Research Evaluation, 27(3), 2018, 222–237
doi: 10.1093/reseval/rvy017
Advance Access Publication Date: 14 June 2018
Article

The characteristics of highly cited researchers in Africa

Hugo Confraria[1,*], Jaco Blanckenberg[2] and Charl Swart[2]

[1]UNU-MERIT, Maastricht University, Boschstraat 24, Maastricht 6211AX, The Netherlands and [2]Centre for Research on Evaluation, Science and Technology (CREST) and DST/NRF, Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy, (SciSTIP), Stellenbosch University, Private Bag X1, Matieland 7602, South Africa

*Corresponding author. Email: confraria@merit.unu.edu

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 article, 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 do their highest qualification in an Anglo-Saxon university (the USA, the UK, Canada, or Australia), have a higher probability of producing highly cited research. We conclude by arguing about the duality of our results. On the 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.

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

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 article, 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 that obtained their PhD 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 allow 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 researchactive scientists that have at least one publication indexed in WoS^{TM} or Scopus^{TM} 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.

2. Background

Africa's scientific production 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 Africa's output 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 African total output (AOSTI 2014). African countries have become focused in agricultural sciences and related areas, such as environmental and ecology sciences, plant and animal sciences, as well as in some specific health sciences (Confraria and Godinho 2015). Scientific areas with higher potential to support innovation, such as engineering, material sciences, and molecular biology have been under-represented in terms of scientific output (Pouris 2010; Juma 2016).

The importance of international collaboration and the legacy of colonial ties are also recognized as playing a pivotal role in Africa's scientific output. Bibliometric studies usually find little scientific co-authorship between African countries with preference given to collaboration with higher-income nations (Narváez-Berthelemot et al. 2002; Mêgnigbêto 2013). 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 (Toivanen and Ponomariov 2011; Adams et al. 2013; Mêgnigbêto 2013). The only exception is South Africa that seems to be playing a key role in coordinating some research networks across Africa (Confraria and 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 and Godinho, 2015). It has been suggested that this may be because some scientific communities in these countries have very high levels of international collaboration; 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 (Confraria et al. 2017).

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, HCRs 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.

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^{TM} by field to identify which were the authors that 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 and 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 over-represented 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. Sinatra et al. 2016; Bornmann et al. 2017). Different thresholds are employed, from the top 1%, 5%, and 10% 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 contributes to scientific progress.

In this study, since our aim is to assess the characteristics of African researchers that produce high-impact work, we will consider researchers that are authors of the top 10% and 5% most cited papers published each year (between 2010 and 2014) for each discipline (normalized by subject area—252 WoS^{TM} categories), and compare them to African researchers that 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 illuminated by existing indicators may not capture science that is not visible through them (Hicks et al. 2015; Lopez Pineiro and Hicks 2015; Chavarro et al. 2017).

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 and actuality. These types of articles have been associated with opening a research field or changing the direction of a field (Aksnes 2003; Aksnes and Rip 2009); therefore, we will investigate whether researchers that were authors of this type of papers possess some characteristics that allow them to produce high-impact work.

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.

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 and Combes 2013; Larivière and Costas 2016). This also holds true for high-impact papers (Abramo et al. 2014). 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.

2.2.2 *Region of highest qualification*

There are also reasons to expect that 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 connect and recombine ideas with people with different backgrounds. 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). Franzoni et al. (2012) also show that migrant scientists have a higher propensity to establish international links, collaborate with co-authors in a large number of countries, and perform better than non-mobile scientists. We will model this 'mover's advantage' (Franzoni et al. 2014) by including three dummy variables for researchers that 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.

2.2.3 *Collaboration patterns*

The positive effects of collaboration on impact of research are also widely accepted (Narin et al. 1991; Glänzel et al. 1995; Katz and Martin 1997). Researchers are likely to develop new and alternative ways of thinking when they interact with other researchers 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).

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.

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 that 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.

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 (Sánchez-Peñas and Willett 2006; van den Besselaar and Sandström 2016). 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 and 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.

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. Data and methodology

This article combines survey data with bibliometric data. Survey data were collected via a self-administered, web-based, structured questionnaire. It was adapted from the questionnaire used for the Global State of Young Scientists precursor study (GLOSYS) (Friesenhahn and 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 maximum of 36 items. It was initially developed in English and then translated into French 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 7,513 answers.

Survey respondents were linked to a WoS™ author identifier based on their email address and the author disambiguation algorithm described in Caron and 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 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 that 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 that reported that they belong to humanities-related fields due to the limitations of bibliometric indicators in this area. We also excluded researchers that published their first article in WoS™ after 2013, to only consider researchers that have at least one full year of experience after producing their first publication in our main period of interest (2010–14). Finally, researchers that 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 top 10% publication, and 95 were authors of top 5% publications (see Table A1 for descriptive statistics).

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 top 10% and 5% highly cited publications[1] 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 and 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 data set 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 is a function of the regressors of interest that follows a log-linear specification:

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

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) collaboration patterns (collaboration intensity with researchers at own institution, other institutions in own country, institutions in other African countries, and institutions outside of Africa)[2]; 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 area[3], 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 A2). The sign of the estimated parameters β in the regression indicates 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 top 10% 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

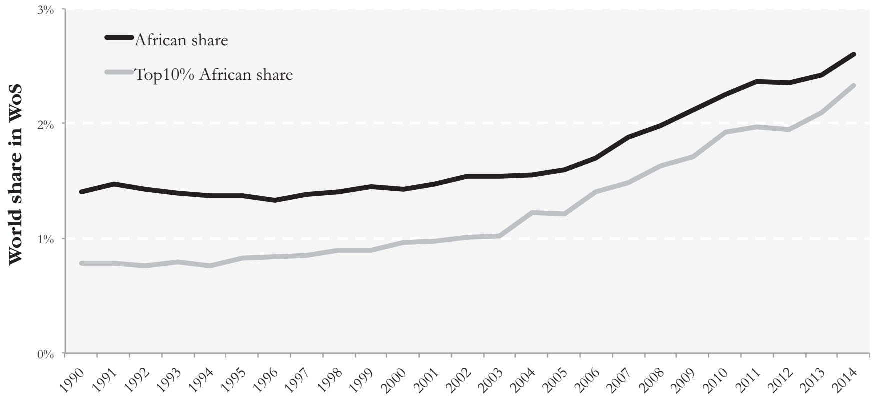

Figure 1. Trends in African output (world share) and top 10% cited papers (world share).
Source: WoS™.

