

Developing scientific capacity in the Global South

Citation for published version (APA):

Confraria, H. (2019). *Developing scientific capacity in the Global South*. [Doctoral Thesis, Maastricht University]. ProefschriftMaken Maastricht. <https://doi.org/10.26481/dis.20190508hc>

Document status and date:

Published: 01/01/2019

DOI:

[10.26481/dis.20190508hc](https://doi.org/10.26481/dis.20190508hc)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Developing scientific capacity in the Global South

Hugo Confraria

© copyright Hugo Confraria, Maastricht 2019

Printing: ProefschriftMaken || www.proefschriftmaken.nl

Cover image: Courtesy of Daniel Zeller and Places & Spaces: Mapping Science, edited by Katy Börner and Julie M. Davis

ISBN 978 94 6380 326 7

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission of the author or the copyright-owning journals for previous published chapters.

Developing scientific capacity in the Global South

DISSERTATION

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus, Prof. Dr. Rianne M. Letschert
in accordance with the decision of the Board of Deans,
to be defended in public
on Wednesday 8 May, 2019, at 10:00 hours

by

Hugo Confraria

Supervisors

Prof. Dr. Bart Verspagen

Prof. Dr. Robin Cowan

Co-supervisor

Dr. Lili Wang

Assessment Committee

Prof. Dr. Fred Gault (**Chairman**)

Prof. Dr. Robert Tĳssen, University of Leiden

Dr. Ismael Ràfols, Universitat Politècnica de València

Dr. Fabiana Visentin

Summary

In this thesis, our objective is to provide a better understanding of how researchers, institutions and countries in low and middle-income regions can advance their scientific capacity and employ it for higher research citation impact, international collaboration, alignment between research priorities and social needs, and university-industry interactions. We raised different research questions and used different quantitative approaches in order to provide insights for policy makers to create and manage policies that promote the absorption, creation, diffusion, application, and retention of scientific knowledge in the Global South.

The second chapter is the first study that creates a framework to analyse the determinants of citation impact at country level. It displays a U-shaped relationship between research citation impact and GDPpc and finds that previous citation impact, level of international collaboration and total publications in a specific scientific field are important determinants of citation impact among all nations. However, specialisation in particular scientific fields seem significantly more important in lower income regions. The third and fourth chapters combine unique survey data with bibliometric data to study the characteristics of highly cited researchers in Africa and which factors influence international research collaboration in Africa. The results are consistent with previous literature but discussed within the African context. In the fifth chapter of the thesis, we change focus, and we investigate to what extent health research capacity in Africa has been aligned with medical needs in the continent. We use text mining to combine bibliometric data from the Web of Science with the estimates of the disability-adjusted life years produced by the World Health Organization. We find that in sub-Saharan Africa most diseases with high disease burden are also the ones with relatively more research effort and that the resources used to fund that research come mostly from international research funders (public non-African and philanthropic). In the last empirical chapter, we study what kind of characteristics make research institutions more prone to interact with the private sector in Latin America. The main finding is that the scientific institutions that have a more diverse set of knowledge sources and are in brokerage positions at the national level, within their scientific discipline, are the ones that are working more intensively with industry.

Acknowledgments

There are many people I would like to thank for their support during this thesis. First of all, I want to thank my supervisors; Prof. Bart Verspagen, Prof. Robin Cowan and Dr. Lili Wang who provided me with continuous support, guidance and intellectual freedom throughout this process. Individually, I want to thank Bart for his rigour, vision and help in turning ideas into coherent research questions and chapters; Robin for inviting me to participate in a workshop in South Africa in 2016 and present some preliminary results of my second chapter. The ideas and intellectual exchanges during that workshop were essential to develop this thesis. I am also profoundly thankful to him for giving me the opportunity to do a 3-month research visit in Stellenbosch, which allowed me to have access to some of the data that I use in the third, fourth and fifth chapter. I would also like to thank him for opening my mind to the social network analysis field; and Lili for always being ready to help at each stage of my research, supporting with all my technical queries and introducing me to several people in our field. I learned a lot from all of you!

I am also very thankful to Prof. Fred Gault, Prof. Robert Tijssen, Dr. Ismael Ráfols and Dr. Fabiana Visentin for having accepted to serve as committee members and giving me great comments. The time and effort you took to review this work is highly appreciated.

My sincere thank you also goes to my chapter co-authors, Manuel Mira Godinho, Charl Swart, Jaco Blanckenberg and Fernando Vargas. Without Manuel, I would not be doing a PhD. My journey in research started because of him and since then he has been a mentor and a friend. The contributions of Charl and Jaco were hugely significant to this thesis. Chapters four and five are done with data they collected and cleaned. At the beginning of my 3-month visit in Stellenbosch I was unsure of what was going to happen. But you were extremely welcoming, and I think my time there was the most productive time of my life due to our continuous interactions. That visit was not possible without the support of Prof. Johann Mouton and Prof. Catherine Beaudry. I am also profoundly thankful to them. Finally, huge thanks to my “cumpadre” Fernando. We did not manage to achieve our objective of having more papers together than Europa League Finals, but I think we can still do it.

The thesis has benefited from comments, suggestions and data from many other people in conferences, workshops, poster sessions and other casual interactions. I want to thank Aldo Geuna, Eddy Szirmai, Pierre Mohnen, Renè Belderbos and Rodrigo Costas for their contributions particularly.

I am especially grateful to my paranymphs and amazing friends; Cho and Mario. We made this journey together, and I hope we can continue to share trips, concerts, drinks and ideas in the future.

I have been lucky to be part of an incredible cohort at UNU-MERIT: Arip, Bruhan, Cho, Clotilde, Davina, Fernando, Francesca, Eli, Janyl, Jemal, Maria, Mario, Nga, Nora, Wondie, thank you so much for all the debates, dinners, drinks, parties and days together. Our first year was memorable! I am also grateful to my other UNU-MERIT colleagues, Allison, Bea, Caio, Charlotte G., Charlotte K., Dachi, Daniel, Danilo, Eleni, Elisa, Elvis, Emmanuel M., Gintare, Giulia, Ibrahima, Jenny, Joseph, Juanca, Julietta, Jussi, Lika, Lorena, Mary, Mira, Mueid, Paula, Racky, Sachin, Sheng, Shivani, Solomon, Stefania, Stephan, Tamara, Tobias, Victor and others for all the lovely time and experiences.

I want to express my gratitude to the research and administrative staff at UNU-MERIT, who have been caring and helpful during the time I was at the institute as a student, tutor, seminar organiser and PhD representative. Special thanks go to Ad, Eveline, Herman, Lutz, Marc, Maty, Michaela, Monique, Neil, Susan, Tatiana and Wilma.

The sweet memories of living in Maastricht will always be with me throughout my life. This is partially thanks to my brother of many battles, Afonso. Thanks also to the Portuguese crew (Cristiana, David, Daniela, Francisco, Miguel, Stacy, etc.) who made me feel at home in many occasions, and other amazing people I came across during my journey (Boye, Charli, Frank, Kay, Carmen, Emmanuel R., Marco, Nello, Sergio, Fabiola, Zarina, etc.)

Would also like to thank some of my friends in Portugal from “Os Varredores” and “República Boa-Bay-Ela” that helped me to forget about the thesis during summer times. Paradoxically, you were crucial contributors to it.

This thesis is an outcome of a voyage that started in my MSc in Economics and Management of Science Technology of Innovation at the University of Lisbon in 2011. I want to use this opportunity also to thank Ana Correia Moutinho, Sandro Mendonça, Vitor Corado Simões and João Caraça for incentivising and helping me in this path.

This voyage continues, and the last revisions of the thesis were done in Brighton while working at SPRU. I want to thank Tommaso Ciarli and Joanna Chataway for letting me use some of my time to do that.

Huge thanks to my mother, father and my family for supporting me unconditionally throughout my life. And, finally, some special words go to Rita who shared with me most of the best moments I had during the last year and a half.

This research was made possible thanks to a scholarship from UNU-MERIT. Chapters three and four use data from the Young Scientists in Africa project funded by Robert Bosch Stiftung and the International Development Research Centre. I had access to that data during my three month visit to CREST/SciSTiP which was very generously supported, both intellectually and financially, by CREST and its director Johann Mouton. Chapter five includes valuable data provided by Rodrigo Costas. The image in the cover of this book is called a "Hypothetical model of the evolution and structure of science" and it was a courtesy of Daniel Zeller and *Places & Spaces: Mapping Science*, edited by Katy Börner and Julie M. Davis.

Contents

Summary	5
Acknowledgments	7
Chapter 1. Introduction	19
Chapter 2. Determinants of citation impact: A comparative analysis of the Global South versus the Global North	25
2.1. Introduction.....	27
2.2. Background.....	28
2.2.1. Can the Global South use the same bibliometric indicators as those used in the North?.....	28
2.2.2. Factors associated with higher levels of citation impact at the country level.....	29
2.3. Methodology.....	31
2.3.1. Data	31
2.3.2. Approach and metrics	32
2.4. Results.....	35
2.4.1. Global Trends.....	35
2.4.2. Wealth Intensity versus research citation impact	37
2.4.3. Regressions analysis	40
2.5. Discussion & Conclusion	47
2.6. Appendix.....	52
Chapter 3. The characteristics of highly cited researchers in Africa	57
3.1. Introduction.....	59
3.2. Background.....	60
3.2.1. Highly cited researchers.....	61
3.2.2. The factors that affect the probability of producing a highly cited paper	62
3.3. Data and methodology	65
3.3.1. Approach.....	66
3.4. Results.....	68
3.4.1. Bibliometric descriptive statistics	68
3.4.2. Survey descriptive statistics.....	69
3.4.3. Regression results	71
3.5. Discussion and conclusions.....	74
3.6. Appendix.....	76
3.7. Questionnaire	83

Chapter 4. Which factors influence international research collaboration in Africa?	89
4.1. Introduction.....	91
4.2. Background.....	92
4.2.1. The main characteristics of internationally collaborating researchers	93
4.3. Data and Methodology.....	94
4.3.1. Approach.....	95
4.4. Results.....	97
4.4.1. Econometric Results.....	100
4.5. Conclusions.....	103
4.6. Appendix.....	106
Chapter 5. Medical research versus medical needs in Africa.....	109
5.1. Introduction.....	111
5.2. Background.....	112
5.2.1. Health research in Africa.....	114
5.2.2. Setting priorities for health research in Africa	115
5.3. Data and Methods.....	116
5.3.1. Health needs.....	116
5.3.2. Research priorities	117
5.3.3. Metrics.....	118
5.3.4. Econometric approach	119
5.4. Results.....	120
5.4.1. Descriptive Analysis.....	120
5.4.2. Medical research funding	124
5.4.3. Econometric analysis	128
5.5. Conclusion	132
5.6. Appendix.....	134
Chapter 6. Scientific systems in Latin America: Performance, networks, and collaborations with industry	143
6.1. Introduction.....	145
6.2. Background.....	146
6.2.1. University – Industry Collaboration in Latin America	146
6.2.2. Measuring University-Industry knowledge transfer	147
6.2.3. Co-publications as a measure of university-industry collaboration	148
6.2.4. Network position as a correlate of performance	149

6.3. Methodology.....	150
6.3.1. Data collection.....	150
6.3.2. Bibliometric analysis	151
6.3.3. Social Network analysis	152
6.3.4. Econometric analysis	153
6.4. Science in LA: Trends and Specialisation.....	156
6.5. Network analysis	161
6.6. Econometric analysis.....	163
6.7. Conclusions.....	169
6.8. Appendix.....	171
6.8.1. Indicators used.....	171
6.8.2. Network graphs of all scientific areas in 2004-2008 and 2009-2013'.....	174
6.8.3. Consistency check of econometric estimation.....	184
Chapter 7. General discussion	185
7.1. Future research.....	189
Bibliography	191
Valorisation Addendum.....	209
About the Author	213

List of tables

Table 2.1. Descriptive Statistics	41
Table 2.2. Determinants of citation impact in the Global South and the Global North	43
Table 2.A.1. Significant determinants of citation impact, based on previous studies (not exhaustive)	52
Table 2.A.2. Descriptive statistics in the Global South and the Global North	53
Table 2.A.3. Determinants of citation impact in all countries	54
Table 2.A.4. Determinants of citation impact in the Global South and the Global North by international collaboration groups.....	54
Table 2.A.5. Determinants of citation impact in four World regions (Africa; Asia; LA&C - Latin America & Caribbean; E&NA&P - Europe & North America & Pacific)	55
Table 2.A.6. Determinants of citation impact in the World (2008-2012) by subject area using the percentile ranking (PPTop10%)	56
Table 3.1. Negative binomial estimation for top10% and top5%.....	73
Table 3.A.1. Descriptive statistics	79
Table 3.A.2. Description of variables used in the model	80
Table 3.A.3. Probit estimation for top10% and top5%	81
Table 3.A.4. Negative binomial estimation for top10% by age level	82
Table 4.1. Ordinal probit regression model.....	101
Table 4.A.1. Descriptive statistics	107
Table 5.1. Regression analysis. Match between disease burden specialisation and research specialisation	129
Table 5.2. Regression analysis. Match between disease burden specialisation of a region and research specialisation of a specific funder group.....	130
Table 5.A.1. Keywords queries for each disease.....	134
Table 5.A.2. Top 5 funding institutions by African region.....	139
Table 5.A.3. Top 5 funding institutions by funding group	140
Table 5.A.4. Top 5 diseases by top 10 Funding institutions (>0.8%total diseases)	140
Table 5.A.5. Disease burden (log) versus research output (log).....	141
Table 6.1. Research performance of LA: Summary statistics (2004-2008 and 2009-2013)	158
Table 6.2. Top 5 subject areas, in the 9 LA with higher scientific output (2009-2013).....	159
Table 6.3. Summary statistics of scientific networks.....	161
Table 6.4. Descriptive statistics of variables used in the econometric analysis.....	164
Table 6.5. Craggit estimation of intensity of collaboration with industry.....	167
Table 6.5. (cont.) Craggit estimation of intensity of collaboration with industry	168
Table 6.A.1. Two-step Heckman of intensity of collaboration with industry	184

List of figures

Figure 2.1. Scientific productivity (publications per million inhabitants): growth rate versus previous situation.....	35
Figure 2.2. Citation Impact (FNCS) versus GDPpc (2008-2012).....	38
Figure 2.3. Share of highly cited publications (PPTop10%) versus GDPpc (2008-2012).....	38
Figure 2.A.1. Distribution of international collaboration levels. South vs North (2008-2012) ...	53
Figure 3.1. Trends in African output (world share) and top10% cited papers (world share)	68
Figure 3.2. Scientific productivity of HCR10 versus the scientific productivity of non-HCR....	69
Figure 3.3. Number of researchers (and HCR) resident in an African country.....	70
Figure 3.A.1. Density distribution of researchers number of publications per academic age – top10% and top5%.....	76
Figure 3.A.2. Density distribution of researchers academic age – top10% and top5%	76
Figure 3.A.3. Number of researchers by area and gender.....	77
Figure 3.A.4. Collaboration patterns, by being or not a HCR	77
Figure 3.A.5. Challenges faced, by being or not a HCR.....	78
Figure 4.1. Trends in the share of international scientific collaboration in Africa	97
Figure 4.2. Number of researchers resident in an African country (by location of highest qualification).....	98
Figure 4.3. Intensity of collaboration with four types of collaborators	99
Figure 4.A.1. Number of researchers by area and gender	106
Figure 4.A.2. Challenges faced during the career	106
Figure 4.A.3. Scatter plot of scientific productivity versus academic age by level of collaboration intensity outside of Africa	107
Figure 5.1. African regions medical scientific production, and disease burden per capita.	121
Figure 5.2. Disease burden specialisation vs research specialisation by disease.	122
Figure 5.3. Share of publications by funding type (2011-2015)	125
Figure 5.4. Funders health research specialisation by African region	127
Figure 5.A.1. Disease burden specialisation (2005) vs research specialisation by disease (2006-2010).....	138
Figure 5.A.2. Scatter plot. Disease burden (log 2010) vs research production (log 2011-2015) by disease.....	139
Figure 5.A.3. Funders health research output (log) by African region	141
Figure 6.A.1. Network structure of Agricultural sciences (2004-2008).....	174
Figure 6.A.2. Network structure of Agricultural sciences (2009-2013).....	175
Figure 6.A.3. Network structure of Engineering sciences (2004-2008).....	176
Figure 6.A.4. Network structure of Engineering sciences (2009-2013).....	177
Figure 6.A.5. Network structure of Environmental sciences (2004-2008).....	178
Figure 6.A.6. Network structure of Environmental sciences (2009-2013).....	179
Figure 6.A.7. Network structure of Geosciences (2004-2008).....	180
Figure 6.A.8. Network structure of Geosciences (2009-2013).....	181
Figure 6.A.9. Network structure of Plant & Animal sciences (2004-2008).....	182
Figure 6.A.10. Network structure of Plant & Animal sciences (2009-2013).....	183

Chapter 1

Introduction

Science, technology and innovation are critical ingredients in the aspiration of all countries to develop knowledge-based economies, where knowledge drives productivity, economic growth and social wellbeing. A strand of literature that has significantly contributed to our understanding of these phenomena is the so-called “catch-up tradition” together with “evolutionary economists”. These theoretical frameworks give a central stage to the role played by technological upgrading in the process of low- and middle-income economies (or the Global South) moving out of the vicious circles of underdevelopment, and the capabilities needed to reduce the technological gap with the leading economies of the world (Dosi, Pavitt, & Soete, 1990; Fagerberg, 1987; Freeman, Clark, & Soete, 1982; Verspagen, 1991).

Rejecting the implicit idea of the basic convergence hypothesis (Solow 1956; Swan 1956), according to which international knowledge spillovers take place automatically, this strand of literature emphasises the importance of the “absorptive capacity” (Goode 1959) of the receiving country to assimilate technological knowledge. According to Abramovitz (1986), the efficiency with which assimilating knowledge from abroad takes place depends on the “social capability” (capacity to absorb more advanced technology) and on “technological congruence” (the idea that knowledge from abroad must be relevant for the local production structure) of each country.

In the same line of thought, scholars like Lall (2000), Bernardes & Albuquerque (2003) and Fagerberg & Godinho (2004) have shown that in recent years, countries that have grown rapidly have tended to invest in their higher education system and have developed indigenous research efforts. In a recent DFID (2014) report about the mechanisms by which scientific research has been hypothesised to contribute to development, four major pathways were proposed: 1) Human capital – Research carried out within national borders builds human capital, and the capabilities, cognitive skills and knowledge acquired during this process can have impact on tertiary education via up-to-date and qualified training for the new generations of university graduates (Marsh & Hattie 2002). Investments in science can also generate a “domestic base of good scientists, which can break into the international networks where new technologies are being hatched” (Nelson 2005). These scientists can act as important conduits of frontier knowledge into the local academic research community (Barnard et al. 2012), which can potentially diffuse that knowledge to students, the economy, and the general public; 2) Economic growth - Research leads to basic discovery and inventions, which in turn lead to the creation of technology and innovations that allows increases in productivity. This pathway is based on the linear model of innovation (Bush 1945; Godin 2006) and endogenous growth theories (e.g. Arrow, 1962; Romer, 1986). However, there is little evidence to suggest that public investment in research was a major contributor to the rapid economic growth of the “Asian tiger” countries, for example (Hobday 1995; Kim 1997).

According to Mazzoleni & Nelson (2007), the research programmes that effectively contributed to economic growth did not operate within “ivory towers”, but were rather oriented towards an actual, or potential user-community. These programmes were projected to help solve problems, and to advance technology, being applicable to a particular economic area; 3) Pro-poor products and technologies – Research carried out within national borders can help provide both effective and focused responses to domestic problems and have a direct beneficial impact on the lives of the poor (e.g. bed nets for malaria (Alonso et al. 1991) or other “frugal innovations” (Bound & Thornton 2012); and 4) Evidence-informed policy/practice – Research may inform decisions on specific interventions, and it may be used in a more subtle way to inform a decision maker’s understanding of a context. As a result, this pathway leads in theory to better policies and collective decisions (Boaz et al. 2009; Goldemberg 1998).

However, scientific systems are only a small component of innovation systems (Lundvall 1992; Nelson 1993), and the innovation studies literature shows that innovations and increases in productivity do not merely emerge because of local scientific research. Rather, research usually becomes significant through the complex interactions and feedback loops (Kline & Rosenberg 1986) between many different players, including researchers, firms, user-communities and policy makers.

The process by which science influences development is complex and the links between progress in science and progress in society are indirect. Basic research in one field has the potential to contribute to a large number of areas. Usually there are long time lags between scientific discovery and application, scientific knowledge needs to be combined with other factors to influence a society’s capacity to solve problems, and since knowledge is intangible, it is difficult to capture the impact of science on society. Those who oppose the idea of research funding in the Global South often argue that it is too difficult to demonstrate how research will lead to benefits and that it wastes funds that could be otherwise spent on more poverty reduction interventions (DFID 2014).

In this thesis, we will assume that all countries should have a certain level of scientific absorptive capacity. While “absorptive capacity” has been generally defined as being “the ability to learn and implement knowledge” (Cohen & Levinthal 1990). Radosevic & Yoruk (2014) defined absorptive capacity in the context of scientific research as being “the ability to recognize the value of new, external information, assimilate it, and apply it in another context”. In accordance with this view, researchers recombine and re-contextualize existing scientific knowledge and are able to generate novelty through their new publications. Our argument is that without this scientific absorptive capacity, the skills and capabilities available in a country are constrained, and therefore the

possibility for socio-economic development is reduced. Following this line of thought, Pavitt (1998), for example, argued that the main practical benefits of scientific research are not easily transmissible information, ideas and discoveries that are available on equal terms to anyone anywhere in the World. Instead, they are various elements of a problem-solving capacity, involving the transmission of often tacit knowledge through personal mobility, training and face-to-face interactions. The benefits, therefore, tend to be geographically localised, generating a need for each country to have their pool of researchers that belong to the international professional networks and exchange new scientific knowledge (Salter & Martin 2001).

If we assume this hypothesis, a significant gap remains in identifying concrete ways to develop scientific capacity in the Global South, where resources are limited, and in linking those research efforts and capabilities acquired to national socio-economic objectives. Our main objective, with this thesis, is to provide insights for policy makers to create and manage policies that improve the scientific capacity in lower income countries. By policy makers we mean government officials, national and international science funders that can influence scientific systems in the Global South. The insights we want to provide are directly related to our specific research questions in each chapter. The overall key research question is how scientific capacity can be developed for four different objectives: research citation impact (Chapter 2 and 3), international collaboration (Chapter 4), alignment between research priorities and social needs (Chapter 5), and university-industry interaction (Chapter 6).

In chapter 2 we start by asking if countries in the Global South are converging with the high-income countries regarding scientific capacity (proxied by scientific publications per capita), and how scientific capacity can be developed in the Global South to achieve high levels of scientific impact. We use normalised measures of citations per year (controlling for research field) to compare citation impact in resource-deprived contexts with high-income countries, and we apply regression analysis to understand which factors influence higher results.

Because highly cited researchers are seen in the literature as key drivers of knowledge production for their countries and as important conduits of frontier knowledge into the local academic research community and society in general (Barnard et al. 2012; Waldinger 2016), a next logical research question, in chapter 3, is to try to understand why some scientists in lower income regions produce highly cited research. We do it by studying the characteristics of researchers working in Africa who have produced highly cited publications and compare them to researchers who have not produced highly cited work in the same period. We use a unique individual survey dataset with

around 2500 valid responses combined with bibliometric data, and we develop a regression model to find which individual features are more important.

Since we found that international research collaboration positively affects researchers citation impact and is seen as one of the most efficient means to build research capacity and to create learning opportunities in Africa (AOSTI, 2014), a subsequent research question in chapter 4 is: what are the main drivers of international research collaboration at the individual level? Using the same survey dataset, we address this question by studying the characteristics of African researchers who collaborate both frequently and infrequently with foreign and non-African researchers.

In the fifth chapter of the thesis, we change focus, and we investigate to what extent health research capacity in Africa has been aligned with medical needs in the continent. We use text mining to link estimates of the disability-adjusted life years by disease produced by the World Health Organization to scientific disease areas in the Web of Science. The main research question in the chapter is to evaluate whether the amount of research produced on various medical conditions by African researchers is related to their regions burden. A secondary objective is to evaluate what kind of health research is being funded by international organisations. With these research questions, we hope to understand if development funders are supporting research that matches the societal needs in lower income regions, or if they are funding research that is mostly applied and used in more developed regions.

In the last empirical chapter (6), we study what kind of scientific capacity makes research institutions more prone to interact with the private sector in Latin America. We define the percentage of publications of research departments that are co-authored with the industry as the dependent variable, and relate it to a set of features that could influence such collaborations, namely: knowledge production capacity; research quality; orientation towards industry; and knowledge diversity. Our primary objective is to understand if the network structure of a scientific system shapes the links between industries and universities. Chapter 7 draws some general conclusions, policy implications and directions for future research.

Chapter 1

The thesis consists of five quantitative empirical chapters that study countries (Chapter 1 and 4), institutions (Chapter 6) and individuals (Chapter 2 and 3). Methodologically, different chapters apply an array of econometric, bibliometric and network analysis techniques applied to four different datasets. Taken together, the results highlight the importance of research interactions and mobility for the development of scientific capacity in the Global South that lead to research impact, industry collaboration and achievement of societal needs.

Chapter 2

Determinants of citation impact: A comparative analysis of the Global South versus the Global North

Hugo Confraria

UNU-MERIT, Maastricht University

Manuel Mira Godinho^{*#}

^{*} ISEG (Lisbon School of Economics and Management), Universidade de Lisboa

[#] UECE (Research Unit on Complexity and Economics)

Lili Wang

UNU-MERIT, Maastricht University

Abstract

The impact of the scientific output produced by different nations in different fields varies extensively. In this chapter, we apply bibliometric and econometric analysis to study how citation impact varies across countries. This chapter differs from previous research in that a cross-section model is put forward to account for such variation. A special focus is given to the Global South, as countries in this group have been converging with the Global North recently. We find that previous citation impact, level of international collaboration and total publications in a specific scientific field are important determinants of citation impact among all nations. However, specialisation in particular scientific fields seems significantly more important in the Global South than in the Global North. These findings imply that most lower- and middle-income countries would better concentrate their resources in generating higher critical masses in specific fields, in addition to pursuing long-lasting international collaboration partnerships, as these actions may lead to higher impact research.

Keywords: Science policy; Development; Global South; Bibliometrics; Scientific impact; Citation impact.

This chapter draws upon:

Confraria, H., Godinho, M. M., & Wang, L. (2017). Determinants of citation impact: A comparative analysis of the Global South versus the Global North. *Research Policy*, 46(1), 265-279.

2.1. Introduction

There is a widely held assumption that scientific research has positive effects on economic development, namely by increasing human capital, by driving productivity growth, or by providing evidence to inform policies and practice (DFID 2014; Salter & Martin 2001). However, the process by which this happens is complex, and there has been extensive debate about the extent to which development funders and governments in the Global South, or more generally in the peripheries, should invest in research.

A crucial aspect for analysing the scientific performance of countries is to understand whether their scientific output is having an international impact or influence. The impact of published articles can be regarded as being one crucial aspect of scientific quality, and is thus a “proxy” for quality, as follows from the bibliometrics literature (Moed 2005). Studies that focus on measuring the scientific impact of countries usually use citation analysis, as this arguably enables international comparisons to be more objective (Garfield 1979). However, there is substantial literature that questions that citation impact is a proxy for research quality since the earliest bibliometric evaluations (Gilbert 1977; Martin & Irvine 1983).

There are numerous studies in this field that assess research at the country level, however only a few try to understand what the determinants of citation impact are. This type of analysis can help to understand why some scientific systems are performing better than others. Overcoming this gap in the literature can be particularly helpful to provide relevant insights for science policy, for furthering the policy learning cycle and ultimately for increasing the accountability of public policies.

Using the *InCites* tool of Web of Science/Thomson Reuters (*WoS*), this chapter applies bibliometric and econometric analysis to evaluate which countries in the world are producing research with higher research citation impact, and to account for those factors that lead to higher results. The ability to estimate the expected number of citations of countries, by taking country characteristics and other variables at the subject category level, can be helpful for policy-makers in low-income and middle-income countries (the Global South), where public funds for financing the research system are scarce.

Our main objectives are: first, to create a comprehensive framework that can be used in the interpretation of different countries’ citation impact, particularly in the Global South; second, to contribute to citation theory by understanding how the citation impact indicators commonly used in high-income countries can be used in lower income contexts, and; third, to provide assistance

to policy-makers by identifying those independent variables that significantly influence the citation impact of countries.

In what follows, we will first focus on the framework aspects of our analysis, then we will describe the data and methodology used, and afterwards we will discuss the results obtained. Finally, conclusions will be put forward.

2.2. Background

2.2.1. Can the Global South use the same bibliometric indicators as those used in the North?

The North–South divide is generally considered based on its political and socio-economic dimensions. Commonly, definitions of the Global North include North America, Western Europe, and developed parts of East Asia, while the Global South is perceived as being made up of Africa, Latin America, and developing Asia, including the Middle East. In this study we define Global North and Global South in two ways: firstly, by using the World Bank definition of low & lower-middle-income countries versus upper-middle & high-income countries;¹ and secondly, by dividing the world between OECD countries² and non-OECD countries.

As discussed in Chapter 1, this division of the world into Global South and Global North exists not only in terms of wealth or human development, but also in terms of scientific capacity. Therefore, governments in the Global South should find ways to monitor and evaluate the various facets of their scientific systems. By measuring the different characteristics of scientific systems, it is possible to create and manage policies for improving the scientific capacity of countries.

The use of bibliometric indicators for assessing the impact of scientific publications has been on the rise in recent years. The ability of the use of such indicators to lower costs and time of assessment, without being invasive, and to enlighten political choices by carrying out international comparisons, as well as their perceived objectivity, have all been some of the main forces behind its growing popularity (Moed 2005). However, the bibliometric assessment of research performance is based on a central assumption: namely that scientists who have to communicate something important, do so by publishing their findings in international peer-reviewed journals. This choice unavoidably introduces a limited view of a complex reality (van Raan 2004). For instance, regionally focused papers in the Global South (e.g. in Agricultural Sciences) may make particularly important contributions to the local economy, yet remain uncited, as researchers

¹ See the list of countries here: <http://data.worldbank.org/about/country-and-lending-groups>

² See the list of countries here: <http://www.oecd.org/about/membersandpartners/list-oecd-member-countries.htm>

elsewhere are indifferent to those topics. Citation patterns can also differ for other reasons: there are considerable database coverage biases (Moed 2005; Rafols et al. 2015); the research focus can be locally or more internationally oriented (van Raan 2003); there is a language bias, as most journals in *WoS* are written in English (Leeuwen et al. 2001), and; finally, countries have different levels of access to some journals, due to their financial constraints, selectivity, or publication policies (Lawrence 2003). This last limitation is particularly relevant in the Global South and may have acted in the past as a stimulus for researchers from those countries to seek publication through other channels, namely through other means that are not registered in *WoS*, or in other similar databases. This problem was challenged recently by the Research4life³ partnership, which intends to provide developing countries with easy access to peer-reviewed content. This initiative, which aims to reduce the “e-gap” between rich and poor countries, could contribute to a “normalization” of access to the international circuit in the future. Yet this is still a limitation that we have to keep in mind when interpreting our results.

At the same time, both *WoS* and other indexing systems have considerably enlarged the database’s coverage of Latin American and Caribbean (LA-C) journals in recent years. According to Collazo-Reyes (2014), the number of LA-C indexed journals in *WoS* has increased from 69 to 248 titles in just a period of four years (2006–2009). This unprecedented growth is mainly related to a change in the editorial policy of *WoS*.

For these reasons, and despite some recognized limitations, the use of bibliometric data and indicators has also been rising in the context of the Global South, where this type of analysis can be particularly relevant to understand successful processes of closing the S&T gap with the most advanced economies (Albuquerque 2004).

2.2.2. Factors associated with higher levels of citation impact at the country level

In line with this framework, one way to assess scientific impact is by citation analysis. According to the seminal work of Merton (1973), when a scientist cites a given article, he or she indicates that the article was somehow relevant to their research. The citing author calls attention to some useful information included in an article, be it a method, a statistic, a result, or other information⁴, and then acknowledges intellectual or cognitive influence. Therefore, when a comparable article is cited

³ <http://www.research4life.org/>

⁴ Authors also write self-citations, cite peers based on personal networks, use flattery (citations of editors and potential referees), and write “negative” citations (contradicting another author). However, it is reasonable to assume that most citations are “positive”, that is to say, they are a sign of the fact that the citing author finds something useful in the material that they cite. Deviating citation patterns, such as negative citations, can affect an analysis of an individual article or author, however this adverse effect tends to disappear in an analysis of a larger aggregations of authors, such as departments, universities, or countries (Moed, 2005).

more times than others, it is considered to have more international scientific influence or impact (Moed, 2005).

Numerous studies assessing research at the individual, institutional, and country level can be found in the literature. Many other studies create and discuss new methods and metrics for evaluating citation impact. However, few try to understand what the determinants of citation impact are. Table A.1 in the Appendix summarises some of the factors that are known to be associated with higher citation rates at the article, author, institutional, and country level. In our study, we focus on those factors that are known to be associated with higher citation impact at country level, namely: level of international collaboration (Glänzel et al., 1995; Katz & Hicks, 1997; Narin et al., 1991; Puuska et al., 2013; van Raan, 1998), wealth intensity⁵ (King 2004) and having English as an official language (Leimu & Koricheva 2005). These determinants centre on ad hoc considerations, and the literature has not, to the best of our knowledge, presented a comprehensive framework that could be used to interpret a country's citation impact, particularly in the Global South. By bringing together the main arguments in this literature, this study aims to fill such a gap in the literature.

In our analysis we will also include, as explanatory variables previous citation impact, logarithmic scientific output, the percentage of publications in collaboration with industry, and we will also control for population size. Our argument regarding previous citation performance is that there might be path dependency, or the “Matthew Effect” (Merton 1968) in science. Research communities, whose work has been highly cited in the past, are more likely to receive citations in the future (Lafond, 2014). Regarding scientific output, the rationale is that a higher scientific production, in the specific subject area, is a sign of higher critical mass and resources applied to the field, which tend to foster quality and impact (Shibayama & Baba 2015). This measure can also be used as a proxy for scientific specialisation, as we are controlling for the total number of publications produced by a country. As for the share of publications in collaboration with industry, we intend to understand whether citation impact is positively associated with a higher intensity of collaboration with industry (Perkmann et al., 2011). This indicator can thus be seen to be a measure of knowledge transfer between industry and academia, and therefore, if a country has a higher percentage of publications with at least one author from a corporation, then we assume that this country is performing more applied research.⁶ At the same time, nations have obvious differences in size, and to control for this we add logarithmic population as an independent variable.

⁵ Gross Domestic Product per capita.

⁶ An industry collaborative publication is one that lists its organization type as being “corporate” for one or more of the co-author's affiliations. However, not all single affiliations of all publications in *InCites* are unified as “university”.

2.3. Methodology

2.3.1. Data

Publication data were extracted from the (InCites™ 2014) platform provided by *Thomson Reuters*, which facilitates national comparisons across time periods. *InCites* provides output and citation metrics from *WoS*, based on a dataset of more than 27 million papers from 1981 to 2014. The metrics for comparisons are created based on address criteria, using the whole-counting method, that is to say, counts are not weighted by number of authors, neither by addresses.

In this study, our main research question is to understand whether there are different determinants of citation impact between the Global South and the Global North across different subject areas. To solve this, we used two different specifications for South and North (GDPpc levels and being an OECD country or not), and we adopted the disciplinary breakdown of the *Essential Science Indicators* (ESI) areas.

The ESI scheme incorporates a selection of journals carried out by Thomson Reuters. Our dataset covers 21 of the 22 ESI categories with a time span of 5 years (2008-2012). The research fields retained are as follows: Agricultural Sciences, Biology & Biochemistry, Chemistry, Clinical Medicine, Computer Science, Economics & Business, Engineering, Environment/Ecology, Geosciences, Immunology, Materials Science, Mathematics, Microbiology, Molecular Biology & Genetics, Multidisciplinary, Neuroscience & Behaviour, Pharmacology & Toxicology, Physics, Plant & Animal Science, Psychiatry & Psychology, Social Sciences (general), and Space Science. The Multidisciplinary area was excluded, as the publications included in this category could not be unambiguously classified into any of the 21 disciplinary areas.

The option to choose the ESI scheme took into account the fact that there are several approaches to defining a research field: on the basis of selected concepts (keywords); selected sets of journals; a database of field-specific publications; or any combination of these. The selection of a specific scheme⁷ for the division of research fields needs to take into account the trade-off between robustness of results and specificity of the subject category. Citation data is characterised by skewed distributions, and hence robust statistics require considerably large sample sizes. This

“research institute”, “corporate”, etc. There are corporate affiliations which have not been unified yet and which do not have an organization type assigned, and, therefore, these are not identified as industrial collaborations. Large multinational corporations (MNE) have a higher probability of being identified and unified. Therefore, publications listed as industry collaborations represent the lower boundary of real co-publication activities. We would expect countries with a lower presence of MNEs to have larger differences between the number of publications authored by the industry captured by *InCites*, and real activity.

⁷ *InCites* provides six further schemes besides the 21 ESI, based on a conglomerated of journals indexed in the *WoS*, e.g. the 251 *WoS* subject categories, or the 6 OECD categories.

favours the use of fewer categories, with more observations per category. However, articles from different subject categories have different citation propensities. Therefore, the use of very broad categories (e.g. the 6 OECD categories scheme) can lead to differences in citation impact levels, which only reflect differences in the research portfolios of countries, as some countries are more specialised in fields within a given category which have a higher citation propensity.

We believe that the choice of the ESI scheme is the most adequate solution for solving this trade-off in this study. A common, although arbitrary, threshold is often a minimum of 50 full count publications for citation analysis. We use this threshold at a country/category level, and we only consider those countries that have at least 400 publications between 2008 and 2012. By applying these thresholds, we dropped the outliers that have extremely low numbers of publications, which occur particularly in countries/subject areas of the Global South. This is markedly the case for observations before 2003-2007, and therefore we are only able to use two periods in our analysis (2003-2007 and 2008-2012).

A common debate in bibliometric studies is the use of social sciences and humanities for analysis (e.g. Hicks et al., 2015; Marx & Bornmann, 2014). The usefulness of citation impact indicators depends on the extent to which the research outputs are covered in bibliometric databases, and this coverage varies by subject category. The coverage tends to be higher for natural sciences, which gives high priority to journal publications. In the case of social sciences and humanities, where the publication of books, book chapters, monographs, etc. is more traditional, the extent of coverage is reduced. The 21 ESI categories include three categories which are related to social sciences, namely: Economics & Business; Psychiatry & Psychology, and; Social Sciences (general), excluding Humanities. Although the exclusive use of *WoS* data might not be appropriate for the analysis of citation impact in the social sciences, we decided that coverage was sufficient to include these three categories in our broad, country-level, analyses.

2.3.2. Approach and metrics

It is well known that different subject areas have different output propensities, and that publications belonging to each field have singular characteristics. Therefore, to be able to explain the different citation performances among countries and subject areas, we do a multivariate regression analysis (OLS) with fixed effects, at the subject area level.⁸

⁸ According to McCaffrey et al. (2012), this method is designed for efficient computation in models with many fixed effects at one level, on the assumption that fixed effects are included as nuisance parameters to control for differences among units that could bias the estimates of interest (our coefficients).

At the same time, ordinary regression assumes that all observations are independent. However, in our case, each country has 21 subject areas.⁹ As these 21 potential observations share specific country characteristics, our observations are not independent of each other, which could potentially lead to a correlation of errors within countries, implying that the findings of statistical significance would be spurious. To tackle this, we had to relax the independence assumption, by clustering the errors at the country level (McCaffrey et al., 2012; Moulton, 1990).

When interpreting the results presented in this study, it should be borne in mind that indicators measuring citation impact capture the influence of journal articles in the scholarly communication system. As a consequence of the partial and one-dimensional nature of these impact indicators, it is recommended to use more than one single indicator in order to obtain more robust conclusions (Bornmann & Leydesdorff 2013; Waltman 2016). Consequently, for this study, our dependent variable will be measured by two different indicators: (1) the share of highly cited publications (PPtop10%), which shows the proportion of publications belonging to the top ten percent most cited documents in a given subject category, year and publication type, and; (2) the field normalized citation score (FNCS), which calculates the mean citation rate of a country's set of publications in a specific subject area, period of time, and document type, divided by the mean citation rate of all publications in that subject area/period/document type¹⁰. Both these variables are normally distributed indexes, with some outliers on the right tale:

$$PPtop10\% (\%) = \frac{PPtop10\% (n)}{P} \quad (1)$$

$$FNCS = \frac{\sum_{i=1}^P c_i}{\sum_{i=1}^P [\mu_f]_i} \quad (2)$$

Currently, there are several ways to calculate citation impact indicators. From basic calculations such as: raw citation counts; citations per publication; the h-index; geometric means (Fairclough & Thelwall 2015); or discretized lognormal and hooked power law distributions (Thelwall 2016), to normalized methods controlled for research field, publication year, and document type as: the “crown indicator”; field normalized citation score (Waltman et al., 2011); percentile-based approaches (Pudovkin & Garfield 2009), and; source normalized indicators (Waltman & Eck

⁹ As we use the threshold of at least 50 publications per subject areas, not all countries have 21 observations. This may lead to selection bias, as in those countries that do not fulfil these threshold, the categories that are being computed are potentially those that the country performs better in. Thus this hypothesis needs more research in order to be fully understood.

¹⁰ We could have also used the mean normalized citation score, that arguably has better statistical properties (Waltman et al., 2011). This method first normalises the citation of each publication, then makes the average over publications. However, at high aggregation levels, such as at the level of large research institutions or the level of countries, the differences between the two mechanisms are minimal (Waltman et al. 2011b). Since we are dealing with high aggregation levels, for the sake of simplicity we used the FNCS.

2012), amongst others. As publications belonging to different subject areas have different propensities for being cited (Bornmann et al., 2012; Peters & van Raan, 1994; van Raan, 2003; Waltman et al., 2011), we use normalized indicators. The use of percentile ranking and field normalized citation score can avoid bias toward a large size of country or field. Both these indicators can be computed by using the consistent *InCites/Thomson Reuters* databases, to which we had access.

In our model (3), I is a measure of citation impact in a certain period t , subject area s , and country c . LI is a lag dependent variable from the previous period, O is the number of articles and reviews in WoS , IC is the percentage of publications of a country in international collaboration, and IND is the percentage of publications of a country in collaboration with industry. C is a set of country controls, including total output, gross domestic product per capita (GDPpc), population size, and English as an official language. Finally, α is the constant, and ε is the unobserved residual.

$$I_{cs} = \alpha + \mu LI_{cs} + \lambda O_{cs} + \eta IC_{cs} + \varphi IND_{cs} + \beta C_c + \varepsilon_c \quad (3)$$

Some of our independent variables have an exponential distribution (GDPpc, population size and number of articles). We applied logarithms in these cases. As for the multicollinearity problem, none of our independent variables is highly correlated with any others ($>60\%$).

Variables such as R&D intensity, or numbers of researchers, were dismissed as typically these are highly correlated with GDPpc or the number of articles. Furthermore, for many countries, the availability or reliability of this type of data is dubious. In an earlier phase we also included in our model the variable “percentage of individuals using the internet” as a proxy for level of access to scientific journals, which is provided by the *International Telecommunications Union*. However, this indicator is also highly correlated with GDPpc. Finally, we also try to include an indicator of civil liberties, which is provided by the *Freedom House*, in order to account for country’s freedom of expression and individual rights. Our argument is that a freer society is more creative, and therefore it is more prone to generate radical ideas. Nonetheless, our preliminary results showed that this variable was insignificant in all model specifications, and therefore we decided to drop it from our model.

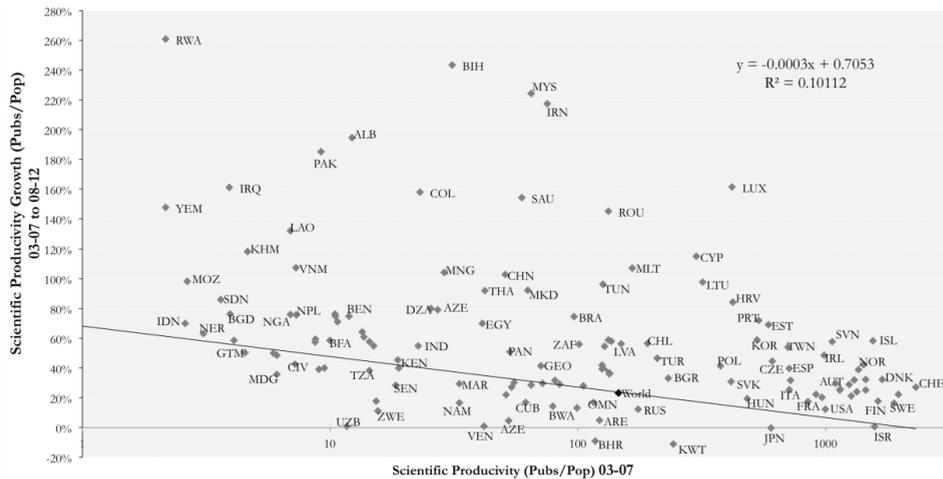
2.4. Results

2.4.1. Global Trends

The world's long-term publication output in *WoS* has increased at an average rate of 3.5% since 1981. This growth rate has increased in the decade between 2004 and 2013 to an average of 5%. In 2013, the EU28 was still the world leader for publications (35%), followed by the US (27%), China (15%), and Japan (6%). Despite these impressive figures, the world shares of the EU28, US, and Japan all fell during the preceding decade. This decline was not due to the reduction of their scientific productivity (number of publications per population size), but rather was due to the higher growth rates of other rising players, such as China or Brazil.

In Fig. 2.1, by comparing the scientific productivity growth rates of 132 countries between 2003-2007 and 2008-2012 with their scientific productivity in 2003-2007, we observe a modest trend of convergence, denoted by the negative slope of the adjusted line.

Figure 2.1. Scientific productivity (publications per million inhabitants): growth rate versus previous situation.



Source: Own calculations based on *InCites*.

Note: In this chart Pubs stands for Publications and Pop for Population (million inhabitants). The vertical axis shows the growth rate of the Pubs/Pop ratio between 2003-2007 and 2008-2012. The horizontal axis shows the number of publications per million people (Pubs/Pop) in 2003-2007 (yearly average).

The Chinese case is impressive. China's scientific publications have more than doubled over the past ten years. In contrast with the explosive growth of publications, the citation impact of that output has been observed to still be at a relatively low level (Jin & Rousseau 2004; Zhou & Leydesdorff 2006). Using the data for the period between 1997 and 2001, Zhou & Leydesdorff (2006) argue that Chinese citation rates do not match the exponential growth of their scientific production. However, based on our more recent dataset, which refers to the 2008-2012 period, our analysis shows that the average Chinese citation impact is very close to the world average, and that China already is performing considerably better than the world average in some scientific areas, such as "Agricultural Sciences"; "Engineering"; "Mathematics"; "Plant & Animal Science", and; "Social Sciences". A similar finding has been highlighted previously by (L. Wang 2016). This rapid growth reflects the coming of age of the Chinese research system, be it in terms of publications, number of researchers, or investment (UNESCO 2010).

As for Brazil, its share of world scientific output has increased at a constant rate from 1993 to 2006, followed by a fast rise in 2007 and 2008 to the level shown by Brazil in 2013. Vargas et al. (2014) argue that, in areas such as Agricultural Sciences, Brazil's output increase since 2006 was mainly due to the expansion of Brazilian journals in *WoS*, and also to an increase in the number of issues published by these journals. This phenomenon may have led to more publications, but fewer citations, as journals edited in Portuguese have less international visibility.

Iran presents another remarkable story. This country more than tripled its number of publications between the two periods analysed. According to Akhondzadeh (2013), "scientific progress in Iran over the past few years was the result of the country's recent policies and programmes to develop knowledge and facilitate researchers' access to the world's top academic resources".

In general, the world figures show a global converging trend in science regarding the quantity of publications from a significant number of world regions. This result may be inflated by changes in the size of the database, although we do not know to what extent this may be the case. *WoS* significantly increased between 2005 and 2010, in order to enlarge its regional coverage (Testa 2011), and also in response to competition from *Scopus*, which entered the market in 2004. Despite these relatively recent expansions of *WoS* possibly being one reason for the convergence that has been noted in scientific publication worldwide, a similar convergence trend has also been observed for R&D investment by the public sector between the Global North and Global South (UNESCO 2015). As Radosevic & Yoruk (2014) argue, these trends are possibly associated with a change in the scientific absorptive capacity of countries in the Global South. While 'absorptive capacity' has been generally defined as being "the ability to learn and implement knowledge" (Cohen &

Levinthal 1990), Radosevic & Yoruk (2014) defined absorptive capacity in the context of scientific research as being “the ability to recognize the value of new, external information, assimilate it, and apply it in another context”. In accordance with this view, researchers recombine and re-contextualize existing scientific knowledge and are able to generate novelty through their new publications.

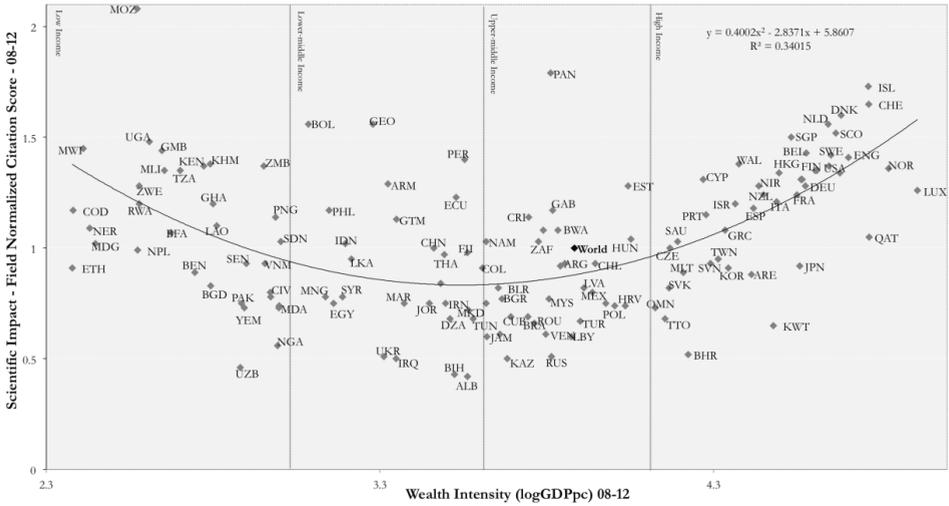
There is still a huge gap to overcome between higher-income and lower-income nations, however the convergence which has been noted over the most recent years has certainly occurred due to the fact that some countries in the Global South are expanding their scientific capabilities, and are increasing their presence in scientific journals that have high international visibility. Such changing trends provide some support to our quest to understand the determinants of citation impact in the Global South, despite the fact that we analyse this by adopting indicators that are normally used to assess science in the Global North.

2.4.2. Wealth Intensity versus research citation impact

In this study, we are particularly interested in understanding whether there are different determinants of citation impact (scientific influence) at different levels of GDPpc. In order to have a general overview of the relation between wealth intensity and citation impact, in Fig. 2.2 and Fig. 2.3 we scatter the relation between average GDPpc and citation impact, as measured respectively by FNCS and PPTop10% between 2008 and 2012, for countries that have more than 400 publications.

In both charts we can observe a U-shaped pattern, with the adjusted lines having their inflexion points close to the *World Bank* borderline which divides “low & lower-middle-income” countries from “upper-middle & higher-income” countries. For “low & lower-middle-income” countries, the citation impact performance seems to follow a downward trend, although with substantial deviations from the curve. For “upper-middle & higher-income” countries, there seems to be a positive relation between the two variables. Such an upward trend has already been revealed by King (2004). In contrast, the U-shaped pattern of our data seems to suggest that a nation’s wealth only correlates positively with citation impact after a certain level of GDPpc.

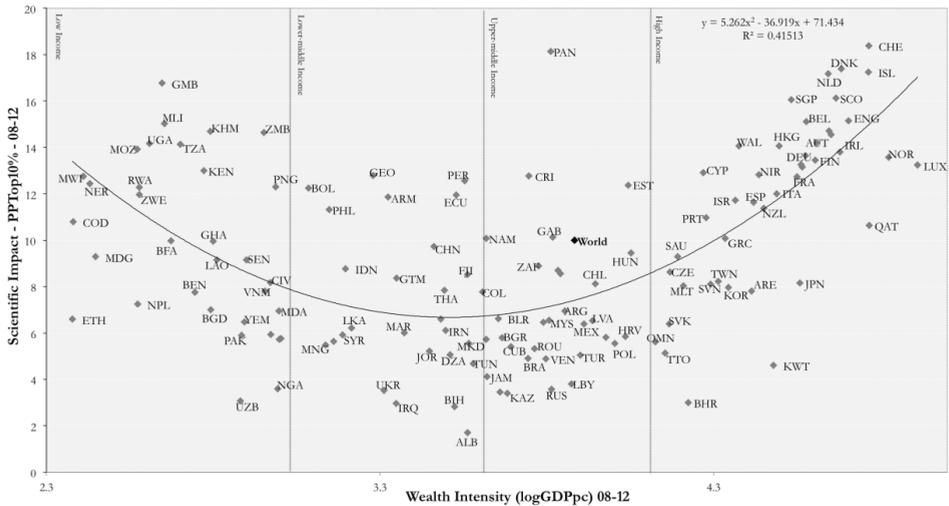
Figure 2.2. Citation Impact (FNCS) versus GDPpc (2008-2012)



Source: Own calculations based on *InCites* & World Bank

Note: The vertical axis shows normalized citation impact in 2008-2012; the horizontal axis shows the logarithm of GDPpc (constant 2005 USD) in 2008-2012 (yearly average).

Figure 2.3. Share of highly cited publications (PPtop10%) versus GDPpc (2008-2012)



Source: Own calculations based on *InCites* & World Bank

Note: The vertical axis shows the share of highly cited publications in 2008-2012; the horizontal axis shows the logarithm of GDPpc (constant 2005 USD) in 2008-2012 (yearly average).

