Building Bridges:  
The impact of return migration by African scientists

Caroline Viola Fry  
Massachusetts Institute of Technology  
Sloan School of Management

October 27, 2019

Abstract

Despite significant interest in the potential for ‘returnee’ scientists moving back to developing countries to connect developed and developing countries, prior work has found limited evidence of success. I shift the focus to the broader network of the returnee, and study the extent to which the return home of American-trained HIV researchers to African institutions impacts publication outcomes of non-migrant scientists in Africa. I find that following the arrival of a returnee in their institution, non-migrants experience increased productivity, mostly in HIV research. I find strong evidence that the mechanism driving this effect is that of the returnee providing a bridge to their central connections and subsequent knowledge and resources thus affecting outcomes. In settings where ‘outsiders’ struggle to access knowledge and resources that are usually reserved for exclusive ‘insiders’, this kind of bridge in the network can help through providing legitimacy to the outsiders. These findings inform a network perspective on the consequences of the mobility of skilled individuals, the development of national innovation ecosystems, and the globalization of knowledge production.

Keywords: economics of science, international migration, social network, knowledge spillovers, global innovation, economics of innovation

*Acknowledgements to Scott Stern, Pierre Azoulay and Ezra Zuckerman for advising on the project. Thanks to Megan MacGarvie, Alexander Kowalski, Wendy Bradley, Hagay Volvowsky and Brittany Bond for comments on the draft manuscript. NIH Fogarty International Center staff, US principal investigators and staff involved with the AITRP program, as well as African trainees for providing input into the research questions and assisting the interpretation of results. The usual disclaimer applies.
1 Introduction

The consequences of the global movement of high-skilled workers has animated much research.\(^1\) In particular, return migration of high-skilled workers from more developed countries to developing countries has received renewed interest in recent years (Borjas and Bratsberg 1996; Zucker and Darby 2007; Dustmann et al 2011; Gaule 2014). Great hope is attached to these return migrants and their role in transforming their home country economy through brokering access to knowledge and resources in more developed countries (Saxenian 2006).

Celebrated cases of returnees contributing towards home country economies (Saxenian 2006), and empirical evidence of returnees bringing back knowledge and resources (Jonkers and Tjissen 2008; Choudhury 2015; Gianetti et al 2015) support the image of return migrants as successful brokers. But another line of research provides evidence that returnees face challenges in their brokerage role due to a variety of individual and interpersonal barriers, including the presence of other returnees, home country xenophobia, and challenges in maintaining ties at home and abroad (Obukhova 2012; Wang 2015). Returnees working in knowledge production may face additional challenges due to the limited availability of resources and collaborators at home, geographic concentration of knowledge flows, and bias based on institutional or geographic affiliation. Indeed, evidence of declining productivity of scientists as they move home to developing countries seems to confirm that benefits to brokerage are limited in this setting (Gibson and McKenzie 2014; Kahn and MacGarvie 2016a).

It may be, however, that this reflects a narrow view of brokerage and diverts our attention from a broader set of causal pathways by which bridges across disparate parts of a system can benefit those who are impacted by the bridge. In particular, and following Burt’s notion of ‘second-hand brokerage,’ it could be that while the returnees themselves benefit relatively little from the connections they make, their associates in developing countries do indeed benefit. Evidence that a broker’s connections, and subsequent access to knowledge and resources, can be shared or ‘borrowed’ by their associates is limited (Burt 2007, 2010). However, there is strong theoretical and empirical reason to expect such a benefit when the bridge allows ‘outsiders’ in a system access to connections, knowledge or resources otherwise restricted to legitimate ‘insiders’ (Burt 1997, 1998, 2010; Stuart et al., 1999). Moreover, recent research (Fry 2019) documents such borrowing in the context of scientific collaboration across the divide between developed and developing countries.

