden is through high-quality targeted research programmes.

The role of research in addressing the burden of disease, improving health and healthcare delivery and in developing evidence-based policy is increasingly being recognised.<sup>5–8</sup> Research publications are an important part of the scientific process, making results available to the wider scientific community and playing a key part in linking knowledge generation to uptake and use. Several methods exist to evaluate research publications.

Bibliometry is the quantitative evaluation of scientific publications in a particular field as opposed to the analysis and interpretation of the content of the publication.<sup>9</sup> Bibliometric analyses commonly produce measures of productivity (eg, number of publications); impact (eg, citation counts, journal impact factor) and collaboration. Bibliometrics provides an understanding of the growth and impact of scientific literature, the trends and flow of knowledge within a specified field of academic research and has contributed to the development of specific public health initiatives.<sup>10 11</sup>

Over the past few decades, there has been considerable scientific activity in cancer in SA; however, there has been very limited analysis of cancer scientific publications.<sup>12</sup> Such analyses could provide insight into research trends, recognise research advances and provide useful information to the scientific, governmental and funding community in terms of addressing the cancer burden and strengthening cancer research capacity in SA.

In this paper, we present the protocol for a study aimed at describing the characteristics and trends in published cancer research conducted in SA in the past 10 years. Specific objectives are to: determine the quantity and quality of cancer research publications; classify cancer research by tumour type and by scientific domain; relate the cancer scientific production to the cancer burden in SA; determine trends in author affiliations and institutional collaborations; and identify gaps in cancer research in SA.

## METHODS

### Bibliographic search

The search for papers to be included in the analysis will be performed using several databases: PubMed (National Library of Medicine, National Institutes of Health), SCOPUS (Elsevier), Web of Science (Thomson Reuter) and EBSCO (Africa-Wide, CINAHL and MEDLINE). The search strategy was developed in consultation with an academic library information specialist. Each database search will contain the same search terms but will be adapted according to the available fields of each database. Key search words that will be used to identify cancer-related articles were selected by reviewing

previous bibliometric analysis studies on cancer research<sup>13 14</sup> to form a set of consensus search words. The search words will include the following terms and Boolean separators “cancer” OR “neoplasm”, “tumour” OR “tumor” OR “carcinoma” OR “adenocarcinoma” OR “leukemia” OR “leukaemia” OR “sarcoma” OR “lymphoma” OR “malignant” OR “oncology” OR “metastasis” OR “oncogene” OR “chemotherapy”. To determine SA’s contribution to cancer research relative to that of Africa, a comprehensive list of African countries will be included to the cancer-specific terms (see online supplementary appendix 1). To identify cancer research conducted in SA, the Boolean separator “AND” was included to the cancer-specific terms followed by the search terms “South Africa” OR “Republic of South Africa”. The search terms also included the names of the SA provinces as follows “Western Cape” OR “Eastern Cape” OR “Northern Cape” OR “Gauteng” OR “KwaZulu-Natal” OR “Natal” OR “Limpopo” OR “Mpumalanga” OR “Free State” OR “North-West”. The names of the metropolitan cities in SA with the largest population size according to the results of the 2011 South African census<sup>15</sup> will be included in the search terms as follows “Johannesburg” OR “Cape Town” OR “Durban” OR “Pretoria” OR “East Rand” OR “Port Elizabeth”.

An additional search that includes terms for SA research organisations will be conducted to ensure that articles that do not clearly report the use of South African data or authors from SA are not omitted. The search terms will include: “South African Medical Research Council” OR “Medical Research Council of South Africa” OR “MRC” OR “Cancer Association of South Africa” OR “CANSAs” OR “National Research Foundation South Africa” OR “NRF”. Results from this additional search will then be filtered using the cancer-specific search terms outlined above and compared with the initial search results to ensure that all cancer research in which SA and South African research organisations have contributed to are included in this study. To further check for completeness of South African cancer publications, recipients of cancer-related South African Medical Cancer Research, NRF and CANSAs grants for the period 2000–2014 will be identified and their publication outputs reviewed.

### Selection of studies

The period of study will be from 2004 to 2014. An article will be considered for inclusion in the study if it falls within the stipulated time period and involves cancer research conducted in SA or using biological or clinical data from South African participants. Primary peer-reviewed cancer research including randomised controlled trials, all observational studies (cross-sectional, cohort, case series, case–control) and quasi-experimental studies (before-after design) will be included. Reviews, editorials, dissertations, conference abstracts and letters to the editor will be excluded. Owing to limited resource, only English articles will be included in the review.

Search results from each database will be exported to an online referencing database (Refworks, <http://www.refworks.com>) and duplicate articles will be identified and removed. The remaining articles will first be screened by title and abstract and those that do not meet the eligibility criteria will be removed. Second, the full texts of the remaining articles will be downloaded and assessed for eligibility. Eligibility screening will be carried out by two independent research team members. Results of both researchers will be compared and agreement between the two research team members will be measured using the  $\kappa$  statistic calculated by the statistical program STATA V.10.1 (STATA Corporation, College Station, Texas, USA). Any disagreements on inclusion/exclusion will be discussed with the entire research team until consensus is reached and only article selected as relevant by all of the team will be included in the final database.