One should be aware that the countries shown in these two charts have different size dimensions. If a small country, in terms of publication output, has a set of publications that is very influential, then the citations received by the articles produced by those researchers will improve its citation intensity score significantly. Mozambique, for example, is one of these cases. Despite its total production normalized by population being very low when compared to the world average (5 vs. 179 yearly publications per million people), its FNCS is two times higher than the world average, and its PPTop10% is close to 14%. In Mozambique, from 2008 to 2012, 95% of the country's publications have a foreign author. The high levels of citation impact in Mozambique may stem from the country having only a small group of scientists who produce scientific publications with highly reputed international co-authors (Confraria & Godinho 2015).

Another outlier in our charts is Panama, and there is also an explanation for this case. Its citation impact (in intensity) is 79% higher than the world average for the FNCS indicator, and 18% of its publications are in the top ten most highly-cited papers. If we take a close look at the most highly cited publications from Panama between 2008 and 2012, we find that most of them come from researchers affiliated to the *Smithsonian Tropical Research Institution*. This organization is a bureau of the *Smithsonian Institution* based outside the United States, which is dedicated to understanding biological diversity. According to its website,¹¹ its “facilities provide a unique opportunity for long-term ecological studies in the tropics, and are used extensively by some 900 visiting scientists from academic and research institutions in the United States and around the world every year”. In a country such as Panama, which had a scientific output close to 1,500 publications during the five years analysed, the presence of this research institute certainly makes a difference. The *Smithsonian Institution* functions as a hub, attracting world leading scientists, and it certainly has a huge influence on the high citation impact of Panama.

On the right edge of the U curve, we find high-income countries such as Switzerland, Denmark, Iceland and The Netherlands. These are all relatively small European nations, which have been leading performers in this indicator for quite some time. For example, in one of the first studies analysing this issue, May (1997) also found that these countries were already leading the world in terms of “citation intensity”.

In summary, our descriptive analysis suggests that higher levels of international collaboration may be extremely relevant for countries that may simultaneously have low GDPpc and relatively smaller scientific communities. It is also perceivable that despite the fact that middle-income countries may enjoy more resources and larger scientific communities, they do not engage in overseas

¹¹ http://www.stri.si.edu/english/about_stri/index.php

collaboration so much, and this is reflected in the lower levels of impact on average. As we progress from the left to the right in our chart, the initial downward trend of the U-shaped pattern indicates that the improvement in GDPpc, from low-income to middle-income levels, leads to lower citation impact, on average. Finally, for high-income countries, both higher levels of GDPpc and country size, which is again negatively correlated with higher levels of international collaboration, seem to be critical factors.

2.4.3. Regressions analysis

We used *Stata* (StataCorp 2013) to compute the multilevel regression (OLS) with fixed effects at the subject area level, and errors clustered at the country level. The determinants of citation impact for publications between 2008 and 2012 were examined for 21 subject areas for countries with at least both 50 publications in a subject area and a total of 400 publications. After applying these restrictions, 126 countries and 1686 observations compose our global sample (see Table 2.1 for descriptive statistics).

Table 2.1. Descriptive Statistics

Variable	N	Mean	SD	Min	Max	Correlation														
						1	2	3	4	5	6	7	8	9	10					
1) PPTop10%_0812	1686	9.39	5.28	0.00	55.51	1.00														
2) PPTop10%_0307	1686	8.45	4.73	0.00	32.81	0.78	1.00													
3) FNCS_0812	1686	1.02	0.44	0.09	5.75	0.90	0.69	1.00												
4) FNCS_0307	1686	0.92	0.34	0.03	3.09	0.78	0.93	0.71	1.00											
5) Pubs Area (log)	1686	3.03	0.72	1.72	5.57	0.24	0.29	0.19	0.29	1.00										
6) Internat. Collab	1686	56.07	19.55	9.66	100	0.34	0.23	0.36	0.25	-0.51	1.00									
7) Industry. Collab	1686	1.75	2.50	0.00	23.38	0.35	0.28	0.32	0.29	0.13	0.09	1.00								
8) Total Pubs (log)	1686	4.40	0.77	2.63	6.24	0.20	0.26	0.13	0.26	0.87	-0.52	0.17	1.00							
9) English Official	1686	0.25	0.43	0.00	1.00	0.20	0.17	0.18	0.19	0.09	0.03	-0.02	0.08	1.00						
10) GDPPc (log)	1686	3.95	0.62	2.38	4.91	0.38	0.36	0.30	0.38	0.47	-0.21	0.31	0.55	-0.02	1.00					
11) Popul. (log)	1686	7.27	0.67	5.50	9.13	-0.20	-0.12	-0.20	-0.14	0.42	-0.27	-0.12	0.46	0.07	-0.39	1.00				

Note 1: Correlation with bold numbers significant at $p < .05$;

Note 2: The numbers 0307 and 0812 in the variables stand for the time periods 2003-2007 and 2008-2012.

Previously, in Fig. 2.2 and Fig. 2.3 we have seen that for different levels of GDPpc (below and above world average level), there are different patterns of citation impact. To understand whether these differences are substantive for the purpose of our analysis, we split our sample into two groups of countries, following the World Bank's definition of "low & lower middle income" countries (Global South), and "upper middle & higher income" countries (Global North). Furthermore, we also introduce another North-South distinction, namely being, or not being an OECD country. In Table 2.A.2, in the Appendix, we provide descriptive statistics of these four groups. In both North-South specifications, citation impact, number of publications, and level of industry collaboration is significantly higher in the North. Level of international collaboration is substantially higher in the South.

Regressions were carried out for each of these groups separately. Generally, the results for both North and South specifications are robust. Table 2.2 reports the effect of the predictor variables on citation outcomes, using two dependent variables for the citation rates, which are respectively, PPTop10% and FNCS. The South samples include 54 "low & lower-middle-income" countries (490 observations), and 89 non-OECD countries (928 observations), while the North samples include 72 "upper middle & higher income" countries (1196 observations), and 37 OECD countries (758 observations).¹² Our model not only identifies those variables that are significant in predicting higher levels of citations, but it also identifies the relative contribution of each independent variable to the citation rates of countries.

¹² We count England, Scotland, Wales and Northern Ireland as separate nations.

Table 2.2. Determinants of citation impact in the Global South and the Global North

Independ. Variables	Dependent variables							
	PPtop10%_0812				FNCS_0812			
	South (GDPpc)	North (GDPpc)	non- OECD	OECD	South (GDPpc)	North (GDPpc)	non- OECD	OECD
PPtop10%_0307	0.581*** (0.061)	0.583*** (0.046)	0.557*** (0.048)	0.692*** (0.053)				
FNCS_0307					0.643*** (0.107)	0.561*** (0.067)	0.576*** (0.076)	0.687*** (0.086)
Pubs Area (log)	2.281*** (0.572)	1.189*** (0.335)	2.049*** (0.404)	0.944** (0.402)	0.156** (0.062)	0.015 (0.035)	0.128*** (0.040)	-0.013 (0.043)
Int. Collab	0.082*** (0.013)	0.090*** (0.011)	0.091*** (0.010)	0.087*** (0.010)	0.007*** (0.002)	0.007*** (0.001)	0.008*** (0.001)	0.007*** (0.001)
Ind. Collab	0.317*** (0.106)	0.186** (0.090)	0.268*** (0.095)	0.188* (0.107)	0.053** (0.023)	0.012** (0.006)	0.031*** (0.011)	0.010 (0.006)
Total Pubs (log)	0.144 (0.414)	1.211*** (0.423)	0.557* (0.315)	0.253 (0.758)	0.047 (0.054)	0.209*** (0.049)	0.104*** (0.036)	0.131 (0.08)
English Official	1.267*** (0.445)	0.651** (0.271)	0.929*** (0.324)	1.001*** (0.338)	0.059 (0.051)	0.043** (0.020)	0.050* (0.029)	0.066*** (0.021)
GDPpc (log)	-0.171 (0.592)	0.634 (0.435)	-0.067 (0.352)	0.146 (0.650)	-0.097 (0.063)	-0.002 (0.036)	-0.075** (0.033)	-0.034 (0.053)
Popul. (log)	-1.053** (0.410)	-1.723*** (0.345)	-1.292*** (0.336)	-0.820 (0.537)	-0.099** (0.041)	-0.176*** (0.030)	-0.122*** (0.031)	-0.099** (0.048)
Constant		-0.277 (2.652)		-0.165 (2.853)		0.342 (0.220)		0.302 (0.239)
Observs.		1,686		1,686		1,686		1,686
R-squared		0.730		0.729		0.649		0.644

Note 1: Robust standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note 2: Linear regression, absorbing indicators. Std. error adjusted for 126 clusters (countries).

These results show that, in both groups of countries, previous citation impact, level of international collaboration, and number of publications in the specific area are strongly associated with higher citation rates.

The first of these results indicates that, despite the fast growth of some countries in recent years, globally a strong path-dependency in citation impact still holds. There are specific reasons for this occurrence. It is well known in the literature that better-known scientists tend to receive more credit than less well-known ones, even if their research is similar (Merton 1968). Frequently cited researchers generally have higher status than those researchers who are cited less frequently. Because status influences perceptions of quality, those with a high reputation are more likely to be cited in the future, which thus further reinforces their status. If we admit that similar self-reinforcing mechanisms exist at a more aggregate level, then we can argue that nations performing better are also more likely to attract more tangible resources, such as research funding and outstanding graduate students, which can result in research of better quality and also perpetuation of higher levels of citation impact. Our models do not suggest big differences between South and

North in this regard, however they do show that previous performance is strongly associated with future citation impact; for example, for the PPTop10% indicator, those countries that have 1% more papers in the top 10% more-cited publications than others, have around 0.6% more papers in that same “excellent” tier during the next period.

As for the scientific output variable, those countries that produce more publications in specific subject areas also have higher citation rates per paper. This is intuitive, as, in theory, these are the subject areas in which countries have a higher scientific capacity. At the same time, given the effect of scale, researchers in the same subject areas probably cite their own compatriots more frequently, which thus increases the number of citations received by their country. This covariate represents not only the scientific output, but also the intensity of involvement in scientific activities of a country in a specific area (as gross expenditure on R&D, and number of researchers are usually highly-correlated with number of publications). As this effect is significantly higher in the Global South, one implication of this result is that the importance of generating a higher critical mass in a specific field, in order to produce research with more influence in the world, seems to be larger in the South. For instance, those countries in the South that have 50% more publications in a subject area than others, have, on average and controlling for all the other factors, 1% more papers in the top 10% most-cited publications in the world. In the North this relation is significantly smaller.

With regards to international collaboration, it is well known in the literature that citation impact is typically greater when research groups collaborate amongst each other, and the benefit strengthens when co-authorship is international (van Raan, 1998). The rationale behind this is that scientists are likely to develop new and alternative ways of thinking when they interact with other scientists with diverse areas of expertise and backgrounds (Hollingsworth 2006). Co-publication allows access to a larger social network, which consequently leads to increased visibility, which in turn is reflected in higher citation rates (Goldfinch et al., 2003). This cross-fertilization is amplified by international collaboration, as scientists who produce co-authored papers with foreign scientists are more likely to belong to elite research groups within their own countries (Adams 2013). As countries in the Global South depend a lot on international scientific networks to produce research that has visibility and impact (see Fig. 2.A.1 in appendix), we would expect that this positive relation to be higher in the countries from the Global South. Yet our results seem to show that the importance of international collaboration is not significantly different in both groups of countries. Specifically, those countries that have 10% more internationally co-authored publications in a subject area have, on average, 0.85% more publications in the top 10% in that subject area.

An interesting finding is that industry collaboration seems to matter for citation impact, especially in the Global South. This is because most co-publications with industry are co-authored by staff from the large R&D-intensive technology companies in science-based industrial sectors, such as biotechnology, pharmaceuticals, electronics, chemicals, and computers (Godin, 1996). This indicator can be seen as being a “knowledge linkage indicator” (R. J. W. Tijssen 2012) between multinational R&D-intensive technology companies and public research organisations. This type of collaboration with industry is very likely to be driven by the need for access to international R&D networks, advanced research facilities, and contributions by scientists and research teams of international repute. Whereas from the industry side, researchers may be tempted to publish because they aspire to be active members of a research community and they want to be regarded as such by their peers, together with the other objectives of making corporate research findings public (Godin, 1996; Tijssen et al., 2009). The industry side may feel a particular appeal to collaborate with scientists from the South as a way of reaching specific resources, or of testing new medicines.

We found that this type of collaboration may be relevant for countries in the Global South, not only for updating their technological capabilities, but also for increasing their visibility and impact in the scientific community of their field. However, it is relevant to acknowledge that co-authorships with industry are far from common in science, which thus represents a case of corner outcomes with an edge at zero, and a continuous distribution for strictly positive values (our sample as mean value of 1.73%). Our results show that, in line with (Tijssen 2012), the intensity of science-industry co-authorship is lower in African and Latin American countries, than in countries in the North. Therefore, we should be cautious when interpreting this result, as the incidence of few publications in collaboration with industry can substantially change this indicator (high sensibility). Further to this, if we add a variable to our model that interacts industry collaboration intensity with international collaboration intensity, then the covariate industry collaboration changes sign and its significance disappears. At the same time, the international collaboration parameter remains positive and significant. Therefore, it is not clear whether the positive and significant effect of industry collaboration intensity in citation impact in our general model is due to the industry “effect”, or whether it just occurs because most industry collaborations are also international collaborations.

For those countries that have English as an official language, our results show that the relation is positive and significant in almost every model specification. As the majority of scientific journals are written in English, and as articles published in a non-English language have less potential readers, this positive relation was an expected result. In the Global South, an Anglophone colonial

history and concomitant opportunities for partnerships with English speaking countries (e. g., by hosting international research institutes) may have a significant effect on their citation impact.

These results also indicate that, contrary to what has been revealed by King (2004), the relation between GDPpc and citation impact is not strictly positive. One would expect wealthier countries to have more resources to apply to science, and that therefore they would perform better in terms of citation impact. However, for countries in the Global South, the coefficients are negative, and in the North they are positive and negative (non-significant), depending on the model specification. We also tried to understand whether the U relationship show in Fig. 2.2 and 2.3 holds in our model with all countries. However, when we include the variable GDPpc squared in the regression (see Table 2.A.3 in the Appendix), the coefficient is positive, but non-significant. These results indicate that there are other elements beyond wealth intensity that matter for research quality in the South, namely: previous performance, a higher level of international collaboration, and more publications in the specific subject area.

Finally, for country size, in terms of total scientific output, there is no clear pattern, as our results differ depending on the model specification. However, countries with higher population seem to have, on average, less citation impact than smaller countries. A possible interpretation for this is that smaller countries are more involved in international collaborations to produce their scientific articles. This may be so, as when we interact country size with level of international collaboration, the negative effect of population size is no longer significant, and the interaction variable seems to capture this effect. For example, Frame & Carpenter (1979) also argued that the scientific size of a nation determines the need for international collaboration. Small countries provide fewer opportunities to find collaborators inside their borders when compared to larger ones, and thus have a greater need for research partners from other countries (Narin et al. 1991). Our results do not show significant difference between the South and North.

To complement this analysis, in the Appendix we carry out three different robustness checks. In Table 2.A.4, in order to explore the performance of countries with different levels of international collaboration, we create two sub-groups (i.e. low international collaboration intensity and high international collaboration intensity) in both Global South and Global North. In Table 2.A.5, we carry out the same analysis as in Table 2.2, but instead of separating the world into South and North, we use four broad world regions to see whether there are significant differences between them. In Table A.6, we computed our model for all countries in our sample, using subject area groups, in order to check whether the results are consistent in most areas of knowledge. In general, these results are consistent with the previous models. In Table 2.A.4, we show that previous

citation impact contributes more in the lower international collaboration group, whereas number of publications, level of international collaboration, and level of collaboration with industry all have a higher effect in the higher international collaboration group. In Table 2.A.5, the main findings are that previous citation impact is more relevant in Africa, and that Latin America & the Caribbean is the region where international collaboration has a higher effect on citation impact. Finally in Table 2.A.6, we found that both level of international collaboration and previous citation impact are positively and significantly associated with higher levels of citation impact in almost all areas. The same also occurs in eight areas for scientific output in the specific area, and in eleven areas for English as an official language.

2.5. Discussion & Conclusion

In this chapter, bibliometric and econometric analysis were used to identify which countries are producing research with higher scientific influence, and also to understand which factors lead to these higher results. We focused particularly on the Global South, as the scientific output of some of these countries has been converging recently with that of the Global North.

We found some evidence suggesting that the determinants of citation impact may not coincide across countries in different wealth intensity levels. While previous citation impact, level of international collaboration, and publication output in a specific scientific field are all important determinants of citation impact among all nations, we observed that the variable number of publications in a specific disciplinary area appears to be substantially more important in the South than in the North. This covariate represents not only the scientific output, but also the intensity of involvement in the scientific activities of a country in a specific area. The agglomeration effects that may arise in some disciplines in scientific communities that are generally much smaller than their counterparts in the North seem therefore to be relevant. This implies that the importance of concentrating resources and of generating higher critical masses in specific fields, in order to produce research with more influence in the world, is seemingly greater in the South.

As for our lagged dependent variable, we confirmed that it has an important effect on citation impact in both groups of countries. Societies vary in their capacity to produce major scientific discoveries over time. This happens because they are influenced in various ways by previous historical processes and institutional settings. This type of path dependence in scientific knowledge arises as a consequence of researchers in different types of organisations in the same country engaging in a great deal of common learning and socialization, which is transmitted across time and across organisations.

Eventually, this path-dependence can be transformed into virtuous or vicious circles of development. Due to the fact that countries in the Global South have, on average, few “excellent” researchers with the know-how and tacit knowledge needed to engage in virtuous circles, one potential implication is that a “brain-drain” may have a severe negative effect on their scientific performance. If their few best “minds” leave to carry out research abroad and do not come back, or do not interact with their national colleagues, then the tacit knowledge will decline and the potential spillovers that they generate will stop being used for their countries’ benefit.

With regard to the level of international collaboration, as has been widely shown by past research, a positive and significant relationship exists with citation impact. With the advances in information and communication technology and institutional changes, scientists can more easily obtain relevant knowledge by collaborating with other peers with diverse areas of expertise and backgrounds. Accessing external complementary knowledge and skills through networking, namely with scientists working in more developed environments, seems to be extremely relevant for performing research with high impact. However, interestingly, our results suggest that, contrary to what could be expected, this covariate does not seem more relevant in the South than in the North. This therefore indicates that the interest in pursuing international collaborations seems to be equally relevant in both environments.

Our analysis also suggests that industry collaboration seems to be positively associated with citation impact, especially in the Global South. However, it is not clear whether the positive and significant effect of industry collaboration intensity on citation impact in our general model is due to the “industry effect”, or just occurs because the relatively few industry collaborations performed by the South also happen to be international collaborations.

In our regressions we also used country controls. We found that smaller countries (population wise) and countries with English as an official language perform on average better than others in some model specifications. A possible interpretation of this finding is that smaller countries rely more on international collaborations to produce their scientific articles. This was confirmed by assessing the interaction of country size with level of international collaboration. When such a possibility was tested, the negative effect of population size was no longer significant, with the interaction variable seemingly capturing this effect. For those countries that have English as an official language, as the majority of scientific journals are written in English, and as articles published in other languages have less potential readers, this positive relation was an expected result. Besides this, those countries that have English as an official language usually have a colonial legacy with Anglo-Saxon countries (US, UK, Canada, Australia), and consequently have more

collaboration with them (Mégnybêto 2013; Pouris 2010). As these are the leading countries in many scientific fields, this positive relation is therefore reinforced. Finally, there is no clear relation between wealth intensity as measured by GDPpc and citation impact. It would be expected that wealthier countries would have more resources to apply to science, and therefore would perform better in terms of citation impact. However, we found that other elements beyond wealth intensity are much more relevant for the research quality of nations.

It is worth noting in relation to the groups of countries that we have assumed in this chapter, that there could be other alternatives for their classification. The division suggested by us allocates countries to one of two groups, respectively Global South and Global North, or alternatively OECD and Non-OECD Countries. In doing this, we mainly took into consideration differences in economic wealth. However, as it has been shown by research on the same topic, countries across the globe may cluster into different groups depending also on geographical, political, ideological, cultural, or demographic lines. For example, Moya-Anegón & Herrero-Solana (2013) established a typology of three main groups of countries worldwide, according to the thematic characterisation of scientific output in journals of international visibility. Their results show that each of these groups accounted for specific behavioural models, reflecting the distinctive characteristics of knowledge production in each country.

Furthermore, scientific performance worldwide may be influenced by the reputation of the affiliation, this being determined by the institution that the authors belong to (Peters & Ceci, 1982), or the country of the address of the submitted publications. Smith et al. (2014) investigated specifically whether the country where an author is based influences the notoriety of manuscripts. Their study found that, generally, international co-authorship enhanced scientific performance, but more specifically, they found that specific combinations of countries for the authorship of papers influence differently the performance of published papers, with the effect of these specific combinations also varying across disciplinary areas. This result suggests that the complexity of the factors determining scientific performance across countries may go well beyond the relationships stipulated by our econometric model.

Another open question has to do with the adequacy of bibliometric indicators in different socio-economic contexts. In this chapter we were aware that potential biases could arise from applying bibliometric indicators to countries belonging to the Global South. It is widely accepted that these types of indicators capture poorly certain types of research and that they encourage certain scientific activities and behaviours, including a shift towards English publications (Hicks et al., 2015), diversion of research away from local or national issues (Hicks et al. 2015), scientific supply

poorly aligned with societal needs (Sarewitz & Pielke, 2007) and bias toward positive reporting (Fanelli 2011), etc. As the Global South has a “lower” status in the scientific enterprise, these effects may be aggravated within this group. Another important issue when measuring citation impact is to be aware that it is a relative indicator. For example, if a country has the same citation impact (measured in intensity) as the US, but it has 1,000 times less publications than the US, then evidently the actual absolute impact (scientific, societal and economical) of its research in the world is completely different. Therefore, even when using thresholds, as we did, indicators measuring citation impact should always be interpreted within their context, as we have done in the examples of Mozambique and Panama.

Another possible limitation stems from this study being mainly carried out in a macro perspective, based on bibliometric indicators. This has certainly impaired our understanding of the specificities of the national scientific systems. The level of knowledge about science in lower income contexts would certainly improve by complementing this analysis with a more qualitative approach, such as researching why specific institutions in the Global South have such high performance levels, and by understanding their interactions. Furthermore, as the relational dimension (who do you collaborate with? What is the strength of the relationships?) seems to matter for citation impact (Gonzalez-Brambila et al. 2013), improving this model by using measures of network centrality, instead of level of international collaboration could give us a better understanding of the role of scientific network co-authorships for citation impact. In chapters 3 and 4 we will partially address these issues by combining bibliometric with survey data at the author level.

With regards to the normative implications, our findings allow us to draw some potentially relevant indications. Lower and middle-income countries with globally small scientific communities would better concentrate their resources in generating higher critical masses in specific fields, in order to produce research with a higher impact. Furthermore, the interest in pursuing international collaborations seems more than justified. International scientific collaborations have been pursued more intensely by smaller countries, which is understandable, given the fact that larger countries may have larger numbers of researchers in every single major discipline, and thus the need to collaborate abroad does not arise as much as in the former case. However, even for the larger countries, there may be good reasons for scientists to seek collaboration abroad, at least in some fields, thus balancing this orientation without jeopardizing the cohesion of their research systems. These recommendations assume that increasing the impact of scientific publication in the South is an important objective, and that such impact is directly related to the quality of the research produced. However, it may be relevant to bear in mind the distinction between academic and

practical impact, and the value for countries to have diverse “research portfolios” (Wallace & Rafols, 2015). Although one may assume that in the long-term, both academic and practical impact may coincide, wise policy-makers in the Global South may recognize these may well not coincide for shorter time-spans and in specific geographic or institutional conditions. Furthermore, diversifying research options can potentially allow to hedge against uncertainty and ambiguity (e.g. the solution for a medical problem may come from a better understanding of public health in a certain context and not from a new vaccine or pharmaceutical).

Finally, the science policy-making process needs to keep in mind the strong path-dependencies that dominate scientific activities globally. Despite the success stories of a few lower and middle-income countries that have forged ahead in scientific matters during the most recent decades, most countries in the Global South remain held back by the chains of path-dependency. Overcoming such path-dependencies implies persistence, continuous investment, and far-reaching institutional change, as these successful cases have confirmed.

2.6. Appendix

Table 2.A.1. Significant determinants of citation impact, based on previous studies (not exhaustive)

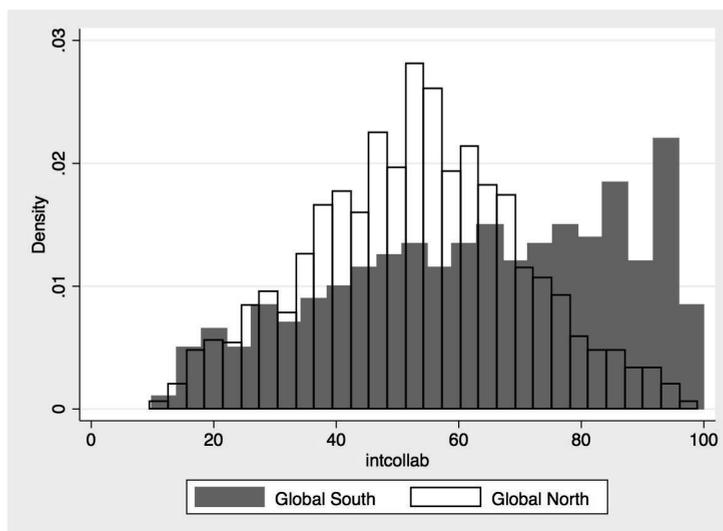
Level and Factors	What associates with higher citation	Prior literature
Article		
Number of authors	Four or more authors	(Leimu & Koricheva 2005)
Length of the abstract	Longer abstract	(Leimu & Koricheva 2005)
Journal impact factor (JIF)	Articles in journals with higher JIF	(Peters & van Raan 1994); (Didegah & Thelwall 2013)
Number of references	More references	(Peters & van Raan 1994)
Impact of references	Higher no. of citations	(Bornmann et al. 2012); (Didegah & Thelwall 2013)
Length of the paper	Longer paper	(Peters & van Raan 1994)
Type of document	Reviews	(Peters & van Raan 1994)
Language	English journal and paper	(Peters & van Raan 1994)
Author		
Country of origin	Native English-speaking authors	Leimu & Koricheva (2005)
Previous performance	More citations in the past	Merton (1968)
Institution		
Size	Universities with a large publication output	(Moed et al. 2011)
Number of institutions	More institutions	(Narin et al. 1991)
Specialisation intensity	Weak negative effect	(Moed et al. 2011)
Country		
Economic development	Higher GDP per capita	(King 2004) Glänzel et al. (1995); Katz & Hicks (1997);
Number of countries of affiliation	More countries	Narin et al. 1991; Puuska et al. (2013); van Raan (1998)
Country of affiliation	English speaking country	(Leimu & Koricheva 2005)

Source: Own elaboration.

Table 2.A.2. Descriptive statistics in the Global South and the Global North

Variables		South (GDPpc)	North (GDPpc)	non-OECD	OECD
	Countries		54	72	89
Obs.		490	1196	928	758
PPtop10%_0812	Mean	7.64	10.11	7.54	11.66
	Std. Dev.	5.22	5.13	5.02	4.67
PPtop10%_0307	Mean	7.02	9.03	6.78	10.49
	Std. Dev.	4.88	4.54	4.62	4.01
FNCS_0812	Mean	0.91	1.06	0.89	1.18
	Std. Dev.	0.52	0.39	0.46	0.35
FNCS_0307	Mean	0.81	0.76	0.79	1.07
	Std. Dev.	0.33	0.96	0.32	0.3
Pubs Area	Mean	2481	5580	2139	7790
	Std. Dev.	10419	16226	8036	19835
Int. Collab (%)	Mean	63.33	53.09	59.01	52.46
	Std. Dev.	22.95	17.11	21.7	15.82
Ind. Collab (%)	Mean	0.96	2.07	1.14	2.51
	Std. Dev.	1.52	2.74	1.81	2.97
Total Pubs	Mean	40023	116912	43185	163870
	Std. Dev.	143143	241811	109194	291501
English Official	Mean	0.27	0.24	0.25	0.25
	Std. Dev.	0.45	0.43	0.43	0.43
GDPpc	Mean	1814	24950	6949	32033
	Std. Dev.	1166	17143	8945	16305
Population	Mean	1.63E+07	3.26E+07	1.00E+08	3.44E+07
	Std. Dev.	3.48E+08	5.38E+07	2.64E+08	5.65E+07

Source: Own elaboration.

Figure 2.A.1. Distribution of international collaboration levels. South vs North (2008-2012)

Source: Own calculations based on *InCites*

Note: Vertical axis shows the density of observations in a specific level of international collaboration in 2008-2012; Horizontal axis shows level of international collaboration.

Table 2.A.3. Determinants of citation impact in all countries

Variables	PPtop10%_0812	FNCS_0812
PPtop10%_0307	0.590*** (0.040)	
FNCS_0307		0.606*** (0.060)
Pubs area (log)	1.507*** (0.324)	0.066* (0.033)
Int. Collab	0.085*** (0.008)	0.007*** (0.001)
Ind. Collab	0.214** (0.084)	0.017*** (0.006)
GDPpc (log)	-2.867 (2.140)	-0.350* (0.205)
GDPpc ² (log)	0.378 (0.291)	0.038 (0.027)
English Official	0.910*** (0.254)	0.053** (0.021)
Total Pubs (log)	0.803** (0.335)	0.134*** (0.037)
Popul. (log)	-1.451*** (0.330)	-0.137*** (0.029)
Constant	6.764 (4.486)	0.992** (0.436)
Observations	1,686	1,686
R-squared	0.726	0.638

Note 1: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Note 2: Linear regression, absorbing indicators. Std. error adjusted for 126 clusters (countries).

Table 2.A.4. Determinants of citation impact in the Global South and the Global North by international collaboration groups

Variables	PPtop10%_0812			
	South (GDPpc)		North (GDPpc)	
	Int. Collab >= average (56%)	Int. Collab < average (56%)	Int. Collab >= average (56%)	Int. Collab < average (56%)
PPtop10%_0307	0.482*** (0.074)	0.526*** (0.066)	0.663*** (0.094)	0.688*** (0.049)
Pubs Area (log)	2.461*** (0.924)	1.188* (0.622)	0.322 (0.432)	0.495 (0.391)
Int. Collab	0.125*** (0.023)	0.153*** (0.024)	0.056*** (0.017)	0.047*** (0.012)
Ind. Collab	0.325** (0.129)	0.249** (0.103)	0.098 (0.204)	0.052 (0.081)
Total Pubs (log)	0.031 (0.667)	2.339*** (0.730)	1.042* (0.568)	0.641 (0.528)
English Official	1.510** -0.58	0.758* -0.387	-0.327 -0.46	0.674** -0.307
GDPpc (log)	0.208 (0.744)	-0.307 (0.696)	-0.278 (0.854)	2.064*** (0.658)
Popul. (log)	-1.504** (0.632)	-2.292*** (0.567)	0.096 (0.428)	-0.956** (0.426)
Constant	-1.069 (3.627)		-5.708 (3.564)	
Observations	815		871	
R-squared	0.682		0.805	

Note 1: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Note 2: Linear regression, absorbing indicators. Std. error adjusted for 126 clusters (countries).

Table 2.A.5. Determinants of citation impact in four World regions (Africa; Asia; LA&C - Latin America & Caribbean; E&NA&P - Europe & North America & Pacific)

Variables	Dependent variables							
	PPtop10%_0812				FNCS_0812			
	Africa	LA&C	Asia	E&NA&P	Africa	LA&C	Asia	E&NA&P
PPtop10%_0307	0.690*** (0.062)	0.477*** (0.104)	0.532*** (0.077)	0.598*** (0.052)				
FNCS_0307					0.811*** (0.171)	0.354** (0.144)	0.558*** (0.068)	0.602*** (0.084)
Pubs Area (log)	1.139** (0.523)	2.349*** (0.490)	1.888*** (0.524)	1.494*** (0.425)	0.098* (0.056)	0.197*** (0.064)	0.074* (0.044)	0.052 (0.048)
Int. Collab	0.069*** (0.012)	0.122*** (0.025)	0.069*** (0.014)	0.106*** (0.012)	0.004** (0.002)	0.013*** (0.003)	0.006*** (0.001)	0.009*** (0.001)
Ind. Collab	0.353*** (0.111)	-0.097 (0.081)	0.109 (0.078)	0.252** (0.100)	0.057 (0.036)	0.005 (0.006)	0.009 (0.006)	0.016** (0.007)
Total Pubs (log)	-0.610 (0.551)	-0.040 (0.669)	0.341 (0.461)	0.110 (0.697)	-0.062 (0.106)	0.034 (0.055)	0.122*** (0.043)	0.102 (0.077)
English Official	0.478 (0.401)	0.340 (0.871)	1.193** (0.509)	1.120*** (0.374)	0.033 (0.066)	0.085 (0.082)	0.049 (0.035)	0.080*** (0.026)
GDPpc (log)	-0.351 (0.650)	-0.892 (0.785)	0.571 (0.406)	0.374 (0.577)	-0.092** (0.043)	-0.263** (0.132)	-0.023 (0.035)	-0.037 (0.054)
Popul. (log)	0.028 (0.401)	-0.566 (0.423)	-0.996** (0.411)	-0.925* (0.511)	-0.007 (0.068)	-0.013 (0.071)	-0.104*** (0.034)	-0.107** (0.051)
Constant			-1.814 (3.051)				0.233 (0.274)	
Observations			1,686				1,686	
R-squared			0.740				0.661	

Note 1: Robust standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Note 2: Linear regression, absorbing indicators. Std. Err. adjusted for 126 clusters (countries).

Note 3: Africa = 221 observations (28 countries); LA&C = 195 observations (16 countries); Asia = 329 observations (34 countries); E&NA&P = 832 observations (48 countries).

Table 2.A.6. Determinants of citation impact in the World (2008-2012) by subject area using the percentile ranking (PPtop10%)

Independ. variables		PPtop10 %_0307	Pubs area (log)	Int. Collab	Ind. Collab	English Official	GDPpc (log)	Total Pubs (log)	Popul. (log)
Subject Areas	Obs.								
Agricultural Sciences	91	0.548*** (0.093)	2.451** (1.029)	0.112*** (0.018)	-0.098 (0.178)	1.273** (0.613)	0.188 (1.138)	1.997* (1.086)	-2.247*** (0.831)
Biology & Biochemistry	79	0.755*** (0.109)	0.998 (2.040)	0.052*** (0.016)	0.043 (0.200)	0.218 (0.489)	1.503 (1.038)	-0.546 (2.255)	-0.365 (0.747)
Chemistry	95	0.597*** (0.116)	0.888 (1.252)	0.014 (0.016)	0.258* (0.150)	1.856*** (0.658)	2.114*** (0.635)	-0.515 (1.418)	0.133 (0.557)
Clinical Medicine	115	0.512*** (0.069)	1.758 (1.116)	0.153*** (0.023)	0.598*** (0.184)	1.240** (0.544)	-0.931 (0.788)	1.871* (1.071)	-2.166*** (0.686)
Computer Science	70	0.328*** (0.103)	3.083** (1.374)	0.135*** (0.036)	-0.156 (0.111)	0.458 (0.611)	-1.678 (1.812)	1.728 (1.633)	-2.227 (1.409)
Economics & Business	53	0.405*** (0.121)	6.723*** (2.446)	0.054** (0.025)	0.650 (0.438)	-0.276 (0.745)	-0.267 (2.499)	-3.571** (1.409)	-0.842 (1.348)
Engineering	87	0.680*** (0.171)	1.968 (1.593)	0.039 (0.028)	-0.045 (0.112)	1.617** (0.684)	3.255*** (1.130)	-3.574*** (1.172)	2.044** (1.027)
Environment/ Ecology	92	0.548*** (0.105)	0.138 (2.307)	0.127*** (0.023)	0.445 (0.436)	2.095*** (0.723)	0.299 (1.048)	2.500 (1.873)	-1.529* (0.921)
Geosciences	87	0.429*** (0.081)	2.476** (1.242)	0.165*** (0.024)	-0.155 (0.112)	1.222** (0.577)	1.908 (1.293)	1.165 (1.273)	-0.866 (0.948)
Immunology	85	0.021 (0.107)	2.715** (1.131)	0.145*** (0.026)	0.052 (0.084)	1.994*** (0.495)	2.000** (0.907)	0.474 (1.620)	-1.381 (0.836)
Materials Science	76	0.769*** (0.087)	5.158*** (1.772)	0.072*** (0.027)	0.041 (0.094)	2.520*** (0.758)	0.341 (1.317)	-2.997 (1.959)	-1.104 (0.897)
Mathematics	76	0.747*** (0.153)	0.349 (2.146)	0.027 (0.045)	-0.460 (0.619)	0.093 (0.756)	-1.892 (1.895)	1.664 (2.627)	-1.169 (1.576)
Microbiology	71	0.545*** (0.121)	2.116 (1.699)	0.109*** (0.027)	0.276 (0.241)	2.578*** (0.778)	-1.552 (1.649)	3.024 (2.024)	-4.137*** (1.322)
Molecular Biology & Neuroscience & Behavior	66	0.521*** (0.171)	-2.434 (3.959)	0.081** (0.033)	1.333*** (0.305)	0.844 (0.608)	0.461 (1.400)	4.430 (4.721)	-1.489 (1.144)
Pharmacology & Toxicology	66	0.625*** (0.096)	-0.796 (1.591)	0.098*** (0.037)	0.136 (0.198)	0.381 (0.742)	1.924 (1.771)	1.262 (1.631)	0.635 (1.243)
Physics	77	0.358*** (0.093)	-0.906 (1.095)	0.012 (0.020)	0.331*** (0.074)	1.826** (0.711)	1.198 (1.093)	2.233** (1.075)	-2.364*** (0.835)
Plant & Animal Science	92	0.676*** (0.227)	3.228 (2.836)	0.156*** (0.046)	-0.030 (0.289)	0.242 (1.140)	-2.246 (2.193)	-0.145 (3.486)	-1.474 (2.119)
Psychiatry/Psychology	103	0.599*** (0.100)	-0.009 (1.042)	0.078*** (0.026)	0.776** (0.363)	0.759* (0.452)	0.941 (0.906)	1.903* (1.061)	-0.391 (0.824)
Social Sciences	53	0.554*** (0.163)	2.959* (1.619)	0.064** (0.029)	1.119*** (0.331)	1.120 (0.737)	-2.062 (1.769)	-1.071 (1.680)	-1.328 (1.313)
Space Science	98	0.406*** (0.149)	2.558 (1.580)	0.157*** (0.041)	0.664 (1.372)	0.498 (0.563)	-0.765 (1.319)	0.976 (1.511)	-1.289 (1.310)
	54	0.574*** (0.114)	3.767** (1.505)	0.117** (0.053)	0.568*** (0.212)	1.364 (0.994)	-7.324*** (2.423)	7.294*** (2.261)	-8.620*** (1.907)
Constant					-0.830 (2.798)				
Observations					1,686				
R-squared					0.800				

Note 1: Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1

Note 2: Linear regression. Std. error adjusted for 126 clusters (countries).

Chapter 3

The characteristics of highly cited researchers in Africa

Hugo Confraria
UNU-MERIT, Maastricht University

Jaco Blanckenberg^{*#}
^{*}Centre for Research on Evaluation, Science and Technology (CREST),
[#]DST/NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation
Policy (SciSTIP), Stellenbosch University

Charl Swart^{*#}
^{*}Centre for Research on Evaluation, Science and Technology (CREST),
[#]DST/NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation
Policy (SciSTIP), Stellenbosch University

Abstract

Very little is known about the characteristics of highly cited scientists in Africa. This is unfortunate as highly cited researchers are seen as key drivers of knowledge production for their countries and as important conduits of frontier knowledge into the local academic research community and society in general. In this chapter, we combined bibliometric and survey data to identify which researchers are producing highly cited research in Africa, and we employed econometric analysis to understand which characteristics are associated with the likelihood of being highly cited. Overall our results suggest that, on average, researchers who produce more scientific publications in a year, collaborate more often with non-African partners and did their highest qualification in an Anglo-Saxon university (USA, UK, Canada or Australia), have a higher probability of producing highly cited research. We conclude by arguing about the duality of our results. On one hand, collaborating with frontier universities seems to be a crucial mechanism that allows researchers to develop scientific capabilities. On the other hand, policy makers should be aware that research assessment in African countries should go beyond measuring scientific impact in the academic community. Otherwise, incentives will be in place to stimulate winners that are already well connected with the global scientific elite.

Keywords: science policy; Africa; highly cited researchers; scientific productivity; scientific capabilities.

This chapter draws upon:

Confraria, H., Blanckenberg, J., & Swart, C. (2018). The characteristics of highly cited researchers in Africa. *Research Evaluation*, 27(3), 222–237.

3.1. Introduction

Very little is known about the characteristics of highly cited scientists in Africa, and other lower and middle-income regions in general. This is unfortunate as studying these researchers can enhance our understanding of the conditions that foster high impact work in regions with less resources, as well as the development of scientific capabilities at the country level (Parker et al. 2012; Waldinger 2016).

In this chapter, we aim to understand why some scientists in Africa are producing highly cited research. We intend to do so by studying the characteristics of researchers working in Africa who have produced highly cited publications indexed in Web of Science (WoS), during a 5-year period (2010-14), and compare them to researchers who did not produce highly cited work in the same period. Four central research questions will be asked:

1. Are more productive scientists more likely to produce highly cited research?
2. Are certain collaboration patterns associated with the likelihood of producing highly cited work?
3. Do researchers who obtained their PhDs outside of Africa perform better?
4. Are there specific career challenges that negatively affect the likelihood of producing highly cited publications?

To answer these questions, we combine bibliometric data with survey data and use descriptive and regression analysis to discern the characteristics that are associated with the likelihood of a researcher having a highly cited publication. The data allows us to control for a large number of characteristics such as academic age, gender, subject area and region. Our sample covers all scientific fields with the exception of humanities and arts, amongst research-active scientists who have at least one publication indexed in Web of Science (WoS) or Scopus between 2005 and 2015.

In what follows, we will first contextualize our analysis; then we will describe the data and methodology employed; next we will present our descriptive results; thereafter we will discuss the econometric results. Finally, a discussion and concluding remarks will be given.

3.2. Background

The scientific production of African researchers comprises a small proportion of global science production. According to Tijssen (2007), between 1980 and 2004, Africa's share in worldwide science declined from 1.3% to 1%. However, a recent UNESCO (2015) report found that in the past decade African output grew more than the world average to around 2.6%. Nevertheless, African scientific productivity relative to population is still far below world average. In 2014, the continent produced 27 publications per million inhabitants compared to the world ratio of 176 publications/million inhabitants (UNESCO 2015).

Another important characteristic of the scientific output in Africa is that it is highly skewed across nations and disciplinary areas. As early as 1973, South Africa and Egypt had the highest scientific output (Garfield 1983). This unequal distribution remains, with these two countries representing around 50% of total African output (AOSTI 2014). African countries have become focused in agricultural sciences and related areas, such as environmental and ecology sciences and plant and animal sciences, as well as in some specific health sciences (Confraria & Godinho 2015). Scientific areas with higher potential to support innovation, such as engineering, material sciences, molecular biology and social science have been underrepresented in terms of scientific output (Juma 2016; Pouris 2010).

The importance of international collaboration and the legacy of colonial ties are also recognized as playing a pivotal role in the scientific output in Africa. Bibliometric studies usually find little scientific co-authorship between African countries but rather a stronger tendency towards to collaboration with higher income nations (Mègnigbèto 2013; Narváez-Berthelemot et al. 2002). More specifically, countries with British (South Africa, Nigeria, Kenya, Uganda, Tanzania, and Ghana) and French (Tunisia, Algeria, Morocco, Cameroon, and Senegal) colonial legacies have more collaboration with Anglo-Saxon countries (the USA, the UK, Australia, and Canada) and with France respectively (Adams et al. 2013; Mègnigbèto 2013; Toivanen & Ponomariov 2011). The only exception is South Africa, which seems to be playing a key role in coordinating some research networks across Africa (Confraria & Godinho 2015).

In conjunction with a lower output, it has also been found that international research impact of African science is low (Tijssen 2007). However, more recently, some East African countries have produced research with a citation impact higher than the world average in fields such as immunology, clinical medicine and microbiology (Confraria & Godinho 2015). According to the findings in chapter two, this may be occurring because some scientific communities in these countries have very high levels of international collaboration; and therefore, a small group of

national researchers are producing scientific publications with international partners of higher reputation, which leads to the country's high levels of scientific impact in those fields.

Yet, we still know very little about the characteristics of such highly cited researchers (HCR) in Africa. There are specific reasons to study this population. First, highly cited scientists are the people who are on the cutting edge in their fields. They are performing and publishing work that their peers recognize as vital to the advancement of their field. Knowing something about their social characteristics provides insights into the conditions that foster high-impact work (Parker et al. 2010). Second, these scientists are usually integrated in international networks where new ideas and technologies are often being discussed. They can act as important conduits of frontier knowledge into the local academic research community (Barnard et al. 2012), which can potentially diffuse that knowledge to peers, students, the economy, and the general public. Third, HCR are often seen as key drivers of knowledge production for their countries (Waldinger 2016). They usually obtain high amounts of international research funding and attract other quality researchers, which can reinforce the accumulation of scientific capabilities.

3.2.1. Highly cited researchers

There are various ways to define an HCR. In the pioneering work of Garfield (1977, 1981), he used absolute number of citations in WoS by field to identify which were the authors who received more citations in a certain period of time. This approach can be problematic, given that publications belonging to different subject areas have different citation patterns (Peters & van Raan 1994). For example, publications in health-related areas receive on average substantially more citations than publications in mathematics. Consequently, we can expect that health related researchers would be overrepresented if a direct comparison is made. Another problem is that the older a publication is, the more time this publication has to be cited. Therefore, it is likely that researchers with a portfolio of older publications receive more citations (on average) than researchers with more recent publications.

Hence, more recent studies have shifted from counting numbers of citations to more qualified types of citations and weighted publications. Instead of counting publications and citations, the decisive difference in this perspective is whether a researcher contributes to a set of very highly-cited papers in a specific field and year (e.g. Bornmann et al. 2017; Sinatra et al. 2016). Different thresholds are employed, from the top1%, top5%, and top10% highly cited papers to other citation classes (e.g. Glänzel et al. 2014). The rationale behind these approaches is that only when researchers produce a paper that reaches a very high citation level, are they able to produce a distinctive result that influences substantially their field.

In this study, since our aim is to assess the characteristics of African researchers who produce high impact work, we will consider researchers who are authors of the top 10% and top 5% most cited papers published each year (between 2010 and 2014) for each discipline (normalized by subject area – 252 WoS categories), and compare them to African researchers who did not produce highly cited papers in the same period.

It is important to keep in mind that the importance of highly cited papers is ambivalent. On the one hand, for the scientists, being highly cited shows impact (through acknowledgements) and builds up reputation (Moed 2005). On the other hand, citations are criticized for being a social construction and not reflecting actual quality (Gilbert 1977; Latour 1987). Moreover, it has been argued that conventional bibliometric indicators are inappropriate in “peripheral” spaces; and that research assessment that is done by existing indicators may not capture science that is not measured through them (Chavarro et al. 2017; Hicks et al. 2015; Lopez Pineiro & Hicks 2015).

In our research, we will assume that highly cited papers are important contributions, either methodologically or epistemologically, and that high citation counts can indicate research with high value. These types of articles have been associated with opening a research field or changing the direction of a field (Aksnes 2003; Aksnes & Rip 2009); therefore, we will investigate whether researchers who were authors of this type of papers possess particular characteristics that allow them to produce high-impact work.

3.2.2. The factors that affect the probability of producing a highly cited paper

There are numerous studies assessing the determinants of citation impact at the individual level. However, few use large-scale survey data to capture characteristics not available through bibliometric data, and none looks at this question from an African perspective. In this section, we will summarize what are the factors that have been identified as influencing the probability of a researcher being highly cited or being able to produce a highly cited paper, keeping in mind the two caveats just mentioned.

3.2.2.1. Scientific productivity

One of the consistent findings in studies focusing on HCR, is that there is a high correlation between the number of papers a researcher has published and the number of citations received (Bosquet & Combes 2013; Larivière & Costas 2016). This also holds true for high impact papers (Abramo et al. 2014). For a paper to be cited the paper must first be noticed. Noticing a paper is random but more likely if it is already cited. To be noticed for the first time is very random, so the more papers a researcher writes the more likely it is that one of the papers is noticed. Only then can the positive feedback start. For example, Sandström and van den Besselaar (2016) suggest that,

for most fields, there are constant or increasing marginal returns. The more papers a researcher produces, the higher the probability of producing high impact papers. Following the literature, our hypothesis is that achieving a breakthrough idea is a low probability event. Therefore, the more publications per year an author produced in his/her career the higher the likelihood of producing a highly cited paper.

3.2.2.2. Region of highest qualification

There are also reasons to expect who a researcher that did his/her PhD in an institution outside of Africa will have a higher probability of being an HCR. Students who move to foreign countries to study or do research usually have the chance to develop new and alternative ways of thinking when they recombine ideas with people with different backgrounds (Hollingsworth 2006). Knowledge that is tacit or otherwise difficult to circulate is easily exchanged if people share the same geographical location. Mobility facilitates access to new knowledge, and more knowledge from distant and atypical sources is associated with greater idea generation and highly cited work (Fleming 2001; Uzzi et al. 2013). We will model this “mover’s advantage” (Franzoni et al. 2014) by including three dummy variables for researchers who did their highest qualification in: Anglo-Saxon countries (the USA, the UK, Australia, and Canada), France, and other non-African countries. We choose these categories based on the colonial legacy of African scientific collaboration patterns.

3.2.2.3. Collaboration patterns

The positive effects of collaboration on impact of research are also widely accepted (Glänzel et al. 1995; Katz & Martin 1997; Narin et al. 1991). Collaboration allows access to a larger social network, which consequently leads to increased visibility, which in turn is reflected in higher citation rates (Goldfinch et al. 2003). This cross-fertilization is amplified by international collaboration, as scientists who produce research with foreign scientists are more likely to belong to elite research groups within their own countries (Adams 2013).

In this study, we will measure collaboration intensity in four dimensions: with researchers at a) their own institution; b) other institutions in their own country; c) at institutions in other African countries; and d) at institutions outside of Africa. We expect collaboration intensity to be positively associated with the likelihood of being an HCR in all dimensions. However, collaboration intensity with researchers at institutions outside of Africa probably has a higher effect.

3.2.2.4. Challenges faced during the career

Scientific institutions in many African countries suffer from specific challenges such as poor conditions for research personnel and lack of funding (Mouton 2008). At the same time, across areas of research, scholars agree that mentoring can be associated with a wide range of positive outcomes such as productive research careers, motivational benefits, better preparation in making career decisions, and increased network opportunities (Allen et al. 2004; Evans et al. 2008). Therefore, in our econometric analysis, we will also include, as explanatory variables, dummy variables for researchers who perceived the lack of funding, lack of mentoring, lack of mobility opportunities, and lack of training opportunities as a challenge they faced during their career.

3.2.2.5. Gender

In terms of gender, the literature is ambiguous. Some research shows that most HCRs are male (Parker et al. 2010). However, with regard to citations per publication, other studies find that no gender differences exist (van den Besselaar & Sandström 2016; Sánchez-Peñas & Willett 2006). In Africa, the only study related to this topic was conducted in South Africa, and it found that higher citation levels are associated with South African men. However, collaboration activity is much more relevant than a scholar's gender in this regard (Prozesky & Boshoff 2012). It is, therefore, unclear if there is a relation between gender and the probability of being an HCR. However, we will control for gender in our model.

3.2.2.6. Year of first publication

Another relevant dimension is the time a researcher has spent doing research. In Sinatra et al. (2016), it is argued that the highest cited paper in a researcher's career is randomly distributed in time within his/her body of work. However, it is well known that older researchers with higher reputation have a higher chance of receiving more citations than younger researchers (Merton 1968). In this study, we will use year of first publication as a proxy for "academic age". We choose year of first publication instead of age because it captures the commencement of a scientific career with relative accuracy. It will be calculated by subtracting the year of first publication from 2017 (the year when the survey was closed).

As we described, there are a variety of factors that may influence the probability of researchers producing highly cited papers. Most research analyses the correlation between specific variables and citations levels. In our work, we will use a model to combine the different independent variables listed above and investigate which ones are significant.

3.3. Data and methodology

This chapter combines survey data with bibliometric data. Survey data were collected via a self-administered, web-based, structured questionnaire¹³. It was adapted from the questionnaire used for the Global State of Young Scientists precursor study (GLOSYS) (Friesenhahn & Beaudry 2014) and for GLOSYS in ASEAN (Gefferes et al. 2017). The questionnaire is divided into 10 sections: educational background, employment, working conditions, research output, funding, career challenges, international mobility, collaboration, mentoring, and demographic characteristics and contains a maximum of 36 items. It was initially developed in English and then translated into French in order to increase the probability of receiving responses from countries that have French as a primary language. The survey was administered between May 2016 and February 2017. Invitations to complete the questionnaire were sent to email addresses obtained from WoS and Scopus data for publications published between 2005 and 2015 that indicated at least one author with an African address. The survey received 7513 answers.

Survey respondents were linked to a WoS author identifier based on their email address and the author disambiguation algorithm described in Caron & van Eck (2014). We used these identifiers along with the WoS accession numbers of publications, to find how many WoS publications (articles and reviews) each survey respondent has, and which of these respondents have authored publications in the top 10% and top 5% most highly cited publications in a field and a year between 2010 and 2014.

When analysing our results, it is important to note that some of the researchers who completed the questionnaire may not have a fixed residence in Africa or may not be African. Any researcher that published one article with an African affiliation between 2005 and 2015 may have completed the form. However, this does not mean that his/her main or host institution is in an African country. In our analysis, we exclude authors that reported that their residence and nationality is not in/from an African country. We made this decision because the conditions and settings of researchers with an African affiliation who are not based or born in an African country may be different from our population of interest.