By developing and applying Burt’s idea of second-hand brokerage in the context of global science, which exhibits a core/periphery network structure (Crane 1965; Cole and Cole 1973; Zuckerman 1988; Kerr 2008; Oetl and Agrawal 2008; Hunt and Gauthier-Loiselle 2010; Agrawal et al 2011; Borjas and Doran 2012; Kogut and Macpherson 2011; Moser et al 2014; Ganguli 2015)
Zelnio 2012), we can more clearly illuminate how brokers impact systems more generally. Core/periphery structures are characterized by densely connected core actors (insiders) and loosely connected peripheral actors (outsiders) (Borgatti and Everett 1999). The selective core represents influential actors and their position is associated with privilege, control and prestige (Clauset et al 2015). An examination of the extent to which periphery actors borrow access to connections, knowledge and resources of a core actor who forms a bridge between the core and the periphery - a “core/periphery bridge” - holds promise in this setting and others.

It is difficult to identify the causal impact of such borrowing. Actors in a network may have features unobservable to the researcher that have both determined their network structure and position as well as their outcomes. This suggests that an examination of a periphery actor’s connection to a core/periphery bridge and their outcomes may conflate the role the network plays with innate qualities of the individual (Manski 1993; Jackson and Wolinsky 1996; Goldsmith-Pinkham and Imbens 2013). Furthermore, network studies generally examine connections that already exist. Therefore identifying a comparable control group is extremely difficult, as those with connections are likely different from those without connections.

To overcome these empirical challenges, I exploit variation in the timing of the formation of a core/periphery bridge through evaluating the impact of the return home of a foreign trained scientist to developing countries on outcomes of non-migrants affiliated with the institutions they return to. Scientists returning from developed countries back home to developing countries can be considered insiders whose return home results in the formation a core/periphery bridge for peripheral non-migrant scientists.

Specifically, I study the effect of the return home of 112 HIV researchers trained in top universities in the United States under the National Institute of Health Fogarty AIDS International Training and Research Program between 1988-2014. I construct a panel dataset of 1,657 non-migrant African scientists who are affected by these return events in that they are working in related fields in the institution to which the American-trained scientist returned. Matching with scientists from other institutions in Africa that do not receive a returning trainee, I am able to control for career, field and temporal trends in research output. Difference-in-differences regressions compare within scientist changes in publication outcomes of active researchers in institutions following the return of an American-trained researcher with changes in publication outcomes of observably similar researchers in other African institutions.

The results reveal increases in the rate at which non-migrant scientists collaborate with scientists from the American training institution of the returning scientist following the return event. Non-migrants also increase citation rates to scientists based in the American training institution of the returning scientist.
following the return event. Furthermore, they experience a persistent increase in publication output following the arrival home of an American-trained scientist, particularly in HIV research. The effect is most pronounced for non-migrants who are not connected to developed country scientists prior to the return event. The findings support the idea that a returning scientist forming a core/periphery bridge benefits periphery actors. Potential mechanisms are explored, and evidence is found in support of two possible drivers of the effect. Non-migrants can both borrow knowledge that the return migrant has access to, as well as assume the network position of the return migrant.

The rest of the paper proceeds as follows. Section 2 discusses the theoretical framework. Section 3 describes empirical setting, the National Institute of Health Fogarty AIDS training and research program. Section 4 describes the data and provides some descriptive statistics. Section 5 presents the results and section 6 concludes and outlines implications of the findings.

2 Theoretical Framework

2.1 Can Returning Scientists Broker?

Skilled migrants moving back to developing countries would seem to be in an ideal position to bridge developed and developing country networks and transfer connections, knowledge and resources back home. In her book ‘The New Argonauts’, Saxenian (2006) places considerable weight on this phenomenon amongst entrepreneurs moving back to Taiwan following experience in Silicon Valley:

‘But these highly skilled emigrants are now increasingly transforming the brain drain into “brain circulation” by returning home to establish business relationships or start new companies while maintaining their social and professional ties to the U.S.’