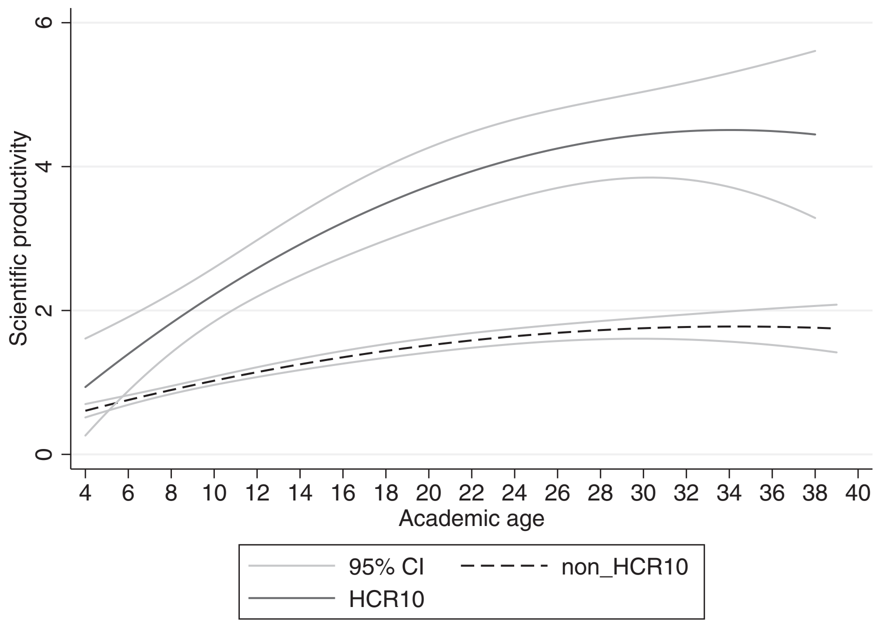

Figure 2. Scientific productivity vs. academic age, by being or not an HCR10.
Note: Scientific productivity = number of publications in WoS™ per academic age; academic age = 2017 minus year of first publication in WoS™.
Source: WoS™ and own elaboration.

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.

4. Results

4.1 Bibliometric descriptive statistics

Africa's scientific output in WoS™ has increased considerably during the past decade. In Fig. 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 (top 10%) 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; Pouris and Ho 2014; Confraria and Godinho 2015), 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).

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. 2, we show the average scientific productivity per academic age, by two groups of researchers HCR and non-HCR (top 10%). On average, HCR produces more publications per year at any stage of their career. In Figs A1 and A2, we also computed the density distributions of scientific productivity and academic age between the two groups of researchers. In both specifications (top 10 and 5%), on average, HCR produces almost three times more papers per academic year than non-HCR. As 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.

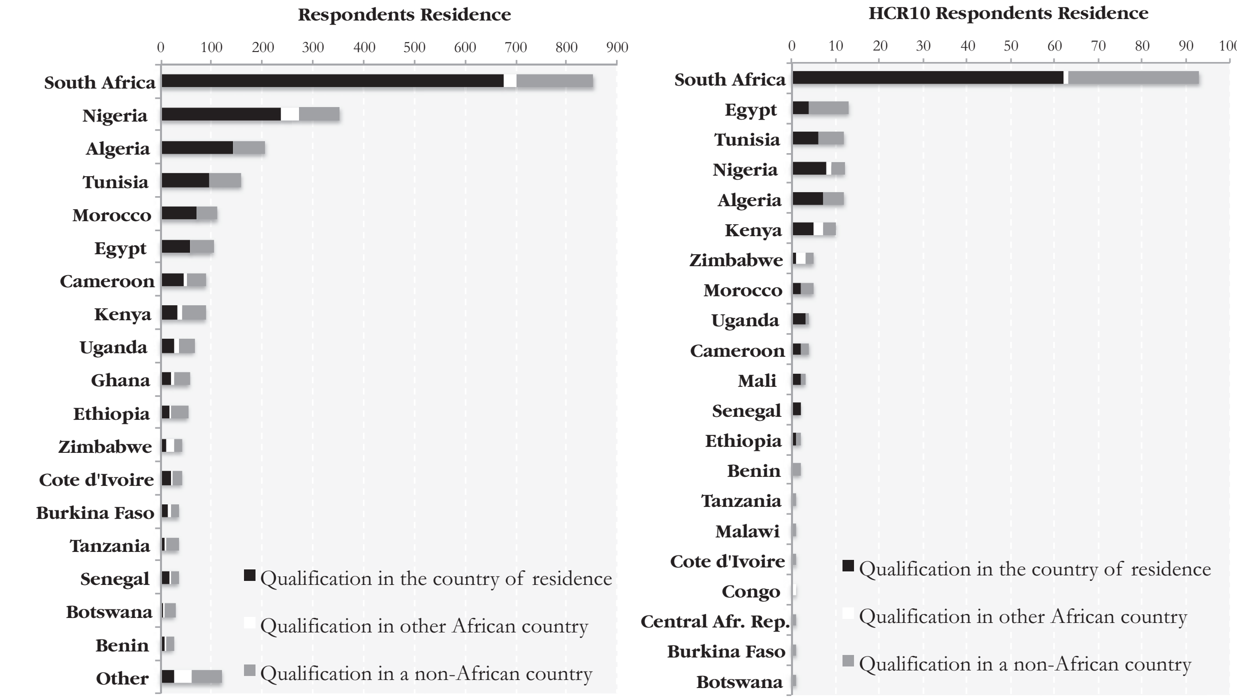

Figure 3. Number of researchers (and HCR) resident in an African country.
Note: In the first graph, we only included African countries that had at least 20 responses. In the graph on the right, we only included countries with at least one researcher that that had one top 10% publication.
Source: Own elaboration.