Our analysis also excluded researchers who reported that they belong to humanities-related fields due to the limitations of bibliometric indicators in this area. We also excluded researchers who published their first article in WoS after 2013, in order to only consider researchers who have at least one full year of experience after producing their first publication in our main period of interest

¹³ An English version of the questionnaire can be found in the appendix (Questionnaire 3.A.1).

(2010-2014). Finally, researchers who did not answer all our questions of interest were also removed from the final sample. After applying these restrictions, 2,490 researchers compose our final sample, of which 183 researchers were authors of at least one top10% publication, and 95 were authors of top5% publications (please see Table 3.A.1 in the appendix for descriptive statistics).

3.3.1. Approach

Our analytical section is composed of two segments. In the first section, we use descriptive statistics to examine trends in African scientific production and to study our sample of researchers. In the second part we use regression analysis to discern the characteristics that are associated with the likelihood of a researcher having a highly cited publication.

We define the number of top10% and top5% highly cited publications¹⁴ authored by a researcher working in Africa between 2010 and 2014 as the dependent variable. Because the outcome variable is count-type data (min=0 and max=8) with a Poisson distribution (right skewed), the Poisson regression model is considered an appropriate type. Nevertheless, the Poisson model is inefficient for overdispersed outcome data, where the variance exceeds the mean (Cameron & Trivedi 2013). When data are overdispersed, the Poisson model generates underestimated standard errors, highly significant parameters, and consequently inefficient parameters. In contrast, negative binomial regression is a model controlling for overdispersion. Since the dataset used here was found to be overdispersed, we used negative binomial regression.

The negative binomial model probability density function is:

$$f(y_i) = \frac{\Gamma(y_i + \theta)}{\Gamma(\theta) * y_i!} * \frac{\mu_i^{y_i} * \theta^\theta}{(\mu_i + \theta)^{y_i + \theta}}$$

in which Γ denotes the gamma function and θ is the model's dispersion parameter, which must also be estimated in the negative binomial regression. The parameterization of μ_i is a function of the regressors of interest that follows a log-linear specification:

$$\ln(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}$$

Based on the literature and insights gained from our descriptive analysis (see section 4.2), we relate our dependent variable to a set of features that could influence the production of highly cited work: (1) scientific productivity (total number of publications per academic age); (2) location of highest qualification (Anglo-Saxon country, France or other non-African country); (3)

¹⁴ Normalized by field (252 WoS categories) and year.

collaboration patterns (collaboration intensity with researchers at own institution, other institutions in own country, institutions in other African countries, and institutions outside of Africa)¹⁵; and (4) challenges faced in the career (mobility, mentorship, funding, and training). The richness of the data allows us to control for other characteristics such as academic age, gender, subject area¹⁶, and region of residence. The variables “year of first publication” and “scientific productivity” are derived from WoS and are relative to the entire researcher career. All the other independent variables are calculated using our survey data (for a complete list of the variables and their definitions see (Table 3.A.2 in the appendix). The sign of the estimated parameters β_i in the regression indicate whether or not the dependent variable increases with the regressor. Incidence rate ratios were also calculated for easier interpretation. They display the ratio of the counts predicted by the model when the variable of interest is one unit above its mean while the other variables are at their mean values.

We also conducted an additional probit regression analysis to investigate the relationship between the characteristics analysed and being an HCR or not. This is conducted by using a derived dichotomous variable taking 1 if the researcher has at least one top10% or 5% publication between 2010 and 2014 and 0 if not.

Given the cross-sectional nature of our data, we cannot claim a causal relationship between these features and the likelihood of a researcher becoming highly cited. We do not have longitudinal data and thus cannot observe changing patterns over time. Nevertheless, we take a first step in analysing what characteristics are associated with producing highly cited work.

¹⁵ In the survey, the question about collaboration patterns is relative to the past three years (2014-16). We will assume that the values reported are a good proxy for the same variables between 2010 and 2014. This may be a strong assumption, but since this is the first survey of this kind, we do not have access to previous information.

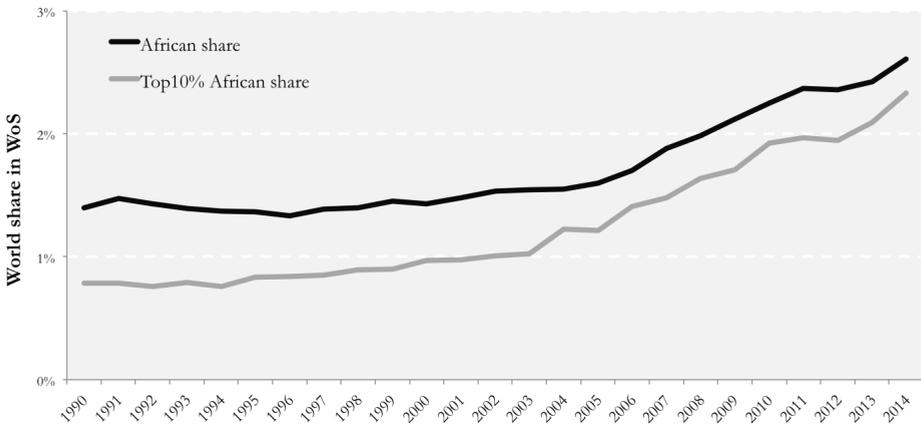
¹⁶ The survey included a question about the scientific field of each researcher highest qualification. We matched each one of those scientific fields to one of the six OECD categories (Frascati Manual): natural sciences, engineering and technology, medical and health Sciences, agricultural sciences, social sciences, and humanities.

3.4. Results

3.4.1. Bibliometric descriptive statistics

Africa’s scientific output in WoS has increased considerably during the past decade. In Fig. 3.1, we can observe that their world share of scientific output in WoS increased from 1.6% in 2005 to 2.6% in 2014. Their world share of highly cited articles (top10%) is slightly inferior but has also increased from 1.2% in 2005 to 2.3% in 2014. This acceleration reveals a trend of convergence with the leading world regions (Adams et al. 2013; Confraria & Godinho 2015; Pouris & Ho 2014), but it may also be related to the addition of scientific journals headquartered in African countries to the Thomson Reuters databases in recent years (Kahn 2011)¹⁷.

Figure 3.1. Trends in African output (world share) and top10% cited papers (world share)



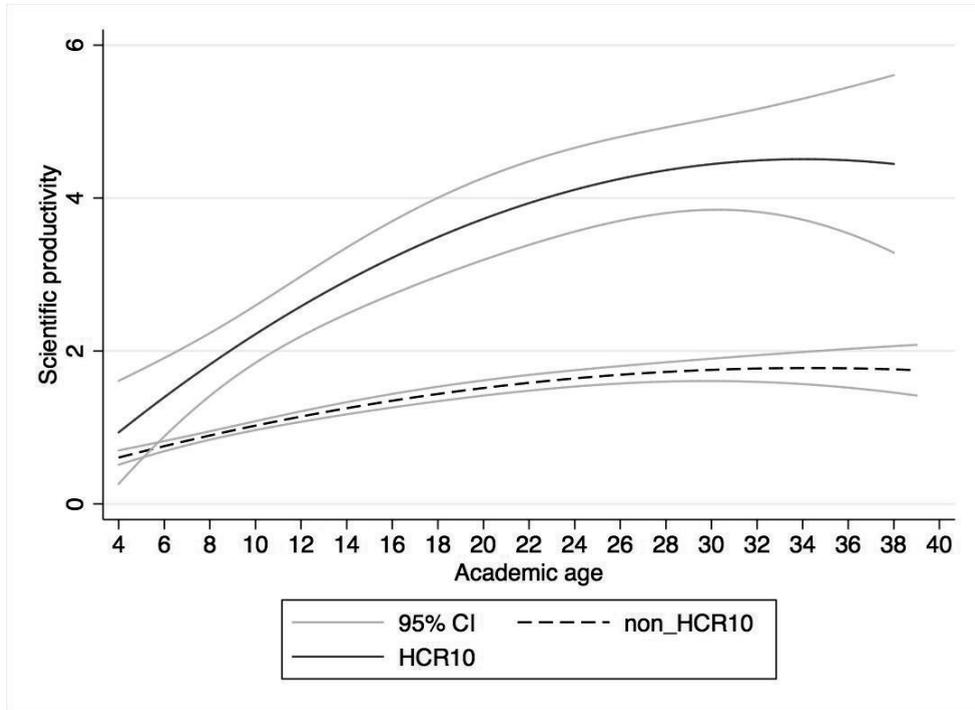
Source: WoS.

After collecting our survey data, we gathered bibliometric information on respondents’ scientific productivity (defined here as number of publications in WoS per academic age) and academic age (2017 minus year of first publication in WoS). In Fig. 3.2 we show the average scientific productivity per academic age, by two groups of researchers HCR and non-HCR (top10%). On average, HCR produces more publications per year at any stage of their career. In the appendix (Figs. 3.A.1 and 3.A.2), we also computed the density distributions of scientific productivity and academic age between the two groups of researchers. In both specifications (top10% and top5%), on average, HCR produces almost three times more papers per academic year than non-HCR. As

¹⁷ An aspect that could be researched in the future is to evaluate whether the adding of scientific journals headquartered in African countries to the Thomson Reuters databases in recent years, may have had an impact upon observable trends.

regards to academic age, the difference is not that large. On average, the academic age of HCR is 1.4 times higher than non-HCR.

Figure 3.2. Scientific productivity of HCR10 versus the scientific productivity of non-HCR



Source: WoS & own elaboration.

Note: Scientific productivity = number of publications in WoS per academic age; Academic age = 2017 minus year of first publication in WoS.

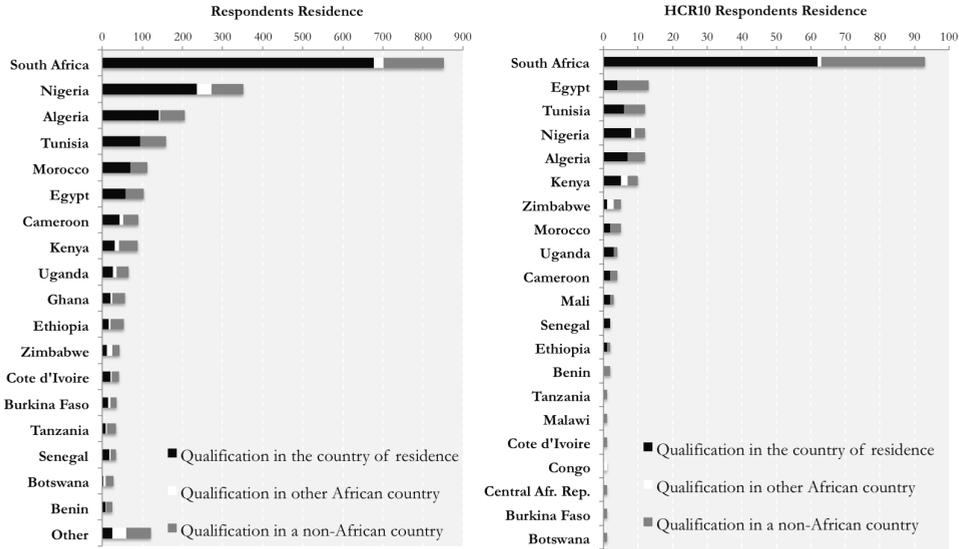
3.4.2. Survey descriptive statistics

Respondents were asked about their demographic characteristics, challenges faced in their career, and collaboration patterns among other questions. In this section, we will describe some of the information that we find relevant to the interpretation of our econometric results.

In Fig. 3.3, we can observe that, geographically, 34% (852) of the African researchers in our sample are based in South Africa. The other three countries with more than 5% of the respondents are Nigeria (14%), Algeria (8.2%), and Tunisia (6.3%). The distribution of HCR is even more skewed. South Africa is home to 51% of the researchers who published at least one top10% article. These results mirror the predominance of the South African research system within Africa. Another interesting finding is that in our sample around 32% of the researchers did their highest qualification in a non-African country. However, the share of researchers with a top10% publication who did

their qualification in a non-African country is substantially higher (40%). In our econometric analysis, we will further analyse this difference by comparing different regions of highest qualification.

Figure 3.3. Number of researchers (and HCR) resident in an African country



Source: Own elaboration.

Note: In the first graph we only included African countries that had at least 20 responses. In the graph on the right we only included countries with at least one researcher that had one top10% publication.

We have reasons to believe that researchers from Egypt are underrepresented in this sample. According to UNESCO (2015), Egypt accounts for more than 20% of the total number of publications with an African author in a similar period of analysis. In our sample, they constitute only 4.2% of the researchers (116). Researchers based in Egypt may have had a more difficult time receiving emails that include surveys or links to surveys. A number of respondents commented that emails of such a nature are blocked by mail servers and firewalls.¹⁸ Due to these regional differences, in order to ensure better representation of our population in our econometric analysis, we control for “regions” by generating a dummy variable for researchers who reside in: South Africa, Egypt, “Northern Africa”¹⁹, and “Central Africa”²⁰.

We also asked respondents about gender and field of highest qualification (see Fig. 3.A.3 in the appendix). About 70% of the researchers are male. The percentage of males is similar amongst HCR. The percentage of males is relatively higher for the subject areas “Engineering and

¹⁸ Furthermore, some respondents mentioned the general suppression of academic freedom and access to information. However, these statements are based on specific comments from only a number of respondents.

¹⁹ Algeria, Libya, Morocco, and Tunisia.

²⁰ All African countries except South Africa, Egypt, and “Northern” African countries.

Technology” and “Agricultural Sciences”. Natural Sciences is the area with a higher percentage of HCR (10.5% for the top10% indicator and 6.5% for the top5%) and “Medical and Health Sciences”, “Social Sciences”, “Engineering and Technology” and “Agricultural Sciences” all have ratios below the top10% and top5% averages. Furthermore, in all the five areas our sample has at least 19 researchers with a top10% publication and 10 researchers with a top5% publication.

Concerning collaborating patterns, in Fig. 3.A.4 in appendix, we can observe that HCR on average collaborate more often with other researchers, in any of our four categories (own institution, other institutions in own country, institutions in other African countries and institutions outside of Africa), than non-HCR. Yet, the main difference is that HCR collaboration intensity with researchers in institutions outside of Africa is substantially higher than that of non-HCR.

Respondents also reported on the major challenges that have impacted negatively on their careers (Fig. 3.A.5 in the appendix). On average, the biggest challenge is lack of funding, and the challenge that they reported as least relevant is “political instability”. The only challenge that HCR reported as more problematic than non-HCR is “balancing work and family demands”. All the other challenges are reported as more problematic for non-HCR. In our econometric analysis we will only include four challenges: lack of research funding, lack of mentoring, lack of mobility opportunities and lack of training opportunities.

3.4.3. Regression results

We used Stata to compute our regressions. The results in Table 3.1 suggest that, on average, researchers who published more articles per year during their career, did their highest qualification (PhD) in an Anglo-Saxon university and collaborate more often with researchers in institutions outside of Africa have a higher probability of producing a highly cited publication.

The characteristic that seems most important is scientific productivity. Researchers who produce one publication per year more than others, have an expected value of highly cited publications 50% higher. This result is consistent with the previous literature and it indicates the cumulative properties of knowledge in science (Merton 1968; Sandström & van den Besselaar 2016). Usually researchers with more publications per year have higher reputation and find it easier to get the resources that facilitate research: grants, equipment, stimulating colleagues, capable students, etc. This can contribute to this self-reinforcing mechanism. In our regression, when we included the term scientific productivity squared, the coefficient of the polynomial becomes negative and significant. This result indicates that, contrary to what Sandström & van den Besselaar (2016) found in the Swedish context, our data shows positive but decreasing marginal returns between scientific productivity and the probability of producing high-impact papers. At the limit, a

researcher that would produce all highly cited papers in his field would need more and more publications to produce one more highly cited publication. What our results seem to indicate is that those at the higher margin of productivity have smaller returns to one more publication per year, than those at lower levels of productivity (see also Kolesnikov et al., 2018).

Having done a PhD outside of Africa also seems important. Researchers who did their PhDs in an Anglo-Saxon university (the USA, the UK, Canada or Australia) have an expected value of highly cited publications two times higher than researchers who did their PhDs in an African country. It is commonly accepted that by doing a PhD abroad, a researcher can increase his collaboration network and learn new skills, which can improve his or her scientific performance. In the same line of thought, recently it has been argued that “scientists have most impact when they are free to move” (Sugimoto et al. 2017). What our results seem to indicate is that going to specific countries (the ones in the global scientific “core”) gives a higher premium to mobility. Since researchers tend to cite more often researchers in their network and since highly cited researchers are usually in the global scientific “core”, in order to get more cited it helps to have a network of the right kinds of “friends” (Colussi, 2018).

At the same time, collaboration intensity only seems to matter if the collaboration is with researchers outside of Africa. Collaboration is a fundamental feature in scientific research. It brings advantages both in scientific and non-scientific terms. By collaborating, researchers not only share knowledge, techniques, and expertise, but can also enhance the visibility of their results. Therefore, one would expect that higher collaboration intensity (in all the four dimensions we examined) would increase the probability of producing a highly cited paper. However, only collaboration intensity outside Africa shows a positive significant association. To test if the effect of collaboration intensity outside of Africa is independent of scientific productivity, we interacted the two in model specification 3 and 6. The non-significant sign of the interaction coefficient and the remaining significant positive sign of the collaboration intensity outside of Africa coefficient indicates that there seems to be an independent effect of collaboration intensity outside of Africa in the probability of producing more highly cited papers.

Finally, regarding the challenges faced by these researchers in general, we find negative but non-significant coefficients. Since some of these challenges are substantially correlated between themselves (between 30% and 50%), we also tried to compute each challenge separately in a specific estimation. The results were identical. This indicates that perceiving a specific challenge during their career is not related with the probability of being an HCR. Other factors seem to be more important.

Table 3.1. Negative binomial estimation for top10% and top5%

Ind. Variables	Dep. variable – Count of top10% pubs			Dep. variable – Count of top5% pubs		
	Nbreg (1)	Nbreg (2)	Nbreg (3)	Nbreg (4)	Nbreg (5)	Nbreg (6)
Scientific Productivity (Pubs/acad. age)	0.41*** (0.055)	0.81*** (0.092)	0.69*** (0.20)	0.32*** (0.052)	0.65*** (0.093)	0.35 (0.22)
Highest qualif. (PhD) in Anglo-Saxon (1 – Yes)	0.64*** (0.23)	0.61*** (0.23)	0.64*** (0.23)	1.09*** (0.30)	1.07*** (0.31)	1.09*** (0.30)
Highest qualif. (PhD) in France (1 – Yes)	0.34 (0.31)	0.36 (0.31)	0.36 (0.31)	0.69* (0.41)	0.68* (0.41)	0.69* (0.41)
Highest qualif. (PhD) in other non-African country (1 – Yes)	0.21 (0.29)	0.13 (0.29)	0.19 (0.29)	0.65 (0.42)	0.58 (0.42)	0.65 (0.42)
Collab. intensity – own institution (1-5)	0.038 (0.073)	0.048 (0.072)	0.031 (0.073)	0.086 (0.10)	0.099 (0.10)	0.086 (0.10)
Collab. intensity – outside Africa (1-5)	0.21*** (0.073)	0.18** (0.073)	0.30*** (0.091)	0.27*** (0.10)	0.25** (0.10)	0.28** (0.13)
Collab. intensity – own country other inst (1-5)	0.035 (0.077)	0.0079 (0.076)	0.040 (0.076)	-0.014 (0.10)	-0.047 (0.10)	-0.013 (0.10)
Collab. intensity – other country ins. Africa (1-5)	-0.043 (0.069)	-0.062 (0.068)	-0.044 (0.069)	-0.013 (0.095)	-0.032 (0.093)	-0.014 (0.094)
Academic age (2017 – Year of first publication)	0.021** (0.0096)	0.014 (0.0094)	0.020** (0.0095)	0.023* (0.013)	0.016 (0.013)	0.023* (0.013)
Gender (1 - Female)	-0.17 (0.19)	-0.20 (0.19)	-0.16 (0.19)	0.044 (0.25)	-0.00020 (0.25)	0.044 (0.25)
Lack of training opport. (1 – Yes “to a large extent”)	0.084 (0.26)	0.041 (0.26)	0.081 (0.26)	-0.57 (0.44)	-0.62 (0.45)	-0.56 (0.44)
Lack of mobility opport. (1 – Yes “to a large extent”)	-0.31 (0.26)	-0.31 (0.26)	-0.32 (0.26)	-0.43 (0.42)	-0.42 (0.42)	-0.43 (0.42)
Lack of mentorship (1 – Yes “to a large extent”)	-0.29 (0.25)	-0.27 (0.25)	-0.29 (0.25)	-0.13 (0.34)	-0.11 (0.35)	-0.13 (0.34)
Lack of research funds (1 – Yes “to a large extent”)	-0.097 (0.18)	-0.029 (0.18)	-0.10 (0.18)	-0.15 (0.25)	-0.058 (0.25)	-0.15 (0.25)
Scientific Productivity^2 (Polynomial)		-0.041*** (0.0087)			-0.028*** (0.0069)	
Productivity x Collab. out. Africa (Interaction)			-0.064 (0.042)			-0.0051 (0.049)
Constant	-3.89*** (0.44)	-4.11*** (0.45)	-4.26*** (0.50)	-4.80*** (0.62)	-4.94*** (0.64)	-4.83*** (0.69)
Indelta	1.02*** (0.19)	0.80*** (0.20)	1.03*** (0.19)	1.07*** (0.30)	0.92*** (0.32)	1.07*** (0.30)
Regional effects	Yes	Yes	Yes	Yes	Yes	Yes
Subject area effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2490	2490	2490	2490	2490	2490
Wald chi2	206.06	320.25	212.49	162.6	212.75	162.56
Pseudo R2	0.14	0.16	0.14	0.16	0.18	0.16

Note 1: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Note 2: The regression model was computed controlling for 5 of the 6 OECD categories: Natural Sciences, Agricultural Sciences, Engineering and Technology, Medical and Health Sciences, and Social Sciences; and 4 regions: South Africa, Egypt, Northern Africa and Central Africa.

Source: Own calculations.

To complement this analysis, we carried out two robustness checks (Table 3.A.3 and 3.A.4 in appendix). First, we carried out the same analysis as in Table 3.1 but instead of using the number of highly cited publications as a dependent variable, we used a dummy that is one for a researcher who has at least one highly cited publication (0 otherwise) and we computed a probit model. Second, we hypothesized that the determinants of producing highly cited papers might be different at different stages of a researcher career. We divided our sample in two groups: younger researchers (academic age <10) and older researchers (academic age ≥ 10); and we regressed our negative binomial model for different age levels. In general, the results are consistent with the previous model. The results from the probit model (Table 3.A.3) are identical to the ones in Table 3.1. In Table 3.A.4, the only difference between age levels seems to be that older researchers are the ones who benefit from having done their highest qualification in an Anglo-Saxon university.

3.5. Discussion and conclusions

In this chapter, we combined bibliometric and survey data to identify which researchers are producing highly cited research in Africa, and we used econometric analysis to understand which characteristics are associated with higher probabilities of being highly cited.

The results from this study highlight that a characteristic that is positively related with the probability of being an HCR is the quantity of publications produced per academic year (scientific productivity). This result is consistent with the previous literature that claims that the more papers a researcher produces, the higher the probability of producing high impact papers. It also shows the cumulative properties of scientific development. In order to increase the scientific impact of a country, previous scientific capabilities should already exist.

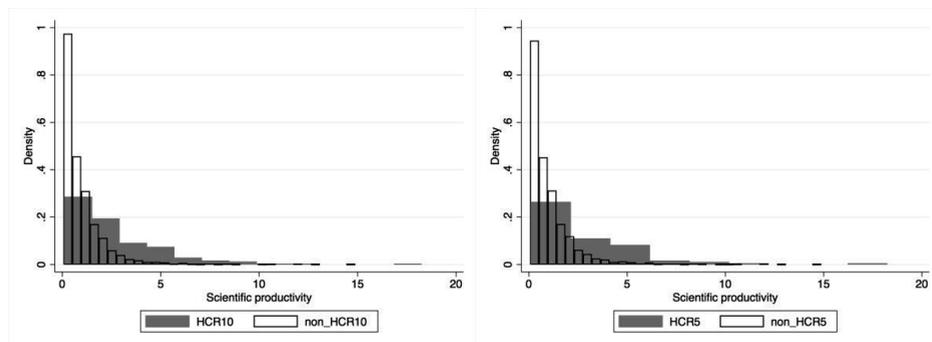
We also found that completing the highest qualification in an Anglo-Saxon university (the USA, the UK, Canada, or Australia) and collaborating more often with researchers outside of Africa are positively and significantly associated with the probability of being an HCR. This implies that in order to increase the scientific capabilities of African researchers, a certain number of African students should continue to go to frontier universities outside Africa in order for them to be better integrated in networks where emergent ideas are being discussed. While there is value to foreign training, recent research by Muller et al. (2018) as shown, that a significant component of the premium effect of PhD training abroad may derive from selection rather than to training effects. Therefore, governments in Africa should not discourage the improvement of their scientific and educational programmes.

These findings may have dual implications for developing regions. It seems that, on average, the scientists in Africa that produce research with high impact and visibility, are the ones who are more integrated in networks of researchers from the global scientific “core”, not those who have fewer challenges during their career or those who collaborate more locally. Consequently, increasing the number of ties and connections to frontier universities seems to be a crucial mechanism that allows researchers to develop scientific capabilities. On the other hand, it is argued that, at the global level, this process of integration may tend to favour the strongest and produce stratification (R. B. Freeman 2005; B. F. Jones et al. 2008), generate “brain-drain” (Hunter et al. 2009; Weinberg 2011), and deviate the focus of research from local or national issues to more internationally oriented topics (Hicks et al. 2015). Therefore, wise policy makers should be aware that research assessment in these contexts should go beyond measuring scientific impact in the academic community (through publications and citations in international journals) and also account for other broader impacts of scientific research in society such as skill formation (teaching and training), knowledge diffusion with other actors in society (talks/presentations, social media and policy advice), fund raising, and innovation activities such as the development of new products or business processes (Tijssen & Kraemer-Mbula 2017). Otherwise, incentives will be in place to stimulate winners that are already well connected with the global scientific elite.

Several caveats must be borne in mind with regards to our study. First, we use a threshold to define what is a highly cited paper is or not. There may be many researchers who produced papers that were close to being top10% papers, but because they did not achieve that status, they are referred in our study as non-HCR. Second, in this study, we assumed that any co-author of a highly cited paper made a significant contribution to that paper. However, it has been suggested that researchers in lower-income contexts are rarely leading authors in international publications and that their role is often still primarily limited to collecting data and linking up with domestic policy debates (Boshoff 2009; Carbonnier & Kontinen 2014). Therefore, this may lead to an identification problem. Nevertheless, since the email addresses in WoS are mostly from corresponding authors, we have reasons to believe that this bias has a limited effect in our results. Third, our variables related to collaboration patterns and challenges faced by the researcher are assumed to be constant during the career of all researchers in this survey. This is a strong assumption one should keep in mind when interpreting the results. Finally, our R-squared is relatively low; therefore, the explanatory power of our model is limited. There may be other factors that are also relevant for our model that are not included. For example, the inherent (childhood) ability or genius of a researcher (Simonton 1999), professional marginality from the discipline they changed (Kuhn 1962), and the “lucky” element in science or serendipity (Roberts 1989) among others.

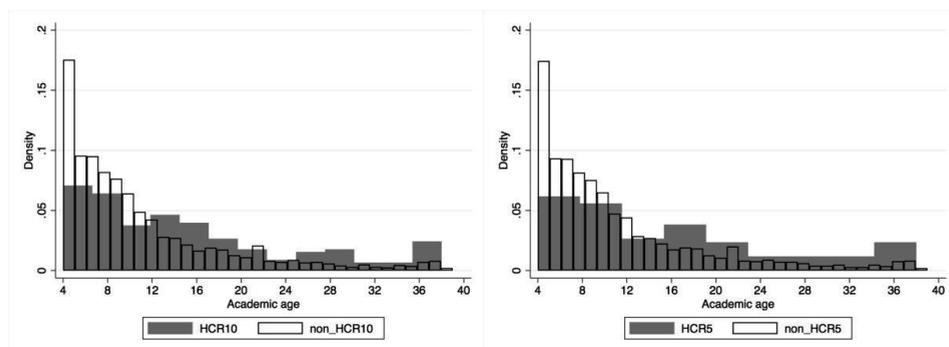
3.6. Appendix

Figure 3.A.1. Density distribution of researchers number of publications per academic age – top10% and top5%



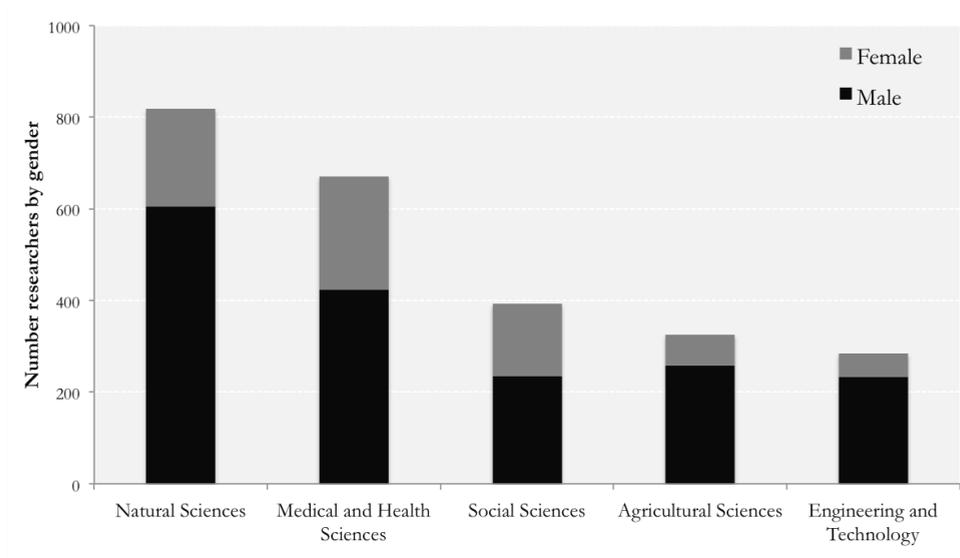
Source: WoS & own elaboration.

Figure 3.A.2. Density distribution of researchers academic age – top10% and top5%



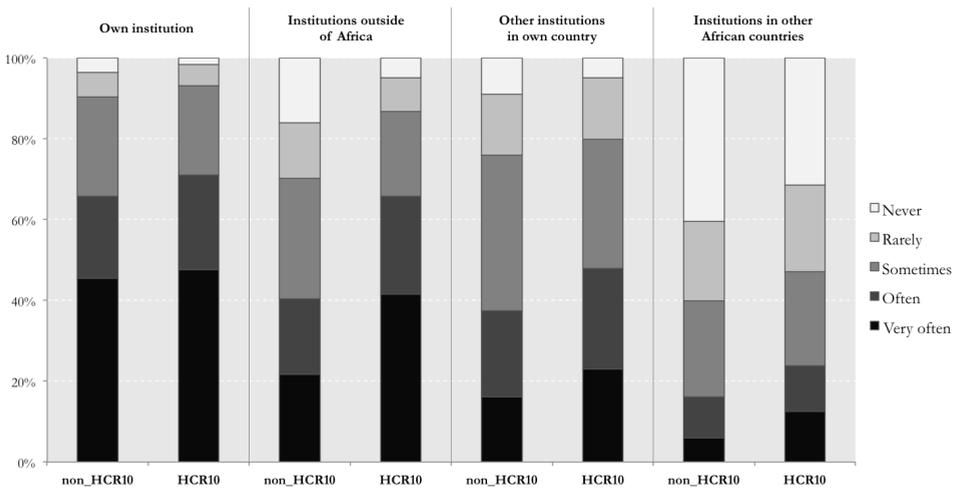
Source: WoS & own elaboration.

Figure 3.A.3. Number of researchers by area and gender



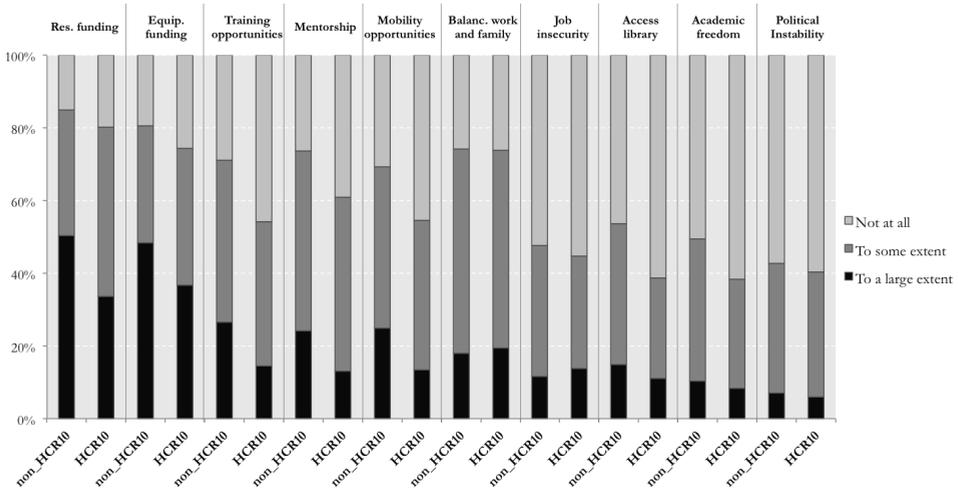
Source: Own elaboration.

Figure 3.A.4. Collaboration patterns, by being or not a HCR



Source: Own elaboration.

Figure 3.A.5. Challenges faced, by being or not a HCR



Source: Own elaboration.

Table 3.A.1. Descriptive statistics.

Variables	Obs	Mean	Std. Dev.	Min	Max
Dummy HCR10	2490	0.07	0.26	0	1
Dummy HCR5	2490	0.04	0.19	0	1
Count of top 10% pubs	2490	0.11	0.49	0	8
Count of top 5% pubs	2490	0.05	0.29	0	6
Scientific Productivity	2490	1.15	1.48	0.06	18.24
Highest qualif. in Africa	2490	0.72	0.45	0	1
Highest qualif. in France	2490	0.09	0.29	0	1
Highest qualif. in Anglo-Saxon	2490	0.09	0.28	0	1
Highest qualif. in other non-African country	2490	0.10	0.30	0	1
Collab. intensity – own institution	2490	4.00	1.12	1	5
Collab. intensity – outside Africa	2490	3.18	1.35	1	5
Collab. intensity – own country other inst	2490	3.21	1.15	1	5
Collab. intensity – other country ins. Africa	2490	2.20	1.23	1	5
Academic age	2490	11.50	7.70	4	39
Gender	2490	0.29	0.46	0	1
Lack of training opport.	2490	0.27	0.45	0	1
Lack of mobility opport.	2490	0.25	0.44	0	1
Lack of mentorship	2490	0.24	0.43	0	1
Lack of research funds	2490	0.51	0.50	0	1
Dummy South Africa	2490	0.34	0.47	0	1
Dummy Egypt	2490	0.04	0.20	0	1
Dummy Central Africa	2490	0.43	0.50	0	1
Dummy North Africa	2490	0.19	0.39	0	1
Dummy natural sciences	2490	0.32	0.47	0	1
Dummy agricultural sciences	2490	0.13	0.34	0	1
Dummy eng. and tech.	2490	0.11	0.32	0	1
Dummy medical sciences	2490	0.27	0.44	0	1
Dummy social sciences	2490	0.16	0.37	0	1

Source: Own calculations.

Table 3.A.2. Description of variables used in the model

Variables	Description
Dummy HCR10	Dummy variable that is 1 for an African author, which is a author in a top10% highly cited paper between 2010 and 2014
Dummy HCR5	Dummy variable that is 1 for an African author, which is a author in a top5% highly cited paper between 2010 and 2014
Count of top 10% pubs	Number of top10% highly cited papers authored between 2010 and 2014
Count of top 5% pubs	Number of top5% highly cited papers authored between 2010 and 2014
Scientific Productivity	Number of publications (articles and reviews) in WoS per academic age
Highest qualif. in Africa	Country of highest qualification (1 – African country; 0 - Otherwise)
Highest qualif. in France	Country of highest qualification (1 – France; 0 - Otherwise)
Highest qualif. in Anglo-Saxon	Country of highest qualification (1 – Australia, Canada, UK or USA; 0 - Otherwise)
Highest qualif. in other non-African country	Country of highest qualification (1 - other non-African country; 0 - Otherwise)
Collab. intensity – own institution	Collaboration intensity with researchers at your own institution (1 - Never or very rarely; 2 - Rarely; 3 - Sometimes; 4 - Often; 5 - Very often/always)
Collab. intensity – outside Africa	Collaboration intensity with researchers at your other institutions in your own country (1 - Never or very rarely; 2 - Rarely; 3 - Sometimes; 4 - Often; 5 - Very often/always)
Collab. intensity – own country other inst	Collaboration intensity with researchers at institutions in other African countries (1 - Never or very rarely; 2 - Rarely; 3 - Sometimes; 4 - Often; 5 - Very often/always)
Collab. intensity – other country ins. Africa	Collaboration intensity with researchers at institutions outside of Africa (1 - Never or very rarely; 2 - Rarely; 3 - Sometimes; 4 - Often; 5 - Very often/always)
Academic age	Number of years since the year of first publication in WoS until 2017
Gender	0 (Male); 1 (Female)
Lack of training opport.	Lack of training opportunities to develop professional skills (1 – Yes “to a large extent”, 0 - Otherwise)
Lack of mobility opport.	Lack of mentoring and support (1 – Yes “to a large extent”, 0 - Otherwise)
Lack of mentorship	Lack of mobility opportunities (1 – Yes “to a large extent”, 0 - Otherwise)
Lack of research funds	Lack of research funding (1 – Yes “to a large extent”, 0 - Otherwise)
Dummy South Africa	Region of residence (1 – South Africa, 0 – Otherwise)
Dummy Egypt	Region of residence (1 – Egypt, 0 – Otherwise)
Dummy Central Africa	Region of residence (1 – Central Africa, 0 – Otherwise)
Dummy North Africa	Region of residence (1 – Northern Africa, 0 – Otherwise)
Dummy natural sciences	Field of highest qualification (1 – Natural Sciences, 0 – Otherwise)
Dummy agricultural sciences	Field of highest qualification (1 – Agricultural Sciences, 0 – Otherwise)
Dummy eng. and tech.	Field of highest qualification (1 – Engineering & applied technologies, 0 – Otherwise)
Dummy medical sciences	Field of highest qualification (1 – Health Sciences, 0 – Otherwise)
Dummy social sciences	Field of highest qualification (1 – Social Sciences, 0 – Otherwise)

Table 3.A.3. Probit estimation for top10% and top5%

Ind. Variables	Dep. Variable – Dummy HCR10			Dep. Variable – Dummy HCR5		
	Probit (1)	Probit (2)	Probit (3)	Probit (4)	Probit (5)	Probit (6)
Scientific Productivity	0.21***	0.46***	0.23**	0.17***	0.33***	0.060
(Pubs/acad. age)	(0.028)	(0.064)	(0.10)	(0.025)	(0.064)	(0.10)
Highest qualification (PhD) in Anglo-Saxon (1 – Yes)	0.26*	0.25*	0.26*	0.51***	0.51***	0.52***
(1 – Yes)	(0.14)	(0.14)	(0.14)	(0.16)	(0.16)	(0.16)
Highest qualification (PhD) in France (1 – Yes)	0.12	0.14	0.12	0.33*	0.34*	0.33*
(1 – Yes)	(0.15)	(0.16)	(0.15)	(0.19)	(0.19)	(0.19)
Highest qualif. (PhD) in other non-African country (1 – Yes)	0.082	0.040	0.081	0.18	0.16	0.19
(1 – Yes)	(0.15)	(0.15)	(0.15)	(0.19)	(0.19)	(0.19)
Collaboration intensity – own institution (1-5 likert)	-0.015	-0.021	-0.015	-0.0015	-0.0039	-0.0012
(1-5 likert)	(0.039)	(0.040)	(0.039)	(0.050)	(0.051)	(0.050)
Collaboration intensity – outside Africa (1-5 likert)	0.097***	0.076**	0.10**	0.11**	0.097**	0.077
(1-5 likert)	(0.037)	(0.038)	(0.045)	(0.048)	(0.048)	(0.057)
Collaboration intensity – own country other inst (1-5 likert)	0.046	0.042	0.046	0.024	0.018	0.023
(1-5 likert)	(0.039)	(0.039)	(0.039)	(0.050)	(0.050)	(0.050)
Collaboration intensity – other country ins. Africa (1-5 likert)	-0.036	-0.043	-0.036	-0.019	-0.023	-0.017
(1-5 likert)	(0.038)	(0.038)	(0.038)	(0.047)	(0.047)	(0.047)
Academic age (2017 - Year of first publication)	0.0058	0.0017	0.0058	0.0065	0.0034	0.0068
(2017 - Year of first publication)	(0.0053)	(0.0054)	(0.0053)	(0.0066)	(0.0067)	(0.0066)
Gender (1 - Female)	-0.086	-0.080	-0.087	0.037	0.034	0.040
(1 - Female)	(0.098)	(0.099)	(0.098)	(0.12)	(0.12)	(0.12)
Lack of training opportunities (1 – Yes “to a large extent”)	0.017	-0.0053	0.017	-0.25	-0.27	-0.25
(1 – Yes “to a large extent”)	(0.13)	(0.13)	(0.13)	(0.18)	(0.18)	(0.18)
Lack of mobility opportunities (1 – Yes “to a large extent”)	-0.17	-0.16	-0.17	-0.20	-0.20	-0.20
(1 – Yes “to a large extent”)	(0.12)	(0.13)	(0.12)	(0.17)	(0.18)	(0.17)
Lack of mentorship (1 – Yes “to a large extent”)	-0.13	-0.12	-0.13	-0.015	-0.0054	-0.015
(1 – Yes “to a large extent”)	(0.12)	(0.12)	(0.12)	(0.14)	(0.15)	(0.14)
Lack of research funds (1 – Yes “to a large extent”)	-0.054	-0.038	-0.055	-0.11	-0.088	-0.10
(1 – Yes “to a large extent”)	(0.098)	(0.099)	(0.098)	(0.12)	(0.12)	(0.12)
Scientific Productivity^2 (Polynomial)		-0.024***			-0.015**	
(Polynomial)		(0.0070)			(0.0062)	
Productivity x COL outside Africa (Interaction)			-0.0051			0.026
(Interaction)			(0.023)			(0.024)
Constant	-2.06***	-2.17***	-2.08***	-2.40***	-2.46***	-2.27***
(Constant)	(0.24)	(0.25)	(0.25)	(0.30)	(0.31)	(0.31)
Regional effects	Yes	Yes	Yes	Yes	Yes	Yes
Subject area effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2490	2490	2490	2490	2490	2490
Wald chi2	131.04	200.39	130.87	120.57	140.22	123.67
Pseudo R2	0.14	0.17	0.14	0.17	0.18	0.17

Note 1: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Note 2: The regression model was computed controlling for 5 of the 6 OECD categories: Natural Sciences, Agricultural Sciences, Engineering and Technology, Medical and Health Sciences, and Social Sciences; and 4 regions: South Africa, Egypt, Northern Africa and Central Africa.

Source: Own calculations.

Table 3.A.4. Negative binomial estimation for top10% by age level

Ind. variables	Dep. variable – Count of top10% pubs (nbreg est)	
	academic age wos<10	academic age wos>=10
Scientific Productivity (Pubs/acad. age)	0.42*** (0.14)	0.44*** (0.064)
Highest qualification (PhD) in Anglo-Saxon (1 – Yes)	0.62 (0.43)	0.75*** (0.26)
Highest qualification (PhD) in France (1 – Yes)	-0.015 (0.58)	0.57 (0.38)
Highest qualif. (PhD) in other non-African country (1 – Yes)	0.30 (0.38)	-0.046 (0.41)
Collaboration intensity – own institution (1-5 likert)	0.014 (0.099)	0.089 (0.090)
Collaboration intensity – outside Africa (1-5 likert)	0.26** (0.12)	0.15* (0.089)
Collaboration intensity – own country other inst (1-5 likert)	0.0027 (0.13)	0.084 (0.098)
Collaboration intensity – other country ins. Africa (1-5 likert)	-0.099 (0.11)	0.0052 (0.089)
Gender (1 - Female)	-0.25 (0.34)	-0.24 (0.21)
Lack of training opportunities (1 – Yes “to a large extent”)	0.18 (0.35)	-0.070 (0.38)
Lack of mobility opportunities (1 – Yes “to a large extent”)	-0.10 (0.18)	-0.25 (0.15)
Lack of mentorship (1 – Yes “to a large extent”)	-0.55 (0.34)	-0.12 (0.32)
Lack of research funds (1 – Yes “to a large extent”)	-0.45 (0.31)	0.15 (0.22)
Constant		-3.40*** (0.47)
Indelta		1.01 (0.20)
Regional effects	Yes	Yes
Subject area effects	Yes	Yes
Observations		2490
Wald chi2		226.33
Pseudo R2		0.15

Note 1: Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Note 2: The regression model was computed controlling for 5 of the 6 OECD categories: Natural Sciences, Agricultural Sciences, Engineering and Technology, Medical and Health Sciences, and Social Sciences; and 4 regions: South Africa, Egypt, Northern Africa and Central Africa.

Source: Own calculations.

3.7. Questionnaire

Educational background

EDU.1 What is your highest qualification?

- Doctoral or equivalent
- Master or equivalent
- Bachelor
- Other (Specify)

EDU.2 When did you obtain your highest academic qualification?

Year []

EDU.3 In which field did you obtain your highest qualification? (e.g. engineering, psychology, virology, agriculture etc.)

Open ended [specify field]

EDU.4 Was your highest qualification conferred by a university in one country?

- Yes
- No

EDU.5 [Only if EDU4=Yes] In which country did you obtain your highest qualification?

Country: [<dropdown list>]

EDU.6 [Only if EDU4=NO] In what countries did you obtain your highest qualification?

Country: [<dropdown list>]

Country: [<dropdown list>]

EDU.7 Are you currently enrolled in further postgraduate studies?

- Yes
- No

EDU.8 [Only if EDU5=Yes] At which institution and in which country?

[<open form.] – University

[<open form.] – country

EDU.9 [Only if EDU7=yes]. Are you receiving a bursary or scholarship for your current studies?

- Yes
- No

Employment

EMP.1 Please specify the sector of employment of your current main job:

- Higher/tertiary education [Explanation: university (public or private), college of technology, polytechnic and other institution providing tertiary education, or other institution directly under control of higher education institution]
- Public research institution
- Private research institution
- Business enterprise
- Non-governmental/non-profit organisation
- Other Please specify: [< open form>]

Chapter 3

EMP.2 What is your current employment status? If you hold more than one job, please answer for your main job.

- Professor, Associate Professor or Reader at a Tertiary Institution
- Senior lecturer at a Tertiary Institution
- Lecturer or equivalent at a Tertiary Institution
- Researcher/scientist
- Postdoctoral fellow
- Self-employed
- Unemployed or inactive
- Other Please specify: [< open form >]

EMP.3 [ONLY IF EMP2 ≠5,6,7] Is this position permanent or contract-based?

- Permanent [Permanent employees are employed on an ongoing basis until the employer or the employee ends the employment relationship]
- Contract-based [Contract employees are employed for a specific period of time or task, for example 6 to 12 months period, and employment ends on the date specified in the contract]

Working Conditions

WOR.1 On average, how many hours do you spend on your main job per week?

- (maximum accepted: 100 hours)

WOR.2 In a typical year, what percentage of your working time do you spend on each of the following tasks?

- % Undergraduate and Postgraduate teaching
- % Training/supervising postgraduate students
- % Research
- % Administration and management
- % Service (counselling of patients, voluntary services within or outside your organisation, article review, editorial duties)
- % Consultancy
- % Raising funds/grants for research
- % Other, please specify [< open form >]

Research OutputRO.1 Please indicate how many of the following research output types you have produced over the last three years:

[Drop down: Options n/a,0-22;21+] Articles published/accepted (including co-authored) in refereed or peer reviewed academic journals

[Same options] Books (i.e. monographs and edited volumes)

[Same option] Book chapters (including co-authored)

[Same option] Conference papers published in proceedings

[Same option] Presentations at conferences to predominantly academic audiences

[Same option] Written input to official public policy documents

[Same option] Research reports (contract/consultation research)

[Same option] Articles in popular journals/magazines, essays, newspaper articles or other public outreach media

[Same option] Patents (applied for and/or granted)

[Same option] Computer programmes (including co-writing)

[Same option] Creative/artistic works of art performed or exhibited (e.g. music, sculpture, paintings, theatre, film)

[] Others, Please specify: [< open form with categories>] x3

RO.2 [Only if RO 1 CAT 1 ≠ 0] When did you publish your first research article in a refereed or peer-reviewed journal?

Year []

RO.3 As far as your research is concerned, which of the following statements best describe the overall value or outcome of your research? Also rate the extent to which you believe that these have been successfully attained where applicable.

	Highly successful	Successful to some extent	Not successful at all	N/A
Advancement of knowledge	[]	[]	[]	
Solving of theoretical problems	[]	[]	[]	
Solving of immediate technical/applied problems	[]	[]	[]	
Solving of environmental or social problems	[]	[]	[]	
Development of skills and competencies	[]	[]	[]	
Change behaviour/attitudes/values	[]	[]	[]	
Influence policy/decision-makers	[]	[]	[]	
Influence practice	[]	[]	[]	
Stimulation of discussion/debate	[]	[]	[]	

RO.4 Please indicate which of the following stakeholders you consider when conceptualising your research:

- Colleagues/scholars/peers in own discipline
- Colleagues/scholars/peers in other discipline
- The contracting agency
- Industry/business/firm(s)
- Ministry/government agency
- Specific interest groups (e.g. farmers, researchers, nurses, doctors, consumers)
- General public/society/community

Funding

FUN.1 Have you received any research funding over the past three years? (Excluding bursaries or scholarships for studying purposes)

No Yes - but I am not the primary recipient/grant holder of the funding

Yes- I am the primary recipient/grant holder of the funding

Yes – In some cases I am the primary recipient and in some cases I am not the primary recipient of the funding

FUN.2 [Only if FUN 1 =Yes] Approximately what percentage of this funding was for infrastructure and equipment? (Don't know, N/A, 0%,10% intervals)

[] %

Chapter 3

FUN.3[Only if FUN 1 =Yes] What proportion of this funding was obtained from national and international sources? (10% intervals)

% National

% International

FUN.4 [Only if FUN 1 =Yes] Which amount best correspond to the total amount of research funding you have received during the past three years?

Dropdown list < Less than US\$10 000; US\$10 000 - 25 000; US\$25 000 - 50 000; US\$50 000 - 75 000;

US\$75 000 - 100 000; US\$100 000 - 250 000; US\$250 000 - 500 000; US\$500 000 - 1 000 000;

More than US\$ 1 000 000>

FUN.5 [Only if FUN 1 =Yes] Please specify the three organisations/agencies from which you have received the most funding over the past three years

[Specify] [< open form>]

[Specify] [< open form>]

[Specify] [< open form>]

Challenges

CHA.1 Indicate, where applicable, which of the factors listed below have impacted negatively on your career as an academic or scientist

	Not at all	To some extent	To a large extent
Lack of mentoring and support	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Job insecurity	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Balancing work and family demands	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Lack of mobility opportunities	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Lack of training opportunities to develop professional skills	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Lack of access to a library and/or information sources	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Lack of research funding	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Lack of funding for research equipment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Limitation of academic freedom	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Political instability or war	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other, please specify	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

International Mobility

MOB.1 In which country do you currently work/reside?

[<dropdown list>]

MOB.2 During the past three years, have you studied or worked in a country other than what you would consider your home country (i.e. abroad)?

Yes

No

MOB.3 [Only if MOB2 = Yes] Compared to the study/working conditions in your home country, how would you rate the study/working conditions abroad?

Researchers from:	Much worse abroad	Somewhat worse abroad	About the same	Somewhat better abroad	Much better abroad
Employment/job security	[]	[]	[]	[]	[]
Work-family balance	[]	[]	[]	[]	[]
Training opportunities	[]	[]	[]	[]	[]
Opportunities for research collaboration	[]	[]	[]	[]	[]
Research resources (personnel, scientific literature, material, etc.)	[]	[]	[]	[]	[]
Research funding opportunities	[]	[]	[]	[]	[]
Others, please specify [< open form>]	[]	[]	[]	[]	[]

MOB.4 [Only if MOB2 = Yes] How would you rate the importance of having studied/worked abroad for your career development?

- Not important
- Somewhat important
- Important
- Very important
- Essential

MOB.5 Have you ever considered leaving the country where you currently work?

- No, never
- Yes, sometimes
- Yes, often

MOB.6 [Only if MOB5 = Yes] List the main considerations for leaving the country:

<open ended form> x3

Collaboration

COL.1 How often do you collaborate, either in joint research or through joint publications, with the following categories of researchers:

	Never or very rarely	Rarely	Sometimes	Often	Very often/always
Researchers at your own institution	[]	[]	[]	[]	[]
Researchers at other institutions in your own country	[]	[]	[]	[]	[]
Researchers at institutions in other African countries	[]	[]	[]	[]	[]
Researchers at institutions outside of Africa (e.g. Europe, North America, Asia, etc.)	[]	[]	[]	[]	[]

Mentoring

MO.1 During your career so far, have you ever received mentoring, support or training in the following:

	Never or very rarely	Yes but it was not valuable	Yes and it was valuable
Career decisions	[]	[]	[]
Introduction to research networks	[]	[]	[]
Attaining a position/job	[]	[]	[]
Research methodology	[]	[]	[]
Fundraising	[]	[]	[]
Scientific writing	[]	[]	[]
Presenting research results	[]	[]	[]

Demographic background

DEM.1 Are you:

[] Male

[] Female

DEM.2 What is your year of birth?

YEAR [] (yyyy)

DEM.3 What is your nationality?

Dropdown list []

DEM.4 How many children or other dependents do you have?

Please enter a number in the relevant boxes.

[] Number of children/dependents aged 0 to 5

[] Number of children/dependents aged 6 to 18

[] Number of adult dependents aged 19 or older (including elderly)

[] I do not have any dependents.

DEM.5 How is the care-work and general housework for all dependents distributed in your family/relationship/household?