### Data extraction

An electronic data extraction form will be developed using Microsoft Access 2010. Variables to be included in the data extraction form are provided in online supplementary appendix 2. The data extraction form will be piloted on a sample of 10 articles to assess completeness, usability and technical issues. All research team members will participate in the piloting process. Once the data form is finalised, data extraction will be conducted by one of the research team members (VS).

Eligible articles will be reviewed independently and relevant bibliographic information including the names and affiliations of the senior authors (first, second, third, last and corresponding authors), year of publication, sources of funding and publication journal will be entered into the data extraction form. The most recently reported journal impact factor and Eigen score will be used to measure the impact of an article.<sup>16–18</sup> Both indices will be obtained from the Journal Citation Reports (Thomson Reuters, New York, USA). The number of citations for each article will be obtained from both Scopus and Web of Science databases.

To assess the nature of cancer research conducted in SA, information regarding the study characteristics such as the aims, study design and main findings will be recorded.

The type of research conducted will be classified into basic, clinical and public health research. Classification will be guided by definitions utilised by the Academy of Science of South Africa (ASSAF).<sup>19</sup> Clinical research will include research primarily conducted on human participants and on materials (eg, tissues, biological specimens) derived from them. Research aimed at understanding fundamental biological systems will be classified as basic research, and research conducted at a population and/or organisational level will be classified as public health research.

The type of cancer being investigated will be classified by disease site according to the International Cancer

Research Partnership (ICRP) guidelines.<sup>20</sup> In addition, the Common Scientific Outline (CSO), a coding system developed by the ICRP, will also be used to classify each study into seven broad areas of scientific interest in cancer research including: (1) biology; (2) aetiology; (3) prevention; (4) early detection, diagnosis and prognosis; (5) treatment; (6) cancer control, survivorship and outcomes research and (7) scientific model systems.<sup>21</sup> The first category (biology) includes research on the biology of cancer initiation and progression as well as normal biology relevant to these processes. Research aimed at identifying all causes of cancer—genetic, environmental and lifestyle, and the interaction between these factors is included in the aetiology category. The prevention category includes research on all interventions aimed at reducing cancer risk either through decreasing exposure or increasing protective factors. Interventions may include lifestyle modification or may involve drugs or vaccines. The early detection, diagnosis and prognosis category includes research such as discovery of cancer markers (eg, proteins, genes) that could be used in detection or diagnosing cancer as well as predicting outcome. The treatment category includes research focusing on locally (eg, surgery, radiotherapy) and systemically administered (chemotherapy), as well as complementary treatment alternatives (eg, herbs). The cancer control, survivorship and outcomes research category is broad and includes: research on patient care, end-of-life care and survivorship, cancer surveillance, patient attitudes and beliefs that affect cancer control, education of patients and healthcare providers, and health services research. The scientific model system category includes the development and application of animal models, cell cultures and computer simulations. Each of the seven broad CSO codes is further subdivided, resulting in a total of 38 more specific CSO codes. However, for the purposes of this analysis, only the seven broad codes will be used to classify the research publications.

### Validation

All research team members will participate in assessment of the articles. The information on 50% of the articles finally selected will be independently assessed by JM and VS. The other 50% will be independently assessed by BMK and LA. Disagreements will be resolved by discussion until consensus is reached. GDH will be consulted in the discussions.

### Data analysis

Data analysis will be conducted using the statistical program STATA V.10.1 (STATA Corporation, College Station, Texas). Descriptive statistics analyses will be based on percentages for categorical variables and medians and IQRs for continuous variables. An annual South African cancer research publication productivity index will be calculated by comparing

South African publications (numerator) to worldwide cancer publications (denominator), that is, world share of publications. SA's cancer publication contribution relative to the cancer publications in Africa will also be calculated. Annual trends in the number of cancer-related articles and the South African cancer research publication productivity index; the number and proportion of articles in each scientific domain; the number and proportion of articles for each cancer disease site and each CSO category; the quality of publications as measured by most recently available impact factor and Eigen score at the time of the bibliometric; the number of citations; and the number of authors per article, proportion of co-authored articles, senior author affiliations and the proportion of papers where senior authors are from SA and from Africa will be determined. Patterns of co-authorship will be examined by developing a co-authorship matrix using web-based software (Coauth.exe, USA). In addition, a separate matrix will be created for institutional collaboration (Intcoll.exe, USA). The UCINET software V.6.421 (UCINET for Windows, Analytic Technologies, USA) and NetDraw (NetDraw Network Visualization, Analytic Technologies, USA) software will be used to draw and analyse the co-authorship networks. To assess scientific production relative to disease burden, a publication to disease type ratio will be used.<sup>22</sup> Scientific publication relative to the top 10 causes of cancer incidence and mortality and to the HIV-associated cancers will be calculated. Data on cancer incidence will be obtained from the South African national cancer registry<sup>3</sup> and mortality data from Statistics South Africa.<sup>2</sup>