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, 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 that published at least one top 10% 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 top 10% publication that 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.

We have reasons to believe that researchers from Egypt are under-represented 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 severs and firewalls[4]. Due to these regional differences, to ensure better representation of our population in our econometric analysis, we control for ‘regions’ by generating a dummy variable for researchers that reside in: South Africa, Egypt, ‘Northern Africa’[5], and ‘Central Africa’[6].

We also asked respondents about gender and field of highest qualification (see Fig. A3). About 70% of the researchers are male. The percentage of males is similar amongst HCRs. The percentage of males is relatively higher for the subject areas ‘Engineering and Technology’ and ‘Agricultural Sciences’. Natural Sciences is the area with a higher percentage of HCR (10.5% for the top 10% indicator and 6.5% for the top 5%) and ‘Medical and Health Sciences’, ‘Social Sciences’, ‘Engineering and Technology’ and ‘Agricultural Sciences’ all have ratios below the top 10% and top 5% averages. Furthermore, in all the five areas, our sample has at least 19 researchers with a top 10% publication and 10 researchers with a top 5% publication.

Concerning collaborating patterns, in Fig. A4, we can observe that HCRs 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. A5). On average, the biggest challenge is lack of funding, and the challenge that they reported as less 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.

Table 1. Negative binomial estimation for top 10% and 5%

Independent variables	Dependent variable—count of top 10% pubs			Dependent variable—count of top 5% pubs		
	Nbreg (1)	Nbreg (2)	Nbreg (3)	Nbreg (4)	Nbreg (5)	Nbreg (6)
Scientific productivity (Pubs/acad. age)	0.41***	0.81***	0.69***	0.32***	0.65***	0.35
	(0.055)	(0.092)	(0.20)	(0.052)	(0.093)	(0.22)
Highest qualif. (PhD) in Anglo-Saxon (1—yes)	0.64***	0.61***	0.64***	1.09***	1.07***	1.09***
	(0.23)	(0.23)	(0.23)	(0.30)	(0.31)	(0.30)
Highest qualif. (PhD) in France (1—yes)	0.34	0.36	0.36	0.69*	0.68*	0.69*
	(0.31)	(0.31)	(0.31)	(0.41)	(0.41)	(0.41)
Highest qualif. (PhD) in other non-African country (1—yes)	0.21	0.13	0.19	0.65	0.58	0.65
	(0.29)	(0.29)	(0.29)	(0.42)	(0.42)	(0.42)
Collab. intensity—own institution (1–5)	0.038	0.048	0.031	0.086	0.099	0.086
	(0.073)	(0.072)	(0.073)	(0.10)	(0.10)	(0.10)
Collab. intensity—outside Africa (1–5)	0.21***	0.18**	0.30***	0.27***	0.25**	0.28**
	(0.073)	(0.073)	(0.091)	(0.10)	(0.10)	(0.13)
Collab. intensity—own country other inst (1–5)	0.035	0.0079	0.040	−0.014	−0.047	−0.013
	(0.077)	(0.076)	(0.076)	(0.10)	(0.10)	(0.10)
Collab. intensity—other country ins. Africa (1–5)	−0.043	−0.062	−0.044	–−0.013	−0.032	−0.014
	(0.069)	(0.068)	(0.069)	(0.095)	(0.093)	(0.094)
Academic age (2017—year of first publication)	0.021**	0.014	0.020**	0.023*	0.016	0.023*
	(0.0096)	(0.0094)	(0.0095)	(0.013)	(0.013)	(0.013)
Gender (1—female)	−0.17	−0.20	−0.16	0.044	−0.00020	0.044
	(0.19)	(0.19)	(0.19)	(0.25)	(0.25)	(0.25)
Lack of training opport. (1—yes ‘to a large extent’)	0.084	0.041	0.081	–0.57	–0.62	–0.56
	(0.26)	(0.26)	(0.26)	(0.44)	(0.45)	(0.44)
Lack of mobility opport.	−0.31	−0.31	−0.32	−0.43	−0.42	−0.43
(1—yes ‘to a large extent’)	(0.26)	(0.26)	(0.26)	(0.42)	(0.42)	(0.42)
Lack of mentorship	–0.29	–0.27	−0.29	−0.13	−0.11	−0.13
(1—yes ‘to a large extent’)	(0.25)	(0.25)	(0.25)	(0.34)	(0.35)	(0.34)
Lack of research funds	−0.097	−0.029	−0.10	−0.15	−0.058	−0.15
(1—yes ‘to a large extent’)	(0.18)	(0.18)	(0.18)	(0.25)	(0.25)	(0.25)
Scientific productivity^2 (polynomial)		−0.041***			−0.028***	
		(0.0087)			(0.0069)	
Productivity × Collab. out. Africa (Interaction)			−0.064			−0.0051
			(0.042)			(0.049)
Constant	−3.89***	−4.11***	−4.26***	−4.80***	−4.94***	−4.83***
	(0.44)	(0.45)	(0.50)	(0.62)	(0.64)	(0.69)
Lndelta	1.02***	0.80***	1.03***	1.07***	0.92***	1.07***
	(0.19)	(0.20)	(0.19)	(0.30)	(0.32)	(0.30)
Regional effects	Yes	Yes	Yes	Yes	Yes	Yes
Subject area effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,490	2,490	2,490	2,490	2,490	2,490
Wald chi^2	206.06	320.25	212.49	162.6	212.75	162.56
Pseudo R^2	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 five of the six OECD categories: natural sciences, agricultural sciences, engineering and technology, medical and health sciences, and social sciences; and four regions: South Africa, Egypt, Northern Africa, and Central Africa. *Source:* Own calculations.