[]% me []% partner []% others (e.g. extended family, paid service)

Chapter 4

Which factors influence international research collaboration in Africa?

Hugo Confraria

UNU-MERIT, Maastricht University

Jaco Blanckenberg**

*Centre for Research on Evaluation, Science and Technology (CREST),

#DST/NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP), Stellenbosch University

Charl Swart**

*Centre for Research on Evaluation, Science and Technology (CREST),

#DST/NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP), Stellenbosch University

Abstract

It is commonly accepted that international research collaboration improves scientists' abilities and performance. In this chapter we investigate which are the characteristics of African researchers who collaborate more often with international partners. Data are taken from Web of Science and a survey that collected detailed information about the individual characteristics of 2954 African researchers in 42 African countries. We use descriptive statistics and an econometric model to discern the characteristics that are associated with higher levels of collaboration with researchers outside Africa. Overall our results suggest that, on average, researchers who did their doctoral studies outside of Africa, had the opportunity to move abroad (over the past three years) and received a higher share of international funding (over the past three years), are more likely to collaborate more frequently with researchers outside of Africa. In our conclusions we discuss that beyond increasing the availability of mobility scholarships and the amount of research funding for African scientists, policy makers and international organisations should also think in incentives to keep long-term research interactions and try to avoid unequal partnerships.

Keywords: Research collaboration, African science, scientific capabilities, research policy

4.1. Introduction

It is widely accepted that international research collaboration positively affects researchers' careers and impact. The benefits include access to expertise, complementary know-how and new techniques, generation of learning opportunities, improving networking activities, better access to funding and equipment, national and international recognition (AOSTI 2014; Arvanitis and Gaillard 2014; Beaver 2001; Bozeman and Corley 2004; Katz and Martin 1997; Wagner et al. 2001). Yet, despite a long history of cross-border cooperation between researchers worldwide, there are few empirical studies on the main drivers of international research collaboration in lower income regions.

This raises the questions of why some researchers collaborate more often with international partners than others; and secondly, what form should collaborative research take to overcome the pitfalls associated with the North-South divide? We will address these issues by studying the characteristics of African researchers who collaborate both frequently and infrequently with non-African researchers. This will be achieved by relying on survey data, bibliometric data and employing ordered probit regression analysis.

Most of the previous research that analyses the intensity of international research collaboration of specific authors relies on the use of co-authorship of scientific publications from different countries. However, co-authorship is seen as a partial indicator of collaboration since scientific collaboration may happen without resulting in a co-authored paper (Katz & Martin 1997; Laudel 2002). In our work, we will use a subjective measure of collaboration that was captured by a large survey sent to all African researchers who (co-) authored an article in WoS (Web of Science) and Scopus between 2005 and 2015. The questionnaire included items that asked respondents how often (1-5 Likert scale) they collaborate, either in joint research or through joint publications, with researchers at "their own institution", "other institutions in their own country", "institutions in other African countries" and "institutions outside of Africa". The last item of this question will be employed as the dependent variable to assess the collaboration intensity outside of Africa of individual researchers.

In our study, we are particularly interested in investigating whether doing the highest qualification (PhD) outside Africa, recent mobility and receiving a higher share of research funding from outside Africa allows African researchers to collaborate more frequently outside Africa. The richness of the data allows us to control for a large number of characteristics including academic age, scientific productivity (number of publications per academic age), challenges faced (lack of mobility opportunities, lack of research funding and lack of mentoring), gender, region and subject area.

In what follows, we first give a background overview of the “African” scientific system and thereafter discuss the main characteristics associated with higher levels of international research collaboration. In section three we describe our data sources and the econometric approach used. Section four presents and analyses our results, and we conclude by highlighting limitations and potential future research.

4.2. Background

As argued in the previous chapter, the total research output from researchers in Africa is a small proportion of World science. It is highly skewed across nations and disciplinary areas, and most countries rely heavily of international funding and international collaboration to sustain their research systems (Confraria & Godinho, 2015; AOSTI, 2014). Therefore the notion of “African science” may be misleading, as previously suggested by Tijssen (2007).

Bibliometric studies usually find little scientific co-authorship between African countries with preference being given to collaboration with higher income nations (Guns & Wang 2017; Mègnigbèto 2013; Narváez-Berthelemot et al. 2002; Onyanha & Maluleka 2011). When African countries do collaborate with one another, frequently those collaborations have been initiated by a non-African country (Boshoff 2009; Toivanen & Ponomariov 2011) or mediated through cooperative health and agricultural programs (Adams et al. 2013). Also it has been suggested that African researchers are rarely leading authors in international publications and that their role is often still primarily limited to collecting data and linking up with domestic policy debates (Boshoff 2009; Carbonnier & Kontinen 2014).

Yet, from a policy perspective, international collaborations are seen as one of the most efficient means to build research capacity and to create learning opportunities for African researchers (AOSTI 2014). With the limited resources that most African universities have, building international collaborations can allow individual researchers to access infrastructures and scientific networks, which they would not have access to when working in isolation. Therefore, governments and international organisations allocate substantial resources to promoting international research.

Several policies can be endorsed to improve links between national researchers and researchers in other countries. Some examples include expanding the number of scholarships that allow PhD students and researchers to go abroad to interact with peers; increasing the amount of research funding provided by governments and international organisations to research projects in African countries, and attracting foreign scholars to work and do research in African institutions. For example, target 4b of the Sustainable Development Goals committed the 193 signatories to the

agreement to “substantially expand globally the number of scholarships available to developing countries”. In this chapter, the question of which set of policies could contribute to improving international research collaboration is one of our driving questions. By investigating the characteristics of researchers who collaborate more often outside of Africa, we aim to contribute to a nuanced policy discussion of this issue.

4.2.1. The main characteristics of internationally collaborating researchers

One consistent finding in studies focusing on the factors that affect international collaboration is mobility. The networking power of doing a PhD, post-doc, visiting or going to conferences abroad are particularly important in initiating long-lasting scientific collaborations (Arvanitis & Gaillard 2014; Jonkers & Tijssen 2008; Marmolejo-Leyva et al. 2015; Scellato et al. 2015). In our study we will measure mobility in three dimensions: 1) Doing the highest qualification in a non-African country; 2) having studied or worked abroad in the past three years; and 3) perceiving that lack of mobility opportunities impacted negatively their career to a large extent.

While mobility is a main concern, other important dimension is the availability and source of research funding. Existing research indicates that international collaboration increases in the long-run in consequence of the use of funding schemes that encourage cross-country collaboration (Defazio et al. 2009) and foreign funding increases the number of South-North collaborations (Zdravkovic et al. 2016). Our survey included three questions about funding: 1) Being or not a primary recipient of research funding in the last three years; 2) the share of international funding received in the last three years (primary recipient or not); and 3) perceiving that lack of research funding impacted negatively their career to a large extent. We will use these three variables to examine how research funding can impact higher levels of international collaboration.

Finally individual characteristics of the researcher may also play a role. Across all areas of research, older researchers tend to have more collaborators (W. Wang et al. 2017), more productive scientists tend to cooperate more (S. Lee & Bozeman 2005); and males tend to collaborate more often with researchers abroad (Abramo et al. 2013). At the same time, scholars agree that mentoring can be associated with a wide range of positive outcomes such as productive research careers, motivational benefits, better preparation in making career decisions, and increased network opportunities (Allen et al. 2004; S. C. Evans et al. 2008). Therefore in our analysis we will also include as independent variables academic age (2017 – year of first publication in WoS), scientific productivity (number of publications in WoS/academic age), gender and perceiving that lack of mentoring was a challenge that they faced during their career.

4.3. Data and Methodology

This chapter combines survey with bibliometric data. Survey data were collected via a self-administered, web-based, structured questionnaire sent to all researchers with an African affiliation that were authors of publications in Web of Science or Scopus between 2005 and 2015²¹. It was adapted from the questionnaire used for the Global State of Young Scientists precursor study (GLOSYS) (Friesenhahn & Beaudry 2014) and for GLOSYS in ASEAN (Geffers et al. 2017). The questionnaire is divided into 10 sections: educational background; employment; working conditions; research output; funding; career challenges; international mobility; collaboration; mentoring; and demographic characteristics and contains a total of 36 items. It was initially developed in English and then translated into French in order to increase the probability of receiving responses from countries that have French as a primary language. The survey was administered between May 2016 and February 2017. The questionnaire received 7513 answers.

Some of the researchers who completed the questionnaire do not have a fixed residence in Africa or may not have a nationality from an African country. Any researcher that published one article with an African affiliation between 2005-2015 may have completed the form. In our analysis we exclude authors that reported that their residence and nationality is not in/from an African country. We made this decision because the conditions and settings of researchers with an African affiliation but are not based in an African country, or were not born in an African country, may be very different from our population of interest.

Our analysis also excluded researchers who reported that they belong to Humanities related fields due to the limitations of bibliometric indicators in this area (Hicks et al. 2015; Marx & Bornmann 2014). Finally researchers who didn't answer all our questions of interest were also removed from the final sample. After applying these restrictions, 2954 researchers compose our final sample²². Despite the high number of responses, the survey observations, based on an uncontrolled sample, cannot be considered as representative of the targeted population.²³ However, the characteristics of the sample show a fair representation among regions, subject areas and gender. Furthermore, several robustness checks were done in order to assess the validity of our model and results.

The source of bibliometric data is WoS. All articles and reviews from researchers with an African affiliation, published between 1980 and 2016, were extracted. After collecting our sample of articles

²¹ Only authors that reported their email address in Web of Science or Scopus were contacted.

²² Our sample is different from the sample in Chapter 3 because in the previous chapter we excluded researchers who published their first article in WoS after 2013.

²³ All African researchers, with an African affiliation, who have a publication in WoS or Scopus between 2005 and 2015.

with an author with an African affiliation we extracted the email addresses, names and affiliations of all African authors involved and we matched that info with the info of the authors that completed our survey.

4.3.1. Approach

Our analytical section is composed of two segments. In the first section, we use descriptive statistics to examine trends in African scientific production and to study our sample of researchers. In the second part we will use an ordered probit model, to allow us to discern the characteristics that are associated with higher levels of collaboration with researchers outside Africa.

The ordered probit model involves a qualitative dependent variable for which the categories have a natural order that reflects the magnitude of some underlying continuous variable (Greene 2012). In our case the dependent variable is expressed in terms of five categories (1 – Never, 2 – Rarely, 3 – Sometimes, 4 – Often, 5 – Very often) which could be viewed as resulting from a continuous variable called “collaboration intensity”.²⁴ All the other characteristics are treated as independent variables that can potentially affect collaboration intensity. The starting point is an index model, with single latent variable:

$$y^* = x'\beta + \varepsilon$$

where x are a set of characteristics of each researcher, that in our case are: (1) location of highest qualification (African or non-African); (2) having studied or worked abroad during the last three years; (3) being a primary recipient of research funding over the past three years; (4) share of international funding received over the past three years; (5) perceiving that lack of mentorship, mobility, and funding was a factor that affected negatively their career to a large extent. To avoid omitted variable bias, controls will be added for individual characteristics such as academic age, scientific productivity, gender, subject area and African region. In the model ε is disturbance and y^* is unobserved. What we do observe is

$$\begin{aligned} y &= 1 \text{ if } y^* \leq \mu_1 \\ y &= 2 \text{ if } \mu_1 < y^* \leq \mu_2 \\ y &= 3 \text{ if } \mu_2 < y^* \leq \mu_3 \\ y &= 4 \text{ if } \mu_3 < y^* \leq \mu_4 \\ y &= 5 \text{ if } \mu_4 < y^* \end{aligned}$$

²⁴ This model permits that the difference between the first and second outcomes may be different from the difference between the third and fourth outcomes, for example.

which is a form of censoring. The μ 's are unknown "threshold" parameters to be estimated with β 's. Estimation is undertaken by maximum likelihood, which in the case of the ordered probit model requires that ε is normally distributed across observations. The probability of obtaining an observation with $y = 1, 2, 3, 4$ and 5 is equal to

$$\begin{aligned} Prob(y = 1|x) &= \Phi(\mu_1 - x'\beta), \\ Prob(y = 2|x) &= \Phi(\mu_2 - x'\beta) - \Phi(\mu_1 - x'\beta), \\ Prob(y = 3|x) &= \Phi(\mu_3 - x'\beta) - \Phi(\mu_2 - x'\beta), \\ Prob(y = 4|x) &= \Phi(\mu_4 - x'\beta) - \Phi(\mu_3 - x'\beta), \\ Prob(y = 5|x) &= 1 - \Phi(\mu_4 - x'\beta), \end{aligned}$$

Where Φ is the normal density function. In terms of interpretation, the sign of the estimated parameters β can be immediately understood as determining whether or not the latent variable y^* increases with the regressor. However, the marginal effects of the regressors on the probabilities are not equal to the coefficients. For marginal effects in the probabilities we write

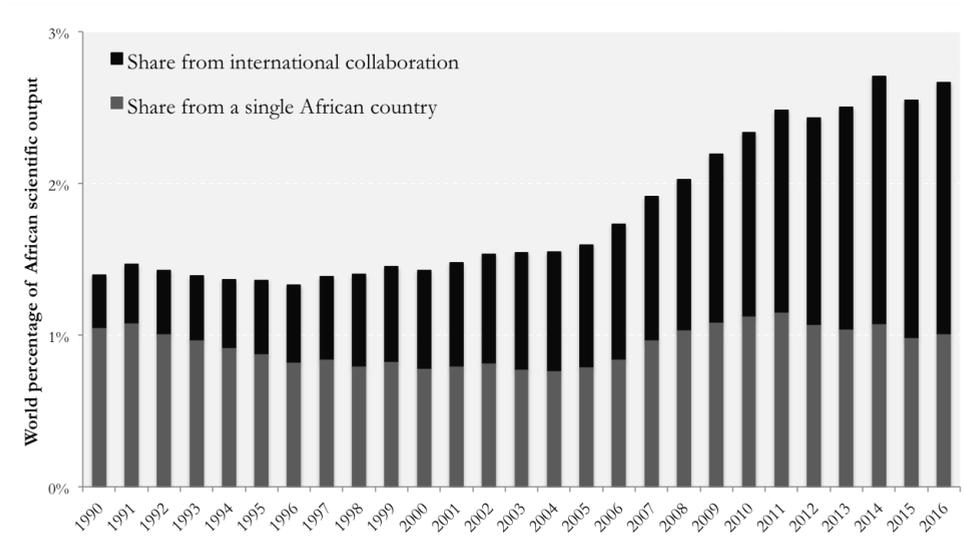
$$\frac{\sigma Pr[y_i = j]}{\sigma x_i} = \{\Phi'(\mu_{j-1} - x_i'\beta) - \Phi'(\mu_j - x_i'\beta)\}\beta$$

where Φ' denotes the derivative of Φ . The term in braces can be positive or negative and it should sum to zero over the j 's (1,2,3,4 and 5), which follows from the requirement that the probabilities sum to one.

4.4. Results

The scientific output in Africa has increased considerably during the last decade. In Fig. 4.1, we can observe that the total world share of articles and reviews in Africa increased from 1.4% in 1990 to 2.7% in 2016. However, as shown in Fig. 4.1, since 2006 this increase has been mainly driven by international collaborations (publications that have at least one foreign author).

Figure 4.1. Trends in the share of international scientific collaboration in Africa



Source: Own elaboration and WoS.

This finding demonstrates the importance of international collaboration for African scientific output and motivates our research question to a certain extent.

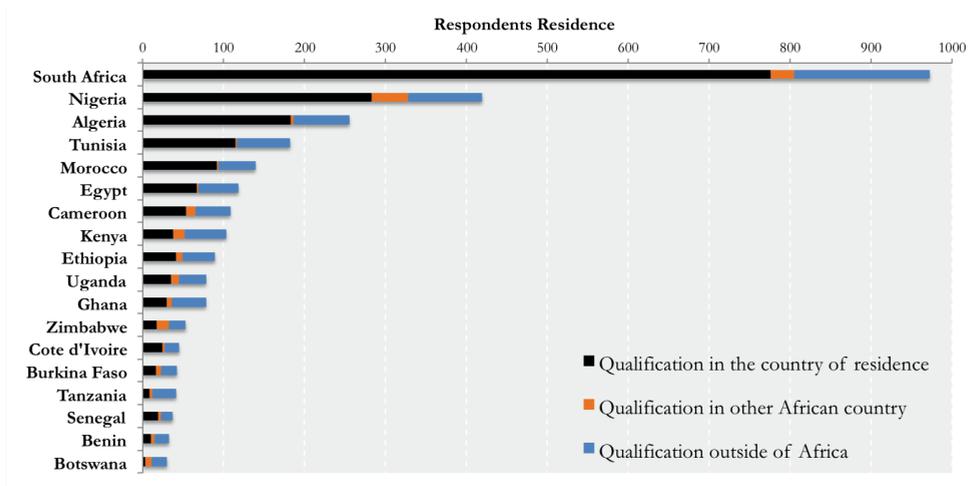
Our survey respondents comprise a fraction of the total number of African researchers. Around 81% have a PhD and 92% reside in the country of their nationality.²⁵ In Fig. 4.2 we can observe that, geographically, 32.4% (973) of the African researchers in our sample are based in South Africa. The other three countries with the most respondents are Nigeria (14%), Algeria (8.7%) and Tunisia (6.2%). At the same time, South Africa is the country with the smallest percentage of researchers who did their highest qualification outside of Africa (17.2%). On the other end, Tanzania (71.4%), Botswana (63.3%), Ghana (53.8%) and Benin (50%) are the countries with highest level of researchers with a non-African highest qualification.

²⁵ We are not counting African researchers who work in a non-African country.

The share of researchers with a qualification in an African country different from their country of residence is relatively small. Northern African countries have very few of those researchers and the country with most researchers who had their highest qualification in a different African country is Nigeria (45).

These results mirror the predominance of the South African research system within Africa. Besides having much more researchers than the other African countries, they produce a bigger share from their own system.

Figure 4.2. Number of researchers resident in an African country (by location of highest qualification).



Source: Own elaboration.

Note: We only included in this graph African countries that have at least 30 responses.

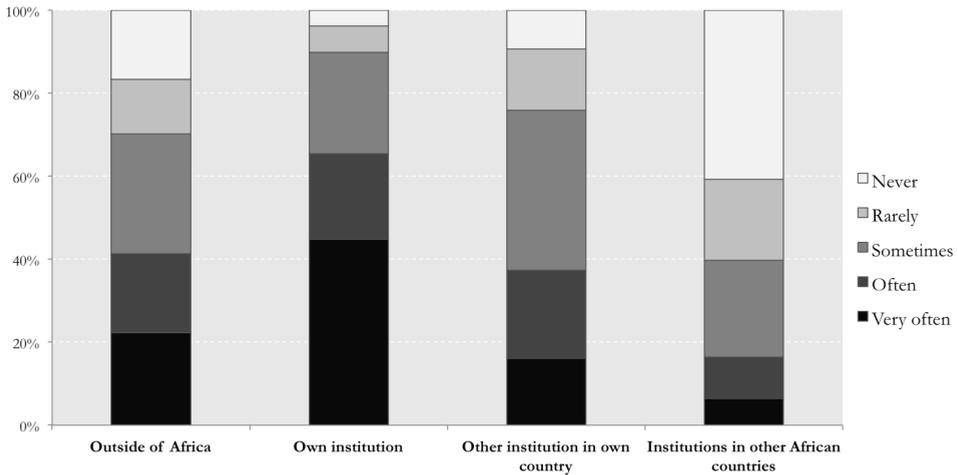
We have reasons to believe that researchers from Egypt are underrepresented in this sample. According to UNESCO (2015), Egypt accounts for more than 20% of the total number of publications with an African author in a similar period of analysis. In our sample, they are only 4% of the researchers (116). Researchers based in Egypt may have had a more difficult time receiving emails that include surveys or links to surveys. A number of respondents commented that emails of such a nature are blocked by mail servers and firewalls²⁶. In our research we will assume that the characteristics of Egyptian researchers are similar to the characteristics of researchers in Northern Africa, and in our econometric model we will control for African regions.

²⁶ Furthermore, some respondents mentioned the general suppression of academic freedom and access to information. However, these statements are based on specific comments from only a number of respondents.

Which factors influence international research collaboration in Africa?

Our main research question is focussed on studying the characteristics of the African researchers who engaged in collaborations with researchers outside of Africa. Therefore, it is important to examine in Fig. 4.3 what is the percentage of researchers who collaborate “very often”, “often”, “sometimes”, “rarely” and “never” with researchers from “outside of Africa”, “own institution”, “other institutions in own country” and “institutions in other African countries”.

Figure 4.3. Intensity of collaboration with four types of collaborators



Source: Own elaboration.

Not surprisingly, on average researchers collaborate most often with researchers from their own institution. Interestingly, it seems that on average the collaboration patterns of African researchers with researchers “outside of Africa” and “other institutions in own country” is very similar. And as expected, researchers in Africa collaborate more often on average with academics outside of Africa than with researchers from their own continent.

We also asked respondents about field of highest qualification and gender (see Fig. 4.A.1). The area with most researchers in our sample is “Natural Sciences” (32%) followed by “Medical and Health Sciences” (26%), “Social Sciences” (17%), “Agricultural Sciences” (13%) and “Engineering and Technology” (12%). About 30% of the researchers are female, with the percentage of females relatively higher for the subject areas “Social Sciences” (39%) and “Medical and Health Sciences” (37%), and smaller for “Engineering and Technology” (18%). Since collaboration patterns may be different between subject areas and gender, in our regression analysis we will control for both dimensions.

Respondents also reported on the major challenges that have impacted negatively on their careers. On average, the biggest challenge is lack of funding, and the challenge that they reported as least relevant is “political instability” (see Fig. 4.A.2). More than 20% of the respondents also reported that “lack of training opportunities to develop professional skills”, “lack of mobility opportunities” and “lack of mentoring and support” have negatively impacted their career to a large extent. In our econometric analysis we will generate three dummy variables that are one for researchers who reported that “mentoring”, “mobility opportunities” and “research funding” were a challenge they faced “to a large extent”.

Finally, we also gathered bibliometric information on respondents’ academic age (defined here as 2017 – year of first publication in WoS) and scientific productivity (defined here as number of publications in WoS per academic age). The average academic age is 10 years and the average scientific productivity is 1.06 publications. In Fig. 4.A.3 we can observe that most researchers in our sample have less than 20 years of academic age (88.5%) and less than 5 publications per year (97.6%). In the same figure we can also see the most highly productive researchers (>5 publications per year) have high levels of collaboration with researchers outside Africa.

4.4.1. Econometric Results

We used Stata to compute the multivariate ordinal probit. The regression model controls for scientific productivity, academic age, gender, subject area (5 of the 6 OECD categories): Natural Sciences, Agricultural Sciences, Engineering and Technology, Medical and Health Sciences, and Social Sciences; and also controls for regions: South Africa, Northern Africa²⁷ and Central Africa²⁸. Please see in appendix Table 4.A.1 for descriptive statistics.

²⁷ Algeria, Egypt, Libya, Morocco and Tunisia.

²⁸ All African countries except South Africa and Northern African countries

Table 4.1. Ordinal probit regression model

Ind. Variables	COL intensity outside Africa		
	oprobit I	oprobit II	oprobit III
Highest qualification (PhD) outside Africa (1 – Yes)	0.41*** (0.044)	0.53*** (0.074)	0.41*** (0.044)
Mobility abroad in the last 3 years (1 – Yes)	0.41*** (0.044)	0.40*** (0.044)	0.39*** (0.051)
Funding recipient in the last 3 years (1 – Yes)	0.14*** (0.045)	0.14*** (0.045)	0.14*** (0.045)
Share of international funding received in the last 3 years	0.0071*** (0.00059)	0.0071*** (0.00059)	0.0072*** (0.00059)
Lack of mentorship (1 – Yes “to a large extent”)	-0.11** (0.048)	-0.11** (0.048)	-0.11** (0.048)
Lack of mobility opportunities (1 – Yes “to a large extent”)	0.092* (0.049)	0.093* (0.049)	0.073 (0.059)
Lack of research funds (1 – Yes “to a large extent”)	-0.043 (0.046)	-0.044 (0.046)	-0.043 (0.046)
Female (1 – Yes)	0.044 (0.046)	0.043 (0.046)	0.043 (0.046)
Scientific productivity WoS	0.16*** (0.021)	0.16*** (0.021)	0.16*** (0.021)
Academic age WoS	0.012*** (0.0030)	0.016*** (0.0037)	0.012*** (0.0030)
PhD abroad x Academic age (Interaction)		-0.011** (0.0054)	
Mobility abroad x Lack of mobility opportunities (Interaction)			0.060 (0.097)
Constant cut1	-0.33*** (0.071)	-0.29*** (0.073)	-0.33*** (0.071)
Constant cut2	0.18** (0.071)	0.21*** (0.073)	0.18** (0.071)
Constant cut3	1.05*** (0.073)	1.09*** (0.076)	1.05*** (0.073)
Constant cut4	1.67*** (0.076)	1.70*** (0.079)	1.66*** (0.077)
Subject area effects	Yes	Yes	Yes
Regional effects	Yes	Yes	Yes
Observations	2,954	2,954	2,954
Pseudo R2	0.077	0.077	0.077
Wald chi2(16)	619		
Wald chi2(17)		622	619

Note 1: Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Note 2: The intercept parameters are significantly different from each other so the five categories should not be combined into one.

Note 3: To complement this analysis, we carried out three different robustness checks. We computed an identical model using ordered logit regression; we computed our model using subject area groups, in order to check whether the results are consistent in all areas of knowledge; and we separated our observations in three African regions (South Africa, Northern Africa and Central Africa). In general, the results were consistent with this model.

Source: Own calculations.

In Table 4.1 we compute our regression using as dependent variable collaboration intensity outside of Africa. Model specification I is the original regression and model specifications II and III include interaction terms.

The first of our results indicates that researchers who did their highest qualification outside of Africa are more likely to be in a higher category of collaboration intensity outside Africa. A possible explanation is that African researchers who studied outside of Africa may expand their collaboration network, and upon returning to the continent, still maintained links with their research groups abroad. Our model confirms this hypothesis. A researcher that did his/her highest qualification outside of Africa is 11% more likely to be in the ‘very often’ category of collaboration outside of Africa, on average, than a researcher that did his/her highest qualification in an African country. However, when we interact this variable with academic age, the interaction term is negative and significant (II). This means that the connections gained during the PhD may lose importance during a researcher career.

Another variable that is positively and significantly associated with a higher category of collaboration intensity outside of Africa is “mobility in the last three years”. A researcher that has studied or worked abroad in the last three years before the survey is 11% more likely to be in the very often category of collaboration outside of Africa.

Being the primary recipient of research funding and receiving a higher share of international funding (being or not a primary recipient) are also positively and significantly associated with the likelihood of collaborating with researchers outside of Africa. Indeed, a researcher that receives 100% of their research funds from international sources is 20% more likely to collaborate very often with people outside of Africa than researchers who receive 0% of funds from international sources (on average). This indicates that collaboration often depends on the resource availability provided via international grants.

An interesting result is that different challenges faced during a career of a researcher seem to impact differently the likelihood of collaborating with researchers outside of Africa. For example, a researcher that reported that lack of mentorship was a challenge they faced “to a large extent” during their career is less likely to be in a higher category of collaboration intensity. A possible explanation is that younger researchers, who usually do not have access to all potential choices for initiating successful and fruitful collaborations, may depend on their mentors/supervisors to establish new collaborations. If their mentors/supervisors do not have access to an extensive research network or do not have the networking skills needed to introduce their students to other researchers, the young researchers may be penalized in the future. We are aware that a chronic

problem in some African universities is the lack of qualified human resources for many teaching and training positions and therefore the availability of experienced researchers for supervision and mentoring is limited (Gaillard 2003). Yet, our results seem to indicate that well targeted pedagogic support may play an important role in expanding a researcher's network.

On the other hand, the positive and significant sign of the coefficient of "lack of mobility opportunities" indicates that researchers who perceive they had lack of mobility opportunities are more likely to be in a higher category of collaborating intensity with researchers outside of Africa. At first glance, this is not intuitive since greater mobility usually is associated with more collaboration. However, this relation may happen because only the researchers who move frequently perceive that they have less mobility opportunities than their peers around the world. To test this hypothesis we interacted the dummy variable "lack of mobility opportunities" with the dummy variable "mobility in the last three years". In specification III we can observe that the coefficient of "lack of mobility opportunities" is no longer significant at 10%. This means that indeed the positive and significant coefficient of "lack of mobility opportunities" in specification I is partially explained by the variable "mobility in the last three years". Many of those who didn't move, may report they don't have lack of mobility opportunities because they do not care about mobility and don't see it as something important for them.

4.5. Conclusions

This chapter contributes to an understanding of the factors that allow some African researchers to collaborate more often with international partners than others. Data are taken from WoS and a survey that collected detailed information about the individual characteristics of 2954 African researchers in 42 African countries. Our research finds that doing the highest qualification (PhD) in a non-African country, recent international mobility and receiving a higher share of international funding is positively and significantly associated with higher collaboration intensity with researchers outside Africa.

One of the main results is that going abroad to do the highest qualification significantly expands your collaboration network outside of Africa. Our research also shows that the positive effect on collaboration of doing a PhD abroad diminishes over time. At the same time, researchers who were recently mobile are more collaborative. PhD scholarships should therefore not be limited to the time-period of the PhD, but should allow for additional travel and visiting opportunities post-PhD, to ensure that the networks are maintained or expanded. This could very simply be in the form of a few conference visits or other forms of research visits in the years following PhD

graduation. What is important is that the gains that a researcher has received from doing a PhD abroad should not be allowed to wither away so easily with time.

While there is value to foreign training, funders need to consider carefully to which universities they send PhD students, as not all foreign PhD training is equally useful. As argued by (Müller et al. 2018) in the South African context, the gains obtained from going to a second tier foreign university may be smaller than going to a first tier local university. Furthermore, in this chapter we don't address the dangers of the "brain drain" or the lack of infrastructure and qualified human resources in many African universities. We are aware that this is a huge problem since some talented PhD students may not return to their home country. Therefore, beyond increasing the number of scholarships for student to go abroad, funders should also continue to develop foundations for research within Africa.

One possibility is for international donors to provide direct funding through research projects. Our work also finds empirical evidence that researchers who received a higher share of international funding are collaborating more often with researchers outside of Africa. This was an expected result since several funding agencies, particularly government agencies, mandate cross-country collaboration as part of their funding conditions. However, what we don't explore in this chapter is if a higher share of international research funding increases the number of projects with societal relevance for the local populations and effectively enhances research capacity. Policies that wish to promote research collaboration should therefore also pay special attention to funding schemes and mechanisms that avoid "unequal partnerships" (Boshoff 2009; Gaillard 1994).

Several caveats must be kept in mind with regards to our study. First, we use a categorical, self-reported collaboration intensity dependent variable. This has some disadvantages in terms of the stability of the construct since certain researchers may perceive that they collaborate more often internationally than others, when objectively (e.g. co-authorships) this is not the case. However, in line with Duque et al. (2005) this approach may have certain advantages: 1) it can include collaborations that that did not involve publication; 2) it might exclude co-authors who achieved that status not by virtue of collaboration but because of influence and hierarchical position.

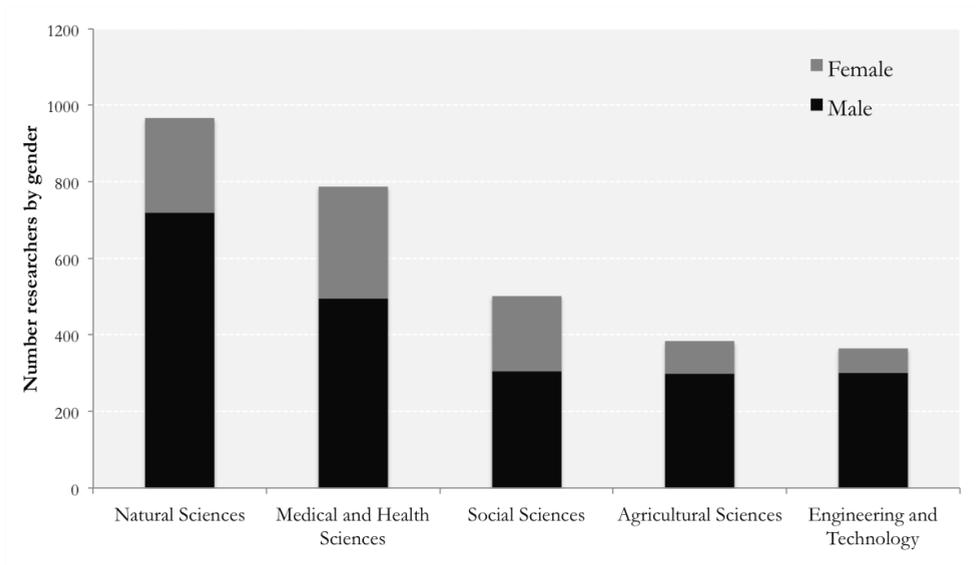
Second, due to the structure of our data, we cannot claim a causal relationship between our independent variables and collaboration intensity with researchers outside of Africa. We do not have longitudinal data since there was only one survey round, and thus cannot observe changing patterns over time. Thus, there may be confounding factors like intrinsic ability that may lead to omitted variable bias. Nevertheless, we take a first step in analysing what characteristics are associated with researchers who collaborate more internationally.

Which factors influence international research collaboration in Africa?

Besides that, the survey didn't include questions about the motives of collaboration or the dynamics of collaboration seeking. We also have no measures of quality of collaboration and the societal impact of those collaborations. Many researchers in African countries, when they do collaborate internationally, tend to participate in projects that have been conceptualized and designed in the "Global North" (Boshoff 2009). Who searches for whom and for what? How many frequent collaborators do you have? Future surveys on this topic should probably include such questions.

4.6. Appendix

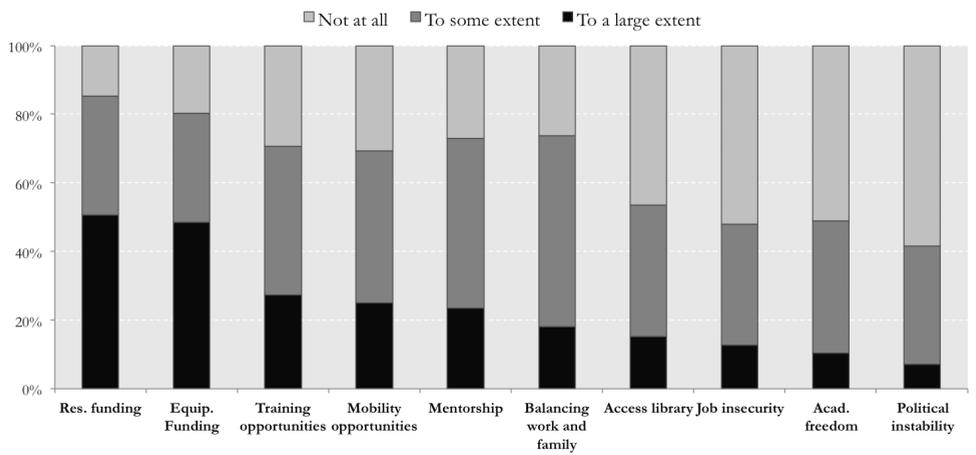
Figure 4.A.1. Number of researchers by area and gender



Source: Own elaboration.

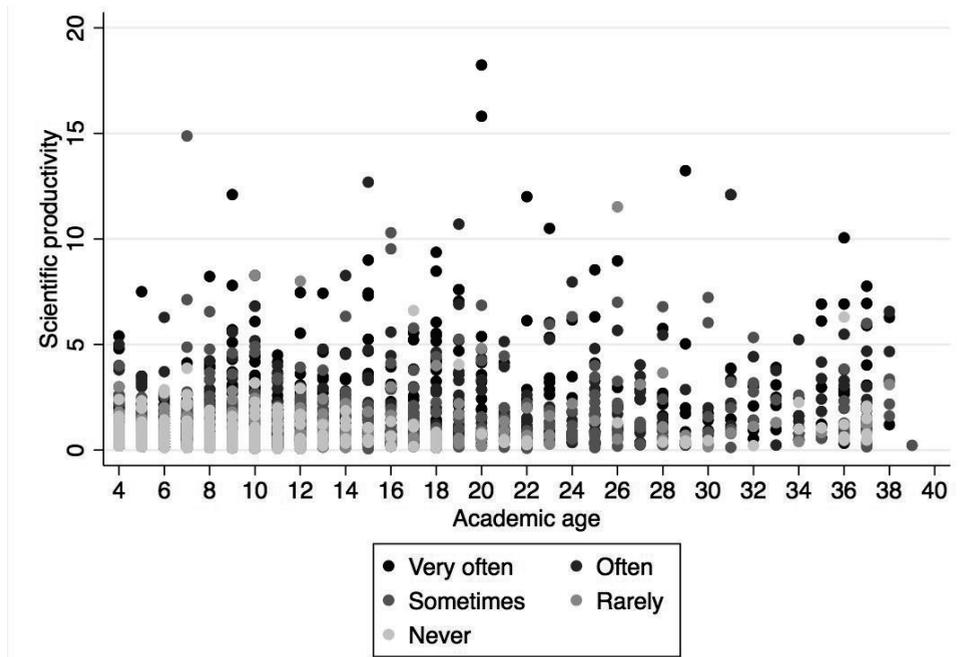
Note: Humanities related fields were excluded due to the limitations of bibliometric indicators in this area.

Figure 4.A.2. Challenges faced during the career



Source: Own elaboration.

Figure 4.A.3. Scatter plot of scientific productivity versus academic age by level of collaboration intensity outside of Africa



Source: Own elaboration and WoS.

Table 4.A.1. Descriptive statistics

Variables	Obs	Mean	Std. Dev.	Min	Max
Collab. intensity with researchers at institutions outside Africa	2954	3.12	1.36	1	5
Collab. intensity with researchers at their own institution	2954	3.97	1.13	1	5
Collab. intensity with researchers at other institutions in own country	2954	3.18	1.16	1	5
Collab. intensity with researchers at institutions in other African countries	2954	2.18	1.23	1	5
Highest qualification (PhD) outside Africa (1 – Yes)	2954	0.31	0.46	0	1
Mobility abroad in the last 3 years (1 – Yes)	2954	0.31	0.46	0	1
Funding recipient in the last 3 years (1 – Yes)	2954	0.39	0.49	0	1
Share of International funding received in the last 3 years	2954	25.33	38.77	0	100
Lack of mentorship (1 – Yes “to a large extent”)	2954	0.24	0.43	0	1
Lack of mob. opportunities (1 – Yes “to a large extent”)	2954	0.26	0.44	0	1
Lack of research funds (1 – Yes “to a large extent”)	2954	0.52	0.50	0	1
Female (1 – Yes)	2954	0.30	0.46	0	1
Scientific productivity WoS	2954	1.06	1.38	0.06	18.24
Academic age WoS	2954	9.99	7.81	0	46

Source: Own elaboration.

Note: No variable is correlated with any other more than 40% (low probability of multicollinearity).

Chapter 5

Medical research versus medical needs in Africa

Hugo Confraria
UNU-MERIT, Maastricht University

Lili Wang
UNU-MERIT, Maastricht University

Abstract

In this chapter we ask if the distribution of medical research priorities and investment in medical research, across diseases in Africa, is related to the medical/health needs of local populations, and to which ones. Research priorities are derived from the mapping of the research publication portfolios of African regions across diseases and the relative specialisation of regions in each of the diseases. Medical needs are proxied by the burden of disease of each disease across African regions. We found that in sub-Saharan Africa most diseases with high disease burden are also the ones with relatively more research effort. The region where there is higher positive association between disease burden and research effort is Eastern Africa. Northern Africa is the region where these two dimensions are less aligned. We also found that most international research funders (public non-African and philanthropic) use their resources to fund specific diseases such as HIV/AIDS, tuberculosis and parasitic and vector diseases. On average, these diseases are also the ones with higher relative disease burden in the regions with higher dependence on international funding. Therefore, what our regression results seem to show is that high levels of dependence in international donors are not necessarily associated with less alignment between local health needs and local medical research effort, as some literature suggests.

Keywords: Research priorities, disease burden, science policy, Africa

5.1. Introduction

Africa is a continent facing severe, urgent, and often unique health challenges. The region has made overall progress during the last decades in reducing mortality and prolonging life but its burden of disease per population continues to be two times higher than that of higher income countries.²⁹ At the same time, most African countries have difficulties in supporting medical research, and the pharmaceutical industry may be reluctant to sponsor research in lower income countries because the prospects of profit are limited, even if effective treatments are developed (Taylor 1986; World Health Organization 2012). Therefore, health research performed in this region should have a special emphasis on their local health problems.

In this study, our main objective is to evaluate whether the amount of research produced on various medical conditions by African researchers is related to their countries' burden of disease and health needs of the local populations. To do that we will combine bibliometric data from the Web of Science with the estimates of the disability-adjusted life years (DALYs) produced by the World Health Organization (WHO).

Previous research has suggested that at the global level there is lack of alignment between research effort and health problems (Agarwal & Searls 2009; Atal et al. 2018; Evans et al. 2014; Rafols & Yegros 2017; Röttingen et al. 2013). In our research we will look specifically at the scientific output of African researchers and also at the funding institutions acknowledged in their publications. This will allow us to evaluate whether international development funders and pharmaceutical companies are supporting research that matches the research needs of African regions.

This approach and research question may be interesting for two main reasons. First, due to the tremendous health challenges the continent faces, improved Africa-relevant health research and well-trained health workers can have a great impact on health outcomes. Second, the improvement in research capabilities in the continent in the health sciences may demonstrate that persistent support and funding from development partners such as the Wellcome Trust, NIH, European Union or Bill and Melinda Gates Foundation, might pay off.

²⁹ <https://vizhub.healthdata.org/gbd-compare/>

5.2. Background

The majority of the gains in human longevity of the past several thousand years have occurred in the last decades. Besides the influence of improved general living conditions, advances in medical scientific knowledge and its application both in creating cures, solutions and in guiding behaviour have contributed significantly to these gains (World Health Organization 1996).

Medical research, perhaps more than any other area of study, holds dearly the idea that science is transformed into health solutions (Sarewitz & Nelson 2008). Scientific research has changed as well as deepened understanding about how the human body works and the nature of disease pathologies, and thus has illuminated the pathways to modes of cure and prevention.

However, health research capacity is distributed unevenly across regions in the world. Regions where disease burden is the greatest are often severely disadvantaged by the small numbers of their scientists and lack of investment. Lack of access to current research findings, low wages, poor career prospects and lack of scientific infrastructure also contribute to the lack of scientific capabilities in the low-income contexts, with many researchers opting to work abroad or forced to devote more time to other activities such as teaching and consultancy (Evans et al. 2014; World Health Organization 1996). From the policy side research has had an image problem, perceived as being an additional demand on overburdened services and taking many years to produce results that have little immediate relevance (Mgone et al. 2010). It has also been argued that, in low-income contexts, many health solutions may appear through frugal innovations³⁰ that do not require basic research, just the application of current knowledge to local problems (Jones et al. 2007). Therefore, since health institutions in higher income countries have more resources to invest in their research priorities (public or private), it is no surprise that the diseases receiving the most research attention in the World are those that are predominant in high-income countries (Atal et al. 2018; Rafols & Yegros 2017; Røttingen et al. 2013).

Nevertheless, it has been well recognized that health research conducted in low-income countries is of great importance. Medical research done and applied in lower income contexts can provide enormous contributions to discovering previously unknown diseases which have substantial social and economic impact on the world (B. Y. Lee et al. 2013). For example Carlos Chagas in Brazil discovered the parasite American trypanosomiasis (also called Chagas disease), which is a type of neglected tropical disease especially common in low-income countries, and Carlos Finlay from Cuba found that a mosquito transmits the yellow fever virus.

³⁰ <https://www.theguardian.com/business-call-to-action-partnerzone/2018/aug/09/how-whatsapp-and-sms-are-being-used-to-save-the-lives-of-babies-in-africa>

Furthermore, the inclusion of local stakeholders at the different stages of health research and innovation is crucial to the adoption of new ideas and technologies emerging from research efforts (Cochrane et al. 2017). Medical research based in low-income countries can help researchers have a clear understanding of the local constraints and barriers to and facilitators of the implementation of research in practice. Researchers in African countries are best placed to identify and address the health challenges of their nations and to provide local and national policy-makers with a broad range of high-quality, relevant evidence to inform policy making. This interaction is especially important in some African communities where target populations are sometimes socially resistant and non-adherent to medical intervention, and it has inspired educational projects to enhance the public understanding of medicine and practitioner understanding of diverse patient cultures (Aizer & Stroud 2010).

Therefore, it has been argued that the capacity of researchers and institutions in low-income regions must be strengthened to address their problems more effectively (Cardoso et al. 2014). Strengthened ability to understand the determinants of health in relation to gender, ethnicity, cohorts, and communities among different African populations, may be crucial to more effective and lasting ways to improve health outcomes and health systems in the region (Ezeh et al. 2010; Juma 2016). Furthermore, the knowledge acquired by local scholars in this process could be diffused through different interactions with peers, students, health professionals and society in general, which can create human capital capable of implementing and creating solutions.

In this chapter we will follow this argument and assume that in order to improve health conditions in a relevant manner in resource-poor settings, the improvement of scientific capabilities of local researchers is critical to have trained professionals, setting priorities and identifying cost-effective interventions. We do not argue that research alone will solve the problems. Commitment, resources, management and knowledge about the context are fundamental to success. What we do claim is that health research is necessary to facilitate health action within communities and nations, and essential to shape new knowledge worldwide. Research informs the attitudes with which people think about themselves and their world. Research fosters a scientific, problem-solving culture, which can be extremely important when new problems arise and cause great suffering before we understand them. Research is also necessary to anticipate problems because, in the complex systems that affect health, common sense can sometimes be misleading (CHRD 1992). Therefore, in our view, building national scientific capacity is essential to identify and prioritise health problems, to adapt and apply new technologies, and to optimise the health benefit of limited resources.

5.2.1. Health research in Africa

Total health research output from African countries is relatively small when compared to high-income contexts. Researchers from South Africa and a few countries in Northern Africa generate the majority of the publications, and in most states the share of international collaboration is relatively high (Confraria & Godinho 2015; Tijssen 2007).

African countries are committed to increasing funding for science, but overall levels of funding are still low (Chataway et al. 2017; Mgone et al. 2010). According to Cardoso et al. (2014), lack of funding for research is the major barrier to the development of clinical research capacity in Africa. However, political, economic or socio-cultural factors like lack of policymakers' understanding of the importance and benefits of research, lack of human resources and lack of infrastructure should also be taken into account. Furthermore, they argue, "the overwhelming majority of clinical research in all 46 countries is based on funding from external donors. In the majority of cases, clinical research typically appears to be conducted in vertical 'silos', with African researchers working closely with their donors and European and US academic partners, while local governments are taking a peripheral role" (Cardoso et al., 2014, p14). On the same line, in an ethnographic study made in Eastern African, (Moyi Okwaro & Geissler 2015) argue that most of their interviewed scientists stated that biomedical research in their institutions would be impossible without northern collaborators. One university director remarked: "Everything you see here has been obtained from donor funding". Except for South Africa, in most African countries, government funding appears to be limited to indirect support such as staff salaries, infrastructure and provision of subsidised equipment rather than funding specific health research programmes (Cardoso et al. 2014).

Since external concerns and criteria for funding often drive research in Africa, an important question is to understand if the high levels of international collaboration are leading to lasting benefits (e.g. capabilities and infrastructure). In this regard, Karim & Karim (2010) for example argue that in South Africa having financial autonomy from the government has had three main benefits for local HIV/AIDS research communities. First, it has reduced dependence on the government, enabling scientists to challenge politicians on their HIV/AIDS denialism without fear of losing research funds. Second, it has raised the quality of local research to international standards. Third, it has enabled South Africa to build research capacity and infrastructure, which potentially allowed the reduction of "brain-drain".

5.2.2. Setting priorities for health research in Africa

Since rich countries produce much more medical research of all kinds than lower-income countries, and since higher and lower income nations have divergent health profiles, there is a general expectation that in resource-poor settings available research funds must respond more directly to community health needs, and therefore be conducted according to recognised priorities (Evans et al. 2014; McGregor et al. 2014; Rafols & Yegros 2017). In the absence of priority setting, there is a risk that research conducted in African countries will follow topics determined by foreign funders for their purposes and consequently fail to respond to specific local health needs (Binka 2005; Gaillard 1994).

In order to maximize the return on investment, funders need tools to help assess and compare disease burden, state of the science, and knowledge and product gaps, as the basis for deciding in which disease and product areas they can best invest. For some diseases, this may mean a stronger focus on basic science rather than product development, while for other diseases, funding may need preferentially be directed to product development.

The methods for setting priorities for health resource allocation range from qualitative methods such as consensus building with health experts and users, to the use of quantitative formulations and prioritisation matrices. Quantitative approaches, as the disability-adjusted life year (DALY) have gained prominence in the research priority-setting process because these measures allow for a cross-comparison among a broad range of diseases, regions and are particularly attractive for cost-benefit analysis.

Using this method, some research argues that there are substantial misalignments, at the global and local levels, between research efforts and WHO estimates of health burden for a given disease (Evans et al. 2014; Rafols & Yegros 2017). However, recent studies have suggested a more nuanced perspective that in Sub-Saharan Africa the conduct of random controlled trials is aligned with the burden of highly prevalent diseases as HIV and malaria, but other major causes of burden remain neglected by research effort, in particular, common infectious diseases and neonatal disorders (Atal et al., 2018; Ndounga et al., 2017).

In this chapter, instead of using random controlled trials as a measure of research supply we will use scientific publications in WoS (articles and reviews). From the demand side, we will use DALYs (disease burden) in each disease and African region as a proxy for societal needs in health. We will focus on four African regions (Eastern Africa, Northern Africa, Southern Africa or West & Central

Africa) as defined by the UN classification³¹, and our central research questions will be:

1. What is the current situation of medical research in Africa?
2. Is the amount of research produced on various diseases by African researchers related to their countries' burden of disease?
3. What kind of health research is being funded by international organisations?
4. What are the main drivers of medical research in Africa?

5.3. Data and Methods

Our analytical section is composed of two segments. In the first section, we use descriptive statistics to display the association between medical research specialisation and disease burden specialisation by African region and by disease field; and examine what type of medical research is relatively more funded by certain types of funders than others. In the second part we further explore the research and disease association using regression analysis.

5.3.1. Health needs

To identify medical priorities, we used the WHO data to measure the burden of disease. The WHO estimates the DALYs for an array of common conditions through its Global Burden of Disease (GBD) project. The DALY is a summary measure that combines time lost through premature death and time lived in states of less than optimal health, loosely referred to as “disability”. One DALY can be thought of as one lost year of healthy life, and the measured disease burden is the gap between a population's health status and that of a normative reference population (World Health Organization 2017). The WHO estimates DALYs for 136 health conditions, which are grouped in three broad cause groups: Group I (communicable, maternal, perinatal and nutritional conditions), Group II (noncommunicable diseases) and Group III (injuries). We excluded injuries (Group III) and we classified all the other conditions into 28 categories³²: Cardiovascular diseases, childhood-cluster diseases, congenital anomalies, diabetes mellitus, diarrhoeal diseases, digestive diseases, encephalitis, endocrine blood immune disorders, genitourinary diseases, hepatitis, HIV/AIDS, intestinal nematode infections, leprosy, malignant neoplasms, maternal conditions, meningitis, mental and substance use disorders, musculoskeletal diseases, neonatal conditions, neurological conditions, nutritional deficiencies, oral conditions, parasitic and vector diseases, respiratory infections & diseases, sense organ diseases, skin diseases, STDs excluding HIV and tuberculosis.

³¹ https://en.wikipedia.org/wiki/United_Nations_geoscheme_for_Africa

³² The categories “Other infectious diseases” and “Sudden infant death syndrome” were also excluded due to ambiguity.

5.3.2. Research priorities

One approach to establishing priorities is to relate research investments to disease burden. While estimates of DALYs by disease have been made, estimates of investments for health problem are usually unavailable at the national level and are only limited and incomplete for international investments. Another approach is to look at scientific research output (published in medical peer review journals) as a proxy for resources applied to a specific disease. In this chapter, we interpret the distribution of medical publications per disease in a given region/period, as the distribution of revealed priorities in medical research for that region/period.

The identification of publication output (articles and reviews) comes from the WoS. We extracted each article that was produced by at least one author from an African institution, and we use the whole counting method (e.g. an article in international collaboration between UK, Kenya and Tanzania biomedical researchers would be credited to both Kenya and Tanzania). We used WoS instead of Scopus or other databases because we also wanted to gather data about funding institutions mentioned in the acknowledgements of every paper and this was only possible with WoS. We are aware that WoS may underrepresent journals from lower-income regions, but it is a database that is otherwise reliable and widely used for bibliometric studies. Although not all R&D efforts are embodied in scientific articles, it is argued that in medical related areas scientific publications tell us more on actual applications of knowledge than publications in other fields of science. This is because, in order to start a new treatment in clinical practice, there must exist scientific evidence that the new drugs, procedures or devices work and are robust. This kind of evidence is systematically published in peer-reviewed journals (Mina et al. 2007).

Publication records were assigned to a specific disease field by searches in abstracts and titles. We built a set of keywords that are strongly associated to a specific disease (or group of diseases) based on the ICD-9 codes³³ and previous research (Cardoso et al. 2014; Chapman et al. 2017; MSF 2016). After building our queries (see Table 5.A.1 in the appendix), two external peer reviewers³⁴ reviewed the keywords for each one of our 28 disease categories.

After cleaning the publication data for entries with missing information, and limiting the analysis to 2006-2015, we are left with 59486 documents that were associated to at least one specific disease (28) and one African region (Eastern Africa, Northern Africa, Southern Africa or West & Central Africa).

³³ <http://icd9.chrisendres.com/>

³⁴ One of the reviewers is a PhD student in international health and development, who worked for five years as a nurse in epidemic contexts in several African countries. The other is a nurse with 15 years of experience and a MSc in health economics.