Yet despite this potential for return migrants to bring connections, knowledge and resources back home, recent work suggests the difficulties in this brokerage role amongst scientists. Kahn and MacGarvie (2016a) study the movement home of Fulbright scholars following American training and find that the returnees to developing countries experience a significant decline in their productivity as compared to carefully matched scientists who remain in the United States. They ascribe this to distance from resources and knowledge production which can limit the ability of scientists to fully transfer connections, knowledge and resources back to developing countries. Accordingly, Gibson and McKenzie (2014) find a productivity decline for scientists trained abroad returning to the Pacific Islands, even after controlling for negative selection of the returnees. Together these results suggest that the benefits
from being a broker between foreign and home locations of returning scientists are limited and that the net effect is negative.

2.2 Second-Hand Brokerage

But these discouraging findings may reflect a conception of brokerage that is overly narrow. In particular, there is reason to think that returnee scientists are especially likely to facilitate what Burt (2007) called ‘second-hand-brokerage’.

Burt introduced the concept of second-hand brokerage to capture the possibility that actors associated with a broker might be able to effectively borrow the broker’s access to connections, knowledge and resources. Figure 1 illustrates this concept, where nodes B, C and D are ‘at risk’ of borrowing A’s access to connections, knowledge and resources - or to become ‘second-hand brokers’.

Figure 1: Illustration of Potential for Second-Hand Brokerage

There are two possible mechanisms whereby second-hand brokers can gain access: (1) borrowing knowledge from the broker, (2) borrowing the network position of the broker.

Borrowing Knowledge The first of the two mechanisms assumes that knowledge can flow freely through indirect ties (Granovetter 1973; Watts and Strogatz 1998), and therefore second-hand brokers have access to the knowledge the broker has access to. As an example, node D in Figure 1 could access information from B and C, via A.

Borrowing Network Position The second mechanism assumes that networks are not static. In particular, assuming that brokerage positions close over time, and second-hand brokers become dis-intermediated (Burt 2002), second-hand brokers can dynamically assume the network position of the broker. This in turn results in the second-hand broker having access to connections, knowledge and
resources that the broker originally had access to. As an example, node D in Figure 1 could assume the position of broker A in a second time period. This mechanism subsumes the first mechanism as both result in access to the knowledge that broker A originally had access to.

Together, these mechanisms suggest that association with a broker can enhance access to connections, knowledge and resources that may lead to subsequent improvements in performance. In practice though, evidence on second-hand brokerage is limited (Burt 2007). Burt has offered two reasons why borrowing connections, knowledge and resources might be difficult. First, if the potential second-hand broker can build their own brokerage position, then there is no additional advantage to second-hand brokerage. If a potential second-hand broker can observe the structure of the network and wants to build relationships such that they are a broker, second-hand brokerage is redundant. Second, the extent to which knowledge is ‘sticky’, or hard to communicate, can overwhelm the possibility of second-hand brokerage. Another possible reason why second-hand brokerage might be limited is that the broker themselves requires an incentive to share their knowledge and connections with potential second-hand brokers. To the extent that the broker can extract rents from their position bridging disparate parts of a system, it is naïve to think that they would give that up and share their position with potential second-hand brokers without receiving something in return (Reagans and Zuckerman 2008).

Whilst each of the foregoing three considerations imply that second-hand-brokerage will have limited benefits, they also imply three contextual factors that can enhance its value. First, whereas it might often be true that potential second-hand brokers can build their own brokerage position, this is sometimes not the case. Outsiders in a system face barriers in taking advantage of their own network position due to a lack of legitimacy, and subsequent mistrust, amongst the community. However, being associated with an insider, or a sponsor can allow them to take advantage of borrowed social capital (Burt 2010). Second, whereas key knowledge and knowhow is often local, there exist settings where organizations and routines exist to transfer knowledge. For example, within organizations (Burt 2010; Choudhury 2015) or science (Mohnen 2016). And third, when the broker has an incentive to share knowledge and connections or can take credit for the success of someone else, or when they are constrained to take full advantage of their brokering position, sharing of knowledge and connections might take place. Examples of such settings are those between mentors and their protégé (Burt 2007; Choudhury 2015), or within teams (Mohnen 2016).