4.3 Regression results

We used Stata™ to compute our regressions. The results in Table 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 that 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 and 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 and van den Besselaar (2016) found in the Swedish context, our data show positive but decreasing marginal returns between scientific productivity and the probability of producing high-impact papers.

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 PhD 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.

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 from scientific productivity, we interacted both in model Specifications 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 indicate 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.

To complement this analysis, we carried out two robustness checks (Tables A3 and A4). First, we carried out the same analysis as in Table 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 that 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 A3) are identical to the ones in Table 1. In Table A4, the only difference between age levels seems to be that older researchers are the ones benefiting more from having done their highest qualification in an Anglo-Saxon university.

5. Discussion and conclusions

In this article, 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 amount 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. 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 to increase the scientific capabilities of African researchers, not only should the scientific institutions in Africa continue to improve and open their scientific and educational programmes, but also 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. This diaspora model seems appealing both for policymakers and researchers in that it appears to present a straightforward way to take advantage of the scientific capabilities of other countries and create a bridge to their own country. However, interaction must be created between these researchers and the home scientific community through exchanges and common projects. Otherwise, the connection between young researchers and their home colleagues may be lost and fruitful interactions may cease to occur. Furthermore, this recommendation may be subject to budget constraint and as a consequence may not be achievable. Other alternative ways such as financial support to African PhD students (or researchers) to international scientific events, organization of international scientific events on a regular basis at a national, regional, or African continent level are also desirable[7].

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 that are more integrated in 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. 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 (Freeman 2005; 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 advise), fund raising, and innovation activities (Tijssen and 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 regard to our study. First, we use a threshold to define what a highly cited paper is or not. There may be many researchers that produced papers that were close to being top 10% 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 and Kontinen 2014). Therefore, this may lead to an identification problem. Nevertheless, since the email addresses in WoS^{TM} 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.

Notes

1. Normalized by field (252 WoS^{TM} categories) and year.
2. 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.
3. 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, and social sciences and humanities.
4. 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.
5. Algeria, Libya, Morocco, and Tunisia.
6. All African countries except South Africa, Egypt, and 'Northern' African countries.
7. We thank a referee for this point.

Acknowledgements

The authors express grateful thanks to Robin Cowan, Lili Wang, René Belderbos, Bart Verspagen, Aldo Geuna, and Johann Mouton for their suggestions and insights. The authors also thank Rodrigo Costas for providing valuable data. This article also benefited from comments made by the participants of the STI2017 conference (Paris, 2017) and WICK2017 workshop (Turin, 2017). The authors acknowledge support from the Young Scientists in Africa project funded by Robert Bosch Stiftung and the International Development Research Centre (IDRC). Any remaining error is the authors.

References

Abramo, G., Cicero, T., and D'Angelo, C. A. (2014) 'Are the Authors of Highly Cited Articles Also the Most Productive Ones?', *Journal of Informetrics*, 8/1: 89–97.

Adams, J. (2013) 'Collaborations: The Fourth Age of Research', *Nature*, 497/7451: 557–60.

Adams, J. et al. (2013) 'International Collaboration Clusters in Africa', *Scientometrics*, 98/1: 547–56.

Aksnes, D. W. (2003) 'Characteristics of Highly Cited Papers', *Research Evaluation*, 12/3: 159–70.

Aksnes, D. W., and Rip, A. (2009) 'Researchers' Perceptions of Citations', *Research Policy*, 38/6: 895–905.

Allen, T. D. et al. (2004) 'Career Benefits Associated with Mentoring for Proteges: A Meta-Analysis', *Journal of Applied Psychology*, 89/1: 127–36.

AOSTI. (2014). 'Assessment of Scientific Production in the African Union Member States 2005-2010'. <http://aosti.org/index.php/publi/66-assessment-of-scientific-production-in-the-african-union-member-states-2005-2010> accessed 9 April 2015.

Barnard, H., Cowan, R., and 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.

Bornmann, L., Bauer, J., and Schlagberger, E. M. (2017) 'Characteristics of Highly Cited Researchers 2015 in Germany', *Scientometrics*, 111/1: 543–5.

Boshoff, N. (2009) 'Neo-Colonialism and Research Collaboration in Central Africa', *Scientometrics*, 81/2: 413–34.

Bosquet, C., and Combes, P. P. (2013) 'Are Academics Who Publish More Also More Cited? Individual Determinants of Publication and Citation Records', *Scientometrics*, 97/3: 831–57.

Cameron, A., and Trivedi, P. (2013). *Regression Analysis of Count Data*, 2nd edn. Cambridge, UK: Cambridge University.

Carbonnier, G., and Kontinen, T. (2014). North-South Research Partnership: Academia Meets Development? <http://www.eadi.org/typo3/fileadmin/Documents/Publications/EADI_Policy_Paper/EADI_policy_paper_Carbonnier_Kontinen_FINAL.pdf> accessed 30 March 2017.

Caron, E., and van Eck, N. J. (2014). 'Large Scale Author Name Disambiguation Using Rule-Based Scoring and Clustering—Research Portal', in E. Noyons (ed.) Proceedings of the Science and Technology Indicators Conference 2014, pp. 79–86. Leiden: Universiteit Leiden.