5.3.3. Metrics

With the hypothesis that health needs in earlier years should drive the research agenda in later years, we compared the number of articles published between 2011-2015 with the disease burden in 2010. First, we will count DALYs in each disease per region/period and number of publications in each disease per region/period. Then, since different diseases have different propensities to affect people and be researched, we will also compute relative specialisation indexes (RSI) to assess the relative specialisation of each disease in a given region. We will do this by computing the revealed comparative index (Balassa 1965):

$$RSI_{rd} = \frac{P_{rd}/\sum_d P_{rd}}{P_d/\sum_d P_d} \quad (1)$$

where P is the number of publications in region r in disease d . This index can be interpreted as a “comparative advantage”. If a region r has a relative specialisation in disease d , it means that r has more scientific research focused on disease d than the world average ($RSI > 1$). Likewise, based on DALYs data, we also calculate the revealed specialisation index for disease types in each region.

The definition of the index implies that its value is necessarily null or positive but is not bound by an upper limit. For this reason, we will standardise this measure by forcing the RSI index to take values between -1 and +1 by computing the ratio of RSI minus one over RSI plus one:

$$NRSI = \frac{(RSI - 1)}{(RSI + 1)} \quad (2)$$

The threshold value of the normalised relative specialisation index (NRSI) remains zero, but the asymptotic limits are now ± 1 . The standardization is implemented for both publication and disease RSI values.

In order to analyse the research being funded by a certain institution in a given disease or region, we use the acknowledgement paratext of scientific publications where authors commonly give thanks to the funding agencies (Costas & Leeuwen 2012; Grassano et al. 2017; Rigby 2011). This source of information has great potential because it allows us to establish a direct link between funding inputs and research outputs on a grand scale, without the need to gain direct access to data via individual funders or researchers. However, making systematic use of the funding information included in the acknowledgement sections of publications must be done with caution because of the lack of standardization in the structure, content and reporting practices of researchers in the acknowledgements by subject area and region (Grassano et al. 2017). Therefore, research using acknowledgement funding data for large-scale bibliometric analyses was limited until, in more recent years, a new wave of studies has emerged with much higher sample sizes (e.g.

Costas & Leeuwen 2012; Gök et al. 2016; Grassano et al. 2017). These studies have been eased by the increasing availability of ready-classified information provided by some citation indexes. WoS, for example, has recently provided access to the funding text of acknowledgements and to the list of funders and grant codes mentioned in these, for publications from August 2008.

In our case, we have focused only on publications from 2009 to 2015, and we separated them into two groups (2009-2010 and 2011-2015)³⁵. We used VantagePoint and manual searching methods to group different name variations for the same funding institution mentioned in the acknowledgements section of our sample of publications. We only included name variations that showed more than 0.05% of times (>30 publications), and we ended up with 141 funding institutions. Besides calculating the number of publications with acknowledgements to a specific funding institution by disease and region, we also grouped each funding institution in five four group types based on the G-finder classification³⁶: 1) African public funding; 2) Non-African public funding; 3) Multilateral funding; 4) Philanthropic funding and 5) Corporation funding. Subsequently, we calculated the normalized relative specialisation index (NRSI), following the steps explained before, for each funding group on 2011-2015 and 2009-2010 by disease and region.

5.3.4. Econometric approach

In this study, our primary research question is to understand whether disease burden relative specialisation is associated with medical research specialisation between different African regions across different diseases. To address this, in our multivariate regression analysis (OLS), we use scientific specialisation (NSRI Pubs) as our dependent variable, and disease burden specialisation as our main independent variable. Since most African countries are highly dependent on international research collaboration, in our model we control for level of international collaboration. We will also control for previous scientific specialisation due to the path dependent nature of scientific production.

In our model (3), $NRSI_{Pub}$ is the scientific specialisation index in a certain region r , disease field d and period t (2011-2015). $NRSI_{DALY}$ is the disease burden specialisation index in period $t-1$ (2010). IC is the percentage of publications of a region and disease in international collaboration (at least one author from a non-African country), $L_{NRSI_{Pub}}$ is a lagged dependent variable from the previous period (2006-2010), and R is a control for each of the four African regions.

³⁵ According to Costas & Leeuwen (2012) two important limitations must be taken into account when working with this source of information: 1) WoS funding information is dependent on the algorithm developed by Thomson Reuters, which may not be applied systematically in all journals, for all publications, for all disciplines, etc. 2) Second, an important conceptual limitation is that acknowledgements are a voluntary activity. Hence, authors can also decide not to acknowledge funding, or forget to do so.

³⁶ <https://gfinder.policycuresresearch.org/PublicSearchTool/>

Finally, α is the constant, and ϵ is the unobserved residual.

$$NRSI_Pub_{rd} = \alpha + \mu NRSI_DALY_{rd} + \varphi IC_{rd} + L_NRSI_Pub_{rd} + R_r + \epsilon_{rd} \quad (3)$$

Since we are also interested in understanding if international funders are supporting medical research that is relevant for the health needs of African regions, we also computed a set of regressions that estimate what the relation between disease burden relative specialisation and research funding relative specialisation by donor category is. We did this analysis by using five different types of donor categories (dependent variables): 1) African public funding; 2) Non-African public funding (includes multilateral funding); 3) Philanthropic funding; 4) Corporation funding; and 5) Non-funded research (or not identified).

$$NRSI_FundCat_{rd} = \alpha + \mu NRSI_DALY_{rd} + \varphi IC_{rd} + L_NRSI_FundCat_{rd} + R_r + \epsilon_{rd} \quad (4)$$

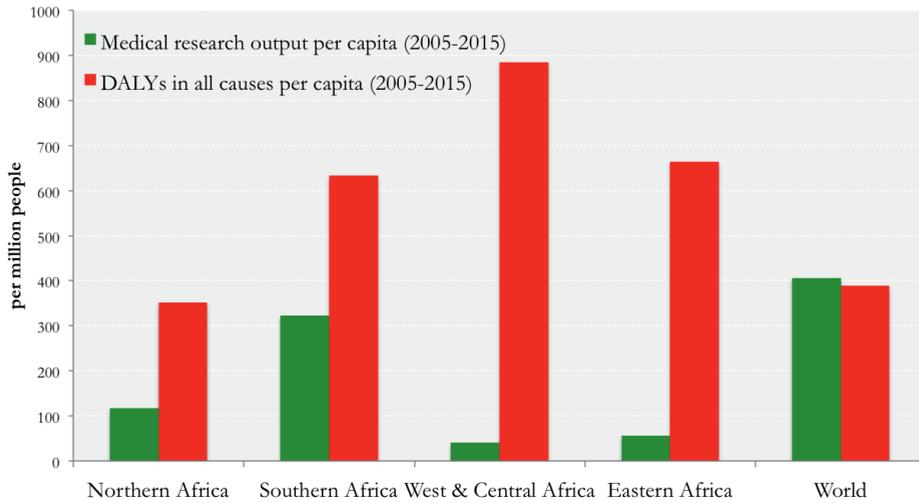
In our model (4), $NRSI_FundCat$ is our specialisation index (for each of the five funding categories) in a certain region r , disease field d and period t (2011-2015). $NRSI_DALY$ is the disease burden specialisation index in period $t-1$ (2010). IC is the percentage of publications of a region in international collaboration (at least one author from a non-African country), $L_NRSI_FundCat$ is a lagged dependent variable from the previous period (2006-2010), and R is a control for each African region. Finally, α is the constant, and ϵ is the unobserved residual.

5.4. Results

5.4.1. Descriptive Analysis

As discussed before, there are vast imbalances in global health between Africa and higher income regions. In 2015, the estimated world share of scientific output in medical and health-related areas by African researchers accounted to only 2.8%, a marked contrast to the fact that almost 26% of the global disease burden in 2015 was in Africa. Fig. 5.1 shows that, different from the situation of world-wide average, the disease burden per capita is much higher than medical research output per capita in all African regions. This mismatch is especially remarkable for Eastern Africa and West & Central Africa.

Figure 5.1. African regions medical scientific production, and disease burden per capita.



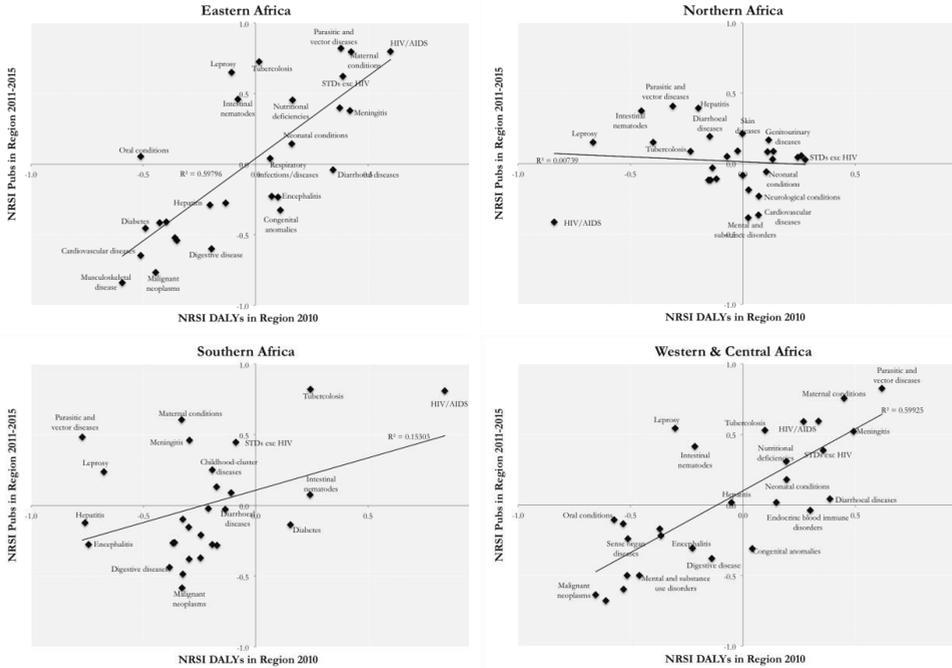
Source: Own calculation based on WoS and WHO

Note: In this chart, the numbers presented are yearly averages. For the medical research output, the calculation only takes to account the 28 diseases identified (Type I and II diseases). For the DALYs, the calculation is for all causes (Type I, II and III diseases).

It is well known that Africa’s scientific output is highly skewed across nations and disciplinary areas, with South Africa and Egypt representing around 50% of total African output, and their specialisation being mostly on agricultural sciences given the needs of the continent and in health-related sciences due to research work on tropical diseases and specific health problems, as well as from the location of international medical research centres on African soil, and the abundance of international cooperation between African researchers and those overseas (AOSTI 2014; Confraria & Godinho 2015; Tijssen 2007; UNESCO 2015). We found that medical research per capita has increased in all regions around 50%, between 2006-2010 and 2011-2015, and that the disease with more publications in each region is “parasitic and vector diseases” in Eastern and Western & Central Africa, “malignant neoplasms” in Northern Africa and “HIV/AIDS” in Southern Africa. As for DALYs we found that the countries with highest incidence by African region are: Egypt (38% of total) for Northern Africa; South Africa (87% of total) for Southern Africa; Nigeria (40% of total) for West & Central Africa and Ethiopia (23% of total) for Eastern Africa. DALYs per capita are decreasing on average in all regions, but the most negative growth rates were in Eastern and Western & Central Africa.

To further assess the association between research output and disease burden, in Fig. 5.2 we display the disease burden specialisation (relative to the world) as a function the medical research specialisation (relative to the world) in 28 diseases (type I and II excluding type III diseases) by African region.

Figure 5.2. Disease burden specialisation vs research specialisation by disease.



Source: Own calculation based on WoS and WHO.

Note 1: DALYs from 2010 and scientific publications from 2011-2015.

Note 2: We also plotted the DALYs from 2005 and scientific publications from 2011-2015 in Fig. 5.A.1 in appendix and the results are very similar.

Eastern Africa and West & Central Africa exhibit a strong positive association between the two dimensions. The interpretation is that the diseases that have a higher disease burden relative to the World, are also the ones that have higher scientific specialisation. On the other hand, this positive association is not so clear for Southern Africa and especially in Northern Africa. In these regions, there are some diseases like “parasitic and vector diseases” and “leprosy” that display a low level of disease burden specialisation but a high level of scientific specialisation. One could argue that these topics are “over-researched” since the disease burden in these regions is not relatively high. However, since in 2015 globally, 4 out of 5 DALYs in “parasitic and vector diseases” are from Africa, due to the high disease burden in the Eastern Africa and West & Central Africa, this high level of scientific specialisation may be justified by the existence of research tradition in these areas,

promotion of intra-African research collaborations and the development of research capacities in other African regions. As for leprosy, since it is considered a neglected disease, with residual impact in high-income countries, the existence of scientific capabilities in these regions may be justified.

Another interesting finding is that in all African regions “tuberculosis” is a topic of high scientific specialisation. Since the risk of developing tuberculosis is estimated to be between 16-27 times greater in people living with HIV than among those without HIV infection³⁷, the high levels of scientific specialisation in “tuberculosis” may be because in most countries HIV/AIDS research is done in conjunction with tuberculosis research. In South Africa, for example, most patients who die from HIV-related causes die from tuberculosis or similar illnesses.

Finally, we didn't find any region where a disease with relatively high burden (NRSI DALYs > 0.5) is not a scientific priority for a region (NRSI Pubs > 0.0). The only diseases that could be seen as “under-researched” are the ones that are between 0 and 0.5 in the x-axis (NRSI DALYs) and between -0.5 and 0 in the y-axis (NRSI Pubs) in every region. Some examples include “diabetes” in Southern Africa and “diarrhoeal diseases” in Eastern Africa.

It is worth noting that the indexes in Fig. 5.2 were calculated based on the comparisons with world-wide average levels, which presents the relative positions of one disease type in one certain region. Besides this method, another way to plot disease burden as a function of research production is to display the total disease burden (x-axis) versus the total number of publications (y-axis) in each disease by African region (in logarithms). In Fig. 5.A.2 in the appendix we do that analysis for our four African regions. All graphs show a slight positive correlation between disease burden and scientific output. This means that, on average, the higher the disease burden of a disease in a region, the higher the scientific production by researchers in that specific region. An outlier in these graphs is HIV/AIDS in Southern Africa with around 43% of the total disease burden of the region in 2010 being related to this disease and 19% of total medical scientific output. South Africa accounts for a third of all new HIV infections in Southern Africa, and it has the most high profile HIV epidemic in the world, with an estimated 7.1 million people living with HIV in 2016 (UNAIDS 2017). The region has made progress in reducing the disease burden associated with this disease since 2005, but it is still a huge issue. It is argued that in South Africa important research is being done on developing HIV vaccines, yet little attention is being paid to planning approaches to reduce the high rates of infection among young women, the primary driver of HIV epidemic in the country (Karim & Karim 2010). Finally, it is important to note the high correlation coefficients between Eastern Africa specialisations and West & Central Africa specialisations. These two

³⁷ <http://www.who.int/hiv/topics/tb/en/>

regions are highly scientifically specialised in diseases where they have relatively high disease burden and have very few publications in diseases that they have relatively low disease burden. Since scientific institutions in many of these African countries suffer from specific challenges such as poor conditions for research personnel, heavy teaching loads, inability to mentor young scholars, inadequate infrastructure and lack of funding (Mouton 2008; Sawyerr 2014), a potential reason for this match between disease burden and medical research may be due to efficient resource allocation or the influence of external funders (Ndounga Diakou et al. 2017).

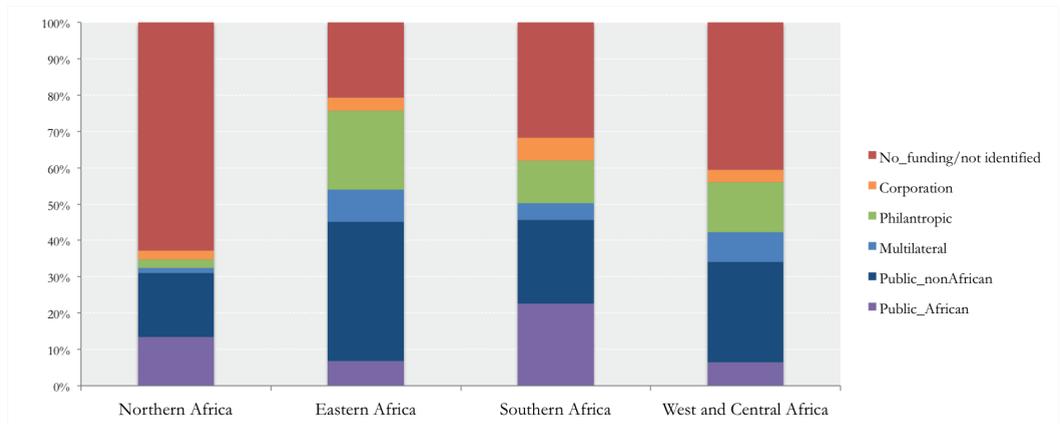
A combination of Fig 5.2 and Fig. 5.A.2 indicates that medical research in South Africa and North Africa has been better associated with their absolute level of disease burden, but not with the relative level compared with the global average.

5.4.2. Medical research funding

In the last thirty years, several international funders of health research and development (R&D) such as the World Health Organization (WHO), National Institutes of Health (NIH), the European Union (EU), the Bill & Melinda Gates Foundation (Gates Foundation) and the Wellcome Trust, as well as institutions supported by government funding, have all embarked on initiatives to help improve the research capacity, research environment and provide institutional support across the continent (Whitworth et al. 2008). In this section, we explore the different contributions of various funding organisations to the medical research in Africa.

The share of the total medical research funded by at least one funding institution in each African region, across the disease categories, ranged from 37% in Northern Africa to 79% in Eastern Africa (Fig. 5.3).

Figure 5.3. Share of publications by funding type (2011-2015)



Note: Own calculation based on WoS

The highest share of research funding in all regions is from public non-African funding institutions (e.g. NIH, EU, USAID, Medical Research Council (UK)), followed by Philanthropic funding institutions (e.g. Wellcome Trust, Gates Foundation) that make particularly relevant contributions in Eastern African countries. Public African funding institutions have higher shares of funding in Southern Africa (e.g. National Research Foundation (ZA), Medical Research Council (ZA)) and Northern African (e.g. Tunisian Government, Egyptian Government). In all regions the contribution of corporations is relatively small but pharmaceuticals like GlaxoSmithKline, Pfizer and Novartis were acknowledged in 328, 302 and 238 publications between 2011 and 2015 respectively. Multilateral funding institutions like WHO, EDCTP and the World Bank are mostly funding medical research in Eastern African countries and West & Central African countries (See Tables 5.A.2 and 5.A.3 in appendix for more info about the major contributors in each funding category and region). While doing this analysis we found a significant overlap between public non-African and multilateral funding (50% of all publications with multilateral funding also acknowledge a public non-African funder). Since theoretically, it is difficult to distinguish what are the different reasons that lead public non-African and multilateral institutions to fund African health research, further analysis in this chapter will consider the two categories the same category (public non-African).

We were also interested to know if each funder (or group of funders) supports research in specific diseases. Table 5.A.4, in the appendix, highlights the top 5 diseases that the top 10 funders in Africa fund more between 2011 and 2015 by number of publications. The key finding is that “Parasitic and vector diseases”, “HIV/AIDS” and “Tuberculosis” are a priority for every top10 funder. These results are in line with Chapman et al. (2017) that also found that three diseases – HIV/AIDS, malaria and tuberculosis – collectively received more than two-thirds (\$2,247m, 70%) of all global funding for neglected disease R&D in 2016.

The only funders that are not so biased towards these three diseases and have a more horizontal distribution of their funding are the National Research Council (ZA), Medical Research Council (ZA) and Tunisian Government. These are all African funders that may have different priorities than international organisations. It is, however, important to notice the absence of Public African funding in Eastern Africa and West & Central Africa. In these regions research is heavily dependent upon foreign funds.

Interestingly, Gates Foundation funds more than 8% of African research on “neonatal conditions” which is the disease with the highest absolute disease burden in Eastern Africa and West & Central Africa. It has been argued that Gates Foundation investment has tried to balance the public sector focus on basic research (Chapman et al. 2017). According to G-finder data, it has provided 55% of all funding to neglected diseases in the world to product development partnerships and 47% of all funding for platform technologies between 2007 and 2016.

The next Fig. 5.4, summarises all this information in four graphs that display the share of medical research funded by each group of funders in each African region. Each radar graph is ordered by disease burden share in each region and only shows diseases that represent more than 0.1% of the disease burden in the region.

includes diseases such as malaria, dengue, trachoma, yellow fever, rabies, chagas disease, among others. Malaria is by far the condition that leads to higher disease burden in this category. According to Head et al., (2017) global research funding for malaria in sub-Saharan Africa are mostly allocated to Tanzania, Uganda, Kenya, Malawi, Ghana, and Nigeria. These are locations with a track record of success in similar projects where it is perceived that investments will make a positive difference and where any research will be feasible. The research supported by corporations is substantially higher in absolute terms in Southern Africa, and in areas such as diabetes, cardiovascular diseases and respiratory infections/diseases.

5.4.3. Econometric analysis

In this section, we present the results of the estimation of model (3).³⁸ We pooled data from the period 2011-2015 for research specialisation and 2010 for disease burden specialisation. After constraining our database to diseases in regions with a minimum of 50 publications (to avoid outliers when computing the NRSI), we end up with 103 observations from our four African regions in the 28 diseases. We find that the Eastern African region is the region where the disease burden is mostly associated with their research specialisation, even when we control for previous research specialisation. The region is highly dependent on international research collaboration (Confraria & Godinho 2015) and, as we have seen in Fig. 5.3, it is also the region which is most dependent on funding from non-African partners and philanthropic institutions. Therefore it is interesting to notice that it is the region where the disease burden and health research specialisation show a higher alignment.

In Table 5.1 we can also observe that the disease burden specialisation in Southern Africa and West & Central Africa are also positively and significantly associated with their research specialisation in model specifications I, II and III. However, when we include in our model, the variable previous scientific specialisation in our equations, the significance disappears. This means that the association between these two dimensions may be derived mostly from the existence of previous scientific capabilities in those areas and not from the awareness to the disease burden in their region.

In this regard, it is important to note that there is a high correlation between NSRI in 2011-2015 and previous scientific specialisation in 2006-2010 in all regions (around 98%). Scientific activities are dominated by strong path-dependencies. If one country as scientists that are involved in a

³⁸ As a consistency check, we also estimated a regression using the number of publications (log) in each disease and region as the dependent variable and number of DALYs in each disease and region as the main independent variable. Those results are similar to the ones presented in Table 5.1, and are available in the appendix (Table 5.A.5).

certain type of research it is very likely that they will continue to do their research in that area.

Finally, we also found that all African regions also seem to be specialised in areas where they have higher levels of international collaboration. Since the research in most African countries is highly dependent on international research collaboration and international research funding, this was an expected result.

Table 5.1. Regression analysis. Match between disease burden specialisation and research specialisation.

Ind. Variables	NSRI_Pubs_11.15				
	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)
NRSI_DALYs_2010	0.76*** (0.10)				
NRSI_DALYs_2010*Eastern_Africa		1.21*** (0.14)	1.20*** (0.14)	1.08*** (0.15)	0.11** (0.046)
NRSI_DALYs_2010*Northern_Africa		-0.016 (0.29)	-0.047 (0.28)	-0.37 (0.29)	-0.042 (0.053)
NRSI_DALYs_2010*Southern_Africa		0.46* (0.25)	0.53** (0.27)	0.56*** (0.21)	0.065* (0.039)
NRSI_DALYs_2010*West&Central_Africa		0.90*** (0.11)	0.91*** (0.11)	0.71*** (0.14)	0.048 (0.055)
Int_collab_11_15 (%)				1.14*** (0.30)	0.16** (0.081)
NSRI_Pubs_06.10					0.89*** (0.022)
Constant	0.076** (0.031)	0.058* (0.029)	0.029 (0.062)	-0.90*** (0.26)	-0.078 (0.064)
Regional dummies	No	No	Yes	Yes	Yes
Observations	103	103	103	103	103
R-squared	0.393	0.510	0.516	0.571	0.967

Note 1: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Note 2: In estimations 3-5, the regression model was computed controlling for four regions: Eastern Africa, Northern African, Southern Africa and West & Central Africa

In Table 5.A.5, in the appendix, we also run a set of regressions that run the same model but, instead of using normalize specialisation indexes as dependent and independent variables, we use the logarithmic number of publications of each funder in a disease and region, as the dependent variable; and logarithmic number of DALYs as the main independent variable. The results are similar. When controlling for all the other variables, the only region where there is a positive and significant association between disease burden and research specialisation is Eastern Africa.

To assess the extent to which a higher share of funding from international donors in specific scientific areas/diseases is associated with higher disease burden specialisation, we computed an additional set of regressions that have as dependent variables the normalized relative specialisation index of a certain funder in a specific disease. We divided all funding institutions into five types: Public African, public non-African (includes public non-African and multilateral), philanthropic, corporation and non-funded (or non-identified), and we used the whole counting method (if a publication has two different types of funding institutions in their acknowledgments we counted one publication for each funding institution/disease).

Table 5.2. Regression analysis. Match between disease burden specialisation of a region and research specialisation of a specific funder group.

Ind. Variables	AF_NRSI_11.15		NAF_NRSI_11.15		PF_NRSI_11.15		CF_NRSI_11.15		NF_NRSI_11.15	
	OLS (1)	OLS (2)	OLS (1)	OLS (2)	OLS (1)	OLS (2)	OLS (1)	OLS (2)	OLS (1)	OLS (2)
NRSI_DALYs_2010*	1.06***	0.57**	1.26***	0.33***	1.40***	0.70***	1.16***	0.79***	0.86***	-0.056
Eastern_Africa	(0.16)	(0.22)	(0.17)	(0.089)	(0.13)	(0.15)	(0.15)	(0.15)	(0.16)	(0.081)
NRSI_DALYs_2010*	-0.44	-0.39	-0.50*	-0.36***	-0.65**	-0.68**	-0.084	-0.49	0.17	0.071
Northern_Africa	(0.41)	(0.27)	(0.29)	(0.12)	(0.32)	(0.28)	(0.29)	(0.30)	(0.26)	(0.11)
NRSI_DALYs_2010*	0.51*	0.29*	0.51	0.19**	0.66**	0.47***	0.60**	0.42***	0.58**	0.055
Southern_Africa	(0.29)	(0.15)	(0.35)	(0.073)	(0.32)	(0.17)	(0.24)	(0.095)	(0.26)	(0.055)
NRSI_DALYs_2010*	0.74***	0.26	1.14***	0.17*	1.25***	0.56***	0.86***	0.35**	0.73***	-0.021
West&Central_Africa	(0.18)	(0.18)	(0.11)	(0.097)	(0.11)	(0.13)	(0.17)	(0.17)	(0.13)	(0.084)
Int_collab_11_15 (%)		0.32		0.44**		0.90***		0.91*		0.24*
African_Funding		0.47***		(0.18)		(0.32)		(0.48)		(0.14)
NRSI_09.10		(0.091)								
Non-African_Funding				0.72***						
NRSI_09.10				(0.053)						
Philanthropic_Funding						0.41***				
NRSI_09.10						(0.083)				
Corporation_Funding							0.38***			
NRSI_09.10							(0.066)			
Non-Funded									0.86***	
NRSI_09.10									(0.049)	
Constant	-0.015	-0.18	-0.035	-0.34**	-0.026	-0.70***	0.029	-0.58	-0.41***	-0.36***
	(0.077)	(0.31)	(0.072)	(0.15)	(0.067)	(0.27)	(0.076)	(0.40)	(0.060)	(0.12)
Regional effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	101	101	101	101	101	101	101	101	101	101
R-squared	0.367	0.584	0.503	0.899	0.544	0.728	0.375	0.624	0.651	0.934

Note 1: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Note 2: The regression model was computed controlling for four regions: Eastern Africa, Northern African, Southern Africa and West & Central Africa

Comparisons of results from the five models in Table 5.2 enable us to observe a set of findings: First, the diseases in which public non-African and philanthropic institutions fund relatively more are the ones, on average, that have higher disease burden specialisation, in Eastern Africa, Southern Africa and Western & Central Africa. As we have seen in Fig. 5.3, these three regions are highly dependent on international research funding. Therefore, this finding is significant since it shows that these international donors, on average, are funding research on diseases that are relevant to these African regions. This result is in line with Breugelmans et al. (2015) that argues that the increasing trend in research output on poverty-related and neglected infectious diseases from sub-Saharan African researchers suggests that sustained support and funding by non-African governments, development partners, private foundations and public partnerships are paying off. In Northern Africa, there is a negative association between disease burden specialisation and research funding specialisation in all funding categories. These results mirror the results found in Table 5.1 that showed that medical research in Northern Africa is not highly associated with their disease burden.

Second, Eastern Africa is the only region where exists a positive association, on average, between disease burden specialisation and public African funding specialisation. Third, unexpectedly publications funded by corporation funding seem to be positively associated on average with disease burden in all regions except for Northern Africa. Fourth, there is no clear association between research that is not funded by a specific institution and disease burden specialisation. The only driver of research that has no funding acknowledgments seems to be previous specialisation on that topic (lagged dependent variable).

Finally, we should also note that we controlled for the level of international collaboration and previous share of research funding (2009-2010) for all the dependent variables. As expected, we only found a significant positive association between the intensity of international collaboration and research on diseases that receive relatively more funding from public non-African and philanthropic research institutions. This may happen because the research that is done in international collaboration usually is supported by external research funding. As for the previous funding specialisation in a specific disease, there is a positive association between our lagged variable and all dependent variables. This indicates the path dependence that exists in research funding in general. Investing requires confidence on the part of the investor that they will see a return on their investment. In environments where the logistics for research might be complex and challenging, the inclination is to fund governments and institutions with a track record of success (Head et al. 2017).

5.5. Conclusion

In this chapter we evaluated the alignment between the medical research effort and the burden of disease across four African regions. Within each region, we estimated the research and disease burden specialisation (compared to the world specialisation levels) across 28 diseases. We found that the region with higher positive association between disease burden and research effort is Eastern Africa. Northern Africa is the region where these two dimensions are less aligned. In the other two regions (West & Central Africa and Southern Africa) our estimations show that the research specialisation is associated with the disease burden specialisation when we don't control for previous scientific specialisation.

All African regions have a relatively small scientific productivity in medical related areas when compared to the rest of the World (specially Eastern and Western & Central Africa). Therefore the common rationale is that these regions should use their limited resources to study diseases that are relevant to their health needs. What we found is that in sub-Saharan Africa most diseases with high disease burden are also the ones with relatively more research effort. On the other side, the only disease field that we found to be more researched than what we would expect is parasitic and vector diseases in Southern Africa.

These findings are interesting for two main reasons:

One, it has been argued that there are substantial misalignments, at the global and local levels, between research efforts and WHO estimates of health burden for a given disease (e.g. Evans et al., 2014; Rafols & Yegros, 2017). What we found is that while this may be true at the global level (high-income countries perform most of their medical research on diseases that are not the ones with higher global disease burden), Sub-Saharan African researchers are performing research that is relevant for their regional health needs.

Second, some authors also argue that researchers from high-income countries secure most of the funding for global health research projects in low-income regions, and often dictate the research agenda which lead to inappropriate projects unrelated to local research needs, and derive conclusions that do not have any direct local benefit (Binka 2005; Gaillard 1994). The results from this chapter seem to contradict this idea. What we found is that most international research funders (public non-African and philanthropic) support research on HIV/AIDS, tuberculosis and parasitic and vector diseases, which are diseases that generate a big share of disease burden in sub-Saharan Africa. What our regression results seem to show is that, on average, high levels of dependence on international donors are not necessarily associated with less alignment between local health needs and local medical research effort.

This does not mean that because international funders seem to be doing the “right” job and funding medical research on diseases that are relevant for the local communities, African countries do not need stronger scientific and institutional capacity. It is well known that substantial advantages exist in investment in local research, particularly with regard to ownership of the results, trust, inter-sector sharing of expertise between researchers and policy makers, and increased contextualisation of findings. Local knowledge from the people most directly affected by an issue is crucial to finding an innovative way of addressing it. Therefore, creating schools of public health and other institutions to train quality scientists in public health should continue to be a priority, as many African countries have few or no institutions that can provide proper training in public health research. Collaborative solutions and establishing funding partnerships between countries can also be a possibility for countries with very little resources.

Our study has limitations and the results must be interpreted with caution since publications in WoS (or DALYs) are imperfect estimates of research efforts (health needs) in a specific disease and country. First, measurement of priorities in medical research with scientific publications associated to certain diseases is not straightforward because there is some health research related to health education approaches, beliefs related to health and prevention, quality and financing of healthcare, that is important for health outcomes and do not necessarily derive from research on certain diseases. Second, since scientific production and disease burden change over time, future studies should conduct a dynamic analysis of DALYs and publications to understand how the two dimensions evolve together. Third, future studies should also analyse the extent to which the research that is funded is actually used to contribute to health action.

5.6. Appendix

Table 5.A.1. Keywords queries for each disease

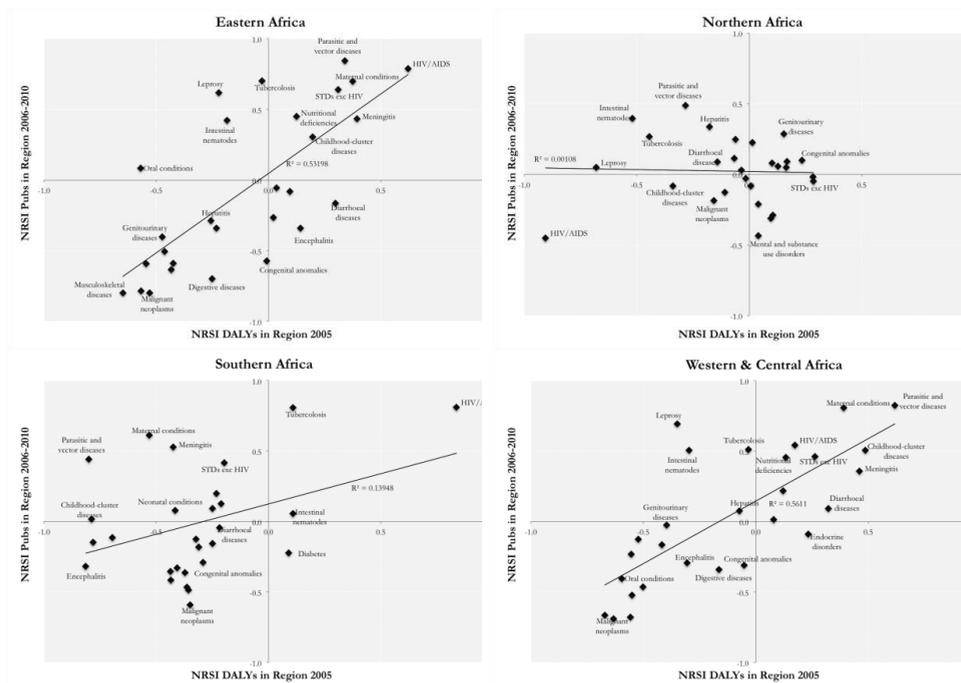
Diseases	Query
Infectious and parasitic diseases	
Tuberculosis	"tuberculosis" OR "tuberculosis" OR "tubercle bacillus" OR "tuberculin" OR "tb infection" OR "pulmonary tb" OR "extrapulmonary tb"
STDs excluding HIV	"Syphilis" OR "Chlamydia" OR "Gonorrhoea" OR "Trichomoniasis" OR "Genital herpes"
HIV/AIDS	("hiv/aids" OR "human immunodeficiency virus" OR "human immuno-deficiency virus" OR "acquired immunodeficiency syndrome" OR "acquired immuno-deficiency syndrome" OR "hiv infection") NOT (feline OR simian)
Diarrhoeal diseases	"Diarrhoeal" OR "Diarrhea" OR "escherichia coli" OR "E. Coli" OR "vibrio cholerae" OR "shigellosis" OR "shigella" OR "Giardia" OR "cryptosporidium" OR "rotavirus"
Childhood-cluster diseases	"Whooping cough" OR "pertussis" OR "Diphtheria" OR "diphtheriae" OR "Measles" OR "rubeola" OR "neonatal tetanus" OR "tetanus neonatal" OR "mumps virus" OR "Polioomyelitis"
Meningitis	"Meningitis" OR "meningitidis" OR "neisseria pneumoniae" OR "cryptococ*" OR "meningococcus"
Encephalitis	"Encephalitis"
Hepatitis	"Hepatitis"
Parasitic and vector diseases	("Malaria" OR "plasmodium" OR "anopheles" OR "black water fever") NOT "physarum") OR "Human african trypanosomiasis" OR "sleeping sickness" OR "trypanosom human" OR "Chagas disease" OR "American Trypanosomiasis" OR "Trypanosoma cruzi" OR "Trypanosoma brucei" OR "Schistosomiasis" OR "bilharzia" OR "Schistosoma mansoni" OR "Schistosoma haematobium" OR "Schistosoma intercalatum" OR "Schistosoma japonicum" OR "Schistosoma mekongi" OR "Leishmaniasis" OR "Leishmania" OR "phlebotomine" OR "psychodidae" OR "kalaazar" OR "kala-azar" OR "kala azar" OR "sand fly" OR "sandflies" OR "sand flies" OR "filariasis" OR "elephantiasis" OR "wuchereria" OR "brugia malayi" OR "Onchocerciasis" OR "Onchocerciasis" OR "river blindness" OR "onchocerca volvulus" OR "Cysticercosis" OR "taeniasis" OR "Taenia solium" OR "Echinococcosis" OR "hydatid disease" OR "echinococcus" OR "dengue" OR "aedes aegypti" OR "aedes albopictus" OR "Trachoma" OR "chlamydia trachomatis" OR "Yellow fever" OR "Rabies" OR "zika virus" OR "Flavivirus" OR "chikungunya" OR "Lassa fever" OR "Ebola" OR "Haemorrhagic Fever" OR "typhoid" OR "loiasis" OR "cestodes"
Intestinal nematode infections	("fasciolosis" OR "fascioliasis" OR "distomatosis" OR "fasciola hepatica" OR "fasciola gigantica" OR "distomatosis") NOT "cattle") OR "dracunculiasis disease" OR "guinea-worm disease" OR "guinea worm disease" OR "dracunculus medinensis" OR "salmonella" OR "paratyphoid fever" OR "ancylostomiasis" OR "strongyloidiasis" OR "Ascariasis" OR "Trichuriasis" OR "Hookworm*" OR "heminth*" OR "hook-worm*" OR "hook worm*" OR "ascaris lumbricides" OR "trichuris trichiura" OR "geohelminth*" OR "necatoramericanus" OR "necator americanus" OR "necatoriasis" OR "ancylostoma duodenale" OR "ancylostoma-duodenale" OR "clonorchiasis" OR "opisthorchiasis" OR "paragonimiasis"
Leprosy	"Leprosy" OR "hansen disease" OR "mycobacterium leprae"
Respiratory infections & diseases	"respiratory infectio*" OR "Asbestosis" OR "rheumatic fever" OR "Haemophilus Influenzae" OR "lung absces" OR "bronchitis" OR "Streptococcus pneumoniae" OR "pneumonia" OR "Moraxella catarrhalis" OR "Klebsiella pneumonia" OR "tonsillitis" OR "rhinitis" OR "sinus infection" OR "sinusitis" OR "rhinosinusitis" OR "rhinopharyngitis" OR "nasopharyngitis" OR "pharynx inflammation" OR "hypopharynx inflammation" OR "uvula inflammation" OR "tonsils inflammation" OR "pharyngitis" OR "epiglottitis" OR "laryngitis" OR "laryngotracheitis" OR "tracheitis" OR "Otitis media" OR "Respiratory diseas*" OR "Chronic obstructive pulmonary disease" OR "Asthma" OR "emphysema" OR "Laryngotracheitis" OR "Epiglottitis" OR "Bacterial tracheitis"
Maternal conditions	"maternal death" OR "maternal mortality" OR "pregnancy infection*" OR "abortion care" OR "unsafe abortion" OR "childbirth severe bleeding" OR "childbirth infection*" OR "Placental abruptio*" OR "placenta praevia"
Neonatal conditions	"Preterm birth" OR "Birth asphyxia" OR "birth trauma" OR "Neonatal sepsis" OR "neonatal infection*" OR "Gastroschisis" OR "Jaundice" OR "Necrotizing enterocolitis" OR "Persistent pulmonary hypertension of the newborn" OR "Intrauterine growth restriction" OR "Bronchopulmonary dysplasia" OR "infant apnea" OR "infant respiratory distress syndrome" OR

	"asphyxia at birth" OR "anaemia in neonates" OR "neonatal alloimmune thrombocytopenia" OR "bronchopulmonary dysplasia" OR "cardiac failure in neonates" OR "hyaline membrane disease" OR "hypocalcaemia in neonates" OR "hypoglycaemia of the newborn" OR "hyponatraemia in neonates" OR "hypothermia in neonates" OR "intestinal obstruction in neonates" OR "pulmonary interstitial emphysema"
Nutritional deficiencies	"Nutritional deficient*" OR "Protein energy malnutrition" OR "Protein-energy malnutrition" OR "Iodine deficiency" OR "Iron-deficiency anaemia" OR "Nutritional deficiency*" OR "Thiamine deficiency" OR "vitamin B-1 deficiency" OR "Niacin deficiency" OR "vitamin B-3 deficiency" OR "vitamin B-9 deficiency" OR "Folate deficiency" OR "Cobalamin deficiency" OR "vitamin B-12 deficiency" OR "Vitamin D deficiency" OR "Calcium deficiency" OR "marasmus" OR "Kwashiorkor" OR "Marasmic-kwashiorkor" OR "nutritional oedema" OR "severe acute malnutrition" OR "moderate acute malnutrition" OR "Vitamin A deficiency"
Malignant neoplasms	"malignant neoplasm*" OR "Mouth cancer" OR "oropharynx cancer" OR "Lip cavity" OR "oral cavity" OR "Nasopharynx" OR "Oesophagus cancer" OR "Stomach cancer" OR "Colon cancer" OR "rectum cancer" OR "Liver cancer" OR "Pancreas cancer" OR "Trachea cancer" OR "bronchus cancer" OR "lung cancer" OR "Melanoma" OR "skin cancer" OR "Breast cancer" OR "Cervix uteri cancer" OR "Corpus uteri cancer" OR "Ovary cancer" OR "Prostate cancer" OR "Testicular cancer" OR "Kidney cancer" OR "Bladder cancer" OR "Brain cancer" OR "nervous system cancer" OR "Gallbladder cancer" OR "biliary tract cancer" OR "Larynx cancer" OR "Thyroid cancer" OR "Mesothelioma" OR "Lymphoma*" OR "multiple myeloma" OR "Leukaemia"
Diabetes mellitus	diabete*
Endocrine, blood, immune disorders	"Endocrine disorder*" OR "blood disorder*" OR "immune disorder*" OR "Glucocorticoid deficiency" OR "Glucose intolerance" OR "Goiter" OR "Hyperparathyroidism" OR "Hypothyroidism" OR "Hypoglycemia" OR "Hypoparathyroidism" OR "Hypothyroidism" OR "Mineralocorticoid deficiency" OR "Pseudohypoparathyroidism" OR "Thyroid cyst" OR "Thyroid nodule" OR "Thyroiditis" OR "Acidosis" OR "Alkalosis" OR "Amyloidosis" OR "Thalassaemias" OR "Sickle cell disorder" OR "trait disorder" OR "haemoglobinopathies" OR "haemolytic anaemia" OR "Cystic fibrosis" OR "Dysmetabolic syndrome" OR "Hemochromatosis" OR "Hyperbilirubinemia" OR "Hypercalcemia" OR "Hypercholesterolemia" OR "Hyperkalemia" OR "Hyperlipidemia" OR "Hyponatremia" OR "Hypertiglyceridemia" OR "Hypocalcemia" OR "Hypokalemia" OR "Hyponatremia" OR "Hypovolemia" OR "Magnesium disorder*" OR "Obesity hypoventilation syndrome" OR "Porphyria" OR "Renal osteodystrophy" OR "Anemia" OR "Coagulation defects" OR "Eosinophilia" OR "Hemophilia" OR "Hypercoagulable state" OR "Idiopathic thrombocytopenic purpura" OR "Leukocytopenia" OR "Leukocytosis" OR "Lymphadenitis" OR "Neutropenia" OR "Polycythemia vera" OR "Sickle cell" OR "Thrombocytopenia"
Mental and substance use disorders	"mental disorder*" OR "substance disorder*" OR "behavioural disorder*" OR "Agoraphobia" OR "Anorexia nervosa" OR "Antisocial personality disorder" OR "Anxiety state" OR "Attention deficit" OR "hyperactivity" OR "Bipolar disorder" OR "Borderline personality disorder" OR "Brixiism" OR "Bulimia nervosa" OR "Conduct disorder" OR "Conversion disorder" OR "Delirium tremens" OR "Dementia" OR "Depression disorder" OR "Depressive disorder" OR "Depressive psychosis" OR "Dyspareunia" OR "Encopresis" OR "Enuresis" OR "Explosive personality disorder" OR "Fluency disorder" OR "Generalized anxiety disorder" OR "Hysteria disorder" OR "Hysterical psychosis" OR "Insomnia" OR "sleep disorder" OR "Intellectual disability*" OR "Neurosis" OR "Neurotic depression" OR "Obsessive-compulsive disorder" OR "Panic disorder" OR "Paranoid reaction" OR "Personality disorder" OR "Post-traumatic stress disorder" OR "Premature ejaculation" OR "Psychosis" OR "Schizoaffective" OR "Schizophrenia" OR "Sleep disorder" OR "Somatization disorder" OR "Somnambulism" OR "Suicidal ideation" OR "Alcohol abuse" OR "Alcoholism" OR "Amphetamine dependence" OR "Cannabis abuse" OR "Cannabis dependence" OR "Cocaine abuse" OR "Cocaine dependence" OR "Drug abuse" OR "Drug withdrawal" OR "Drug-induced paranoia" OR "Opioid abuse" OR "Opioid dependence" OR "Tobacco abuse" OR "dysthymia" OR "opioid disorder" OR "cocaine disorder" OR "amphetamine disorder" OR "cannabis disorder" OR "panic attack" OR "Social anxiety disorder" OR "separation anxiety disorder" OR "selective mutism" OR "eating disorder" OR "Anorexia Nervosa" OR "Bulimia Nervosa" OR "Binge Eating Disorder" OR "Muscle dysmorphia" OR "autism" OR "asperger syndrome" OR "autistic" OR "Attention deficit" OR "hyperactiv* syndrome" OR "Conduct disorder" OR "Idiopathic intellectual disability" OR "mental retardation"
Neurological conditions	"Bell's palsy" OR "Blepharospasm" OR "Carpal tunnel" OR "Cerebral aneurysm" OR "Cerebral artery occlusion" OR "Cerebral edema" OR "Cerebral palsy" OR "Cognitive impairment" OR

	"Encephalopathy" OR "Epilepsy" OR "Guillain-Barré" OR "Hemiplegia" OR "Hydrocephalus" OR "Migraine" OR "Morton's neuroma" OR "Multiple sclerosis" OR "Myasthenia gravis" OR "Narcococpsy" OR "Neuralgia" OR "Neuropathy" OR "Parkinsonism" OR "Phantom limb" OR "Post-concussion syndrome" OR "Postherpetic neuralgia" OR "Pseudotumor cerebri" OR "Reflex sympathetic" OR "Restless legs syndrome" OR "Reye's syndrome" OR "Sciatica" OR "Subarachnoid hemorrhage" OR "Subdural hemorrhage" OR "Thoracic outlet syndrome" OR "Tic disorder" OR "Tourette's disorder" OR "Trigeminal neuralgia" OR "Alzheimer*" OR "dementia" OR "chronic neurodegenerative" OR "Neurodegeneration" OR "Parkinson disease" OR "parkinsonian syndrome" OR "epileptic" OR "motor neuron disease" OR "huntington's disease"
Sense organ diseases	"otitic barotrauma" OR "Cerumen impaction" OR "Eustachian tube dysfunction" OR "Hearing loss" OR "viral Labyrinthitis" OR "Ménière's disease" OR "Nystagmus" OR "Otaglia" OR "Otitis externa" OR "Otitis media" OR "Presbycusis" OR "Tinnitus" OR "Vertigo" OR "Anisocoria" OR "Blepharitis" OR "eye cataract*" OR "Chalazion" OR "Conjunctivitis" OR "Corneal abrasion" OR "Corneal edema" OR "Corneal ulcer" OR "Diplopia" OR "Dry eye syndrome" OR "Esotropia" OR "Glaucoma" OR "Hyphema" OR "Iritis" OR "cyclitis" OR "Lid lag" OR "Macular degeneration" OR "Papilledema" OR "Pterygium" OR "Retinal detachment" OR "Retinopathy" OR "Scotoma" OR "hordeolum" OR "Subconjunctival hemorrhage" OR "Visual disturbance" OR "Visual field defect" OR "Visual loss" OR "Uncorrected refractive errors" OR "sense organ disorder*" OR "cholesteatoma"
Cardiovascular diseases	"Cardiovascular diseases*" OR "Atrial fibrillation" OR "Atrial flutter" OR "Atrioventricular block" OR "Bundle branch block" OR "Long QT syndrome" OR "Sick sinus syndrome" OR "Sinoatrial heart block" OR "Sinus bradycardia" OR "paroxysmal tachycardia" OR "Angina pectoris" OR "artery bypass graft" OR "autologous vein bypass graft" OR "native coronary artery" OR "Cardiac arrest" OR "Cardiac contusion" OR "Cardiomyopathy" OR "Chronic ischemic heart disease" OR "Endocarditis" OR "Heart failure" OR "Heart valve" OR "Kawasaki disease" OR "Myocarditis" OR "Pericarditis" OR "Prinzmetal angina" OR "Pulmonary heart disease" OR "Rheumatic heart disease" OR "Aortic aneurysm" OR "Aortic dissection" OR "Carotid sinus syndrome" OR "Deep vein thrombosis" OR "Esophageal varices" OR "heart hypertension" OR "Hypertensive heart" OR "Hypotension" OR "Intermittent claudication" OR "Peripheral vascular disease" OR "Phlebitis" OR "Polyarteritis nodosa" OR "Postmastectomy lymphedema" OR "Raynaud's syndrome" OR "Thrombophlebitis" OR "Transient ischemic attack" OR "Varicose veins" OR "Venous embolism" OR "Venous insufficiency" OR "Wegener's granulomatosis" OR "Ischaemic heart disease" OR "Ischaemic stroke" OR "Haemorrhagic stroke" OR "myocardial infarction"
Digestive diseases	"Digestive diseases*" OR "Achalasia" OR "cardiospasm" OR "Anal spasm" OR "Angiodysplasia" OR "Aphthous ulcer" OR "Appendicitis" OR "Barrett's esophagitis" OR "Cholangitis" OR "Cholecystitis" OR "Cholelithiasis" OR "Cirrhosis" OR "Crohn's disease" OR "Diverticulitis of colon" OR "Duodenal ulcer" OR "Dyspepsia" OR "Edentulism" OR "Esophageal stricture" OR "Esophageal stenosis" OR "Esophagitis" OR "Fatty liver" OR "Gallbladder disease" OR "Gastric ulcer" OR "Gastritis" OR "Gastroenteritis" OR "Gastroesophageal reflux" OR "Gastroparesis" OR "Glossitis" OR "Hemorrhoids" OR "Impaction of intestine" OR "colostomy" OR "enterostomy" OR "Irritable bowel syndrome" OR "Ischemic bowel disease" OR "Leukoplakia" OR "Liver disease" OR "Mechanical complication of ostomy" OR "Pancreatitis" OR "Parotitis" OR "Peptic ulcer" OR "Periodontitis" OR "Ulcerative colitis" OR "duodenitis" OR "Paralytic ileus" OR "intestinal obstruction" OR "Inflammatory bowel disease"
Genitourinary diseases	"Genitourinary" OR "Breast lump" OR "Fibroadenosis" OR "Fibrocystic disease" OR "Galactorrhea" OR "Gynecomastia" OR "Mastitis" OR "Mastodynia" OR "Amenorrhea" OR "Menopausa*" OR "Metrorrhagia" OR "Mittelschmerz" OR "Premenstrual tension syndrome" OR "postmenopausal atrophic" OR "vulvo atrophy" OR "vaginal atrophy" OR "Bartholin abscess" OR "Bartholin cyst" OR "Cervical polyp" OR "Cervicitis" OR "Corpus luteum cyst" OR "Cyst of ovary" OR "Cystocele" AND "midline" OR "Dyspareunia" OR "Endometrial hyperplasia" OR "Endometriosis" OR "Fibroid uterus" OR "leiomyoma" OR "Leukorrhea" OR "Ovarian failure" OR "Pelvic inflammatory disease" OR "uterine prolapse" OR "Rectocele" OR "Urethrocele" OR "Uterus hypertrophy" OR "Vaginismus" OR "Vaginitis" OR "vulvitis" OR "Vulvodynia" OR "Atrophy of testis" OR "Balanitis" OR "BPH/LUTS" OR "Hematospermia" OR "Hydrocele" OR "Orchitis" OR "epididymitis" OR "Phimosis" OR "Priapism" OR "Prostatitis" OR "Spermatocele" OR "Testicular hypofunction" OR "Torsion of testis" OR "nongonococcal Urethritis" OR "Varicocele" OR "Atony of bladder" OR "Bladder hypertonicity" OR "Bladder neck obstruction" OR "kidney Calculus" OR "ureter Calculus" OR "urinary Calculus" OR "Cystitis" OR "Glomerulonephritis" OR "Hematuria" OR "Hydronephrosis" OR "Kidney disease" OR "Nephrotic syndrome" OR "Proteinuria" OR "Pyelonephritis" OR "Renal

	failure" OR "urethral stricture" OR "Urethral syndrome" OR "Urinary obstruction" OR "Urinary tract infection" OR "Vesicoureteral" OR "prostatic hyperplasia" OR "Urolithiasis" OR "Gynecologic* disease" OR "infertility"
Skin diseases	"skin diseases*" OR "Acne" OR "Actinic keratosis" OR "Alopecia" OR "Cellulitis" OR "Contact dermatitis" OR "Cradle cap" OR "Dermatitis" OR "Dermatophytosis" OR "Diaper rash" OR "Eczema" OR "Erythema multiforme" OR "Erythema nodosum" OR "Hidradenitis suppurativa" OR "Hirsutism" OR "Impetigo" OR "Ingrown nail" OR "Keloid scar" OR "Lichen planus" OR "Lymphadenitis" OR "Onychomycosis" OR "Paronychia" OR "Pityriasis rosea" OR "Pressure ulcer" OR "Pruritus" OR "Psoriasis" OR "Sebaceous cyst" OR "Seborrheic dermatitis" OR "Seborrheic keratosis" OR "Solar radiation dermatitis" OR "Stevens-Johnson syndrome" OR "Tinea cruris" OR "Tinea pedis" OR "Tinea versicolor" OR "Urticaria" OR "Vitiligo"
Musculoskeletal diseases	"Musculoskeletal disease" OR "Musculoskeletal disorder" OR "Musculoskeletal pain" OR "Arthropathy" OR "Dermatomyositis" OR "Eosinophilia myalgia syndrome" OR "Fibromyalgia" OR "Myositis ossificans" OR "Osteoarthritis" OR "Osteochondritis" OR "Osteomyelitis" OR "Osteoporosis" OR "Polymyalgia rheumatica" OR "Polymyositis" OR "Rhabdomyolysis" OR "Sjögren's disease" OR "Synovitis" OR "tenosynovitis" OR "Systemic lupus erythematosus" OR "Temporomandibular arthralgia" OR "Aseptic necrosis" OR "Baker's cyst" OR "Bunion" OR "Calcaneal spur" OR "Chondromalacia of patella" OR "knee* derangement" OR "Hallux rigidus" OR "Hallux valgus" OR "Hammer toe" OR "Iliotibial band syndrome" OR "Knee effusion" OR "Metatarsalgia" OR "Pes anserinus tendinitis" OR "Plantar fasciitis" OR "Prepatellar bursitis" OR "Tendinitis" OR "Tenosynovitis" OR "Ankylosing spondylitis" OR "Cervical spondylosis" OR "Coccygodynia" OR "Costochondritis" OR "Degenerative disc disease" OR "Diastasis recti" OR "Kyphosis" OR "Lumbosacral spondylosis" OR "Postlaminectomy syndrome" OR "Sacroiliitis" OR "Scoliosis" OR "Somatic dysfunction" OR "Spinal stenosis" OR "Spondylolisthesis" OR "Thoracic spondylosis" OR "Torticollis" OR "Adhesive capsulitis" OR "Bicipital tenosynovitis" OR "Boutonniere deformity" OR "de Quervain's disease" OR "Dupuytren's contracture" OR "Lateral epicondylitis" OR "Mallet finger" OR "Medial epicondylitis" OR "Olecranon bursitis" OR "Swan-neck deformity" OR "Tenosynovitis" OR "Rheumatoid arthritis" OR "Osteoarthritis" OR "Gout" OR "Back pain" OR "neck pain" OR "Osteomyelitis"
Congenital anomalies	"Arteriovenous malformation" OR "Atrial septal defect" OR "Hirschsprung's disease" OR "Hydrocephalus" OR "Hypospadias" OR "Imperforate anus" OR "Imperforate hymen" OR "Limb anomaly" OR "Marfan syndrome" OR "Meckel's diverticulum" OR "Microcephalus" OR "Osteogenesis imperfecta" OR "Pectus excavatum" OR "Pyloric stenosis" OR "Spina bifida" OR "Talipes equinovarus" OR "Tongue tie" OR "Congenital Muscular Torticollis" OR "Congenital Torticollis" OR "Undescended testis" OR "Ventricular septal defect" OR "Neural tube defects" OR "Cleft lip" OR "cleft palate" OR "Down syndrome" OR "trisomy" OR "Down's syndrome" OR "Congenital heart anomal*" OR "Congenital anomal*"
Oral conditions	"Oral disorder" OR "oral disease" OR "mouth disease" OR "oral cancer" OR "Gingivitis" OR "Thrush" OR "Mouth Ulcer" OR "dental carie*" OR "periodontal disease" OR "edentulism" OR "tooth decay"

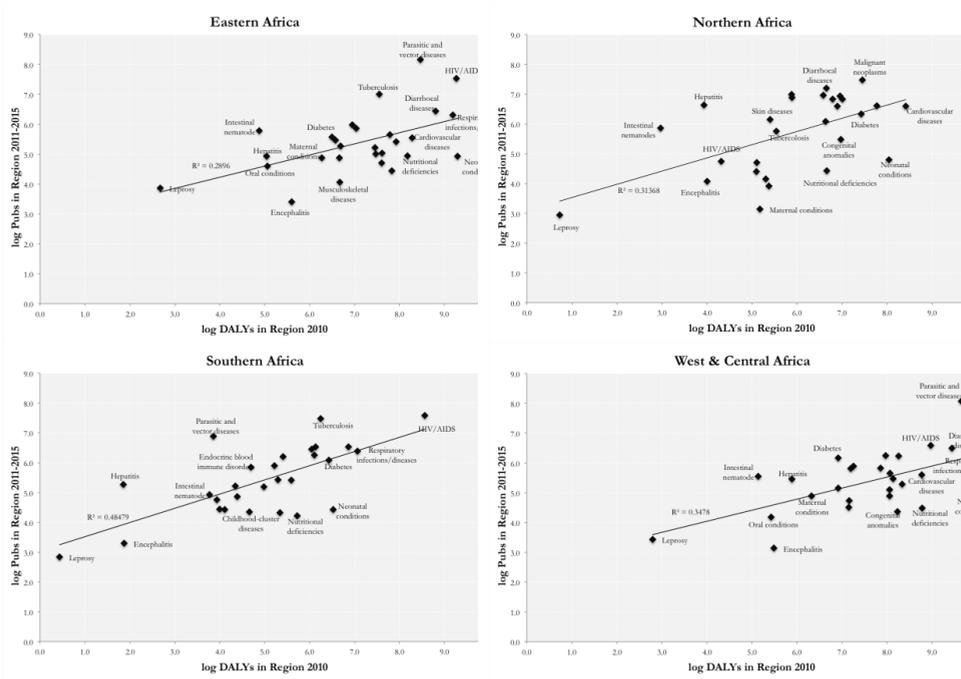
Figure 5.A.1. Disease burden specialisation (2005) vs research specialisation by disease (2006-2010).