Burt (2010) provides a case of female managers in an electronics manufacturing firm to illustrate the potential benefit to second-hand brokerage in a setting that meets the three conditions described above. Women, who he considers outsiders here, in the firm who had male sponsors who were invested

---

2The idea that insiders have some privileged access to knowledge within the community of knowledge producers was formalised in Merton (1972), who attributes differences in status to trust.
in their careers and central to other individuals’ networks were promoted sooner. Burt attributes this difference to borrowing the social capital of male sponsors. This sponsorship from an insider reduced legitimacy problems that the women in the firm originally had and enabled them to take advantage of the borrowed social capital.

Further evidence supports this interpretation of second-hand brokerage for outsiders. Entrepreneurial firms affiliated with more prominent exchange partners have a faster rate of initial public offering and earn greater valuations at IPO than similar firms without such connections (Stuart et al 1999). Graduate students with prominent advisors tend to find better initial job placements than comparable students without such prominent advisors (Long et al 1979), and West African scientists experiencing a random shock to their ties with developed country scientists as a result of the ebola outbreak experience improvements in their publication output (Fry 2019).

Although not framed in terms of second-hand brokerage, two recent studies provide support that association with a brokering individual in knowledge production can impact outcomes. Choudhury (2015) finds that returnee managers moving back to India within a multinational enterprise exert positive benefits onto their R&D employees through forming a bridge between knowledge generated in the headquarters in the United States to R&D employees located in India. Mohnen (2016) finds that the impact of the death of star scientists on their collaborators is more negative if the star is a broker in the network, and if the collaborator is younger or more isolated.

While this prior literature has established an empirical relationship between second-hand brokerage and performance of outsiders, the scope of this effect, particularly in global science, may still be quite limited. Within an organizational or team context there are routines and incentives to transfer knowledge and motivation to see others succeed. Amongst scientists there aren’t always natural routines for sharing knowledge, and there is a great deal of competition between individuals. That being said, science is highly codified and there are incentives for scientists to distribute their knowledge, and so one might expect knowledge to flow through the network more easily than in other settings. Furthermore, there are some relationships amongst scientists - for example, the mentor-protégé relationship - in which incentives do exist to share knowledge and connections with the outsider. Together with the existence of insiders and outsiders, global science provides an interesting case in which to explore the possibilities of second-hand brokerage.
2.3 Core/Periphery Bridges in Global Science

Global science demonstrates a network structure with insiders and outsiders crudely classified by their geographic location. Classic accounts of global science networks describe the structure as core/periphery, with the majority of citations, collaborations, publications and patents occurring in more developed countries in the world, with the gap widening over time (Wagner and Leydesdorff 2005; Leydesdorff and Wagner 2008; Hwang 2008; Zelnio 2012). Peripheral actors in this system suffer from a lack of access to resources, central knowledge and other benefits.

In light of the discussion above on the relevance of second-hand brokerage for insiders and outsiders in a system, a core scientist forming a bridge to periphery scientists (a “core/periphery bridge”) provides an opportunity to test ideas about sponsorship. Periphery scientists associated with the bridge can borrow a core scientist’s access to knowledge, connections and resources. Additional features of the scientific setting give further insight as why the presence of a core/periphery bridge could result in borrowed access, and as to the dominant mechanism(s) driving any observed changes.

Given that scientific knowledge is codified and incentives exist to distribute it widely (Stephan 1996; DasGupta and David 1987), the presence of a core/periphery bridge should facilitate flows of knowledge across disparate parts of the system. Additionally, scientists act as mentors to others as a way to build a legacy. With the assumption that scientists would like to leave behind a legacy (or that they are helpful), periphery scientists associated with the core/periphery bridge could borrow their knowledge and connections. In light of this, and given previous evidence demonstrating the positive impact of access to the core on periphery scientist performance (Fry 2019), performance of the periphery scientist associated with a core/periphery bridge should be positively impacted.