Chavarro, D., Tang, P., and Rafols, I. (2017) 'Why Researchers Publish in Non-Mainstream Journals: Training, Knowledge Bridging, and Gap Filling', *Research Policy*, 46/9: 1666–80.

Confraria, H., and Godinho, M. M. (2015) 'The Impact of African Science: A Bibliometric Analysis', *Scientometrics*, 102/2: 1241–68.

Confraria, H., Godinho, M. M., and Wang, L. (2017) 'Determinants of Citation Impact: A Comparative Analysis of the Global South versus the Global North', *Research Policy*, 46/1: 265–79.

Evans, S. C., Ng, T., and 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.

Fleming, L. (2001) 'Recombinant Uncertainty in Technological Search', *Management Science*, 47/1: 117–32.

Franzoni, C., Scellato, G., and Stephan, P. (2012) 'Foreign-Born Scientists: Mobility Patterns for 16 Countries', *Nature Biotechnology*, 30/12: 1250–3.

Franzoni, C., Scellato, G., and Stephan, P. (2014) 'The Mover's Advantage: The Superior Performance of Migrant Scientists', *Economics Letters*, 122/1: 89–93.

Freeman, R. B. (2005). Does Globalization of the Scientific/Engineering Workforce Threaten U.S. Economic Leadership? <http://www.nber.org/papers/w11457> accessed 28 January 2016.

Friesenhahn, I., and Beaudry, C. (2014). *The Global State of Young Scientists – Project Report and Recommendations*. Berlin: Akademie Verlag.

Garfield, E. (1977) 'The 250 Most-Cited Primary Authors, 1961-1975. Part 1. How the Names Were Selected', *Current Contents*, 3/49: 326–36.

Garfield, E. (1981) 'The 1000 Contemporary Scientists Most-Cited 1965–1978. Part 1. The Basic List and Introduction', *Current Contents*, 5/41: 269–78.

Garfield, E. (1983) 'Mapping Science in the Third World', *Science and Public Policy*, 10/3: 112–27.

Geffers, J., Beaudry, C., Yang, H.-C., Huang, F., Phanraksa, O., Dominik, M., Lin, Y.-C., Huang, M.-C., Komai, S., Lorimer, K., Piyawattanametha, W., Saengchantr, P., Saleh, H., Tagg, B., and Veerakumarasivam, A. (2017). *Global State of Young Scientists in ASEAN. Creativity and Innovation of Young Scientists in ASEAN*. Halle (Saale) Germany: Global Young Academy.

Gilbert, G. N. (1977) 'Referencing as Persuasion', *Social Studies of Science*, 7/1: 113–22.

Glänzel, W., Rinia, E. J., and 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., and 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.

Goldfinch, S., Dale, T., and DeRouen, K., Jr., (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.

Hicks, D. et al. (2015) 'Bibliometrics: The Leiden Manifesto for Research Metrics', *Nature*, 520/7548: 429–31.

Hollingsworth, J. R. (2006). 'The Dynamics of American Science: An Institutional and Organizational Perspective on Major Discoveries', in *Transformationen des Kapitalismus: Festschrift für Wolfgang Streeck zum sechzigsten Geburtstag*, pp. 361–80. Frankfurt, NY: Campus Verlag.

Hunter, R. S., Oswald, A. J., and Charlton, B. G. (2009) 'The Elite Brain Drain', *The Economic Journal*, 119/538: F231–51.

Jones, B. F., Wuchty, S., and Uzzi, B. (2008) 'Multi-University Research Teams: Shifting Impact, Geography, and Stratification in Science', *Science*, 322/5905: 1259–62.

Juma, C. (2016). 'Education, Research, and Innovation in Africa: Forging Strategic Linkages for Economic Transformation', Discussion Paper, 2016-01, Science, Technology, and Globalization Project, Belfer Center, February. <https://www.belfercenter.org/publication/education-research-and-innovation-africa-forging-strategic-linkages-economic>

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.

Katz, J. S., and Martin, B. R. (1997) 'What Is Research Collaboration?', *Research Policy*, 26/1: 1–18.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press.

Lariviére, V., and Costas, R. (2016) 'How Many Is Too Many? On the Relationship between Research Productivity and Impact', *PLoS One*, 11/9: e0162709.

Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.

Lopez Pineiro, C., and Hicks, D. (2015) 'Reception of Spanish Sociology by Domestic and Foreign Audiences Differs and Has Consequences for Evaluation', *Research Evaluation*, 24/1: 78–89.

Mêgnigbêto, E. (2013) 'International Collaboration in Scientific Publishing: The Case of West Africa (2001–2010)', *Scientometrics*, 96/3: 761–83.

Merton, R. K. (1968) 'The Matthew Effect in Science the Reward and Communication Systems of Science Are Considered', *Science*, 159/3810: 56–63.

Moed, H. F. (2005). *Citation Analysis in Research Evaluation. Information Science and Knowledge Management*, 9. Berlin; Heidelberg: Springer-Verlag.

Mouton, J. (2008) 'Africa's Science Decline', *Harvard International Review*, 30/3: 46–51.

Narin, F., Stevens, K., and Whitlow, E. S. (1991) 'Scientific Co-Operation in Europe and the Citation of Multinationally Authored Papers', *Scientometrics*, 21/3: 313–23.

Narváez-Berthelemot, N. et al. (2002) 'Science in Africa: An Overview of Mainstream Scientific Output', *Scientometrics*, 54/2: 229–41.

Parker, J. N., Allesina, S., and 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.

Parker, J. N., Lortie, C., and 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.

Peters, H. P. F., and 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.

Pouris, A. (2010) 'A Scientometric Assessment of the Southern Africa Development Community: Science in the Tip of Africa', *Scientometrics*, 85/1: 145–54.