Source: Own calculation based on WoS and WHO.

Note: DALYs from 2005 and scientific publications from 2006-2010.

Figure 5.A.2. Scatter plot. Disease burden (log 2010) vs research production (log 2011-2015) by disease



Source: Own calculation based on WoS and WHO.

Note: DALYs from 2010 and scientific publications from 2011-2015

Table 5.A.2. Top 5 funding institutions by African region

Region	Pubs diseases % funded pubs	Top 5 funders (Num pubs) 2011-2015				
		1	2	3	4	5
Northern Africa	12020 37%	Tunisian_Gov 402	NIH_US 302	King_Saud_University 288	Egyptian_Gov 221	EU-EC-ERC 180
Southern Africa	8938 68%	NRF_ZA 1389	NIH_US 1163	MRC_ZA 722	Wellcome_Trust 623	Gates_Foundation 329
Eastern Africa	9061 79%	NIH_US 1496	Wellcome_Trust 1053	Gates_Foundation 827	USAID 359	EU-EC-ERC 331
West and Central Africa	8058 59%	NIH_US 684	Gates_Foundation 481	Wellcome_Trust 380	EU-EC-ERC 368	WHO 322

Table 5.A.3. Top 5 funding institutions by funding group

Funding groups	Pubs diseases % pubs	Top 5 funders (Num pubs) 2011-2015				
		1	2	3	4	5
Public non-African	9922 28%	NIH_US 3317	EU-EC-ERC 1034	WHO 710	USAID 701	CDC_US 180
Public African	4406 12%	NRF_ZA 1404	MRC_ZA 722	Tunisian_Gov 402	DST_ZA 305	Egyptian_Gov 223
Philanthropic	4003 11%	Wellcome_Trust 1756	Gates Foundation 1381	Doris Duke Foundation 146	Howard Hughes Health Institute 117	Institut Pasteur 112
Corporation	1306 4%	GlaxoSmithKline 328	Pfizer 302	Novartis 238	Merck 206	Sanofi Aventis 198

Table 5.A.4. Top 5 diseases by top 10 Funding institutions (>0.8%total diseases)

Funder	Pubs diseases % pubs	Top 5 diseases (Num pubs) 2011-2015				
		1	2	3	4	5
NIH_US	3311 5.6%	Parasitic and vector diseases 1003	HIV/AIDS 850	Tuberculosis 705	Mental and substance use disorders 180	Endocrine blood immune disorders 180
Wellcome_Trust	1828 3.1%	Parasitic and vector diseases 842	Tuberculosis 350	HIV/AIDS 285	Respiratory infections/diseases 137	Endocrine blood immune disorders 100
NRF_ZA	1402 2.4%	Parasitic and vector diseases 250	Tuberculosis 238	HIV/AIDS 194	Diarrhoeal diseases 164	Cardiovascular diseases 115
Gates_Foundation	1379 2.3%	Parasitic and vector diseases 756	HIV/AIDS 224	Tuberculosis 143	Diarrhoeal diseases 109	Respiratory infections/diseases 97
EU-EC-ERC	1032 1.7%	Parasitic and vector diseases 419	Tuberculosis 187	HIV/AIDS 84	Mental and substance use disorders 47	Endocrine blood immune disorders 44
MRC_ZA	922 1.5%	Tuberculosis 161	HIV/AIDS 108	Parasitic and vector diseases 102	Cardiovascular diseases 101	Mental and substance use disorders 81
WHO	709 1.2%	Parasitic and vector diseases 355	HIV/AIDS 80	Tuberculosis 64	Respiratory infections/diseases 53	Diarrhoeal diseases 51
USAID	700 1.2%	HIV/AIDS 223	Parasitic and vector diseases 210	Tuberculosis 188	Respiratory infections/diseases 35	Diarrhoeal diseases 31
CDC_US	499 0.8%	HIV/AIDS 186	Parasitic and vector diseases 123	Tuberculosis 104	Diarrhoeal diseases 45	Respiratory infections/diseases 40
MRC_UK	450 0.8%	Parasitic and vector diseases 163	Tuberculosis 91	HIV/AIDS 76	Respiratory infections/diseases 35	Diabetes mellitus 32

Figure 5.A.3. Funders health research output (log) by African region

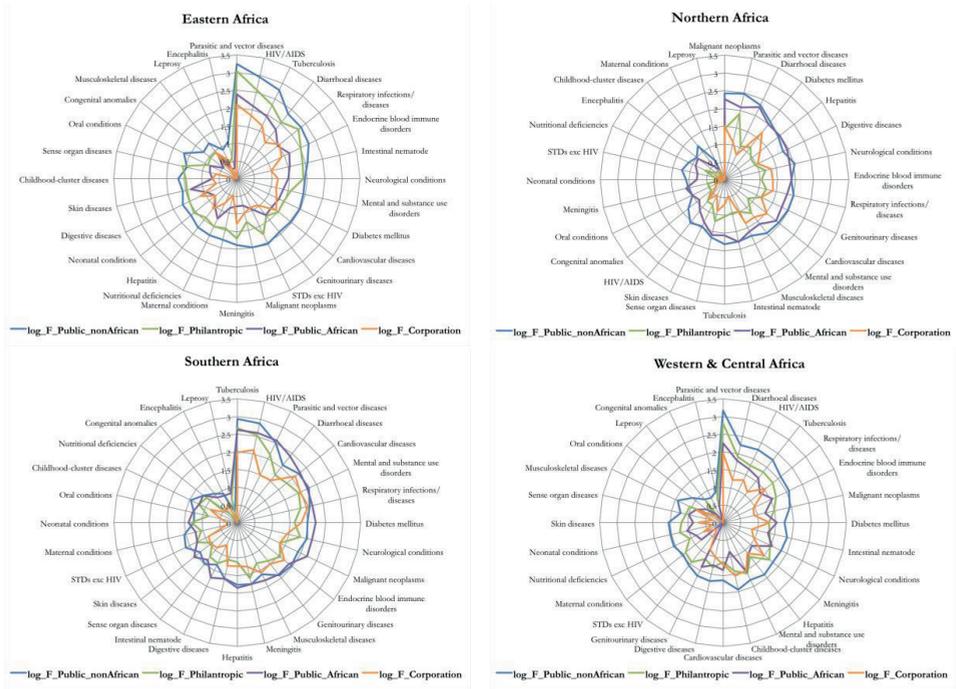


Table 5.A.5. Disease burden (log) versus research output (log)

Ind. Variables	log_Pubs_11.15				
	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)
log_DALYs_2010	0.17*** (0.057)				
log_DALYs_2010*Eastern_Africa		0.26*** (0.071)	0.32** (0.15)	0.30** (0.15)	0.077*** (0.026)
log_DALYs_2010*Northern_Africa		0.38*** (0.077)	0.31** (0.15)	0.28* (0.16)	0.037 (0.027)
log_DALYs_2010*Southern_Africa		0.38*** (0.089)	0.39*** (0.13)	0.40*** (0.12)	0.031 (0.026)
log_DALYs_2010*West&Central_Africa		0.24*** (0.065)	0.27* (0.15)	0.24* (0.13)	0.015 (0.046)
Int_collab_11_15 (%)				1.26 (0.91)	0.032 (0.21)
log_Pubs_06.10					0.92*** (0.026)
Constant	4.52*** (0.38)	3.61*** (0.47)	3.15*** (1.03)	2.32* (1.20)	0.50** (0.25)
Regional dummies	No	No	Yes	Yes	Yes
Observations	103	103	103	103	103
R-squared	0.082	0.222	0.226	0.242	0.958

Note 1: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Note 2: In estimations 2-4, the regression model was computed controlling for four regions: Eastern Africa, Northern African, Southern Africa and West & Central Africa

Chapter 6

Scientific systems in Latin America: Performance, networks, and collaborations with industry

Hugo Confraria
UNU-MERIT, Maastricht University

Fernando Vargas
UNU-MERIT, Maastricht University

Abstract

In this chapter, we use a combination of bibliometric, social network and econometric approaches to increase our knowledge of how research institutions interact with the private sector in Latin America (LA). We first study recent trends in scientific output and specialisation. On average, LA countries have been reducing the gap with the world leading regions. They have also tended to specialise in fields related to economic activities based on natural resources, such as Agricultural and Plant & Animal Sciences. However, collaborations with the private sector remain scarce. In this chapter, we have built scientific networks composed by what we define as Research Departments (RD). These RDs belong to universities, research institutes and government agencies. We model the intensity of collaboration of a RD with industry as a function of its size, previous performance, and its position in the LA and national scientific networks. Our results show that the RDs with higher diversity of research partners in their national scientific network work more intensively with industry. Additionally, collaborations with industry are influenced by previous interactions with the private sector.

Keywords: University-industry collaborations, Co-publishing, Social networks, Bibliometrics, Technology transfer, Latin America

This chapter draws upon:

Confraria, H., & Vargas, F. (2017). Scientific systems in Latin America: performance, networks, and collaborations with industry. *The Journal of Technology Transfer*, 1-42.

6.1. Introduction

Understanding the main traits of scientific institutions that engage in collaborative work with industry is critical for improving policies aimed to increase science-industry linkages. This is even more important in Latin American (LA) countries, where public support for science, technology, and innovation (STI) has increased significantly since the early 2000s, but the study of science-industry linkages has been relatively neglected. In this study, we show that scientific systems in LA countries have improved in performance while specialising in scientific fields that are related to the main economic activities of the region. Furthermore, by analysing scientific organisations characteristics and collaboration networks we find that organisations that have, or that have access to, diversified sources of knowledge work more intensively with the private sector.

One of the main motivations behind the increasing relevance of policies that promote science-industry linkages in the LA region is the potential benefit in innovation and technological capacities in the private sector. Indeed, Crespi (2012) summarises the main results of impact evaluations conducted in LA finding that public policies that promote collaboration between universities and industry increase the level of investments in innovation and labour productivity in firms. Marotta et al. (2007) show that both in Chile and Colombia firms that collaborate with universities are more likely to introduce product innovations and to apply for patents. Despite these positive findings, the magnitude of STI policies is still too modest to pull LA countries towards knowledge-based economies. Evidence from innovation surveys shows that universities and research centres tend to be less relevant partners for product and process innovation in LA firms in comparison to the non-LA OECD countries (OECD, 2015).

Part of the reason behind the limited importance of scientific institutions in the typical LA country has been attributed to prominence that the exploitation of natural resources has in LA economies. Conventional views see natural resources-based industries as activities with slow technological progress, where innovation is mostly driven by the suppliers of machinery and equipment, and that has a reduced potential of producing knowledge spillovers to other sectors (Lall 2000b; Pavitt 1984). Nonetheless, some scholars argue that there are certain specificities in the current context that creates a demand for local knowledge for the exploitation of natural resources which open a 'window of opportunity' for the development of local knowledge providers (Kaplan, 2012; Marin et al., 2015; Urzúa, 2011).

Marin et al. (2015), for example, argue that the intensification of the challenges in the exploitation of natural resources together with changes in volume and requirements of global demand would favour the development of a domestic knowledge-intensive industry built upon local scientific

capacities. Following this rationale, economic growth would be grounded in the capabilities acquired by each country in its specific area of resource endowment. It would then advance along the new technological trajectories being opened by research made in natural resources related fields. In this study, we provide insights about the scientific development in natural resources related areas in LA.

After studying scientific systems at the country level, we focus our analysis on scientific organisations and their patterns of collaboration with the private sector, measured by co-publications. There is an extensive body of literature on university-industry collaborations, and some of these studies examine cross-country and disciplinary differences in the patterns of co-authored scientific publications between university and industry (Godin 1996; Hicks et al. 1996; Tijssen 2004). Nonetheless, not much is known about the characteristics of scientific institutions that favour research collaborations with industry. The lack of evidence is even more noticeable in the LA countries. The results of this study help to close this gap.

In what follows, we first update and discuss current literature related to our research question; then in section three, we describe our data sources and the methodology. Section four shows the scientific production evolution and specialisation of LA countries. Section five describes LA scientific networks in five selected disciplines: Agriculture, Engineering, Environmental, Geosciences, and Plant and Animal Science. Section six presents and analyses the main econometric results, and lastly, we present conclusions in the final section.

6.2. Background

6.2.1. University – Industry Collaboration in Latin America

University-Industry collaboration in LA has largely been built from a top-down perspective as a result of S&T policies based on a supply-push focus (Crespi and Dutrénit 2014; Dutrénit and Arza 2010). Although LA universities differ across countries about their origins, a common feature is that they were initially oriented to undergraduate teaching. As research activities became increasingly common, postgraduate programs were gradually provided. Crespi and Dutrénit (2014) describe how several of the most important public research centres in LA were created during the period of supply-push science policies (1950s-1980s). These centres have focused on supporting sectors considered relevant by the policymakers (for example, coffee in Costa Rica, aeronautics and oil in Brazil, oil in Mexico, nuclear technology in Argentina, and agriculture in most countries).

During the same period, private sector evolved in economic activities that remained fairly protected from international pressures (either naturally or through intervention) and consequently

managed to survive and sometimes to grow without engaging in technological learning or in building innovation capabilities (Crespi & Dutrénit 2014). Consequently, the incentives to engage in time-consuming and challenging activities such as University-Industry collaborations were perhaps smaller. These singularities led LA firms to cooperate less frequently with universities than what would be ideal (Crespi et al. 2010).

However, despite their relative scarcity, university-industry interactions have been vital for successful historical experiences in some industries in LA. For instance, Arza and Vazquez (2012) highlight the importance of research by scientific institutions for agricultural technological upgrading in Argentina, while Casas et al. (2000) discuss the role of research institutions in successful experiences in biotechnology and other industries in Mexico. In a similar line, Suzigan and Albuquerque (2011) argue for the importance of university research for the development of the aircraft, steel and agricultural industry in Brazil. Evidence from quantitative studies show that manufacturing firms that collaborate with local universities increase their investments on R&D, are more likely to innovate and to apply for patents, and reach higher levels of labour productivity (G. A. Crespi 2012; Marotta et al. 2007).

6.2.2. Measuring University-Industry knowledge transfer

There are a wide variety of channels through which tacit and codified knowledge is being transferred between universities and industry. Some mechanisms include the mobility of students, personnel exchange, informal exchanges of information, public conferences, consulting, collaborative and contract R&D projects, joint ventures, scientific publications and patents (Cohen et al. 2002; Gray et al. 2013; Link et al. 2007; Meyer-Krahmer & Schmoch 1998; Francis Narin et al. 1997). According to Bekkers and Bodas Freitas (2008) the relative importance of these different channels in different contexts is explained, to a large degree, by the basic characteristics of the knowledge in question (tacitness, systemicness, expected breakthroughs), the disciplinary origin of the knowledge involved, and to a lesser degree the individual and organisational characteristics of those involved in the knowledge transfer process.

Due to this variety of channels and mechanisms, there are methodological challenges in measuring and assessing University-Industry collaborative research. The impacts of University-Industry collaborations are usually spread in space and time, can be numerous, and they are almost always difficult to separate from other parts of organisational life (Bozeman 2000). This methodological challenge is compounded by problems of data availability and measurability. For this reason, to focus on research collaboration between university and industry we adopt co-authored publications as a measure of occurrence and intensity of collaboration (Godin 1996; Tijssen 2012).

6.2.3. Co-publications as a measure of university-industry collaboration

The analysis of co-authorship has become one of the standard ways of measuring research collaborations between organisations (Lundberg et al. 2006). Co-authored publications indicate the achievement of access to an often-informal network, and can be viewed as successful scientific collaboration in themselves. They can also foster diffusion of knowledge and skills. Moreover, co-authorship as an indicator is quantifiable and invariant, while the measurement is not invasive (Abramo et al. 2009).

However, it should be emphasised that joint publications are just one type of the different channels of knowledge transfer. Academic databases do not capture research financed by industry, co-patenting, or even research collaborations that do not involve scientific publications. Furthermore, as argued in chapter 4, some co-authored articles do not reflect real collaboration. A publication co-authored by two institutions could suggest a collaboration that has not taken place, for example, if an author has the two affiliations. Also, most scientific publications are about a specific topic or research question, and interdisciplinary research may be left out of the publication system (Porter & Rafols 2009). Hence, co-authorship can never be more than a somewhat imperfect or partial indicator of research collaboration (Katz & Martin 1997; Laudel 2002). In the LA context, it has been argued that co-authored publications are one of the most important channels of knowledge transfer for researchers and firms (Dutrénit & Arza 2010). Therefore, for our study, we start from the assumption that companies need to perform research to absorb and appropriate codified scientific and technical knowledge (Aristei et al. 2016; Rosenberg 1990). Although “the traditional motivation of the technologist is not to publish, but to produce his artifact or process without disclosing material that may be helpful to his peers” (Price 1963), industrial researchers involved in scientific production activities act strategically. They publish in order to build their reputations, increase their visibility, reorient R&D agendas, establish intellectual claims and legal rights, signal capabilities to attract potential partners, and remain effectively plugged in to scientific networks where new ideas are emerging (Godin 1996; Lee 2000; Li et al. 2015; Tijssen, 2004).

Many of these papers are likely to be co-authored with researchers in the public sector. These researchers, on the other hand, have a different set of motives to collaborate with industrial researchers, namely to generate additional research funds, gain insights in the area of research, look for business opportunities, increase the output of commercialization activities and further the university’s outreach mission (Belkhdja & Landry 2007; Bozeman & Gaughan 2007; D’Este & Patel 2007; Lee 2000; Wong & Singh 2013).

Consequently, university-industry co-authorships can constitute a strategic way of acting that gives the researchers involved valuable insights in comparison with peers who are not participating in such collaborations. In this study, we assume that scientific institutions are always available for co-authorships with industry researchers, i.e. these institutions are not actively selecting their potential private sector partners. On the other hand, we assume that researchers in industries prefer to collaborate with institutions that exhibit specific structural characteristics. These include a quality dimension and a measurement of the diversity of knowledge. The latter can be studied by analysing collaboration network structures.

6.2.4. Network position as a correlate of performance

Scholars of social networks have consistently shown a significant association between network position and performance. One line of research indicates that actors with a higher number of direct ties will have access to additional sources of knowledge, ideas, and resources, thereby enhancing performance (Ahuja 2000; Reagans and McEvily 2003). Other research emphasises the benefit of brokerage. Actors brokering between otherwise disconnected actors are characterised by having a timing advantage, being in an advantageous position for identifying arbitrage opportunities, having higher chances of creating new knowledge or products, and being better able to capitalise on their existent capabilities (Burt 2004; Burt 2005; Zaheer and Bell 2005). The benefits of both types of network positions have also been suggested (Reagans and McEvily 2003; Fleming et al. 2007). Despite these different perspectives, the consensus has been that network positions correlate significantly with actor performance in different areas.

If we consider scientific collaboration network studies, most of the previous work focuses on the individual/researcher level. Some studies highlight the importance of structural collaboration network positions as a driver of preferential attachments (Barabasi et al. 2002; Moody 2004; Abbasi et al. 2012). Others try to understand if the location of a researcher in a network can bring some advantages, for instance, a higher level of citations, better access to knowledge sources, awareness of potential projects or access to more funding (Abbasi et al. 2011; Ebadi and Schiffauerova 2015). However, to the best of our knowledge, there is no evidence yet on the impact of structural collaboration network positions on the level of collaboration with industry.

In this study, we use a mixed set of methodologies and metrics to analyse the LA scientific system, its interactions, and its proximity to industry. Taking into consideration the productive structure of the region, we focus on five natural resource related fields: Agricultural, Engineering, Environmental, Geosciences, and Plant & Animals sciences. Our contribution is to update trends and specialisations patterns of scientific production in LA countries using bibliometric analysis and

descriptive statistics. Furthermore, we will assess to what extent the specialised knowledge diversity of scientific institutions, proxy by its position in the scientific networks, affect collaborations between science and industry. Understanding the determinants of these collaborations will provide useful information for the design of policies aimed at fostering industry-university linkages.

6.3. Methodology

6.3.1. Data collection

We used the *InCites* (2017) tool of *Thomson Reuters*, which is a web-based research evaluation tool that facilitates national and institutional comparisons across long time periods using indicators of publication output, productivity, specialisation and normalized citation impact. *InCites* provided output and citation metrics from the WoS (Web of Science, *Thomson Reuters*), which in turn allowed us to access data and metrics from a dataset of 22 million WoS papers from 1981 to 2013. All articles and reviews from researchers with a LA affiliation, published between 2004 and 2013, were analysed. The metrics for comparisons between countries are created based on address criteria, using the whole-counting method, that is, counts are not weighted by number of authors or addresses.

InCites classifies author addresses (affiliations) as “university”, “research institute” or “corporate”. In our work, an industry collaborative publication is one that has at least one author with a “corporate” affiliation, and at least one author with an affiliation from a LA “university” or “research institute”. It is important to keep in mind that not all single affiliations of all publications in *InCites* are unified as “university”, “research institute” or “corporate”.³⁹ There are corporate affiliations that have not been identified or unified yet; hence, they have not been classified as industrial publications. Multinational enterprises (MNEs) are more likely to have been identified and unified as “corporate”. Therefore, publications listed as industry (co)publications are a lower boundary of the real private sector research output. We would expect that countries with a lower presence of MNEs have larger differences between the number of publications authored by industry captured by *InCites* and the real activity.

Another important caveat in our analysis is that LA’s research output may be underestimated because its researchers often publish in journals that are not indexed in major citation databases, such as WoS or Elsevier’s Scopus.

³⁹ There were few cases where the name of the institution was available but the classification type was labeled as “unknown”. We assigned those cases to its correspondent type of affiliation after checking their information on the web.

6.3.2. Bibliometric analysis

In this section, we analyse the evolution of the science systems in LA, focusing mainly on its output, productivity, specialisation, quality, and linkages with industry. In addition to publication output (number of articles and reviews) and research performance (publication output relative to GDP and population), we calculate the percentage of publications of each country that were co-authored with industry, and the share of total publication output co-authored with international institutions. We also compute standard specialisation indexes to depict the relative specialisation of each country in a given area (Balassa 1965). This also serves to assess the overall level of specialisation of each country (Laursen 2000). Finally, to study the quality of the research output, we use two normalized measures of citation impact. These are values which evaluate the scientific influence or visibility of a set of publications in a given period. For the Quality Citation Index (Bornmann & Leydesdorff 2013), a country value of 1.2 indicates that the citation impacts of papers published by researchers in this country are, on average, 20 percentage points above the worldwide average. For the Quality Top 10% Index a country value of “10” indicates that ten percent of the publications of that country are in the top ten percent of the world, regardless of subject, year and document type (Pudovkin & Garfield 2009). Therefore, that country can be considered as performing at the same level as the world average. A value higher than “10”, indicates a higher performance relative to the world average (see appendix for more details).

6.3.3. Social Network analysis

In this section, we describe the structure and patterns of collaborations of the LA scientific network. We focus this part of our analysis on what we defined as the Research Department level (RD).⁴⁰ This unit of analysis is defined by an output measure. We assume that all publications from one institution, in a determined scientific field, were produced by a specific RD. For example, we treat all publications from one institution in two scientific fields as research output from two different RDs that belong to the same institution. In addition, we assume that the research performed in each area faces its specific conditions and it is embedded in a particular scientific network, independent from other scientific topics. Although these assumptions could be debatable, scientific research in each field demands high levels of specialisation and knowledge, which makes it very costly to get involved in research in other disciplines (Jeffrey 2003). Hence, we expect that this definition may include some errors but not a consistent bias.

As it was mentioned above, we define institutions conducting research in more than one field as having different RDs operating separately in each one of them. To extract the relevant scientific networks, we define a threshold to select the most prolific RDs in LA. For each field studied, we select RDs with more than 50 publications in each of the two five-years periods analysed. Afterwards, for each of these “elite” RD, we gather all partners with 5 or more collaborations in the same field, in the same period. Thus, two RDs are going to be linked if they have 5 or more co-authorships in the field and period⁴¹. It is worth mentioning that collaboration partners are not necessarily part of the “elite” RDs group, given that they only need to satisfy the minimum of 5 co-publications with one “elite” RD. This group of collaboration partners also includes RDs that are not from LA institutions; however, we do not consider in our calculations those “foreign” institutions that are linked only with one LA RD.

We perform this analysis in two periods of 5 years each (2004 to 2008, and 2009 to 2013). Besides the graphical description of networks of both periods, we obtain information at the RD (node) level, such as centrality indicators (degree, betweenness, and closeness), and network features, namely the number of nodes, number of communities and average path length.

⁴⁰ RDs can belong to universities, research institutes (public or private) and governmental agencies.

⁴¹ We had to decide a certain output threshold because with few publications our dependent variable would be very sensible to an extra U-I co-authorship. We decided that 50 would be an acceptable number since it represents 10 publications per year on average. We decided on 5 co-authorships to represent at least one co-authorship per year.

6.3.4. Econometric analysis

In this section, we describe a model that allows for gathering new evidence of the characteristics of the RDs working most closely with industry. We define the percentage of publications of RDs that are co-authored with the industry as the dependent variable, and we relate it to a set of RD features that could influence such collaborations: (1) knowledge production capacity; (2) research quality; (3) orientation towards industry; and (4) knowledge diversity.

Co-authorships with industry are far from common in science. The occurrence of these events can be represented as a case of corner outcomes with a corner at zero percent and a continuous distribution for strictly positive values (upper-censored at 100 percent). Wooldridge (2002) suggests addressing these cases implementing “hurdle” or “two-tiered” models. This allows explanatory variables to differently affect the participation decision, i.e. the co-authorship of at least one publication, and the intensity of these collaborations, measured as the percentage of the total publications of a RD that were produced jointly with firms. Therefore, we firstly follow the specification of the two-tiered model developed by Cragg (1971). In the “first-tier” of the model, we estimate the probability of participation in co-publication with industry using a probit model. In the “second-tier” a truncated normal model is used to estimate the intensity of the collaborations with industry, formally:

$$f(w, y | x_1 x_2) = \{1 - \Phi(x_1 \gamma)\}^{1(w=0)} \left[\Phi(x_1 \gamma) (2\pi)^{-\frac{1}{2}} \sigma^{-1} \exp\left\{-\frac{(y - x_2 \beta)^2}{2\sigma^2}\right\} / \Phi\left(\frac{x_2 \beta}{\sigma}\right) \right]^{1(w=1)} \quad (4)$$

Where w is a dichotomous variable equal to 1 if the RD has at least one co-publication with industry and 0 otherwise, and y is the percentage of publications of the RD co-authored with the private sector. When w is equal to 0, then y also takes the value of 0. While $w = 1$, then $y > 0$. Variables x_1 and x_2 are sets of characteristics of the RDs that affect the likeliness to co-publish with industry and the intensity of these activities, respectively. Hence, γ captures the effects on the participation and β those associated with the intensity of co-publication. This specification assumes conditional independence between the two tiers of the model. In this case, that means to assume that after controlling the observable characteristics of the RDs, there is no correlation between the decision to participate and the intensity of co-publications. We are aware that the latter assumption could be debatable. Therefore, we also use the approach developed by Heckman (1979) as a consistency check. Although this model is aimed to address the selectivity problem that arises when an interval of the outcome variable is not observable, statistically is very similar to Cragg’s model and its flexibility allows for correlation between the participation and intensity

equations. However, a variable that affects the participation but not the intensity of collaborations with industry needs to be included to identify the model.

As we mentioned above, we model the participation of the RDs in collaboration with industry and the intensity of co-publication as a direct function of the main RDs characteristics. The first independent variable is the total number of scientific publications during the period, which depicts the capacities of knowledge production of the RDs. This variable is expected to have a positive effect on the relationships with the private sector since the capability of a university to attract private enterprise collaboration is influenced by the size of the group of academic researchers and their output (Abramo et al. 2010). Furthermore, this variable is also a proxy for the size of the RD. Larger organisations may have more resources available to assign for relationships with the private sector.

We also include a measurement of the scientific quality of the research output of the RD, in the form of a citation impact index. In principle, we expect an ambiguous effect of quality in co-publications between science and industry. On the one hand, highly cited institutions enjoy reputational benefits that make them perceived as more desirable partners for research by the private sector, increasing the likeliness of this type of collaboration. On the other hand, institutions that produced highly cited publications may be mainly focused on academic research, leaving few resources available to create linkages with industry. The empirical evidence is also mixed. Some studies have shown a correlation between universities' citation impact and their intensity of collaboration with industry (Abramo et al. 2010; Balconi and Laboranti 2006; Giunta et al. 2016). However, further analysis, examining specifically the Italian situation, showed that enterprises do not necessarily choose partners with higher scientific influence (Abramo et al. 2009).

The orientation of a RD towards working closely with the private sector certainly will affect the share of co-publications (Bozeman & Gaughan 2007; Giunta et al. 2015). We proxy this factor, using a variable that calculates the previous record of science-industry collaborations of the RD. By measuring prior partnership, we are also able to control for the pre-existent linkages with industry that could have been developed at the institutional or personal⁴² (researchers) level.

Finally, we consider that the diversification of knowledge within each RD positively affects its closeness to the private sector. In particular, we assume that industrial research projects in which companies involve RDs are significantly more complex and uncertain than the common ones (Hall et al. 2003). Hence, RDs that possess or have access to diverse but complementary expertise, even

⁴² The choice of a university partner by an enterprise often develops not only based on objective information but also through personal contacts. Selection and maintenance of relationships is strongly conditioned by social proximity (Granovetter 1985).

within the same scientific field, are going to be working more intensively with the industry. Unfortunately, the level of disaggregation of the publication data by scientific field does not allow us to test this directly. Nevertheless, we make use of the social network features of each RD to proxy their internal knowledge diversification. Specifically, we assume that RDs that have a more varied set of research partners are more likely to have higher knowledge sources.⁴³ Accordingly, we include variables that provide information about the linkages of the RDs and its relevance in the scientific network. We rely on three commonly used measures of network centrality (Freeman 1978): degree, betweenness, and closeness (see the appendix for more details).

By including this type of variables in our estimation, we are also controlling for other mechanisms that are taking place in parallel. Namely, RDs that are relatively better connected in their scientific network could be given preference in work collaborations, since they have earlier access to sources of knowledge and ideas (Burt 2005). Higher centrality can also lower the cost of screening other RDs for future partnerships, help to diffuse the scientific challenges in which companies are interested and increase its scientific reputation thereby attracting top researchers (Godin 1996; Lee 2000; Li et al. 2015; Tijssen 2004). On the other hand, working with highly connected RDs can be risky for companies because it increases the potential damages of leakages of relevant information of the firms. Finally, we can also expect that geographical proximity plays a role in shaping these collaborations (Bozeman & Corley 2004; Giunta et al. 2015; Pinch et al. 2003). Hence, we include RDs information regarding their position in both LA and their national scientific networks⁴⁴.

Usually, quantitative studies assessing causality based on statistics and data from networks are subject to endogeneity biases. In our case, it would be in the causal direction of the relationship between the linkages of a RD within their scientific network and the intensity of collaborations with industry. We try to address this potential problem by using information from two separate periods of time. This enables us to analyse RDs characteristics and position in the network in one period and the collaborations with the private sector in the following period. Furthermore, from the management literature, we know that previous alliances tend to remain or to be repeated because routines decrease asymmetries of information among partners and facilitate the estimation of future returns of joint activities (Gulati 1995). At the same time, processes of path dependence

⁴³ Unlike the alliances between firms, driven by complementary knowledge for learning purposes (Baum et al., 2010), we assume that co-publications between researchers from different scientific institutions can also be driven by the opportunity to increase their publication productivity levels, therefore joint research could be conducted between researchers and institutions with fairly similar specialisation patterns and levels of knowledge.

⁴⁴ We do not analyse the centrality measures of each RD in their global subject area network for simplicity. However, there may be LA RDs that are collaborating a lot internationally, compared to others, and in those cases we are potentially underestimating their (normalised) degree. However, this issue shouldn't affect (normalized) betweenness in the same way.

induced by the influence of initial conditions on future developments may also occur here (Thune & Gulbrandsen 2014). The choice of the 5-year time span is a compromise between robustness of results and timeliness.

We control for differences in the intrinsic degree of proximity to industry of different scientific fields by using field dummies. We also include a set of country dummies to control for idiosyncratic characteristics and specific science-industry policies. Also, we include a dummy variable that controls characteristics of RDs that are part of universities, relative to other types of institutions. Finally, we allow errors to be correlated among RDs that belong to the same institution⁴⁵.

6.4. Science in LA: Trends and Specialisation

LA's long-term world percentage of publication output in WoS has increased from 1.32% in 1981 to 5.03% in 2013. In 2013, all LA countries accounted for 71,391 publications in WoS. Brazil's share of publication output is particularly high when compared with other countries of the region (around 55% of LA output in 2013), reflecting differences in the size of the economies. According to our analysis, the share of world scientific output from Brazil increased at a constant rate from 1993 to 2006, when publications skyrocketed to the levels seen in Brazil in 2013.⁴⁶ Other countries that show higher average shares of scientific output than LA in the last decade are Mexico, Argentina, and Chile.

Table 6.1 provides data adjusting scientific output by other characteristics of the countries. This allows for an assessment of the scientific "productivity" per billion of USD and per million inhabitants.

LA countries are ranked in Table 6.2 by aggregate scientific production from 2004-2013. Although Brazil has the highest number of publications, it has the lowest scientific impact. This may happen due to a significant percentage of articles being published in national journals that had recently been included in the databases (Collazo-Reyes, 2013; Vargas, 2014). Countries with smaller scientific systems tend to rely more intensively on international collaborations. The average LA country has 75% of its scientific outputs co-published with a foreign institution, while that figure goes down to 42% when considering the top 4 largest science systems in the region (Brazil, Mexico, Argentina and Chile).

⁴⁵ There is heterogeneity in size across institutions. Since we only collected and cleaned data for five subject areas we cannot control for institutions size. However, we are confident that the size of each RD is highly correlated with the size of the institution where the RD is.

⁴⁶ Vargas et al. (2014) argue that, in areas such as Agricultural Sciences the increase of output since 2006 was due to the expansion of Brazilian journals in WoS and an increase in the number of issues published by these journals.

In general, although LA's scientific impact is growing, it remains relatively low when compared to the world average. In the region, Peru and Panama perform best in this indicator, probably because more than 85% of their publications are co-authored with researchers outside their country and that influences the visibility of the research being produced by local researchers (van Raan 1998). Chile, by far the most productive country in the region, has also increased its research output and maintained a medium level of scientific impact. With regard to the levels of collaboration with industry, we can observe a relatively low percentage, mainly in the countries that are not so dependent on international collaboration or have larger science systems.⁴⁷ These results are in line with Tijssen (2012), who showed that LA and North Africa are the regions in the world with the lowest intensity of science-industry co-authorship.

Countries often try to invest strategically in research areas critical to their economic development (see chapter 5). Creation of specific local knowledge may increase innovation capacities of incumbents, but also promote the birth of start-ups or spin-offs. These trends run in parallel with others that do not necessarily operate in the same direction. Historical and cultural influences, strengths of scientific establishments, as well as incentives and government funding for scientific research play a relevant role in defining the revealed scientific specialisation of a country. The size of the scientific system also matters, since larger science systems have the capacity for more diversity and greater coverage of the full scope of sciences.

In contrast, smaller systems may be limited in their ability to invest in specific domains. We explore the outcome of these trends through a specialisation analysis based on the 22 Essential Science Indicators (ESI) areas.⁴⁸ Table 6.2 contains the five subject areas of highest specialisation for the nine countries in LA with more than 1% of LA total scientific output over the 2009-2013 period. Table 6.2 also provides information on aggregate specialisation level (given by the SII index) for each of these nine countries.

⁴⁷ In comparison, high-income countries like the United States or Germany have levels of industry collaboration higher than 2%.

⁴⁸ The Essential Science Indicators schema (*Thomson Reuters*) comprises 22 subject areas in science and social sciences and is based on journal assignments. Arts & Humanities journals are not included. Each indexed journal (11,000+) is found in only one of the 22 subject areas and there is no overlap between categories.

Table 6.1. Research performance of L.A.: Summary statistics (2004-2008 and 2009-2013)

Country	Total publications output		Productivity		International Collaboraition		Industry Collaboration		Quality Citation Index	
	04-08	09-13	04-08	09-13	04-08	09-13	04-08	09-13	04-08	09-13
Brazil	106692	178313	23	32	28%	27%	0.8%	0.9%	0.61	0.65
Mexico	38729	52715	9	11	43%	43%	1.2%	1.3%	0.70	0.79
Argentina	28440	39992	24	26	44%	44%	0.9%	0.9%	0.77	0.93
Chile	17185	28632	26	37	54%	55%	0.9%	0.7%	0.90	0.92
Colombia	6436	15169	8	16	60%	59%	1.2%	1.1%	0.67	0.80
Venezuela	5948	5801	8	6	52%	53%	1.6%	1.5%	0.62	0.71
Cuba	3630	4093	16	14	62%	70%	0.6%	0.3%	0.59	0.72
Uruguay	2388	3553	26	29	66%	67%	0.6%	0.9%	0.83	0.95
Peru	2100	3562	5	6	83%	85%	2.2%	2.4%	0.94	1.32
Costa Rica	1679	2160	16	17	72%	78%	1.7%	2.4%	0.99	1.15
Ecuador	1136	1875	5	7	83%	87%	0.7%	1.2%	1.08	1.17
Panama	1007	1576	12	12	87%	90%	1.4%	1.9%	1.48	1.48
Bolivia	768	1028	15	16	89%	90%	1.0%	0.1%	0.81	1.17
Guatemala	326	583	2	3	86%	89%	1.8%	2.0%	0.81	0.97
Nicaragua	242	342	7	9	94%	94%	1.6%	2.5%	0.78	0.99
Paraguay	175	307	4	5	85%	89%	0.0%	0.6%	1.00	1.07
Dominican Rep.	145	258	1	1	93%	93%	4.1%	4.3%	1.00	0.80
Honduras	143	254	3	4	94%	97%	2.1%	3.3%	1.23	0.96
Haiti	94	208	4	9	95%	96%	3.1%	2.8%	0.94	1.32
El Salvador	102	196	1	2	88%	87%	1.6%	0.9%	0.57	1.11
Latin America	209756	325069	14	18	73%	75%	1.5%	1.6%	0.71	0.70
World	5017977	6493637	21	24	-	-	-	-	1.00	1.00

Source: Own calculations based on InCites and World Bank. ^a Docs/GDP - Publications per billion constant 2005 US\$ of GDP (yearly average). ^b Docs/Population - Publications per million inhabitants (yearly average).

Note: The international and industry collaboration value for Latin America represents the average of the countries of the sample.

Research specialisation is quite similar across these LA countries. In aggregate terms, the top 5 areas with the largest output from LA, relative to the world are Agricultural Sciences (15.7%), Plant & Animal Science (12.3%), Space Science (9.3%), Environment/Ecology (7.7%) and Microbiology (7.3%). The higher LA specialisations are in Agricultural Sciences and Plant & Animal Sciences, which is in line with the high importance of agricultural, livestock and agro-industrial activities in the region.

Table 6.2. Top 5 subject areas, in the 9 LA with higher scientific output (2009-2013)

Country	Relative Specialisation Intensity (Rank)										SII ^c					
	1		2		3		4		5							
	RSI ^b	Docs ^a	RSI ^b	Docs ^a	RSI ^b	Docs ^a	RSI ^b	Docs ^a	RSI ^b	Docs ^a						
BRAZIL	Agricultural sciences	4.00	20033	Plant & Animal science	2.42	21557	Microbiology	1.36	3428	Pharmacology & Toxicology	1.18	5557	Environment & Ecology	1.17	6082	0.43
MEXICO	Space science	2.51	1366	Plant & Animal science	2.50	6600	Environment & Ecology	2.12	3258	Agricultural sciences	1.92	2851	Microbiology	1.47	1095	0.28
ARGENTINA	Plant & Animal science	2.89	5791	Microbiology	2.25	1273	Space science	2.09	861	Agricultural sciences	2.06	2318	Geosciences	1.90	2176	0.39
CHILE	Space science	9.70	2863	Plant & Animal science	1.86	2669	Agricultural sciences	1.82	1467	Environment & Ecology	1.71	1430	Mathematics	1.45	1206	0.95
COLOMBIA	Agricultural sciences	2.23	949	Plant & Animal science	1.88	1425	Engineering	1.73	2048	Immunology	1.48	390	Physics	1.35	1706	0.21
VENEZUELA	Agricultural sciences	2.44	397	Immunology	2.41	244	Plant & Animal science	2.41	700	Environment & Ecology	2.24	379	Economics & Business	1.85	192	0.35
CUBA	Immunology	2.58	184	Pharmacology & Toxicology	2.43	263	Agricultural sciences	2.24	258	Plant & Animal science	2.13	437	Microbiology	1.59	92	0.34
URUGUAY	Plant & Animal science	3.37	600	Agricultural sciences	3.32	331	Microbiology	2.71	136	Environment & Ecology	1.77	183	Immunology	1.54	95	0.64
PERU	Immunology	5.64	350	Plant & Animal science	2.52	450	Microbiology	2.33	117	Environment & Ecology	2.30	239	Social sciences	2.21	463	0.96

Source: Own calculations; InCites

^a Docs = Scientific publications.

^b RSI = Share of a country's papers in a given field, relative to the share of world papers in that field.

^c SII = Specialisation Intensity Index. This measure provides a ratio to assess whether a country is “specialised” or “not specialised.” It grows with the specialisation intensity of a country.

The cases of Peru and Chile are interesting because they revealed high specialisation⁴⁹ in subject areas different from the other countries of the sample. The specialisation of Peru is related to issues in public health (prevention of HIV, tuberculosis, and lupus) in which they also have a high scientific impact (Van Noorden 2014). Chile's high specialisation in Space Science is related to its excellent infrastructure of giant telescopes housed in the Atacama Desert. According to Catanzaro et al. (2014), funding for astrophysics has increased from \$2 million in 2006 to \$6.8 million in 2010. Over the same period, the number of faculty positions has almost doubled. This has led not only to an increase in the number of publications in this field but also to an increase in quality. In contrast, Economics & Business, Materials Science, Computer Science, Psychiatry/Psychology and at a certain level Engineering seem to be neglected research disciplines across LA countries.

In summary, scientific activity has been growing in LA countries during the last decade but not at a pace that allowed it to catch-up with the rest of the world. Only four countries show productivity levels closer to the world averages. Co-publications with international institutions are frequent and highly relevant for the scientific impact (quality) of smaller scientific systems. On the other hand, collaborations with industry are scarce even when research specialisation seems to be influenced by economic specialisation.

In what follows, we will focus on the study of five main scientific fields: Agricultural, Engineering, Environmental, Geosciences, and Plant & Animals sciences. We choose Agricultural, Geosciences and Plant & Animals as they are closely related to the natural resources-based economic activities in which LA countries are more intensive. We also include engineering and environmental sciences because we assume that this type of knowledge needs to be consistently applied across the main economic activities of the countries analysed.

⁴⁹ If a country has a scientific output structure equal to the world, the value of the indicator will be zero. The size of SII is an indication of how strongly each country is specialised.

6.5. Network analysis

The data requirements for extracting the scientific networks explained in section 3.3 only allow us to include in our analysis RDs from the following LA countries: Argentina, Brazil, Chile, Colombia, Costa Rica, Cuba, Mexico, Panama, Peru, Uruguay, and Venezuela. Table 6.3 gives some network summary statistics from the five scientific fields that we are analysing. Network graphs are available in the Appendix (Figures 6.A.1 – 6.A.10). The nodes are the RDs, and the links are number of collaborations (co-authorships).

Table 6.3. Summary statistics of scientific networks

Subject Area	Number nodes (RD)		Number Communities		Average Path Length		% of RD collaborating with industry		Average % of co-publications with industry	
	2004/08	2009/13	2004/08	2009/13	2004/08	2009/13	2004/08	2009/13	2004/08	2009/13
Agricultural sciences	96	138	11	25	3	2.9	13.5%	16.7%	0.21%	0.11%
Engineering	77	139	10	12	3.4	3	29.9%	30.9%	0.57%	0.79%
Environmental sciences	119	224	8	10	3	2.5	10.9%	7.6%	0.34%	0.16%
Geosciences	98	165	4	8	2.6	2.5	18.4%	14.6%	1.53%	1.32%
Plant & Animal sciences	164	253	8	8	2.7	2.5	6.1%	5.9%	0.21%	0.04%

Source: Own calculations. InCites

The 55% growth of LA scientific production between 2004-2008 and 2009-2013 is roughly proportional to the increase in the number of RDs in all subject areas (networks). This may suggest that the growth in the number of publications might be an artefact of the data becoming more complete (more journals indexed and more RDs included in the analysis). In both periods, the average path length is less than 4, implying that knowledge that is created in one node has the potential to be diffused in few steps to the rest of the network.

In this section, we also analysed the number of communities in each subject area. In network analysis, communities are a group of nodes that are densely connected between them and more sparsely connected with nodes from other communities. In our case, communities are interpreted as groups of RDs that tend to collaborate more intensively in research with each other. Interestingly, in our analysis, we found that the change in the number of communities⁵⁰ does not follow a common trend. Engineering, Plant & Animal, and to some extent, Environmental Sciences shows a remarkable increment in the number of RDs in the LA network. However, there

⁵⁰ Communities were detected applying the 'Leading eigenvector method' available in the R software package.

are limited changes in the number of communities, suggesting that newcomers were rapidly attached to well-established groups of collaborators.

On the other hand, Agricultural and Geosciences at least double the number of knowledge communities. It can be interpreted that evolving networks are creating new niches of knowledge, either with new local actors or increasing diversification of knowledge sources through new international collaborations. Nevertheless, it should be borne in mind that geographical proximity may also be playing a role in the creation and evolution of these research communities.

Shortest average paths together with an increasing number of communities are signals that a network structure is evolving towards a structure that facilitates both knowledge creation and knowledge diffusion. However, attention needs to be paid to the fact that in almost all scientific fields studied it is common to observe that two neighbouring countries (in geographic terms) are only connected to each other through a RD that is based in a third country. Even when this situation gives potential brokerage power to the external RD, it is not clear what the impact is for the performance of LA scientific networks. This is a topic that requires further research.

We also found that the increases in the number of RDs, between the periods, are not reflected in significant changes in the percentage of RDs collaborating with the industry. Therefore, we can assume that the share of RDs collaborating with industry among the new incumbents is the same as the proportion of RDs connected to the industry in the previous period. On the other hand, the average percentage of co-publications with industry fell in all scientific fields, except for engineering. The latter suggests that the new RDs that co-publish with the industry are doing it less intensively than the average RD of the previous period.

As we mentioned before, for the econometric implementation we also estimate centrality measures for local/national networks. For each country, these networks are formed by all elite national RD (same threshold defined before) and its research partners. Foreign institutions are also included in the network. However, those that have collaborations with only one local RD are considered peripheral and are subsequently dropped from the network. After application of these filters, we are left with data only from Argentina, Brazil, Chile, and Mexico.⁵¹

⁵¹ We could get the national network of Venezuela, but data requirements for the econometric estimations left these observations out of the final dataset.

6.6. Econometric analysis

In this section, we present the results of the estimation of Cragg's (1971) model described in subsection 3.4, run using the user-developed *craggit* routine in the *Stata* software.⁵² We pooled data from the LA scientific networks presented in section 5. After the application of data requirements to the nodes gathered from these networks and dropping outliers on the outcome variable⁵³, we end up with a database of 324 observations from four LA countries (Argentina, Brazil, Chile and Mexico) in the five selected scientific topics.

Table 6.4 summarises descriptive statistics of the main variables used in the econometric analysis. Overall, 70% of the RD in the sample did not have a single publication co-authored with industry in the period 2009-2013, a lower share than the average of 78% in period 2004-2008. The average RD published approximately 179 papers between 2004 and 2008. In the same period, the indexes of quality of publications show that these publications tend to underperform in relation to the rest of the world, having 21% fewer citations (citation index of 0.79) than the average paper in the same field. Furthermore, only 6.1% of the publications of these RDs are in the top 10% of their field.

Brazil accounts for almost 50% of the RDs considered in this sample. Argentina, Chile, and Mexico account for the other half of the observations. Plant & Animal Sciences account for 27% of the RDs here considered. Agricultural, Engineering, and Environmental Sciences represent roughly 20% each, while Geosciences accounts for the remaining 14% of the cases. Most the RDs in the sample are located in universities (86%), and the remaining 14% belong to research institutes or government agencies.