To this point, the discussion has assumed that core/periphery bridges are built somehow - an assumption to which I will return momentarily. For now, assuming that such a bridge is built, two implications follow: periphery scientists associated with a core/periphery bridge can borrow access to connections, knowledge and resources of the broker, resulting in improved performance of these scientists. And further, superior performance due to association with a core/periphery bridge is driven by both borrowed knowledge, and borrowed network position of the broker.

2.4 Returning Scientists as Core/Periphery Bridges

But can core/periphery bridges be built? And if they are, will they have a causal impact? These implications are challenging to test because the network surrounding an individual is rarely randomly
determined (Manski 1993; Jackson and Wolinsky 1996; Goldsmith-Pinkham and Imbens 2013). Actors
in a network have qualities that determine both their network and their outcomes, conflating the role of
the network in outcomes with innate qualities. Furthermore, because we generally observe actors with a
given network, it is very difficult to define a comparable control group, as by the time they are observed
they have already experienced different paths.

I exploit a setting in which I can isolate the timing of the formation of a core/periphery bridge, and
then examine the outcomes accruing to the peripheral scientist associated with the bridge before and after
the shock. This longitudinal contrast removes omitted variable bias that cross-sectional comparisons
face. However, if one expects that the timing of the formation of a core/periphery bridge is endogenous
to expected performance improvements, then this difference-in-differences estimate could be biased.
That is, connections to particular periphery scientists could be made if core scientists expect that those
periphery scientists will perform well. If the best selected scientists are also those who are subject to the
formation of a core/periphery bridge, estimation of the change could reflect positive selection as opposed
to a causal effect of the core/periphery bridge. To remedy this problem, I pair each treated scientist
with a control scientist who exhibits almost-identical performance prior to the potential formation of a
core/periphery bridge, and analyse the data at the individual level of analysis in a difference-in-difference
framework. This provides a reliable way to evaluate the effect of the formation of a core/periphery bridge.

The return home of scientists to developing countries provides an ideal setting for this study. Past
research has demonstrated that mobile scientists take their networks with them (Azoulay 2012; Scellato
et al 2015). In addition, co-location increases the probability of forming a social relationship (Boudreau
et al 2017; Catalini 2018), particularly within the same institution (Azoulay et al 2012). Together, this
suggests that a scientist who returns to a developing country from a developed one occupies a position
in the core/periphery network whereby they are considered an insider, and form a bridge between two
disparate parts of the system. For periphery non-migrant scientists, the return home of a scientist from
more developed countries to their institution implies the formation of a core/periphery bridge.

To illustrate this idea, consider Figure 2. This figure, in which nodes represent scientists and ties their
connections presents a core/periphery network structure. The darker nodes at the center of the figure
(individual B and A) represent core scientists, and the lighter colored node (individual C) represent
periphery scientists (without loss of generality, core scientists are those based in the United States
and periphery scientists are those based in Zimbabwe). Panel A represents the pre-return state, where
periphery scientists are unconnected with the core. Panel B represents the network after the return
home of the scientist (scientist a) from the United States. Taking her connection to core scientist B
with her, and forming a new relationship with periphery scientist C, the return home of scientist A
to Zimbabwe from the United States provides a bridge between the core and periphery. Specifically -
scientist C, the second-hand broker, who is connected to returnee A can borrow access to knowledge, connections and resources in the core.

Figure 2: Core/Periphery Bridge Formation

The return home of a scientist from a more developed country implies a shift in opportunity to periphery non-migrant scientists in their institution to borrow their connections, knowledge and resources. As such, it provides a lens to examine how performance is affected by second-hand brokerage. Specifically, the presence of a core/periphery bridge in the form of a returning scientist may result in improved access to knowledge, collaborators and resources from the core, which in turn may result in improvements in publication outcomes, particularly in the research area of the returning scientist.

Whilst reflecting on how the return home of American-trained scientists to developing countries have impacted their home countries, one American-based mentor I interviewed stated:

‘...there is an impact of the returning trainee. One is the contribution of that individual is from teaching and publishing, and one is the bridging....this bridging is particularly important when you have minimal resources for research.’