Pouris, A., and Ho, Y.-S. (2014) 'Research Emphasis and Collaboration in Africa', *Scientometrics*, 98/3: 2169–84.

Prozesky, H., and 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.

Roberts, R. M. (1989). *Serendipity: Accidental Discoveries in Science*. New York: Wiley.

Sánchez-Peñas, C., and 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.

Sandström, U., and van den Besselaar, P. (2016) 'Quantity and/or Quality? The Importance of Publishing Many Papers', P. Dorta-González (ed.)', *PLoS One*, 11/11: e0166149.

Simonton, D. K. (1999). *Origins of Genius: Darwinian Perspectives on Creativity*. Oxford University Press.

Sinatra, R. et al. (2016) 'Quantifying the Evolution of Individual Scientific Impact', *Science*, 354/6312: aaf5239.

Sugimoto, C. R. et al. (2017) 'Scientists Have Most Impact When They're Free to Move', *Nature*, 550/7674: 29–31.

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.

Tijssen, R., and Kraemer-Mbula, E. (2017) 'Research Excellence in Africa: Policies, Perceptions, and Performance', *Science and Public Policy*. <https://academic.oup.com/spp/advancearticle/doi/10.1093/scipol/scx074/4600842>

Toivanen, H., and Ponomariov, B. (2011) 'African Regional Innovation Systems: Bibliometric Analysis of Research Collaboration Patterns 2005–2009', *Scientometrics*, 88/2: 471–93.

UNESCO. (2015). *UNESCO Science Report: Towards 2030*. Paris: UNESCO Publishing.

Uzzi, B. et al. (2013) 'Atypical Combinations and Scientific Impact', *Science*, 342/6157: 468–72.

van den Besselaar, P., and Sandström, U. (2016) 'Gender Differences in Research Performance and Its Impact on Careers: A Longitudinal Case Study', *Scientometrics*, 106/1: 143–62.

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.

Weinberg, B. A. (2011) 'Developing Science: Scientific Performance and Brain Drains in the Developing World', *Journal of Development Economics*, 95/1: 95–104.

Appendix

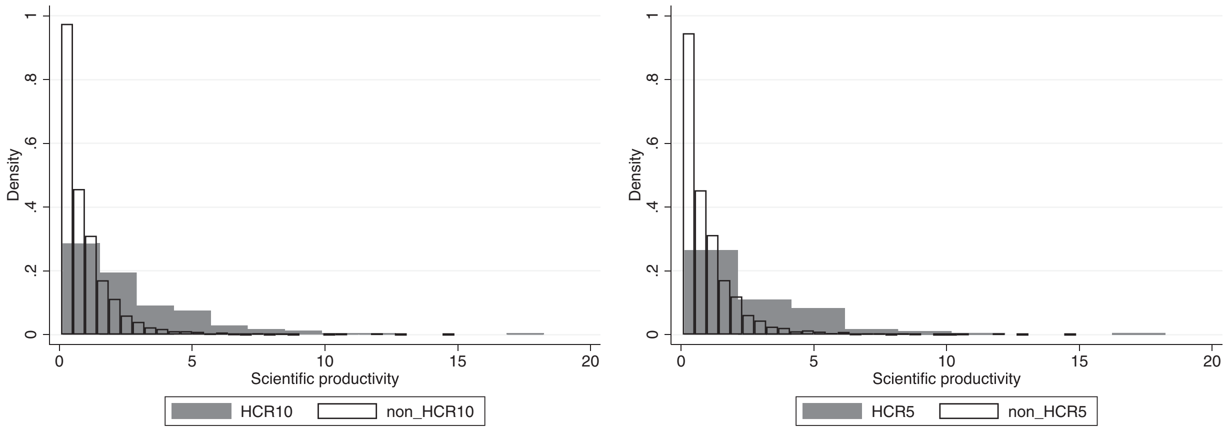

Figure A1. Density distribution of researcher's number of publications per academic age—top 10% and top 5%.

Source: WoS™ and own elaboration.

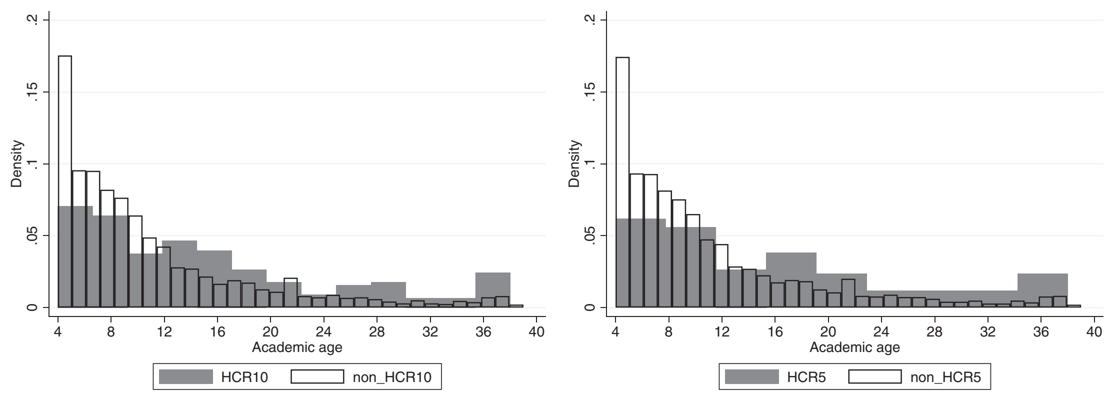

Figure A2. Density distribution of researcher's academic age—top 10% and top 5%.

Source: WoS™ and own elaboration.