Tables 6.5 and 6.6 show findings regarding determinants of the participation of co-publication with industry and its intensity. As expected, past collaborations with industry are revealed as a strong predictor of collaborations in the subsequent period. RDs with a higher percentage of co-publications with industry are more prone to keep engaging in these collaborations. Across the different specifications of the model, the sign and statistical significance of this effect remains. The positive impact of past collaborations in the intensity equation is not statistically significant in the *craggit* estimation. However, the Heckman-model not only confirms the positive relation but in this specification the coefficients are significant.

⁵² As a consistency check, we also estimated a two-step Heckman selection model, using the same software. Those results are available in the appendix.

⁵³ We define as an outlier a RD for which the outcome variable is more than 3 standard deviations above/below the mean.

Interestingly, there is no clear relationship between academic quality and engagement in research with the private sector. The signs of the coefficients (positive for participation and negative for intensity) may suggest that higher academic quality favours participation in research with the industry, but for RDs with higher levels of citation impact, the collaboration intensity with industry is relatively smaller. One possible explanation for the latter is that RDs that produce highly cited research are mainly focused on academic research and not so much on generating linkages with industry.

Table 6.4. Descriptive statistics of variables used in the econometric analysis

Variable	Observations	Mean	Standard Deviation	Min	Max
Industry collaboration (0/1) (2004-2008)	324	0.22	0.41	0	1
Industry collaboration (0/1) (2009-2013)	324	0.30	0.46	0	1
Industry collaboration intensity (2004-2008)	324	0.45	1.42	0	10.87
Industry collaboration intensity (2009-2013)	324	0.36	0.98	0	6.44
Normalized LA degree (2004-2008)	324	0.06	0.08	0.01	0.47
Normalized national degree (2004-2008)	324	0.16	0.20	0.01	1
Normalized LA betweenness (2004-2008)	324	0.03	0.06	0	0.37
Normalized national betweenness (2004-2008)	324	0.09	0.17	0	1
Normalized LA closeness (2004-2008)	324	0.04	0.01	0.02	0.07
Normalized national closeness (2004-2008)	324	0.06	0.02	0.02	0.11
Number of publications (2004-2008)	324	178.73	284.35	5	2368
Number of publications (log) (2004-2008)	324	4.45	1.19	1.61	7.77
Quality Citation Index (2004-2008)	324	0.79	0.32	0.21	2.49
Quality Top 10% Index (2004-2008)	324	6.10	5.43	0	36.36
Argentina	324	0.15	0.35	0	1
Brazil	324	0.47	0.50	0	1
Chile	324	0.15	0.36	0	1
Mexico	324	0.23	0.42	0	1
Agricultural	324	0.20	0.40	0	1
Engineering	324	0.18	0.39	0	1
Environmental	324	0.20	0.40	0	1
Geosciences	324	0.14	0.35	0	1
Plant and Animal	324	0.27	0.44	0	1
Universities	324	0.86	0.35	0	1

Country dummies show that there are no significant differences at this level on the likeliness of RDs to engage in research collaboration with the private sector. However, the Brazilian RDs that do participate in industry collaborations are doing it more intensively than their counterparts in Argentina, Chile, and Mexico. Finally, engineering sciences are consistently the research field with most collaboration with industry in both, participation and intensity, a result that we expected

given the applied orientation of the engineering activities. The result in the intensity equation also holds for Geosciences, probably due to the importance of mining operations in the sample of countries included in our analysis.

The position of the RDs in the LA scientific network does not seem to be related to its relationship with the private sector. Indeed, none of the network centrality measurements tested (degree, estimations 2 and 3; betweenness, estimations 4 and 5; and closeness, estimations 6 and 7) at the LA level show statistically significant coefficients. The unimportance of these RDs features, in the global LA context, contrast with the results observed when we consider the node characteristics of the RDs in the national/local network. Indeed, our most important finding is that two of the centrality measurements of the national/local scientific networks (estimations 3 and 5) show positive and significant effects on the intensity of the collaboration with industry.

Although not relevant in the participation equation, RDs with higher values of local degree and betweenness engage more intensively in research activities with the private sector. These results suggest that the RDs that have or have access to a more diversified set of knowledge sources⁵⁴ in their countries are more prone to engage intensively in research with industry. The mechanism behind this finding may be related to the fact that RDs which are better connected can provide different strands of specialised knowledge that allow them to tackle the type of challenges proposed by the industry adequately. At the same time, these RDs can provide benefits to firms by lowering the costs of screening other RDs for future partnerships, decreasing the risk of knowledge lock-in, attracting high-qualified researchers, and providing a more effective diffusion of the scientific challenges of the company.

Furthermore, nodes in brokerage positions (higher betweenness) are characterised by having a timing advantage. They are not only more likely to be the first recipients of information from diverse groups but also occupy a privileged position from which they can assess the relevance of new information (Burt 2005). Therefore, in a competitive process in which timing is rewarded, a brokerage position of RDs in national borders may be providing a crucial advantage for collaboration with industry. However, as previous research has also suggested (Liao & Phan 2016) since the participation equation is not significant in the local network, we should be careful when arguing that these two types of network positions will lead deterministically to more university-industry collaborations.

⁵⁴ Here we are assuming that having relatively more collaborations with RDs (higher normalized degree), one organisation is more diverse. However, there may be cases of RDs with many collaborative partners with similar expertise, and one RD with few partners that are very diverse.

Larger RDs can cover a broader spectrum of scientific topics, and they also have more resources that could be used to establish relations with the private sector (e.g. TTOs), making them more prone to engage in collaboration with industry. Our results confirm this showing that even after controlling for networks centrality features, the size of the RD is revealed as a strong predictor of the likeliness of performing research with the private sector. On the other hand, the intensity of these collaborations decreases with the size of the RD. We are aware that the relation between size and centrality may raise a multicollinearity problem. The availability of more resources in larger RDs can also increase its centrality. However, we are confident that theoretically, both variables are not measuring the same characteristics of the RDs, that is, more co-publications do not necessarily imply more diversity in co-publications partners. Therefore, both size and centrality must be included in the estimations. Excluding these aspects will give rise to a problem of omitted variables. Nevertheless, this potential problem needs to be considered when interpreting results.

Table 6.5. Craggit estimation of intensity of collaboration with industry

Industry collaboration intensity (2009-2013)	No Centrality		Degree			
	(1)		(2)		(3)	
	Tier1	Tier2	Tier1	Tier2	Tier1	Tier2
Industry collaboration intensity (2004-2008)	0.4458*** (0.1165)	0.3579 (0.2484)	0.4563*** (0.1200)	0.3259 (0.2530)	0.4561*** (0.1199)	0.2802 (0.2025)
Number of pubs (log) (2004-2008)	0.7133*** (0.1122)	-1.3882** (0.5444)	0.5874*** (0.1813)	-0.7842 (0.6362)	0.5877*** (0.1811)	-0.3753 (0.6986)
Quality (2004-2008)	0.0089 (0.0231)	-0.1093 (0.1186)	0.0076 (0.0236)	-0.0797 (0.1142)	0.0075 (0.0237)	-0.0438 (0.1011)
Brazil	0.4351* (0.2496)	4.0722** (1.8118)	0.4063 (0.2535)	4.1814** (1.8458)	0.4014 (0.2557)	3.4978** (1.5916)
Chile	-0.1100 (0.3198)	-0.1071 (1.7456)	-0.1391 (0.3101)	0.2295 (1.8384)	-0.1373 (0.3105)	0.4643 (1.6820)
Mexico	0.1539 (0.3484)	2.6570 (1.6268)	0.1092 (0.3522)	2.7322* (1.6432)	0.1014 (0.3568)	2.0342 (1.5811)
Agricultural	0.6438** (0.3053)	2.0409 (3.2737)	0.5804* (0.3154)	2.1106 (3.3053)	0.5738* (0.3139)	1.5893 (2.9091)
Engineering	1.3783*** (0.3138)	9.8622** (4.5963)	1.3203*** (0.3249)	9.6726** (4.5390)	1.3133*** (0.3232)	8.1659** (3.6974)
Environmental	0.5307* (0.2940)	5.6520 (3.9083)	0.4536 (0.3191)	5.7809 (3.9234)	0.4492 (0.3198)	4.7709 (3.3091)
Geosciences	0.7910** (0.3977)	8.9886** (4.4545)	0.6281 (0.4755)	9.2261** (4.4834)	0.6270 (0.4747)	8.1466** (3.6893)
University	0.5258** (0.2332)	-1.3041 (1.6462)	0.5694** (0.2288)	-1.3348 (1.5327)	0.5643** (0.2287)	-1.0294 (1.3925)
Normalized LA degree (2004-2008)			2.2307 (2.2971)	-9.6152 (7.6304)	2.1664 (2.2777)	-11.7398 (7.1588)
Normalized national degree (2004-2008)					0.1155 (0.4597)	2.7692** (1.2593)
Constant	-5.4481*** (0.8585)	-4.1878 (6.7205)	-4.9338*** (1.0633)	-6.5906 (7.2772)	-4.9380*** (1.0650)	-7.2720 (6.5955)
Sigma	1.8431*** (0.3091)		1.8317*** (0.3028)		1.6888*** (0.2678)	
N	324		324		324	

Notes: Clustered errors at the institution level. Standard errors in parentheses. * Coefficient is statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level; no asterisk means the coefficient is not different from zero with statistical significance.

Table 6.5. (cont.) Cragg estimation of intensity of collaboration with industry

Industry collaboration intensity (2009-2013)	Betweenness				Closeness			
	(4)		(5)		(6)		(7)	
	Tier1	Tier2	Tier1	Tier2	Tier1	Tier2	Tier1	Tier2
Industry collaboration intensity (2004-2008)	0.4547*** (0.1195)	0.3620 (0.2493)	0.4543*** (0.1193)	0.3286* (0.1964)	0.4747*** (0.1201)	0.3283 (0.2756)	0.4651*** (0.1199)	0.3285 (0.2761)
Number of pubs (log) (2004-2008)	0.6213*** (0.1436)	-1.608** (0.6502)	0.6213*** (0.1437)	-1.1141 (0.7083)	0.5950*** (0.1283)	-1.0672* (0.6322)	0.6152*** (0.1311)	-1.1270 (0.7439)
Quality (2004-2008)	0.0068 (0.0242)	-0.1238 (0.1282)	0.0067 (0.0243)	-0.0628 (0.1155)	0.0125 (0.0240)	-0.0901 (0.1258)	0.0116 (0.0248)	-0.0914 (0.1260)
Brazil	0.4261* (0.2538)	3.9814** (1.7920)	0.4242* (0.2555)	3.2640** (1.5256)	0.3280 (0.2498)	4.3527** (1.7759)	0.2728 (0.2660)	4.4197** (1.8297)
Chile	-0.1557 (0.3134)	-0.2540 (1.7903)	-0.1540 (0.3139)	-0.1488 (1.6503)	-0.2142 (0.3024)	0.3227 (1.8479)	-0.2593 (0.3078)	0.3378 (1.8600)
Mexico	0.1037 (0.3549)	2.5558 (1.6259)	0.0976 (0.3595)	1.8905 (1.5557)	0.1463 (0.3475)	2.4304 (1.6192)	0.0985 (0.3636)	2.5040 (1.7214)
Agricultural	0.5909* (0.3092)	1.9563 (3.1977)	0.5875* (0.3081)	1.6884 (2.8596)	0.7687** (0.3171)	1.5155 (3.2933)	0.7614** (0.3166)	1.5376 (3.2565)
Engineering	1.2916*** (0.3322)	9.6009** (4.4741)	1.2859*** (0.3315)	8.0995** (3.8140)	1.6320*** (0.3806)	8.7661* (5.0669)	1.6008*** (0.3745)	8.8415* (5.0038)
Environmental	0.4138 (0.3210)	5.2310 (3.8456)	0.4085 (0.3229)	4.3991 (3.4247)	0.5016* (0.2946)	5.7184 (3.8693)	0.4992* (0.3005)	5.6818 (3.8683)
Geosciences	0.6367 (0.4544)	8.6376* (4.4138)	0.6302 (0.4531)	7.6083** (3.7590)	0.7193* (0.4157)	8.8090** (4.4590)	0.7683* (0.4186)	8.8182** (4.4389)
University	0.5589** (0.2328)	-1.3637 (1.6437)	0.5536** (0.2333)	-1.2150 (1.4357)	0.5463** (0.2252)	-1.0862 (1.4468)	0.5246** (0.2227)	-1.1351 (1.4860)
Normalized LA betweenness (2004-2008)	2.5164 (2.2906)	3.9438 (9.1984)	2.4597 (2.3112)	-1.4633 (7.5636)				
Normalized national betweenness (2004-2008)			0.1604 (0.5978)	3.2971** (1.4762)				
Normalized LA closeness (2004-2008)					23.2792 (15.6393)	-59.7806 (92.7442)	20.9869 (15.2277)	-57.3279 (93.0904)
Normalized national closeness (2004-2008)							6.6198 (6.3685)	-4.4030 (23.1923)
Constant	-5.0130*** (0.9727)	-2.7913 (6.4384)	-5.0142*** (0.9723)	-3.7046 (6.2620)	-5.9531*** (0.9233)	-3.0014 (7.4359)	-6.2486*** (0.9928)	-2.6051 (7.9800)
Sigma	1.8343*** (0.3038)		1.6925*** (0.2636)		1.8249*** (0.3175)		1.8282*** (0.3178)	
N	324		324		324		324	

Notes: Clustered errors at the institution level. Standard errors in parentheses. * Coefficient is statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level; no asterisk means the coefficient is not different from zero with statistical significance.

In Table 6.A.1 (in the Appendix) we present the results of the estimations of a two-step Heckman model for specifications 3, 5 and 7 of the model. Based on the results of the *Craggit* estimations we use the size of the RD as the exclusion variable, i.e. affecting the participation decision but not the intensity equation. Most of the results of the previous estimations hold. However, in this set of estimations, the previous collaboration with industry increases not only the likeliness of participation in co-publication but also the intensity of these collaborations. An extra 1% of co-publications with industry in one period increase these activities in 0.35-0.45% in the next period. Despite this change, the degree and betweenness values of the RDs in their national scientific networks have a positive effect on the intensity of collaborations with industry.

6.7. Conclusions

In this chapter, we use a combination of bibliometric, social network and econometric techniques to increase the understanding of LA scientific systems and its relationship with the private sector. We studied recent trends in the scientific outcome, the linkages that exist between RDs within and between LA countries, and RDs collaboration activities with industry.

We found that the LA share of global scientific publications started to increase at a higher rate since 1993, thus revealing a trend for convergence with the world leading regions. This increase has been mainly driven by Brazil and most notably in subject areas such as Agricultural, and Plant & Animal Sciences. Moreover, when analysing the relative scientific output normalized by GDP (Docs/GDP) and population (Docs/Pop), the results show that in the most recent years Chile, Uruguay, Argentina, and Brazil have levels of scientific productivity higher than the world average. Furthermore, specialisation of scientific systems in LA tends to follow economic specialisation, focusing on scientific fields related to natural resources. However, in the last decade, most LA countries have an average industry collaboration percentage below 1%. This is a low number when compared to the rest of the world. There are differences between fields (Engineering and Geosciences show higher levels than other sciences) but in general, collaborations between science and industry, measured as co-publications, are scarce.

The growth of scientific production can also be appreciated by the increasing number of RDs embedded in the LA scientific networks. However, the structures of these networks are not evolving in the same way. We find preliminary evidence that suggests that LA Geosciences and Agricultural Sciences networks are evolving towards structures that facilitate both knowledge creation and diffusion. It is worth noting that collaborations between RDs of different LA countries remain low. In most of the fields studied, linkages between LA countries are scarce even when these countries tend to specialise in similar scientific fields. Understanding if this lack of

integration between LA scientific institutions is harming potential gains of complementary knowledge is a matter of further research.

The main finding is that the RDs that have a more diverse set of knowledge sources, within their scientific discipline, are the ones that are working more closely with industry. Besides possessing different sources of complementary knowledge within the same discipline, that can tackle more effectively private sector challenges; firms may perceive these RDs as having a higher reputation and stronger research capabilities. Furthermore, by being in brokerage positions, RDs are not only more likely to be early recipients of information from diverse groups but also occupy a privileged position from which they can assess the relevance of new information. This timing advantage may be a crucial element for collaboration with industry. Although interesting, we cannot determine which of these is dominating.

Complementing this analysis with qualitative approaches and primary data which consider other types of technology transfer activities and sources of funding for research would certainly improve the understanding of LA knowledge production, transfer and diffusion systems. Furthermore, splitting the publication analysis of science-industry linkages between basic and applied science, or at the level of technologies, rather than scientific fields, is a matter of further research.

6.8. Appendix

6.8.1. Indicators used

6.8.1.1. Specialisation and citation impact measures

i. Revealed Specialisation Intensity

This index assesses the relative specialisation of each country in a given area. It is an adaptation of the Revealed Comparative Advantage Index, proposed by Balassa (1965), which compares the specialisation intensity of a subject area s in country i with the equivalent relative specialisation intensity of that subject area for all countries worldwide:

$$RSI = \frac{P_{is}/P_i}{P_s/P}$$

Where P_{is} accounts for the number of publications in subject area s in country i , P_i accounts for the total number of publications in that same country i , P_s accounts for the total number of publications in subject area s worldwide, and finally P accounts for the total number of publications in the world.

ii. Specialisation Intensity Index

This measure provides a ratio which in the numerator displays the square of the difference between specialisation intensity of class s in country i and specialisation intensity of that class in the world, while the same denominator shows the sum of the weighting of all subject areas in country i , with this ratio summed up across all s subject areas. This Chi-square of sectoral specialisation is adapted from Laursen (2000) and provides a concentration measure that grows with the specialisation intensity of a country:

$$SII = \sum_s \left(\frac{[(X_{si}/\sum_s X_{si}) - (\sum_i X_{si}/\sum_s \sum_i X_{si})]^2}{(\sum_i X_{si}/\sum_s \sum_i X_{si})} \right)$$

Quality Citation Index

This score calculates the mean citation rate of a country’s set of publications in a specific subject area, period, and document type, divided by the mean citation rate of all publications in that subject area/period/document type:

$$QCI = \frac{\sum_{i=1}^P c_i}{\sum_{i=1}^P [\mu_f]_i}$$

iii. Quality Top 10% Index

This index shows the proportion of publications belonging to the top ten percent most cited documents in a given subject category, year and publication type:

$$PPtop10\% (\%) = \frac{PPtop10\% (n)}{P}$$

6.8.1.2. *Centrality measures*

i. Degree:

This measure of centrality accounts for the total number of links that a node has in a network. In the case of the networks that we are studying it will account for the total number of different research partners with whom each RD collaborates. RDs with higher degree number could be considered popular among their peers, enjoying benefits from reputation. Furthermore, they also hold what could be regarded as a more diversified set of research partners. However, regularly, maintaining links is a costly endeavor, and then we would expect to find limits on the utility of getting new linkages. We use the normalized version of the indicator implemented by the *igraph* package of the R software. Formally:

$$C_D(i) = \frac{1}{(n-1)} \sum_{j=1}^n l(i,j) \quad (1)$$

Where $l(i,j) = \begin{cases} 1 & \text{if there is an edge between } i \text{ and } j \\ 0 & \text{otherwise} \end{cases}$, and n is the number of nodes of the network.

ii. Betweenness:

This index accounts for the total number of shortest paths⁵⁵ in which a node is involved. Under the assumption that shortest paths are preferred in the diffusion of knowledge in a network, RT with higher betweenness values may be connecting knowledge from two very distant RD, broadening the scope of potential sources of information and allowing them to play a role of broker of knowledge. We use the normalized version of the indicator implemented by the *igraph* package of the R software. Formally:

$$C_B(i) = \frac{1}{(n(n-3)(n+2))} \sum_{j \neq k} \frac{g_{jk}(i)}{g_{jk}} \quad (2)$$

Where n is the number of nodes of the network, $g_{jk}(i)$ is the number of shortest paths that pass through node i , and g_{jk} is the total number of shortest paths.

iii. Closeness:

This index is defined by the inverse of the average shortest path to all other nodes in the network. An RD with higher values of closeness would require less effort to reach any other source of information. At the same time, at least theoretically, it could access new knowledge more quickly than others. We use the normalized version of the indicator implemented by the *igraph* package of the R software. Formally:

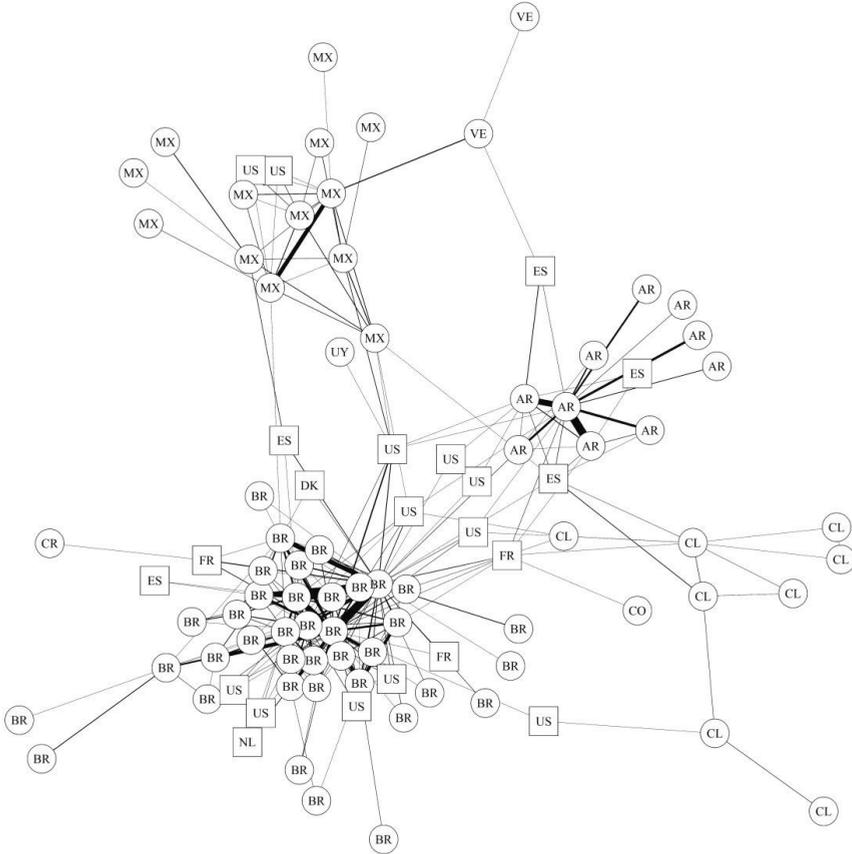
$$C_C(i) = (n - 1) \left[\sum_{j=1}^n d(i, j) \right]^{-1} \quad (3)$$

Where n is the number of nodes of the network, $d(i, j)$ is the length of the shortest path between nodes i and j .

⁵⁵ The shortest path is the minimum distance, accounted by links, between two nodes of a network.

6.8.2. Network graphs of all scientific areas in 2004-2008 and 2009-2013^{56,57}

Figure 6.A.1. Network structure of Agricultural sciences (2004-2008)



⁵⁶ Networks are visualized using the Fruchterman-Reingold algorithm. Circle nodes are for LA RD. Square nodes are for non-LA RD. Edge thickness represent the normalized number of co-publications.

⁵⁷ AR=Argentina, BR=Brazil, CL=Chile, CO=Colombia, CR=Costa Rica, CU= Cuba, MX=Mexico, PA=Panama, PE=Peru, UR=Uruguay, VE=Venezuela. AT=Austria, AU=Australia, BE=Belgium, CA=Canada, CH=Switzerland, CN=China, CZ=Czech Republic, DE=Germany, DK= Denmark, EG=Egypt, ES=Spain, FI=Finland, FR=France, IL=Israel, IT=Italy, JP=Japan, NG=Nigeria, NL=Netherlands, NO=Norway, NZ=New Zealand, PL=Poland, PT=Portugal, RU=Russia, SE=Sweden, UK=United Kingdom, US=United States, ZA=South Africa.

Figure 6.A.2. Network structure of Agricultural sciences (2009-2013)

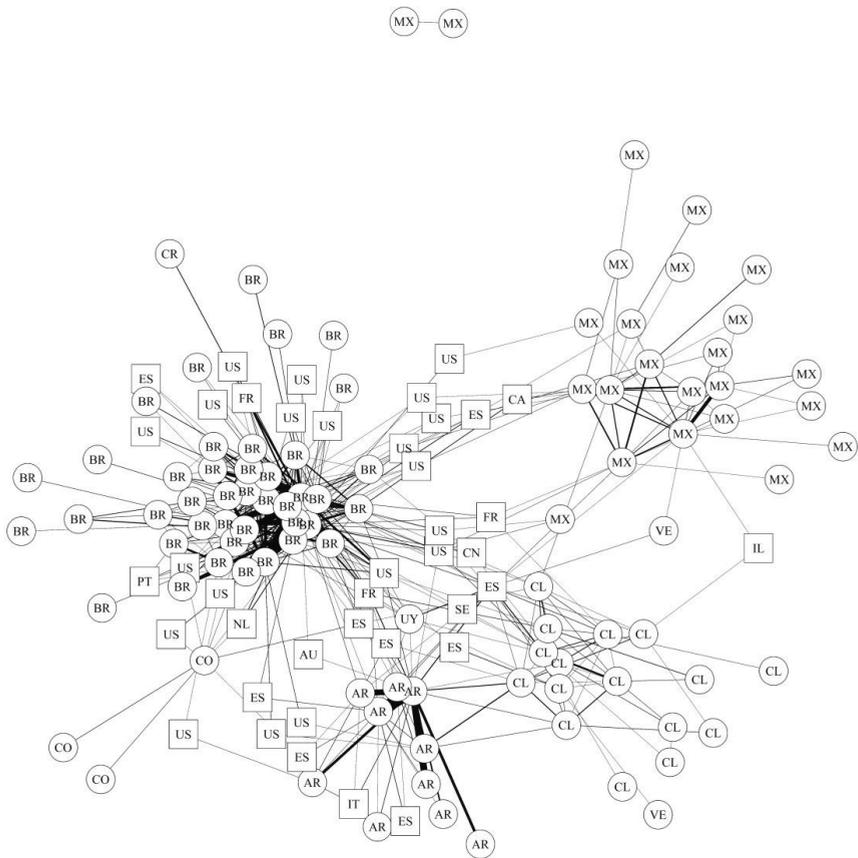


Figure 6.A.3. Network structure of Engineering sciences (2004-2008)

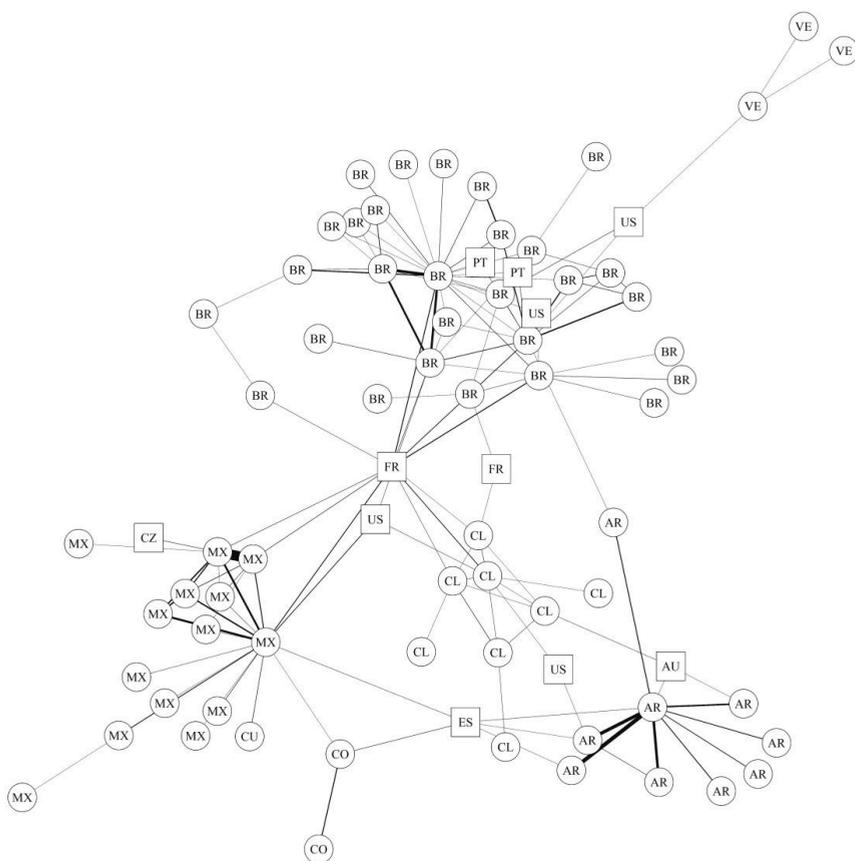


Figure 6.A.4. Network structure of Engineering sciences (2009-2013)

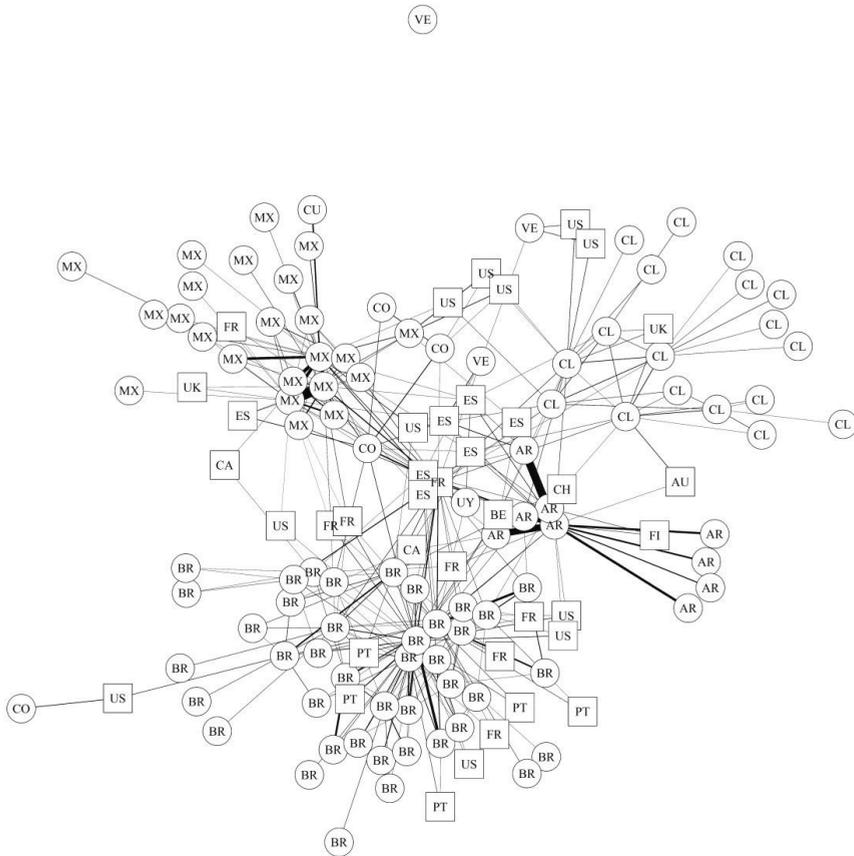


Figure 6.A.5. Network structure of Environmental sciences (2004-2008)

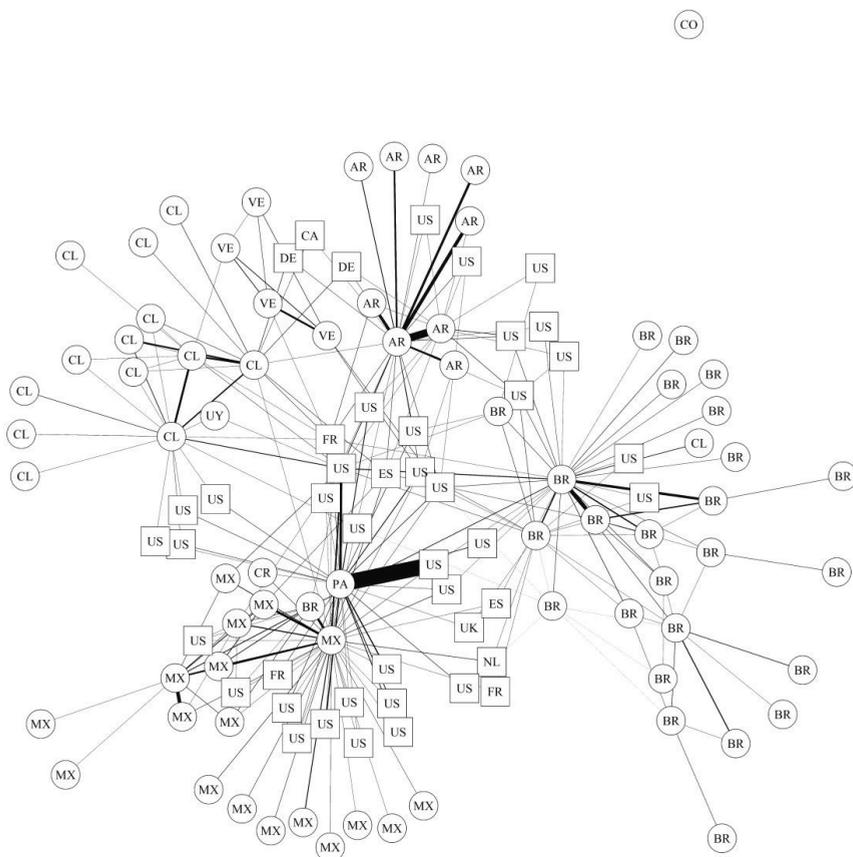
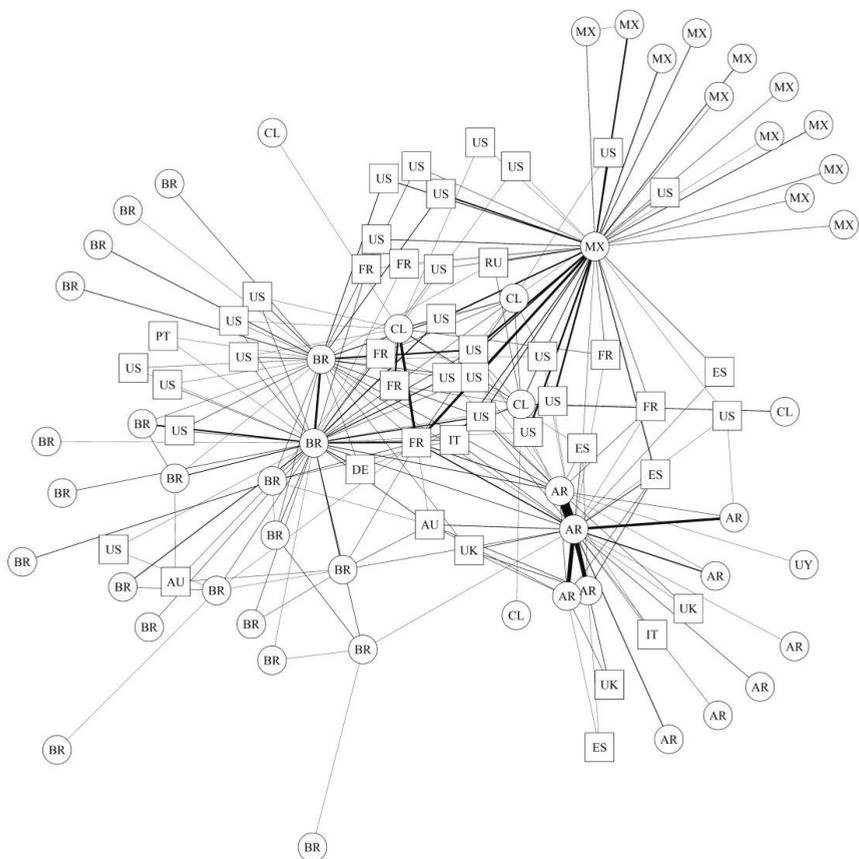


Figure 6.A.7. Network structure of Geosciences (2004-2008)



6.8.3. Consistency check of econometric estimation

Table 6.A.1. Two-step Heckman of intensity of collaboration with industry

	No Centrality (8)		Degree (9)		Betweenness (10)		Closeness (11)	
	Participation	Intensity	Participation	Intensity	Participation	Intensity	Participation	Intensity
Industry collaboration intensity (2004-2008)	0.4458*** (0.1696)	0.3541** (0.1476)	0.4561*** (0.1449)	0.3358* (0.1788)	0.4543*** (0.1664)	0.3361** (0.1661)	0.4651*** (0.1630)	0.3432* (0.2035)
Number of publications (log) (2004-2008)	0.7133*** (0.1440)	--	0.5877*** (0.2073)	--	0.6213*** (0.1600)	--	0.6152*** (0.1620)	--
Quality (2004-2008)	0.0089 (0.0286)	-0.0303 (0.0427)	0.0075 (0.0295)	-0.0239 (0.0414)	0.0067 (0.0279)	-0.0223 (0.0447)	0.0116 (0.0283)	-0.0277 (0.0450)
Brazil	0.4351 (0.3095)	0.7875** (0.3311)	0.4014 (0.3009)	0.7260** (0.3467)	0.4242 (0.2990)	0.7159** (0.3154)	0.2728 (0.3819)	0.3133** (0.3047)
Chile	-0.1100 (0.3629)	-0.2022 (0.4267)	-0.1373 (0.3637)	-0.1445 (0.4296)	-0.1540 (0.4596)	-0.2293 (0.4695)	-0.2593 (0.4360)	-0.2942 (0.5341)
Mexico	0.1539 (0.4529)	0.2080 (0.5426)	0.1014 (0.4657)	-0.0575 (0.5920)	0.0976 (0.4310)	-0.0574 (0.4742)	0.0985 (0.4906)	0.0851 (0.4725)
Agricultural	0.6438** (0.3241)	0.1663 (0.1977)	0.5738* (0.3031)	0.0272 (0.2647)	0.5875 (0.3590)	0.1271 (0.2287)	0.7614** (0.3181)	0.1090 (0.4931)
Engineering	1.3783*** (0.3007)	1.5513*** (0.1977)	1.3133*** (0.3031)	1.3846*** (0.4184)	1.2859*** (0.3590)	1.3639*** (0.3449)	1.6008*** (0.4287)	1.4349 (0.9541)
Environmental	0.5307 (0.3376)	0.3684 (0.2536)	0.4492 (0.3295)	0.2370 (0.2591)	0.4085 (0.3745)	0.2347 (0.3745)	0.4992 (0.3130)	0.3665 (0.2797)
Geosciences	0.5017 (0.5017)	1.4312* (0.7351)	0.6270 (0.4670)	1.4089** (0.6332)	0.6302 (0.5233)	1.4013** (0.6258)	0.7683 (0.4904)	1.3811** (0.6776)
University	0.5258 (0.3260)	-0.1325 (0.5253)	0.5643* (0.3079)	-0.1951 (0.6095)	0.5536* (0.2936)	-0.1885 (0.4921)	0.5246* (0.2819)	-0.1599 (0.6050)
Normalized LA degree (2004-2008)								
Normalized national degree (2004-2008)								
Normalized LA betweenness (2004-2008)								
Normalized national betweenness (2004-2008)								
Normalized LA closeness (2004-2008)								
Normalized national closeness (2004-2008)								
Constant	-5.4481*** (1.1140)	-0.8432 (0.7192)	-4.9380*** (1.1832)	-0.8648 (1.0783)	-5.0142*** (1.0536)	-0.8211 (0.7752)	-6.2486*** (1.1215)	-1.0981 (2.8895)
lambda	0.8369*** (0.2314)		0.8634* (0.4931)		0.8539** (0.3790)		0.8087 (0.5537)	
N	324		324		324		324	
sigma	1.1598		1.1307		1.1195		1.1389	
chi2	61.8592		70.3405		67.2307		68.6761	
p	0.0000		0.0000		0.0000		0.0000	
rho	0.7215		0.7636		0.7628		0.7100	

Notes: Coefficients reported are marginal effects. Bootstrapped clustered errors at the institution level (100 repetitions). *Coefficient is statistically significant at the 10 percent level; ** at the 5 percent level; *** at the 1 percent level; no asterisk means the coefficient is not different from zero with statistical significance.

Chapter 7

General discussion

Science offers humanity the promise of a better life. Scientific advances throughout history have helped humans to understand the natural world, fight diseases, develop technologies and increase productivity. However, there have been vast inequalities in scientific output across nations. According to the most recent UNESCO Science Report (2015) the European Union still leads the world for publications in Web of Science (34%), followed by the USA with 25%, and China with its recent meteoric rise over the past years to 20% of the world total. African and Latin American scientific systems, which are the main focus of this thesis, represent 2.6% and 5.1% of the world share respectively.

In this work, we hope to provide a better understanding of how researchers, institutions and countries in low and middle-income regions can advance their scientific capacity⁵⁸ and employ it for higher research citation impact, international collaboration, alignment between research priorities and social needs, and university-industry interactions. We raised different research questions and used different quantitative approaches in order to provide insights for policy makers to create and manage policies that promote the absorption, creation, diffusion, application, and retention of scientific knowledge in the Global South.

In the second chapter, we observed a U-shaped relationship between research citation impact and GDPpc. We argue that a possible explanation is that, as countries progress from low-income to middle-income levels they enjoy more resources and larger scientific communities, but do not engage in overseas collaboration so much. As a consequence, this is reflected in the lower levels of impact on average. In the econometric section, we found that previous citation impact, level of international collaboration and total publications in a specific scientific field are important determinants of citation impact among all nations. However, specialisation in particular scientific fields seems significantly more important in the Global South than in the Global North. Overall, if we assume that high impact research should be a key objective for low- and middle-income countries, these findings imply that these countries would better concentrate their resources in generating higher critical masses in specific fields, in addition to pursuing long-lasting international collaboration partnerships.

In the third chapter, we move from the macro to the micro level, and we study the characteristics of researchers working in Africa that have produced highly cited publications and compare them to researchers that didn't produce highly cited work in the same period. Overall our results suggest that, on average, researchers who produce more scientific publications in a year, collaborate more often with non-African partners and did their highest qualification in a North American or UK

⁵⁸ In the first chapter, it is discussed what is meant by scientific capacity.

university, have a higher probability of producing highly cited research. The central insight is that, in Africa, high impact researchers are the ones that are more integrated with networks of researchers from the global scientific “core”, not the ones that have fewer challenges during their career or the ones that collaborate more locally. Therefore, a policy implication from this chapter is that individual research assessment in these contexts should go beyond evaluating research by the impact it has in the overall scientific community (through publications and citations in international journals). Otherwise, incentives will be in place to stimulate winners that are already well connected with the global scientific elite.

Since international research collaboration positively affects researchers impact and is seen as one of the most efficient means to build research capacity and to create learning opportunities, a subsequent research question in the thesis is what are the main drivers of international research collaboration at the individual level. Using the same survey dataset, we address this question by studying the characteristics of African researchers that collaborate both frequently and infrequently with foreign and non-African researchers. Our results suggest that researchers who did their highest qualification (PhD) outside of Africa, had the opportunity to move abroad (over the past three years), were a primary recipient of research funding (over the past three years), received a higher share of international funding (over the past three years), and didn't report they had lack of mentorship are more likely to collaborate more often with researchers outside Africa. The most important policy insight from this chapter relates to the importance of mobility and adequate mentorship for developing international collaborations. Therefore, actions from international research funders to increase the number of scholarships available to students in the Global South to do PhDs abroad (and comeback), research visitings or go to international conferences, seem essential to augment the intensity of knowledge diffusion between regions.

In the fifth chapter, the main objective of the study is to evaluate whether the amount of research produced on various medical conditions by African researchers is related to their countries burden of disease of the local populations. We found that in sub-Saharan Africa most diseases with high disease burden are also the ones with relatively more research effort. The region where there is a higher positive association between disease burden and research effort is Eastern Africa. Northern Africa is the region where these two dimensions are less aligned. We also found that most international research funders (public non-African and philanthropic) use their resources to fund specific diseases such as HIV/AIDS, tuberculosis and parasitic and vector diseases. On average, these diseases are also the ones with higher relative disease burden in the regions with higher dependence on international funding. These results are relevant because they seem to contradict some literature (Binka 2005; Gaillard 1994) that argues that global health research projects in low-

income regions often lead to inappropriate projects, unrelated to local research needs, and derive conclusions that do not have any direct local benefit. The major policy implication from this chapter is that high levels of dependence on international donors for health research capacity is not necessarily associated with less alignment between regional disease burden and regional medical research effort.

Finally, in the last empirical chapter, we use a combination of bibliometric, social network and econometric techniques to increase our knowledge of how research institutions interact with the private sector in Latin America. The main finding is that the scientific institutions that have a more diverse set of knowledge sources at the national level, within their scientific discipline, are the ones that are working more intensively with industry. Besides possessing different sources of complementary knowledge within the same discipline that can tackle more effectively private sector challenges; firms may perceive these scientific institutions as having a higher reputation and stronger research capabilities. Furthermore, by being in brokerage positions, scientific institutions are not only more likely to be early recipients of information from diverse groups but also occupy a privileged position from which they can assess the relevance of new information. This timing advantage may be a crucial element for collaboration with industry. A direct policy implication is that by incentivising research departments to do more research collaborations between national universities, governments in Latin America can also increase the attractiveness of those research departments for university-industry interactions.

7.1. Future research

Research is a voyage into the unknown. However, the uncertainty does not vitiate what economics and bibliometrics can contribute to the science of science policy (Fealing et al., 2011; Fortunato et al., 2018). In chapter two, we confirmed that international research collaboration is crucial for citation impact. Since in the Global South, international research collaboration is usually driven via major research programmes (chapter five), it seems essential to try to develop impact evaluation settings for future initiatives, and see what lessons can be taken forward to better inform the design and implementation of new programmes aimed at strengthening research capacity.

In chapter three and four we used data from a single survey sent to “all” researchers in Africa that authored at least one publication between 2005 and 2015. Repeating this exercise periodically and systematically, like for example the “community innovation surveys”, would potentially allow following the career evolution of the same researchers and give insights that bibliometric indicators cannot.

In the third, fourth and sixth chapter it was also clear that more research is needed on the effects of knowledge exchange through mobility between sectoral, national and international borders. People can be seen as the most important agents of knowledge diffusion. Therefore tracking their activities and movements can shed light on major issues such as collective learning processes (Müller et al. 2018), the effects of brain-drain (Docquier & Rapoport 2012), and university-industry knowledge transfer.

Lastly, as shown in chapter five, economic analysis can lay out ways for policy makers to think about how much and in what ways to support scientific research directed towards collective goals or societal needs (Sarewitz & Pielke Jr. 2007). A potential area for future research is to analyse how research activities have been prioritized in the past in certain regions, and how such priorities were aligned with the main economic and societal needs of those regions (e.g. Ciarli & Ràfols, 2018). This kind of approach can help governments, funding bodies and international organisations around the world to analyse how science can better address societal challenges such as access to food and energy, and climate change.

Bibliography

- Abbasi, A., Altmann, J., & Hossain, L. (2011). 'Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures', *Journal of Informetrics*, 5/4: 594–607. DOI: 10.1016/j.joi.2011.05.007
- Abbasi, A., Hossain, L., & Leydesdorff, L. (2012). 'Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks', *Journal of Informetrics*, 6/3: 403–12. DOI: 10.1016/j.joi.2012.01.002
- Abramo, G., Cicero, T., & D'Angelo, C. A. (2014). 'Are the authors of highly cited articles also the most productive ones?', *Journal of Informetrics*, 8/1: 89–97. DOI: 10.1016/j.joi.2013.10.011
- Abramo, G., D'Angelo, C. A., & Costa, F. Di. (2011). 'University-industry research collaboration: A model to assess university capability', *Higher Education*, 62/2: 163–81. DOI: 10.1007/s10734-010-9372-0
- Abramo, G., D'Angelo, C. A., Di Costa, F., & Solazzi, M. (2009). 'University-industry collaboration in Italy: A bibliometric examination', *Technovation*, 29/6–7: 498–507. DOI: 10.1016/j.technovation.2008.11.003
- Abramo, G., D'Angelo, C. A., & Murgia, G. (2013). 'Gender differences in research collaboration', *Journal of Informetrics*, 7/4: 811–22. Elsevier. DOI: 10.1016/j.joi.2013.07.002
- Abramovitz, M. (1986). 'Catching Up, Forging Ahead, and Falling Behind', *The Journal of Economic History*, 46/2: 385–406.
- Adams, J. (2013). 'Collaborations: The fourth age of research', *Nature*, 497/7451: 557–60. DOI: 10.1038/497557a
- Adams, J., Gurney, K., Hook, D., & Leydesdorff, L. (2013). 'International collaboration clusters in Africa', *Scientometrics*, 98/1: 547–56. DOI: 10.1007/s11192-013-1060-2
- Agarwal, P., & Searls, D. B. (2009). 'Can literature analysis identify innovation drivers in drug discovery?', *Nature Reviews Drug Discovery*, 8/11: 865–78. DOI: 10.1038/nrd2973
- Ahuja, G. (2000). 'The duality of collaboration: Inducements and opportunities in the formation of interfirm linkages', *Strategic management journal*, 21/3: 317–343.
- Aizer, A., & Stroud, L. (2010). *Education, Knowledge and the Evolution of Disparities in Health*. National Bureau of Economic Research Working Paper Series. Cambridge, MA.
- Akhondzadeh, S. (2013). 'Iranian science shows world's fastest growth: ranks 17th in science production in 2012', *Avicenna Journal of Medical Biotechnology*, 5/3: 139.
- Aksnes, D. W. (2003). 'Characteristics of highly cited papers', *Research Evaluation*, 12/3: 159–70. DOI: 10.3152/147154403781776645
- Aksnes, D. W., & Rip, A. (2009). 'Researchers' perceptions of citations', *Research Policy*, 38/6: 895–905. DOI: 10.1016/j.respol.2009.02.001
- Albuquerque, E. da M. e. (2004). 'Science and Technology Systems in Less Developed Countries'. Moed H. F., Glänzel W., & Schmoch U. (eds), pp. 759–78. Springer Netherlands.
- Allen, T. D., Eby, L. T., Poteet, M. L., Lentz, E., & Lima, L. (2004). 'Career Benefits Associated With Mentoring for Proteges: A Meta-Analysis.', *Journal of Applied Psychology*, 89/1: 127–36. DOI: 10.1037/0021-9010.89.1.127
- Alonso, P. L., Lindsay, S. W., Armstrong, J. R., Conteh, M., Hill, A. G., David, P. H., Fegan, G., et al. (1991). 'The effect of insecticide-treated bed nets on mortality of Gambian children.', *Lancet (London, England)*, 337/8756: 1499–502.