## ETHICS AND DISSEMINATION

As this bibliometric analysis draws on publicly available data and does not directly involve human participants, ethical review is not required. Results of this analysis will identify the trends in South African cancer research publications. We anticipate the study will identify the relative growth in cancer research productivity and the extent to which cancer research is aligned to the local burden of disease, in terms of cancer site being studied as well as the type of research being conducted. This study will also highlight strengths, weaknesses and opportunities in areas such as the quality and type of research being conducted and could be used to guide the allocation of research budgets. Collaborative research is recognised as important in providing innovative solutions to complex problems such as cancer. This review will provide information of the extent and pattern of collaboration with researchers in Africa and internationally. Opportunities to improve future cancer research within SA will be identified. Results will be published in peer-reviewed journals and presented in a user-friendly format to relevant policymakers and funders.

**Acknowledgements** The authors wish to acknowledge the thoughtful contributions of Mrs Mary Shelton from the Faculty of Health Sciences Library, University of Cape Town.

**Contributors** JM and GDH conceived the study. All authors were involved in development of the study protocol. JM prepared the first draft of the manuscript and all authors critically reviewed, revised and approved the subsequent and final version.

**Competing interests** None.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

## REFERENCES

1. International Agency for Research on Cancer. Globocan 2012: estimated cancer incidence, mortality and prevalence worldwide in 2012. 2013. <http://globocan.iarc.fr> (accessed Apr 2014).
2. Statistics South Africa. Mortality and causes of death in South Africa, 2013: findings from death notification. 2014. <http://www.statssa.gov.za> (accessed Jan 2015).
3. National Cancer Registry. *Cancer in South Africa 2008*. Full Report National Cancer Registry. 2014. <http://www.nioh.ac.za> (accessed Jan 2015).
4. Bradshaw D, Dorrington R, Laubscher R. *Rapid mortality surveillance report 2011*. Cape Town: South African Medical Research Council, 2012.
5. Lavis JN, Posada FB, Haines A, *et al*. Use of research to inform public policymaking. *Lancet* 2004;364:1615–21.
6. Momen H. The role of journals in enhancing health research in developing countries. *Bull World Health Organ* 2004;82:163.
7. Wiysonge CS, Lavis JN, Volmink J. Make the money work for health in sub-Saharan Africa. *Lancet* 2009;373:1174.
8. Lewison G, Purushotham A, Mason M, *et al*. Understanding the impact of public policy on cancer research: a bibliometric approach. *Eur J Cancer* 2010;46:912–19.
9. Pritchard A. Statistical bibliography or bibliometrics. *J Doc* 1969;25:348–9.
10. Arunachalam S, Gunasekaran S. Tuberculosis research in India and China: from bibliometrics to research policy. *Curr Sci* 2002;82:933–47.
11. Subbiah A, Subbiah G. Diabetes research in India and China today: from literature-based mapping to health-care policy. *Curr Sci* 2002;82:1086–97.
12. Albrecht C. A bibliometric analysis of research publications funded partially by the Cancer Association of South Africa (CANSA) during a 10-year period (1994–2003). *S Afr Fam Pract* 2009;51:73–6.
13. Glynn RW, Chin JZ, Kerin MJ, *et al*. Representation of cancer in the medical literature—a bibliometric analysis. *PLoS ONE* 2010;5:e13902.
14. Ortiz AP, Calo WA, Suárez-Balseiro C, *et al*. Bibliometric assessment of cancer research in Puerto Rico, 1903–2005. *Rev Panam Salud Publica* 2009;25:353–61.
15. Statistics South Africa. Statistics South Africa: Consensus 2011. 2012. <http://beta2.statssa.gov.za/> (accessed Aug 2014).
16. The Thomson Reuters Impact Factor. 2014. <http://wokinfo.com/essays/impact-factor/> (accessed Oct 2014).
17. Bergstrom CT, West JD, Wiseman MA. The Eigenfactor metrics. *J Neurosci* 2008;28:11433–4.
18. Franceschet M. Ten good reasons to use the Eigenfactor™ metrics. *Inform Process Manag* 2010;46:555–8.
19. Academy of Science of South Africa (ASSAf). *Consensus report on revitalising clinical research in South Africa*. Pretoria: ASSAf. Chapter 1. 2009:1–16.
20. International Cancer Research Partnership (ICRP). ICRP cancer type list. 2011. <https://www.icrpartnership.org/CancerTypeList.cfm> (accessed Oct 2014).
21. International Cancer Research Partnership (ICRP). Common Scientific Outline. 2012. <https://www.icrpartnership.org/CSO.cfm> (accessed Aug 2014).
22. Al-Shahi R, Will RG, Warlow CP. Amount of research interest in rare and common neurological conditions: bibliometric study. *BMJ* 2001;323:1461–2.