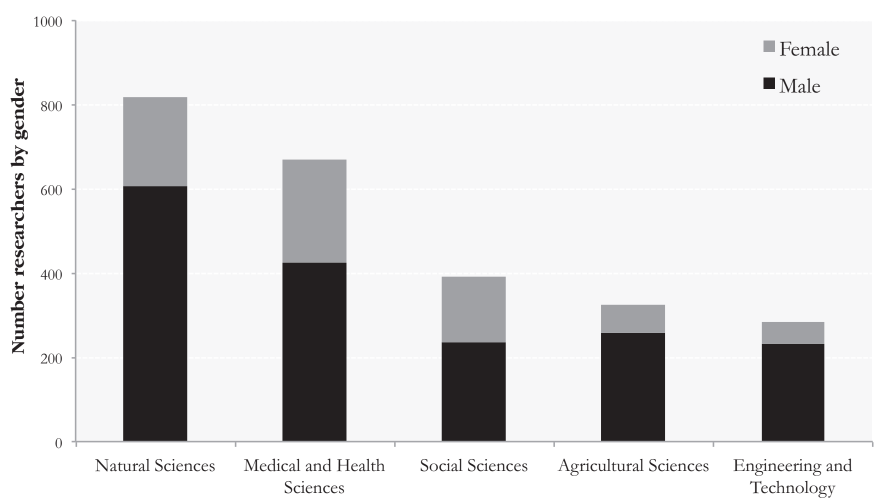

Figure A3. Number of researchers by area and gender.

Source: Own elaboration.

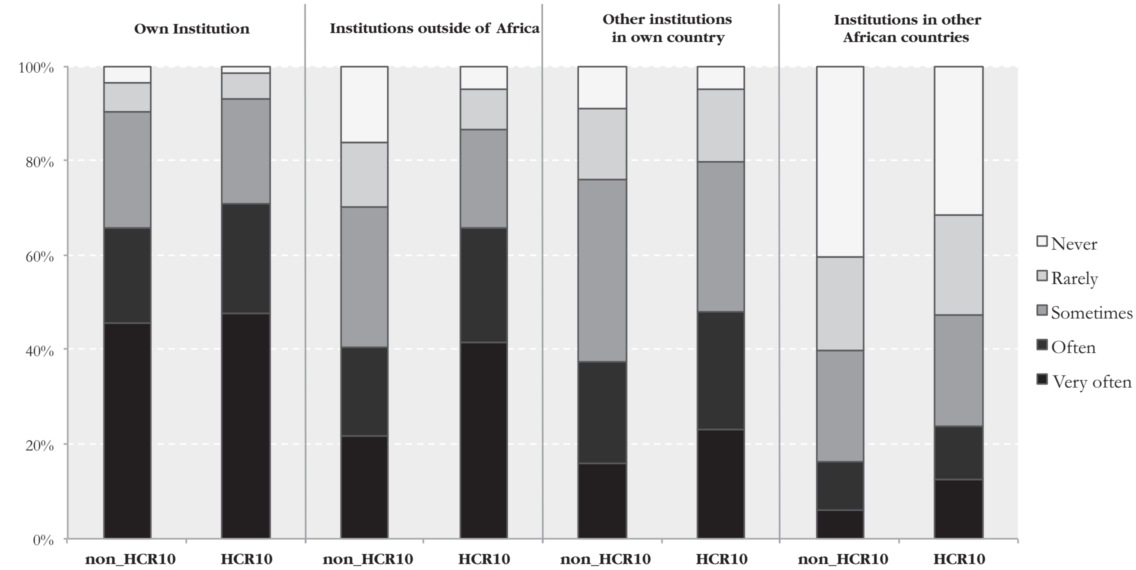

Figure A4. Collaboration patterns, by being or not an HCR.

Source: Own elaboration.

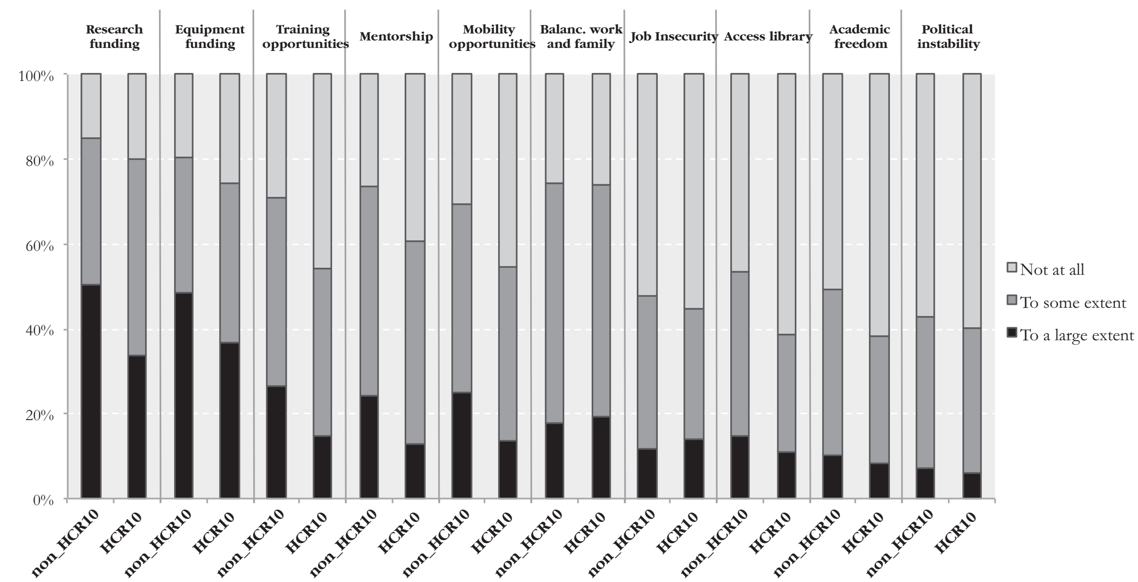

Figure A5. Challenges faced, by being or not an HCR.

Source: Own elaboration.

Table A1. Descriptive statistics.