- AOSTI (African Observatory for Science, T. and I. (2014). *Assessment of Scientific Production in the African Union Member States 2005-2010*. Retrieved April 9, 2015, from <<http://aosti.org/index.php/publi/66-assessment-of-scientific-production-in-the-african-union-member-states-2005-2010>>
- Aristei, D., Vecchi, M., & Venturini, F. (2016). 'University and inter-firm R&D collaborations: propensity and intensity of cooperation in Europe', *The Journal of Technology Transfer*, 41/4: 841–71. Springer US. DOI: 10.1007/s10961-015-9403-1
- Arrow, K. J. (1962). 'The Economic Implications of Learning by Doing', *The Review of Economic Studies*, 29/3: 155. Oxford University Press. DOI: 10.2307/2295952
- Arvanitis, R., & Gaillard, J. (2014). *Research Collaboration between Europe and Latin America: Mapping and Understanding partnership*. Editions des archives contemporaines.
- Arza, V., & Vazquez, C. (2012). 'Firms' linkages with universities and public research institutes in Argentina: factors driving the selection of different channels', *Prometheus*, 30/1: 47–72. DOI: 10.1080/08109028.2012.671287
- Atal, I., Trinquart, L., Ravaud, P., & Porcher, R. (2018). 'A mapping of 115,000 randomized trials revealed a mismatch between research effort and health needs in non-high-income regions', *Journal of Clinical Epidemiology*, 98: 123–32. DOI: 10.1016/j.jclinepi.2018.01.006
- Balassa, B. (1965). 'Trade Liberalisation and "Revealed" Comparative Advantage', *The Manchester School*, 33/2: 99–123. DOI: 10.1111/j.1467-9957.1965.tb00050.x
- Balconi, M., & Laboranti, A. (2006). 'University–industry interactions in applied research: The case of microelectronics', *Research Policy*, Triple helix Indicators of Knowledge-Based Innovation Systems, 35/10: 1616–30. DOI: 10.1016/j.respol.2006.09.018
- Barabasi, A. L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). 'Evolution of the social network of scientific collaborations', *Physica A: Statistical Mechanics and its Applications*, 311/3–4: 590–614. DOI: 10.1016/S0378-4371(02)00736-7
- Barnard, H., Cowan, R., & Müller, M. (2012). 'Global excellence at the expense of local diffusion, or a bridge between two worlds? Research in science and technology in the developing world', *Research Policy*, 41/4: 756–69. DOI: 10.1016/j.respol.2011.12.002
- Baum, J. a. C., Cowan, R., & Jonard, N. (2010). 'Network-Independent Partner Selection and the Evolution of Innovation Networks', *Management Science*, 56/11: 2094–110. DOI: 10.1287/mnsc.1100.1229
- Beaver, D. D. (2001). 'Reflections on Scientific Collaboration (and its study): Past, Present, and Future', *Scientometrics*, 52/3: 365–77. DOI: 10.1023/A:1014254214337
- Bekkers, R., & Bodas Freitas, I. M. (2008). 'Analysing knowledge transfer channels between universities and industry: To what degree do sectors also matter?', *Research Policy*, 37/10: 1837–53. DOI: 10.1016/j.respol.2008.07.007
- Belkhodja, O., & Landry, R. (2007). "'The Triple-Helix collaboration: Why do researchers collaborate with industry and the government? What are the factors that influence the perceived barriers?'"', *Scientometrics*, 70/2: 301–32. DOI: 10.1007/s11192-007-0205-6
- Bernardes, A. T., & Albuquerque, E. da M. e. (2003). 'Cross-over, thresholds, and interactions between science and technology: lessons for less-developed countries', *Research Policy*, 32/5: 865–85. DOI: 10.1016/S0048-7333(02)00089-6
- van den Besselaar, P., & Sandström, U. (2016). 'Gender differences in research performance and its impact on careers: a longitudinal case study', *Scientometrics*, 106/1: 143–62. Springer Netherlands. DOI: 10.1007/s11192-015-1775-3

- Binka, F. (2005). 'Editorial: North-South research collaborations: a move towards a true partnership?', *Tropical Medicine and International Health*, 10/3: 207–9. Wiley/Blackwell (10.1111). DOI: 10.1111/j.1365-3156.2004.01373.x
- Boaz, A., Fitzpatrick, S., & Shaw, B. (2009). 'Assessing the impact of research on policy: a literature review', *Science and Public Policy*, 36/4: 255–70. Oxford University Press. DOI: 10.3152/030234209X436545
- Bornmann, L., Bauer, J., & Schlagberger, E. M. (2017). 'Characteristics of highly cited researchers 2015 in Germany', *Scientometrics*, 111/1: 543–5. Springer Netherlands. DOI: 10.1007/s11192-017-2248-7
- Bornmann, L., & Leydesdorff, L. (2013). 'Macro-Indicators of Citation Impacts of Six Prolific Countries: InCites Data and the Statistical Significance of Trends', *PLoS ONE*, 8/2. DOI: 10.1371/journal.pone.0056768
- Bornmann, L., Schier, H., Marx, W., & Daniel, H.-D. (2012). 'What factors determine citation counts of publications in chemistry besides their quality?', *Journal of Informetrics*, 6/1: 11–8. DOI: 10.1016/j.joi.2011.08.004
- Boshoff, N. (2009). 'Neo-colonialism and research collaboration in Central Africa', *Scientometrics*, 81/2: 413–34. Springer Netherlands. DOI: 10.1007/s11192-008-2211-8
- Bosquet, C., & Combes, P.-P. (2013). 'Are academics who publish more also more cited? Individual determinants of publication and citation records', *Scientometrics*, 97/3: 831–57. Springer Netherlands. DOI: 10.1007/s11192-013-0996-6
- Bound, K., & Thornton, I. W. (2012). *Our frugal future: Lessons from India's innovation system*. London.
- Bozeman, B. (2000). 'Technology transfer and public policy: a review of research and theory', *Research Policy*, 29/4–5: 627–55. DOI: 10.1016/S0048-7333(99)00093-1
- Bozeman, B., & Corley, E. (2004). 'Scientists' collaboration strategies: Implications for scientific and technical human capital', *Research Policy*, 33/4: 599–616. DOI: 10.1016/j.respol.2004.01.008
- Bozeman, B., & Gaughan, M. (2007). 'Impacts of grants and contracts on academic researchers' interactions with industry', *Research Policy*, 36/5: 694–707. DOI: 10.1016/j.respol.2007.01.007
- Breugelmans, J. G., Makanga, M. M., Cardoso, A. L. V., Mathewson, S. B., Sheridan-Jones, B. R., Gurney, K. A., Mgone, C. S., et al. (2015). 'Bibliometric Assessment of European and Sub-Saharan African Research Output on Poverty-Related and Neglected Infectious Diseases from 2003 to 2011', (D. J. Diemert, Ed.) *PLoS Neglected Tropical Diseases*, 9/8: e0003997. Public Library of Science. DOI: 10.1371/journal.pntd.0003997
- Burt, R. S. (2004). 'Structural holes and good ideas', *American Journal of Sociology*, 110/2: 349–399.
- . (2005). *Brokerage and Closure: An Introduction to Social Capital*. Clarendon Lectures in Management Studies.
- Bush, V. (1945). *Science the Endless Frontier*. Washington.
- Cameron, A., & Trivedi, P. (2013). *Regression analysis of count data*. Cambridge, UK: Cambridge Univ, Second edi.
- Carbonnier, G., & Kontinen, T. (2014). *North-South Research Partnership: Academia Meets Development?* Retrieved March 30, 2017, from <http://www.eadi.org/typo3/fileadmin/Documents/Publications/EADI_Policy_Paper/EADI_policy_paper_Carbonnier_Kontinen_FINAL.pdf>
- Cardoso, A. L., Breugelmans, G., Manville, C., Chataway, J., Cochrane, G., James, S., Chataway,

- M., et al. (2014). *Africa Mapping - Current state of health research on poverty-related and neglected infectious diseases in sub-Saharan Africa*. The Hague, Netherlands.
- Casas, R., de Gortari, R., & Santos, M. J. (2000). 'The building of knowledge spaces in Mexico: a regional approach to networking', *Research Policy*, 29/2: 225–41. DOI: 10.1016/S0048-7333(99)00062-1
- Catanzaro, M., Miranda, G., Palmer, L., & Bajak, A. (2014). 'South American science: Big players', *Nature*, 510/7504: 204–6. DOI: 10.1038/510204a
- Chapman, N., Doubell, A., Oversteegen, L., Chowdhary, V., Rugarabamu, G., Zanetti, R., Ong, M., et al. (2017). *Neglected disease research and development: reflecting on a decade of global investment*.
- Chataway, J., Ochieng, C., Byrne, R., Daniels, C., Dobson, C., Hanlin, R., & Hopkins, M. (2017). *Case Studies of the Political Economy of Science Granting Councils in Sub-Saharan Africa*.
- Chavarro, D., Tang, P., & Rafols, I. (2017). 'Why researchers publish in non-mainstream journals: Training, knowledge bridging, and gap filling', *Research Policy*, 46/9: 1666–80. North-Holland. DOI: 10.1016/J.RESPOL.2017.08.002
- CHRD. (1992). *Health research, essential link to equity in development. Report of the Commission on Health Research for Development*. New York, NY.
- Ciarli, T., & Rafols, I. (2018). 'The relation between research priorities and societal demands: The case of rice', *Research Policy*. North-Holland. DOI: 10.1016/J.RESPOL.2018.10.027
- Cochrane, G., Morgan Jones, M., Marjanovic, S., MacLure, C., Varnai, P., Jongh, T. de, Rosemberg, C., et al. (2017). *Evaluation of the impact of the European Union's Research Funding for Poverty-Related and Neglected Diseases*.
- Cohen, W. M., & Levinthal, D. A. (1990). 'Absorptive Capacity: A New Perspective on Learning and Innovation', *Administrative Science Quarterly*, 35/1: 128–52. DOI: 10.2307/2393553
- Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2002). 'Links and Impacts: The Influence of Public Research on Industrial R&D', *Manage. Sci.*, 48/1: 1–23. DOI: 10.1287/mnsc.48.1.1.14273
- Collazo-Reyes, F. (2014). 'Growth of the number of indexed journals of Latin America and the Caribbean: the effect on the impact of each country', *Scientometrics*, 98/1: 197–209. DOI: 10.1007/s11192-013-1036-2
- Confraria, H., & Godinho, M. M. (2015). 'The impact of African science: a bibliometric analysis', *Scientometrics*, 102/2: 1241–68. DOI: 10.1007/s11192-014-1463-8
- Costas, R., & Leeuwen, T. N. (2012). 'Approaching the “reward triangle”: General analysis of the presence of funding acknowledgments and “peer interactive communication” in scientific publications', *Journal of the American Society for Information Science and Technology*, 63/8: 1647–61. Wiley-Blackwell. DOI: 10.1002/asi.22692
- Cragg, J. G. (1971). 'Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods', *Econometrica*, 39/5: 829–44. DOI: 10.2307/1909582
- Crespi, G. A. (2012). 'Incentivos fiscales a la innovación empresarial'. *Las instituciones fiscales del mañana*, pp. 137–75. Inter-American Development Bank: Washington DC.
- Crespi, G. A., Navarro, J. C., & Zuñiga, P. (2010). *Science, Technology, and Innovation in Latin America and the Caribbean: a Statistical Compendium of Indicators*. Washington DC: Inter-American Development Bank.
- Crespi, G., & Dutrénit, G. (2014). 'Introduction to Science, Technology and Innovation Policies for Development: The Latin American Experience'. *Science, Technology and Innovation Policies for Development*, pp. 1–14. Springer International Publishing: Cham. DOI: 10.1007/978-3-319-

04108-7_1

- D'Este, P., & Patel, P. (2007). 'University–industry linkages in the UK: What are the factors underlying the variety of interactions with industry?', *Research Policy*, 36/9: 1295–313. DOI: 10.1016/j.respol.2007.05.002
- Defazio, D., Lockett, A., & Wright, M. (2009). 'Funding incentives, collaborative dynamics and scientific productivity: Evidence from the EU framework program', *Research Policy*, 38/2: 293–305. North-Holland. DOI: 10.1016/j.respol.2008.11.008
- DFID. (2014). *What is the Evidence on the Impact of Research on International Development?* London.
- Didegah, F., & Thelwall, M. (2013). 'Determinants of research citation impact in nanoscience and nanotechnology', *Journal of the American Society for Information Science and Technology*, 64/5: 1055–64. DOI: 10.1002/asi.22806
- Docquier, F., & Rapoport, H. (2012). 'Globalization, Brain Drain, and Development', *Journal of Economic Literature*, 50/3: 681–730. DOI: 10.1257/jel.50.3.681
- Dosi, G., Pavitt, K., & Soete, L. (1990). *The economics of technical change and international trade*. New York University Press.
- Duque, R. B., Ynalvez, M., Sooryamoorthy, R., Mbatia, P., Dzorgbo, D.-B. S., & Shrum, W. (2005). 'Collaboration Paradox', *Social Studies of Science*, 35/5: 755–85. Sage Publications/Sage CA: Thousand Oaks, CA. DOI: 10.1177/0306312705053048
- Dutrénit, G., & Arza, V. (2010). 'Channels and benefits of interactions between public research organisations and industry: comparing four Latin American countries', *Science and Public Policy*, 37/7: 541–53. DOI: 10.3152/030234210X512043
- Ebadi, A., & Schiffauerova, A. (2015). 'How to Receive More Funding for Your Research? Get Connected to the Right People!', *PLoS ONE*, 10/7: e0133061. DOI: 10.1371/journal.pone.0133061
- Emiel Caron, & Nees-Jan van Eck. (2014). 'Large scale author name disambiguation using rule-based scoring and clustering - Research portal'. E. Noyons (ed.) *Proceedings of the Science and Technology Indicators Conference 2014*, pp. 79–86. Leiden: Universiteit Leiden.
- Evans, J. A., Shim, J.-M., & Ioannidis, J. P. A. (2014). 'Attention to Local Health Burden and the Global Disparity of Health Research', (M. Shamji, Ed.) *PLoS ONE*, 9/4: e90147. Public Library of Science. DOI: 10.1371/journal.pone.0090147
- Evans, S. C., Ng, T., & DuBois, D. L. (2008). 'Does mentoring matter? A multidisciplinary meta-analysis comparing mentored and non-mentored individuals', *Journal of Vocational Behavior*, 72/2: 254–67. DOI: 10.1016/j.jvb.2007.04.005
- Ezeh, A. C., Izugbara, C. O., Kabiru, C. W., Fonn, S., Kahn, K., Manderson, L., Undieh, A. S., et al. (2010). 'Building capacity for public and population health research in Africa: the consortium for advanced research training in Africa (CARTA) model', *Global Health Action*, 3/0. DOI: 10.3402/gha.v3i0.5693
- Fagerberg, J. (1987). 'A technology gap approach to why growth rates differ', *Research Policy*, 16/2–4: 87–99. North-Holland. DOI: 10.1016/0048-7333(87)90025-4
- Fagerberg, J., & Godinho, M. M. (2004). 'Innovation and catching-up', Fagerberg, J., pp. 514–543. Oxford University Press: Oxford.
- Fairclough, R., & Thelwall, M. (2015). 'More precise methods for national research citation impact comparisons', *Journal of Informetrics*, 9/4: 895–906. DOI: 10.1016/j.joi.2015.09.005
- Fanelli, D. (2011). 'Negative results are disappearing from most disciplines and countries',

- Scientometrics*, 90/3: 891–904. DOI: 10.1007/s11192-011-0494-7
- Fealing, K. H., Lane, J., Marburger III, J. H., & Shipp, S. (2011). *The science of science policy : a handbook*. Stanford Business Books.
- Fleming, L. (2001). 'Recombinant Uncertainty in Technological Search', *Management Science*, 47/1: 117–32. INFORMS. DOI: 10.1287/mnsc.47.1.117.10671
- Fleming, L., Mingo, S., & Chen, D. (2007). 'Collaborative Brokerage, Generative Creativity, and Creative Success', *Administrative Science Quarterly*, 52/3: 443–75. DOI: 10.2189/asqu.52.3.443
- Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., et al. (2018). 'Science of science.', *Science*, 359/6379. American Association for the Advancement of Science. DOI: 10.1126/science.aao0185
- Frame, J. D., & Carpenter, M. P. (1979). 'International Research Collaboration', *Social Studies of Science*, 9/4: 481–97. DOI: 10.1177/030631277900900405
- Franzoni, C., Scellato, G., & Stephan, P. (2014). *The mover's advantage: The superior performance of migrant scientists*. *Economics Letters*, Vol. 122. DOI: 10.1016/j.econlet.2013.10.040
- Freeman, C., Clark, J., & Soete, L. (1982). *Unemployment and technical innovation : a study of long waves and economic development*. Greenwood Press.
- Freeman, L. C. (1978). 'Centrality in social networks conceptual clarification', *Social Networks*, 1/3: 215–39. DOI: 10.1016/0378-8733(78)90021-7
- Freeman, R. B. (2005). *Does Globalization of the Scientific/Engineering Workforce Threaten U.S. Economic Leadership?* Retrieved January 28, 2016, from <<http://www.nber.org/papers/w11457>>
- Friesenhahn, I., & Beaudry, C. (2014). *Global state of young scientists: project report and recommendations*.
- Gaillard, J. (2003). 'Overcoming the scientific generation gap in Africa : an urgent priority', *Interdisciplinary Science Reviews*, 28/1: 15–25. DOI: 10.1179/030801803225010359
- Gaillard, J. F. (1994). 'North-South research partnership: Is collaboration possible between unequal partners?', *Knowledge and Policy*, 7/2: 31–63. Springer Netherlands. DOI: 10.1007/BF02692761
- Garfield, E. (1977). 'The 250 Most-Cited Primary Authors, 1961-1975. Part 1. How the Names Were Selected.', *Current Contents*, 3/49: 326–36.
- . (1979). *Citation Indexing: Its Theory and Application in Science, Technology, and Humanities*. Isi Press.
- . (1981). 'The 1000 contemporary scientists most-cited 1965–1978. Part 1. The basic list and introduction.', *Current Contents*, 5/41: 269–78.
- . (1983). 'Mapping science in the Third World', *Science and Public Policy*, 10/3: 112–27. Oxford University Press. DOI: 10.1093/spp/10.3.112
- Geffers, J., Beaudry, C., Yang, H.-C., Huang, F., Phranska, O., Dominik, M., Lin, Y.-C., et al. (2017). *Global State of Young Scientists in ASEAN. Creativity and Innovation of Young Scientists in ASEAN*. Saale.
- Gilbert, G. N. (1977). 'Referencing as Persuasion', *Social Studies of Science*, 7/1: 113–22.
- Giunta, A., Pericoli, F. M., & Pierucci, E. (2015). 'University–Industry collaboration in the biopharmaceuticals: the Italian case', *The Journal of Technology Transfer*, 41/4: 818–40. Springer US. DOI: 10.1007/s10961-015-9402-2
- Glänzel, W., Rinia, E. J., & Brocken, M. G. M. (1995). 'A Bibliometric Study on Highly-cited European Physics Papers in the 80s Based on "Physics Briefs" Source Publications and "SCI"

- Citation Data', *Research evaluation*, 5/2: 113–22.
- Glänzel, W., Thijs, B., & Debackere, K. (2014). 'The application of citation-based performance classes to the disciplinary and multidisciplinary assessment in national comparison and institutional research assessment', *Scientometrics*, 101/2: 939–52. Springer Netherlands. DOI: 10.1007/s11192-014-1247-1
- Godin, B. (1996). 'Research and the practice of publication in industries', *Research Policy*, 25/4: 587–606. DOI: 10.1016/0048-7333(95)00859-4
- . (2006). 'The Linear Model of Innovation', *Science, Technology, & Human Values*, 31/6: 639–67. Sage Publications/Sage CA: Thousand Oaks, CA. DOI: 10.1177/0162243906291865
- Gök, A., Rigby, J., & Shapira, P. (2016). 'The impact of research funding on scientific outputs: Evidence from six smaller European countries', *Journal of the Association for Information Science and Technology*, 67/3: 715–30. DOI: 10.1002/asi.23406
- Goldemberg, J. (1998). 'What Is the Role of Science in Developing Countries?', *Science*, 279/5354: 1140–1. DOI: 10.1126/science.279.5354.1140
- Goldfinch, S., Dale, T., Jr, K. D., & DeRouen Jr., K. (2003). 'Science from the periphery: Collaboration, networks and "Periphery Effects" in the citation of New Zealand Crown Research Institutes articles, 1995-2000', *Scientometrics*, 57/3: 321–37. Kluwer Academic Publishers. DOI: 10.1023/A:1025048516769
- Gonzalez-Brambila, C. N., Veloso, F. M., & Krackhardt, D. (2013). 'The impact of network embeddedness on research output', *Research Policy*, 42/9: 1555–67. DOI: 10.1016/j.respol.2013.07.008
- Goode, R. B. (1959). 'Adding to the Stock of Physical and Human Capital', *The American Economic Review*, 49/2: 147–55.
- Granovetter, M. (1985). 'Economic Action and Social Structure: The Problem of Embeddedness', *American Journal of Sociology*, 481–510.
- Grassano, N., Rotolo, D., Hutton, J., Lang, F., & Hopkins, M. M. (2017). 'Funding Data from Publication Acknowledgments: Coverage, Uses, and Limitations', *Journal of the Association for Information Science and Technology*, 68/4: 999–1017. Wiley-Blackwell. DOI: 10.1002/asi.23737
- Gray, D. O., Boardman, C., & Rivers, D. (2013). 'The New Science and Engineering Management: Cooperative Research Centers as Intermediary Organizations for Government Policies and Industry Strategies'. *Cooperative Research Centers and Technical Innovation*, pp. 3–33. Springer New York: New York, NY. DOI: 10.1007/978-1-4614-4388-9_1
- Greene, W. H. (2012). *Econometric analysis*. Pearson.
- Gulati, R. (1995). 'Does familiarity breed trust? The implications of repeated ties for contractual choice in alliances', *Academy of Management Journal*, 38/1: 85–112. DOI: 10.2307/256729
- Guns, R., & Wang, L. (2017). 'Detecting the emergence of new scientific collaboration links in Africa: A comparison of expected and realized collaboration intensities', *Journal of Informetrics*, 11/3: 892–903. Elsevier. DOI: 10.1016/j.joi.2017.07.004
- Hall, B. H., Link, A. N., & Scott, J. T. (2003). 'Universities as Research Partners', *Review of Economics and Statistics*, 85/May: 485–91. DOI: 10.1162/rest.2003.85.2.485
- Head, M. G., Goss, S., Gelister, Y., Alegana, V., Brown, R. J., Clarke, S. C., Fitchett, J. R. A., et al. (2017). 'Global funding trends for malaria research in sub-Saharan Africa: a systematic analysis.', *The Lancet. Global health*, 5/8: e772–81. Elsevier. DOI: 10.1016/S2214-109X(17)30245-0

- Heckman, J. J. (1979). 'Sample Selection Bias as a Specification Error', *Econometrica*, 47/1: 153–61.
- Hicks, D. M., Isard, P. A., & Martin, B. R. (1996). 'A morphology of Japanese and European corporate research networks', *Research Policy*, 25/3: 359–78. DOI: 10.1016/0048-7333(95)00830-6
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). 'Bibliometrics: The Leiden Manifesto for research metrics', *Nature*, 520/7548: 429–31. DOI: 10.1038/520429a
- Hobday, M. (1995). *Innovation in East Asia : the challenge to Japan*. Edward Elgar.
- Hollingsworth, J. R. (2006). 'The dynamics of American Science : an institutional and organizational perspective on major discoveries'. Transformationen des Kapitalismus : Festschrift für Wolfgang Streeck zum sechzigsten Geburtstag. - Frankfurt [u.a.] : Campus-Verl., ISBN 978-3-593-38281-4. - 2006, p. 361-380, pp. 361–80. Campus Verlag: Frankfurt, New York.
- Hunter, R. S., Oswald, A. J., & Charlton, B. G. (2009). 'The Elite Brain Drain', *The Economic Journal*, 119/538: F231–51. Blackwell Publishing Ltd. DOI: 10.1111/j.1468-0297.2009.02274.x
- InCites™. (2014). 'InCites™. This data is reproduced under a license from Thomson Reuters to the University of Lisbon.: Thomson Reuters.
- Ioannidis, J. P. A. (2004). 'Global estimates of high-level brain drain and deficit.', *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 18/9: 936–9. Federation of American Societies for Experimental Biology. DOI: 10.1096/fj.03-1394ife
- Jeffrey, P. (2003). 'Smoothing the Waters: Observations on the process of cross-disciplinary research collaboration', *Social Studies of Science*, 33/4: 539–62. DOI: 10.1017/CBO9781107415324.004
- Jin, B., & Rousseau, R. (2004). 'Evaluation of Research Performance and Scientometric Indicators in China'. *Handbook of Quantitative Science and Technology Research*, pp. 497–514. Kluwer Academic Publishers: Dordrecht. DOI: 10.1007/1-4020-2755-9_23
- Jones, B. F., Wuchty, S., & Uzzi, B. (2008). 'Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science', *Science*, 322/5905.
- Jones, N., Bailey, M., & Lyytikäinen, M. (2007). 'Research capacity strengthening in Africa: trends, gaps and opportunities', *Overseas Development Institute*, 44/November: 1–17.
- Jonkers, K., & Tijssen, R. (2008). 'Chinese researchers returning home: Impacts of international mobility on research collaboration and scientific productivity', *Scientometrics*, 77/2: 309–33. DOI: 10.1007/s11192-007-1971-x
- Juma, C. (2016). *Education, Research, and Innovation in Africa Forging Strategic Linkages for Economic Transformation* (No. 2016-01).
- Kahn, M. (2011). 'A bibliometric analysis of South Africa's scientific outputs - Some trends and implicatons', *South African Journal of Science*, 107/1–2: 1990–4. DOI: 10.4102/sajs.v107i1/2.406
- Kaplan, D. (2012). 'South African mining equipment and specialist services: Technological capacity, export performance and policy', *Resources Policy*, 37/4: 425–33. DOI: 10.1016/j.resourpol.2012.06.001
- Karim, S. S. A., & Karim, Q. A. (2010). 'AIDS research must link to local policy', *Nature*, 463/7282: 733–4. Nature Publishing Group. DOI: 10.1038/463733a
- Katz, J. S., & Hicks, D. (1997). 'How much is a collaboration worth? A calibrated bibliometric

- model', *Scientometrics*, 40/3: 541–54. DOI: 10.1007/BF02459299
- Katz, J. S., & Martin, B. R. (1997). 'What is research collaboration?', *Research Policy*, 26/1: 1–18. DOI: 10.1016/S0048-7333(96)00917-1
- Kim, L. (1997). *Imitation to innovation: the dynamics of Korea's technological learning*. Harvard Business School Press.
- King, D. A. (2004). 'The scientific impact of nations', *Nature*, 430/6997: 311–6. DOI: 10.1038/430311a
- Kline, S. J., & Rosenberg, N. (1986). 'An Overview of Innovation'. Landau R. & Rosenberg N. (eds) , pp. 275–304. National Academy Press: Washington D.C.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.
- Lall, S. (2000a). 'Technological change and industrialization in the Asian newly industrializing economies: achievements and challenges'. Kim L. & Nelson R. R. (eds) , pp. 13–68. Cambridge University Press: Cambridge.
- . (2000b). 'The Technological Structure and Performance of Developing Country Manufactured Exports, 1985-98', *Oxford Development Studies*, 28/3: 337–69. DOI: 10.1080/713688318
- Larivière, V., & Costas, R. (2016). 'How many is too many? On the relationship between research productivity and impact', *PLoS ONE*, 11/9: 1–10. DOI: 10.1371/journal.pone.0162709
- Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge, Mass: Harvard University Press.
- Laudel, G. (2002). 'What do we measure by co-authorships?', *Research Evaluation*, 11/1: 3–15.
- Laursen, K. (2000). 'Do export and technological specialisation patterns co-evolve in terms of convergence or divergence? Evidence from 19 OECD countries, 1971–1991', *Journal of Evolutionary Economics*, 10/4: 415–436.
- Lawrence, P. A. (2003). 'The politics of publication', *Nature*, 422/6929: 259–61. DOI: 10.1038/422259a
- Lee, B. Y., Bacon, K. M., Bottazzi, M. E., & Hotez, P. J. (2013). 'Global economic burden of Chagas disease: a computational simulation model', *The Lancet Infectious Diseases*, 13/4: 342–8. DOI: 10.1016/S1473-3099(13)70002-1.Global
- Lee, S., & Bozeman, B. (2005). 'The Impact of Research Collaboration on Scientific Productivity', *Social Studies of Science*, 35/5: 673–702. DOI: 10.1177/0306312705052359
- Lee, Y. S. (2000). 'The Sustainability of University-Industry Research Collaboration: An Empirical Assessment', *The Journal of Technology Transfer*, 25/2: 111–33. DOI: 10.1023/A:1007895322042
- Leeuwen, T. Van, Moed, H., & Tijssen, R. (2001). 'Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance', *Scientometrics*, 51/1: 335–46. DOI: 10.1023/A:1010549719484
- Leimu, R., & Koricheva, J. (2005). 'What determines the citation frequency of ecological papers?', *Trends in Ecology & Evolution*, 20/1: 28–32. DOI: 10.1016/j.tree.2004.10.010
- Li, Y., Youtie, J., & Shapira, P. (2015). 'Why do technology firms publish scientific papers? The strategic use of science by small and midsize enterprises in nanotechnology', *The Journal of Technology Transfer*, 40/6: 1016–33. Springer US. DOI: 10.1007/s10961-014-9391-6
- Liao, Y.-C., & Phan, P. H. (2016). 'Internal capabilities, external structural holes network positions, and knowledge creation', *The Journal of Technology Transfer*, 41/5: 1148–67. Springer US. DOI:

- Link, A. N., Siegel, D. S., & Bozeman, B. (2007). 'An empirical analysis of the propensity of academics to engage in informal university technology transfer', *Industrial and Corporate Change*, 16/4: 641–55. DOI: 10.1093/icc/dtm020
- Lopez Pineiro, C., & Hicks, D. (2015). 'Reception of Spanish sociology by domestic and foreign audiences differs and has consequences for evaluation', *Research Evaluation*, 24/1: 78–89. Oxford University Press. DOI: 10.1093/reseval/rvu030
- Lundberg, J., Tomson, G., Lundkvist, I., Skår, J., & Brommels, M. (2006). 'Collaboration uncovered: Exploring the adequacy of measuring university-industry collaboration through co-authorship and funding', *Scientometrics*, 69/3: 575–89. DOI: 10.1007/s11192-006-0170-5
- Lundvall, B.-Å. (1992). *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. London: Pinter Publishers.
- Marín, A., Navas-Alemán, L., & Perez, C. (2015). 'Natural Resource Industries As a Platform for the Development of Knowledge Intensive Industries.', *Tijdschrift voor Economische en Sociale Geografie (Journal of Economic & Social Geography)*, 106/2: 154–68. DOI: 10.1111/tesg.12136
- Marmolejo-Leyva, R., Perez-Angon, M. A., & Russell, J. M. (2015). 'Mobility and International Collaboration: Case of the Mexican Scientific Diaspora', (V. Larivière, Ed.) *PLOS ONE*, 10/6: e0126720. Public Library of Science. DOI: 10.1371/journal.pone.0126720
- Marotta, D., Mark, M., Blom, A., & Thorn, K. (2007). *Human Capital and University-Industry Linkages' Role in Fostering Firm Innovation: An Empirical Study of Chile and Colombia* (No. 4443). Policy Research Working Papers. Washington DC. DOI: <http://dx.doi.org/10.1596/1813-9450-4443>
- Marsh, H. W., & Hattie, J. (2002). 'The Relation Between Research Productivity and Teaching Effectiveness: Complementary, Antagonistic, or Independent Constructs?', *The Journal of Higher Education*, 73/5: 603–41. DOI: 10.1353/jhe.2002.0047
- Martin, B. R., & Irvine, J. (1983). 'Assessing basic research: Some partial indicators of scientific progress in radio astronomy', *Research Policy*, 12/2: 61–90. North-Holland. DOI: 10.1016/0048-7333(83)90005-7
- Marx, W., & Bornmann, L. (2014). 'On the causes of subject-specific citation rates in Web of Science', *Scientometrics*, 102/2: 1823–7. DOI: 10.1007/s11192-014-1499-9
- May, R. M. (1997). 'The Scientific Wealth of Nations', *Science*, 275/5301: 793–6.
- Mazzoleni, R., & Nelson, R. R. (2007). 'Public research institutions and economic catch-up', *Research Policy*, 36/10: 1512–28. DOI: 10.1016/j.respol.2007.06.007
- McCaffrey, D. F., Mihaly, K., Lockwood, J. R., & Sass, T. R. (2012). 'A review of Stata commands for fixed-effects estimation in normal linear models', *Stata Journal*, 12: 406–32.
- McGregor, S., Henderson, K. J., & Kaldor, J. M. (2014). 'How are health research priorities set in low and middle income countries? A systematic review of published reports.', *PloS one*, 9/9: e108787. Public Library of Science. DOI: 10.1371/journal.pone.0108787
- Mégniqbêto, E. (2013). 'International collaboration in scientific publishing: the case of West Africa (2001–2010)', *Scientometrics*, 96/3: 761–83. DOI: 10.1007/s11192-013-0963-2
- Merton, R. K. (1968). 'The Matthew Effect in Science The reward and communication systems of science are considered', *Science*, 159/3810: 56–63. DOI: 10.1126/science.159.3810.56
- . (1973). *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago Press.

- Meyer-Krahmer, F., & Schmoch, U. (1998). 'Science-based technologies: university-industry interactions in four fields', *Research Policy*, 27/8: 835–51.
- Mgone, C., Volmink, J., Coles, D., Makanga, M., Jaffar, S., & Sewankambo, N. (2010). 'Linking research and development to strengthen health systems in Africa', *Tropical Medicine & International Health*, 15/12: 1404–6. Blackwell Publishing Ltd. DOI: 10.1111/j.1365-3156.2010.02661.x
- Moed, H. F. (2005). *Citation Analysis in Research Evaluation*. Information Science and Knowledge Management, Vol. 9. Berlin/Heidelberg: Springer-Verlag.
- Moed, H. F., de Moya-Anegón, F., López-Illescas, C., & Visser, M. (2011). 'Is concentration of university research associated with better research performance?', *Journal of Informetrics*, 5/4: 649–58. DOI: 10.1016/j.joi.2011.06.003
- Moody, J. (2004). 'The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999', *American Sociological Review*, 69/2: 213–38. DOI: 10.1177/000312240406900204
- Moulton, B. R. (1990). 'An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Unit', *The Review of Economics and Statistics*, 72/2: 334–8.
- Mouton, J. (2008). 'Africa's Science Decline', *Harvard International Review*, 30/3: 46–51. DOI: Article
- Moya-Anegón, F., & Herrero-Solana, V. (2013). 'Worldwide Topology of the Scientific Subject Profile: A Macro Approach in the Country Level', (G. E. Derrick, Ed.) *PLoS ONE*, 8/12: e83222. Public Library of Science. DOI: 10.1371/journal.pone.0083222
- Moyi Okwaro, F., & Geissler, P. W. (2015). 'In/dependent Collaborations: Perceptions and Experiences of African Scientists in Transnational HIV Research.', *Medical anthropology quarterly*, 29/4: 492–511. DOI: 10.1111/maq.12206
- MSF. (2016). *Clinical guidelines - Diagnosis and treatment manual*. *Clinical guidelines Diagnosis and treatment manual*.
- Müller, M., Cowan, R., & Barnard, H. (2018). 'On the value of foreign PhDs in the developing world: Training versus selection effects in the case of South Africa', *Research Policy*, 47/5: 886–900. North-Holland. DOI: 10.1016/j.respol.2018.02.013
- Narin, F., Hamilton, K. S., & Olivastro, D. (1997). 'The increasing linkage between U.S. technology and public science', *Research Policy*, 26/3: 317–30. DOI: 10.1016/S0048-7333(97)00013-9
- Narin, F., Stevens, K., & Whitlow, E. S. (1991). 'Scientific co-operation in Europe and the citation of multinationally authored papers', *Scientometrics*, 21/3: 313–23. DOI: 10.1007/BF02093973
- Narváez-Berthelemot, N., Russell, J. M., Arvanitis, R., Waast, R., & Gaillard, J. (2002). 'Science in Africa: An overview of mainstream scientific output', *Scientometrics*, 54/2: 229–241.
- Ndounga Diakou, L. A., Ntoumi, F., Ravaud, P., & Boutron, I. (2017). 'Published randomized trials performed in Sub-Saharan Africa focus on high-burden diseases but are frequently funded and led by high-income countries', *Journal of Clinical Epidemiology*, 82. Elsevier Ltd. DOI: 10.1016/j.jclinepi.2016.10.008
- Nelson, R. R. (1993). 'National Innovation Systems: A Comparative Analysis'.
- . (2005). 'The Roles of Research in Universities and Public Labs in Economic Catch-up'. Santangelo G. (ed.) *Technological Change and Economic Catch-up*. Edward Elgar Publishing.
- Van Noorden, R. (2014). 'The impact gap: South America by the numbers', *Nature*, 510/7504: 202–3. DOI: 10.1038/510202a

- OECD. (2015). *OECD Science, Technology and Industry Scoreboard 2015: Innovation for growth and society*. Paris: OECD Publishing. DOI: http://dx.doi.org/10.1787/sti_scoreboard-2015-en
- Onyancha, O. B., & Maluleka, J. R. (2011). 'Knowledge production through collaborative research in sub-Saharan Africa: how much do countries contribute to each other's knowledge output and citation impact?', *Scientometrics*, 87/2: 315–36. Springer Netherlands. DOI: 10.1007/s11192-010-0330-5
- Parker, J. N., Allesina, S., & Lortie, C. J. (2012). 'Characterizing a scientific elite (B): publication and citation patterns of the most highly cited scientists in environmental science and ecology', *Scientometrics*, 94/2: 469–80. DOI: 10.1007/s11192-012-0859-6
- Parker, J. N., Lortie, C., & Allesina, S. (2010). 'Characterizing a scientific elite: the social characteristics of the most highly cited scientists in environmental science and ecology.', *Scientometrics*, 85/1: 129–43. Springer. DOI: 10.1007/s11192-010-0234-4
- Pavitt, K. (1984). 'Sectoral patterns of technical change: Towards a taxonomy and a theory', *Research Policy*, 13/6: 343–73. DOI: 10.1016/0048-7333(84)90018-0
- Perkmann, M., King, Z., & Pavelin, S. (2011). 'Engaging excellence? Effects of faculty quality on university engagement with industry', *Research Policy*, 40/4: 539–52. DOI: 10.1016/j.respol.2011.01.007
- Peters, D. P., & Ceci, S. J. (1982). 'Peer-review practices of psychological journals: The fate of published articles, submitted again', *Behavioral and Brain Sciences*, 5/02: 187. Cambridge University Press. DOI: 10.1017/S0140525X00011183
- Peters, H. P. F., & van Raan, A. F. J. (1994). 'On determinants of citation scores: A case study in chemical engineering', *Journal of the American Society for Information Science*, 45/1: 39–49. DOI: 10.1002/(SICI)1097-4571(199401)45:1<39::AID-ASIS>3.0.CO;2-Q
- Pinch, S., Henry, N., Jenkins, M., & Tallman, S. (2003). 'From "industrial districts" to "knowledge clusters": a model of knowledge dissemination and competitive advantage in industrial agglomerations', *Journal of Economic Geography*, 3/4: 373–88. DOI: 10.1093/jeg/lbg019
- Porter, A. L., & Rafols, I. (2009). 'Is science becoming more interdisciplinary? Measuring and mapping six research fields over time', *Scientometrics*, 81/3: 719–45. Springer Netherlands. DOI: 10.1007/s11192-008-2197-2
- Pouris, A. (2010). 'A scientometric assessment of the Southern Africa Development Community: science in the tip of Africa', *Scientometrics*, 85/1: 145–54. DOI: 10.1007/s11192-010-0260-2
- Pouris, A., & Ho, Y.-S. (2014). 'Research emphasis and collaboration in Africa', *Scientometrics*, 98/3: 2169–84. Springer Netherlands. DOI: 10.1007/s11192-013-1156-8
- Price, D. J. de S. (1963). *Little science, big science*. Columbia University Press.
- Prozesky, H., & Boshoff, N. (2012). 'Bibliometrics as a tool for measuring gender-specific research performance: an example from South African invasion ecology', *Scientometrics*, 90/2: 383–406. Springer Netherlands. DOI: 10.1007/s11192-011-0478-7
- Pudovkin, A. I., & Garfield, E. (2009). 'Percentile rank and author superiority indexes for evaluating individual journal articles and the author's overall citation performance', *CollNet Journal of Scientometrics and Information Management*, 3/2: 3–10.
- Puuska, H.-M., Muhonen, R., & Leino, Y. (2013). 'International and domestic co-publishing and their citation impact in different disciplines', *Scientometrics*, 98/2: 823–39. DOI: 10.1007/s11192-013-1181-7
- van Raan, A. F. J. (2003). 'The use of bibliometric analysis in research performance assessment

- and monitoring of interdisciplinary scientific developments', *Technikfolgenabschätzung-Theorie und Praxis/Technology Assessment-Theory and Practice*, 1/12: 20–9.
- Raan, A. F. J. van. (2004). 'Measuring Science'. Moed H. F., Glänzel W., & Schmoch U. (eds), pp. 19–50. Springer Netherlands.
- Van Raan, A. F. (1998). 'The influence of international collaboration on the impact of research results', *Scientometrics*, 42/3: 423–428.
- Radosevic, S., & Yoruk, E. (2014). 'Are there global shifts in the world science base? Analysing the catching up and falling behind of world regions', *Scientometrics*, 101/3: 1897–924. DOI: 10.1007/s11192-014-1344-1
- Rafols, I., Ciarli, T., & Chavarro, D. (2015). 'Under-reporting research relevant to local needs in the global south. Database biases in the representation of knowledge on rice'. Havana: 13th Globelics International Conference.
- Rafols, I., & Yegros, A. (2017). 'Is Research Responding to Health Needs?', *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3106713
- Reagans, R., & McEvily, B. (2003). 'Network Structure and Knowledge Transfer: The Effects of Cohesion and Range', *Administrative Science Quarterly*, 48/2: 240–67. DOI: 10.2307/3556658
- Rigby, J. (2011). 'Systematic grant and funding body acknowledgement data for publications: new dimensions and new controversies for research policy and evaluation', *Research Evaluation*, 20/5: 365–75. Oxford University Press. DOI: 10.3152/095820211X13164389670392
- Roberts, R. M. (1989). *Serendipity: accidental discoveries in science*. Wiley.
- Romer, P. M. (1986). 'Increasing Returns and Long-Run Growth', *Journal of Political Economy*, 94/5: 1002–37. The University of Chicago Press. DOI: 10.1086/261420
- Rosenberg, N. (1990). 'Why do firms do basic research (with their own money)?', *Research Policy*, 19/2: 165–174.
- Rottingen, J. A., Regmi, S., Eide, M., Young, A. J., Viergever, R. F., Ardal, C., Guzman, J., et al. (2013). 'Mapping of available health research and development data: what's there, what's missing, and what role is there for a global observatory?', *Lancet*, 382/9900: 1286–307. DOI: 10.1016/S0140-6736(13)61046-6
- Salter, A. J., & Martin, B. (2001). 'The economic benefits of publicly funded basic research: a critical review', *Research Policy*, 30/3: 509–32.
- Sánchez-Peñas, C., & Willett, P. (2006). 'Brief communication: Gender differences in publication and citation counts in librarianship and information science research', *Journal of Information Science*, 32/5: 480–5. DOI: 10.1177/0165551506066058
- Sandström, U., & van den Besselaar, P. (2016). 'Quantity and/or Quality? The Importance of Publishing Many Papers', (P. Dorta-González, Ed.) *PLOS ONE*, 11/11: e0166149. Oxford University Press. DOI: 10.1371/journal.pone.0166149
- Sarewitz, D., & Nelson, R. (2008). 'Three rules for technological fixes', *Nature*, 456/7224: 871–2. DOI: 10.1038/456871a
- Sarewitz, D., & Pielke Jr., R. A. (2007). 'The neglected heart of science policy: reconciling supply of and demand for science', *Environmental Science & Policy*, Reconciling the Supply of and Demand for Science, with a Focus on Carbon Cycle Research, 10/1: 5–16. DOI: 10.1016/j.envsci.2006.10.001
- Sawyer, A. (2014). 'African Universities and the Challenge of Research Capacity Development'. *Journal of Higher Education in Africa / Revue de l'enseignement supérieur en Afrique*. CODESRIA.

DOI: 10.2307/24486132

- Scellato, G., Franzoni, C., & Stephan, P. (2015). 'Migrant scientists and international networks', *Research Policy*, 44/1: 108–20. DOI: 10.1016/j.respol.2014.07.014
- Shibayama, S., & Baba, Y. (2015). 'Impact-oriented science policies and scientific publication practices: The case of life sciences in Japan', *Research Policy*, 44/4: 936–50. DOI: 10.1016/j.respol.2015.01.012
- Simonton, D. K. (1999). *Origins of genius : Darwinian perspectives on creativity*. Oxford University Press.
- Sinatra, R., Wang, D., Deville, P., Song, C., & Barabási, A.-L. (2016). 'Quantifying the evolution of individual scientific impact', *Science*, 354/6312.
- Smith, M. J., Weinberger, C., Bruna, E. M., Allesina, S., Peters, D., Ceci, S., Akre, O., et al. (2014). 'The Scientific Impact of Nations: Journal Placement and Citation Performance', (V. Larivière, Ed.) *PLoS ONE*, 9/10: e109195. Public Library of Science. DOI: 10.1371/journal.pone.0109195
- Solow, R. M. (1956). 'A Contribution to the Theory of Economic Growth', *The Quarterly Journal of Economics*, 70/1: 65. Oxford University Press. DOI: 10.2307/1884513
- StataCorp. (2013). 'Stata Statistical Software'. College Station, TX.
- Sugimoto, C. R., Robinson-Garcia, N., Murray, D. S., Yegros-Yegros, A., Costas, R., & Larivière, V. (2017). 'Scientists have most impact when they're free to move', *Nature*, 550/7674: 29–31. DOI: 10.1038/550029a
- Suzigan, W., & Albuquerque, E. da M. e. (2011). 'The underestimated role of universities for the Brazilian system of innovation', *Revista de Economia Política*, 31/1: 03-30. DOI: 10.1590/S0101-31572011000100001
- Swan, T. W. (1956). 'Economic Growth and capital accumulation', *Economic Record*, 32/2: 334–61. John Wiley & Sons, Ltd (10.1111). DOI: 10.1111/j.1475-4932.1956.tb00434.x
- Taylor, D. (1986). 'The pharmaceutical industry and health in the third world', *Social Science & Medicine*, 22/11: 1141–9. Pergamon. DOI: 10.1016/0277-9536(86)90180-2
- Testa, J. (2011). *The Globalization of Web of Science: 2005-2010*. Philadelphia. Retrieved from <<http://wokinfo.com/media/pdf/globalwos-essay.pdf>>
- Thelwall, M. (2016). 'Are the discretised lognormal and hooked power law distributions plausible for citation data?', *Journal of Informetrics*, 10/2: 454–70. DOI: 10.1016/j.joi.2016.03.001
- Thune, T., & Gulbrandsen, M. (2014). 'Dynamics of collaboration in university–industry partnerships: do initial conditions explain development patterns?', *The Journal of Technology Transfer*, 39/6: 977–93. Springer US. DOI: 10.1007/s10961-014-9331-5
- Tijssen, R. J. . W. (2004). 'Is the commercialisation of scientific research affecting the production of public knowledge?: Global trends in the output of corporate research articles', *Research Policy*, 33/5: 709–33. DOI: 10.1016/j.respol.2003.11.002
- Tijssen, R. J. W. (2007). 'Africa's contribution to the worldwide research literature: New analytical perspectives, trends, and performance indicators', *Scientometrics*, 71/2: 303–27. DOI: 10.1007/s11192-007-1658-3
- . (2012). 'Co-authored research publications and strategic analysis of public–private collaboration', *Research Evaluation*, 21/3: 204–15. DOI: 10.1093/reseval/rvs013
- Tijssen, R. J. W., Leeuwen, T. N. van, & Wijk, E. van. (2009). 'Benchmarking university–industry research cooperation worldwide: performance measurements and indicators based on co-

- authorship data for the world's largest universities', *Research Evaluation*, 18/1: 13–24. DOI: 10.3152/095820209X393145
- Tijssen, R., & Kraemer-Mbula, E. (2017). 'Research excellence in Africa: Policies, perceptions, and performance', *Science and Public Policy*. DOI: 10.1093/scipol/scx074
- Toivanen, H., & Ponomarev, B. (2011). 'African regional innovation systems: bibliometric analysis of research collaboration patterns 2005–2009', *Scientometrics*, 88/2: 471–93. DOI: 10.1007/s11192-011-0390-1
- UNAIDS. (2017). 'Ending Aids Progress Towards the 90-90-90 Targets', *Global Aids Update*, 198. DOI: UNAIDS/JC2900E
- UNESCO. (2010). *UNESCO science report 2010*. Paris: United Nations Educational, Scientific and Cultural Organization.
- (Ed.). (2015). *UNESCO science report: towards 2030*. Paris: UNESCO Publishing.
- Urzúa, O. (2011). *The emergence and development of knowledge intensive mining service suppliers in the late 20th century*. University of Sussex.
- Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). 'Atypical Combinations and Scientific Impact', *Science*, 342/6157: 468–72. DOI: 10.1126/science.1240474
- Vargas, R. A., de Souza Vanz, S. A., & Stumpf, I. R. (2014). 'The role of National journals on the rise in Brazilian Agricultural Science Publications in Web of Science', *Journal of Scientometric Research*, 3/1: 28.
- Verspagen, B. (1991). 'A new empirical approach to catching up or falling behind', *Structural Change and Economic Dynamics*, 2/2: 359–80. North-Holland. DOI: 10.1016/S0954-349X(05)80008-6
- Wagner, C. S., Brahmakulam, I. T., Jackson, B. A., Wong, A., & Yoda, T. (2001). 'Science & Technology Collaboration'. Retrieved August 11, 2015, from <http://www.rand.org/pubs/monograph_reports/MR1357z0.html>
- Waldinger, F. (2016). 'Bombs, Brains, and Science: The Role of Human and Physical Capital for the Creation of Scientific Knowledge', *Review of Economics and Statistics*, 98/5: 811–31. The MIT Press. DOI: 10.1162/REST_a_00565
- Waltman, L. (2016). 'A review of the literature on citation impact indicators', *Journal of Informetrics*, 10/2: 365–91. DOI: 10.1016/j.joi.2016.02.007
- Waltman, L., & Eck, N. J. van. (2012). 'Source normalized indicators of citation impact: an overview of different approaches and an empirical comparison', *Scientometrics*, 96/3: 699–716. DOI: 10.1007/s11192-012-0913-4
- Waltman, L., van Eck, N. J., van Leeuwen, T. N., Visser, M. S., & van Raan, A. F. J. (2011a). 'Towards a new crown indicator: Some theoretical considerations', *Journal of Informetrics*, 5/1: 37–47. DOI: 10.1016/j.joi.2010.08.001
- (2011b). 'Towards a new crown indicator: an empirical analysis', *Scientometrics*, 87/3: 467–81. Springer Netherlands. DOI: 10.1007/s11192-011-0354-5
- Wang, L. (2016). 'The structure and comparative advantages of China's scientific research: quantitative and qualitative perspectives', *Scientometrics*, 106/1: 435–52. Springer Netherlands. DOI: 10.1007/s11192-015-1650-2
- Wang, W., Yu, S., Bekele, T. M., Kong, X., & Xia, F. (2017). 'Scientific collaboration patterns vary with scholars' academic ages', *Scientometrics*, 112/1: 329–43. Springer Netherlands. DOI: 10.1007/s11192-017-2388-9

- Weinberg, B. A. (2011). 'Developing science: Scientific performance and brain drains in the developing world', *Journal of Development Economics*, 95/1: 95–104. DOI: 10.1016/j.jdeveco.2010.05.009
- Whitworth, J. A. G., Kokwaro, G., Kinyanjui, S., Snewin, V. A., Tanner, M., Walport, M., & Sewankambo, N. (2008). 'Strengthening capacity for health research in Africa.', *Lancet (London, England)*, 372/9649: 1590–3. Elsevier. DOI: 10.1016/S0140-6736(08)61660-8
- Wong, P. K., & Singh, A. (2013). 'Do co-publications with industry lead to higher levels of university technology commercialization activity?', *Scientometrics*, 97/2: 245–65. DOI: 10.1007/s11192-013-1029-1
- Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.
- World Health Organization. (1996). 'Summary of Investing in Health Research and Development. The Report of the Ad Hoc Committee on health Research Relating to Future Intervention Options'.
- . (2012). *Research and Development to Meet Health Needs in Developing Countries : Strengthening Global Financing and Coordination Report of the Consultative Expert Working Group on Research and Development : Financing and Coordination Briefing to Representatives of*.
- . (2017). *WHO methods and data sources for global burden of disease estimates. Who*, Vol. 1. Retrieved from <http://www.who.int/gho/mortality_burden_disease/en/index.html>
- Zaheer, A., & Bell, G. G. (2005). 'Benefiting from network position: firm capabilities, structural holes, and performance', *Strategic Management Journal*, 26/9: 809–25. DOI: 10.1002/smj.482
- Zdravkovic, M., Chiwona-Karltun, L., & Zink, E. (2016). 'Experiences and perceptions of South–South and North–South scientific collaboration of mathematicians, physicists and chemists from five southern African universities', *Scientometrics*, 108/2: 717–43. Springer Netherlands. DOI: 10.1007/s11192-016-1989-z
- Zhou, P., & Leydesdorff, L. (2006). 'The emergence of China as a leading nation in science', *Research Policy*, 35/1: 83–104. DOI: 10.1016/j.respol.2005.08.006

Valorisation Addendum

In accordance with article 23.5 of the Regulation governing the attainment of doctoral degrees at Maastricht University, this section discusses the valorisation opportunities of this doctoral thesis. The main ambition of this thesis is to provide insights to create and manage policies that stimulate the development of scientific capacity in lower income countries. A core target group of this research consists of policy makers in developing countries and international agencies that give advice on science, technology and innovation policies for low and middle-income countries.

In this regard, the sixth chapter of this thesis, about the scientific system in Latin America and collaborations with industry, was shared with some of the directors of *Inter-American Development Bank (IDB)* and the *Ibero-American Programme on Science and Technology for Development (CYTED)*. After some interaction with them, my co-author (Fernando Vargas) and I produced an addendum to the paper, requested by *IDB*, that tried to understand if two specific technologies (Plant Breeding and Bioleaching) have the potential for knowledge-based growth in Latin America.

Currently, I am also using some of the knowledge and techniques acquired and developed during the thesis to work on two research projects at SPRU (Science Policy Research Unit, University of Sussex): 1) A Department for International Development (DFID) project that seeks to develop an understanding of knowledge systems and what works to promote science, technology and innovation in Kenya, Rwanda and Tanzania; 2) An UK Research & Innovation (UKRI) / United Nations Development Programme (UNDP) joint report that aims to study how science, research, innovation can best contribute to meeting the Sustainable Development Goals for developing countries.

Other target group are corporations that create and provide services related with science and technology indicators (e.g. *Clarivate*, *Elsevier*, *Times Higher Education World University Rankings*, *ARWU Ranking* and consultancy companies that create policy/research reports using S&T indicators). During my thesis, some interactions were done with *Clarivate* in order to discuss the usage of a citation-based evaluation tool they own (*InCites*).

All original research in this dissertation was (is to be) presented in international conferences and published in peer-reviewed journals or book chapters. Chapter two was presented in a workshop in Stellenbosch (2016) entitled “Science and development: Growth, expansion, and the role of the university system in South Africa”, DRUID16 conference (Copenhagen), STI Conference 2016 (Valencia), and is published in *Research Policy*, a scientific journal well known for examining empirically and theoretically the interaction between innovation, technology and research with a policy focus.

The third chapter was presented in STI2017 conference (Paris), WICK2017 workshop (Turin) and is published in *Research Evaluation*, a journal focused on the evaluation of activities concerned with scientific research, technological development and innovation.

The first version of chapter four was presented in Cape Town on December 11, 2017, and a more advanced version in Globelics 2018 (Accra). This chapter is to be included as a book chapter in a book published by *Springer* on “Africa and the Sustainable Development Goals”. This book is part of the agenda of the *World University Network Global Africa Group*, which aims to enhance the opportunities for building global research collaborations, innovations and impact in support of Africa’s development agenda.

The sixth chapter was presented in Globelics 2015 (Havana), and it is published in the *Journal of Technology Transfer*, a scientific journal that emphasizes research on strategies of knowledge and technology transfer between academia and industry.

Finally, the fifth chapter was presented in two internal workshops in UNU-MERIT and SPRU, and is yet to be presented in other international conferences.

The research also targets the readership in journals and blogs. As part of the dissemination efforts, a section of the first chapter of this thesis was discussed in a short video interview, available on *Youtube*, and was also transformed in a research report published in the *UNU-MERIT* website. I intend to follow a similar approach for my fifth chapter on health research in Africa.

About the author

Hugo was born in Portugal in 1986. He holds a bachelor degree in Economics (University of Coimbra) and a master degree in Economics and Management of Science, Technology and Innovation (University of Lisbon). Since 2010 he has worked as an economist and consultant on research management, knowledge transfer, business development and science, technology and innovation (STI) indicators.

In September 2014, he joined UNU-MERIT and Maastricht University. During his PhD, Hugo worked as a tutor in data science, econometrics, STI policy (master level) and macroeconomics (bachelor level); and he also worked for two years as a co-coordinator of the Joint UNU-MERIT/School of Governance Seminars. Between March and May 2017, he was a visiting PhD student at CREST, University of Stellenbosch (South Africa). Since September 2018, Hugo has been working as a Research Fellow at the Science Policy Research Unit, University of Sussex (United Kingdom) in two projects that aim to study how STI can best contribute to achieve sustainable development in low- and middle-income countries.