Variables	Obs	Mean	Std. Dev.	Min	Max
Dummy HCR10	2,490	0.07	0.26	0	1
Dummy HCR5	2,490	0.04	0.19	0	1
Count of top 10% pubs	2,490	0.11	0.49	0	8
Count of top 5% pubs	2,490	0.05	0.29	0	6
Scientific Productivity	2,490	1.15	1.48	0.06	18.24
Highest qualif. in Africa	2,490	0.72	0.45	0	1
Highest qualif. in France	2,490	0.09	0.29	0	1
Highest qualif. in Anglo-Saxon	2,490	0.09	0.28	0	1
Highest qualif. in other non-African country	2,490	0.10	0.30	0	1
Collab. intensity – own institution	2,490	4.00	1.12	1	5
Collab. intensity – outside Africa	2,490	3.18	1.35	1	5
Collab. intensity – own country other inst	2,490	3.21	1.15	1	5
Collab. intensity – other country ins. Africa	2,490	2.20	1.23	1	5
Academic age	2,490	11.50	7.70	4	39
Gender	2,490	0.29	0.46	0	1
Lack of training opport.	2,490	0.27	0.45	0	1
Lack of mobility opport.	2,490	0.25	0.44	0	1
Lack of mentorship	2,490	0.24	0.43	0	1
Lack of research funds	2,490	0.51	0.50	0	1
Dummy South Africa	2,490	0.34	0.47	0	1
Dummy Egypt	2,490	0.04	0.20	0	1
Dummy Central Africa	2,490	0.43	0.50	0	1
Dummy North Africa	2,490	0.19	0.39	0	1
Dummy natural sciences	2,490	0.32	0.47	0	1
Dummy agricultural sciences	2,490	0.13	0.34	0	1
Dummy eng. and tech.	2,490	0.11	0.32	0	1
Dummy medical sciences	2,490	0.27	0.44	0	1
Dummy social sciences	2,490	0.16	0.37	0	1

Source: Own calculations.

Table A2. Description of variables used in the model

Variables	Description
Dummy HCR10	Dummy variable that is 1 for an African author, which is an author in a top 10% highly cited paper between 2010 and 2014
Dummy HCR5	Dummy variable that is 1 for an African author, which is an author in a top 5% highly cited paper between 2010 and 2014
Count of top 10% pubs	Number of top 10% highly cited papers authored between 2010 and 2014
Count of top 5% pubs	Number of top 5% 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 A3. Probit estimation for top 10% and top 5%

Independent variables	Dependent variable—dummy HCR10			Dependent variable—dummy HCR5		
	Probit (1)	Probit (2)	Probit (3)	Probit (4)	Probit (5)	Probit (6)
Scientific productivity (Pubs/acad. age)	0.21***	0.46***	0.23**	0.17***	0.33***	0.060
	(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***
	(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*
	(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
	(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
	(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
	(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
	(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
	(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
	(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
	(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
	(0.13)	(0.13)	(0.13)	(0.18)	(0.18)	(0.18)
Lack of mobility opportunities	−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	−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	−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**	
		(0.0070)			(0.0062)	
Productivity×COL outside Africa (interaction)			−0.0051			0.026
			(0.023)			(0.024)
Constant	−2.06***	−2.17***	−2.08***	−2.40***	−2.46***	−2.27***
	(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	2,490	2,490	2,490	2,490	2,490	2,490
Wald chi^2	131.04	200.39	130.87	120.57	140.22	123.67
Pseudo R^2	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 five of the six OECD categories: natural sciences, agricultural sciences, engineering and technology, medical and health sciences, and social sciences; and four regions: South Africa, Egypt, Northern Africa, and Central Africa. *Source:* Own calculations.

Table A4. Negative binomial estimation for top 10% by age level

Independent variables	Dependent variable—count of top 10% pubs (nbreg est)	
	Academic age wos <10	Academic age wos ≥10
Scientific productivity (pubs/acad. age)	0.42***	0.44***
	(0.14)	(0.064)
Highest qualification (PhD) in Anglo-Saxon (1—yes)	0.62	0.75***
	(0.43)	(0.26)
Highest qualification (PhD) in France (1—yes)	−0.015	0.57
	(0.58)	(0.38)
Highest qualif. (PhD) in other non-African country (1—yes)	0.30	−0.046
	(0.38)	(0.41)
Collaboration intensity—own institution (1–5 Likert)	0.014	0.089
	(0.099)	(0.090)
Collaboration intensity—outside Africa (1–5 Likert)	0.26**	0.15*
	(0.12)	(0.089)
Collaboration intensity—own country other inst (1–5 Likert)	0.0027	0.084
	(0.13)	(0.098)
Collaboration intensity—other country ins. Africa (1–5 Likert)	−0.099	0.0052
	(0.11)	(0.089)
Gender (1—female)	−0.25	−0.24
	(0.34)	(0.21)
Lack of training opportunities (1—yes 'to a large extent')	0.18	−0.070
	(0.35)	(0.38)
Lack of mobility opportunities	−0.10	−0.25
(1—yes 'to a large extent')	(0.18)	(0.15)
Lack of mentorship	−0.55	−0.12
(1—yes 'to a large extent')	(0.34)	(0.32)
Lack of research funds	−0.45	0.15
(1—yes 'to a large extent')	(0.31)	(0.22)
Constant	−3.40***	
	(0.47)	
Lndelta	1.01	
	(0.20)	
Regional effects	Yes	Yes
Subject area effects	Yes	Yes
Observations	2,490	
Wald chi^2	226.33	
Pseudo R^2	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 five of the six OECD categories: natural sciences, agricultural sciences, engineering and technology, medical and health sciences, and social sciences; and four regions: South Africa, Egypt, Northern Africa, and Central Africa. *Source:* Own calculations.