Another American-based mentor I interviewed described how returnees bridge his own institution with their network back at home:

‘Just last year we were give some money by a private corporation to do some capacity building in low income countries – I was looking at who could we recruit to come here for training, and I relied on my network who had been here [in the United States] who are back in country...’
As described earlier, there are two main ways that the return home of a scientist from developed to developing countries can impact non-migrant scientists in developing countries. The first possibility is that the non-migrant can borrow access to the knowledge that the return migrant has access to, subsequently affecting the performance of the non-migrant (Hoenen and Kolympiris 2018; Ganguli 2015; Kahn and MacGarvie 2016b; Singh 2005; Agrawal et al. 2006). The second is that in addition to access to their knowledge, the non-migrant can borrow the network position of the returning scientist, which can affect subsequent performance due to new collaborative relationships (Azoulay et al 2010; Wuchty et al. 2007) and/or benefits from elevated status (Azoulay et al 2013).

Although these mechanisms are very hard to separate, because by definition the existence of the second mechanism obscures the existence of the first within an individual, direct measurement of knowledge flows and collaborations (which imply the existence of each mechanism, respectively) and contextual factors may help to distinguish them. There is reason to think that adoption of a network position, which is based on reputation inferred from association with the broker, is particularly important for those with which there is the greatest uncertainty (Stuart et al 1999). Furthermore, ceiling effects to reputation have been found in sciences (Azoulay et al 2013). Both of these imply that if the non-migrants borrow the network position of the return migrant, non-migrant scientists with previous connections to central actors - who already provide a signal of quality to the community - would not benefit as much from the arrival home of a return migrant from more developed countries. As such, it will be important to examine which non-migrant scientists are most impacted by the return migrant.

The remainder of the paper tests these propositions through examining a program that systematically supports low income country scientists to study in the United States and return home following their studies. The next section provides details of the program.

3 Empirical Setting

The empirical setting for this paper is that of life sciences research in African institutions, and the return home of African scientists who took part in long-term training in the United States supported by the National Institute of Health (NIH) Fogarty International Center AIDS International Training and Research Program (FIC AITRP).

AIDS International Training and Research Program

Established in 1968, the NIH Fogarty International Center (FIC) funds around 500 research and training projects across 100 American universities, and 120 countries. With a budget of just over USD
$75 million in 2018, they boast having contributed towards major advances in global health and Low and Medium Income Country (LMIC) scientific capacity development.

The flagship program when it comes to human scientific capacity building LMICs is the AIDS International Training and Research Program (AITRP) (now known as the HIV Research Training Program). Started in 1988, this program was developed in response to the HIV epidemic and the perceived need for strengthened scientific capacity in AIDS endemic countries around the world.

‘But it really changed with the AIDS epidemic, and the realization that to address this particular epidemic we had to change our style of conducting research internationally. We had to overtly move away from the ‘colonial’ research, or the ‘parachute’ approach, and really get into collaborative research and capacity building on site.’ Gerald Keusch, MD, Director of FIC 1998-2003

AITRP provides grants to principal investigators (PIs) in American universities to work with LMIC sites (universities, hospitals and research centers) in strengthening their human capacity through training of scientists, clinicians and allied health workers in research. This training is offered as short or long-term (graduate and non-degree studies, generally over 6 months in duration) programs, usually with a combination of American and field site location. The American-based PIs apply for, and receive, grants in five year cycles, of around USD $500,000 a year, renewable upon re-competition. While the first cycle in 1988 involved eight American institutions, this has now expanded to include around thirty American institutions offering a variety of HIV related research (with TB added in later on) training programs across the world. The American universities involved in the program are some of the leading educational and research institutes: including Harvard University, Johns Hopkins University, Brown University and many more. Between 1988 and 2010 FIC claims to have contributed towards training 1,559 LMIC researchers in long-term AITRP programs with a cost of just over USD $200 million under the AITRP.

The United States based PIs are at liberty to design the training program, across any HIV/AIDS related fields. Most run a variety of short-term programs (workshops, summer courses usually at the LMIC site), although the long-term, degree level (Ph.D, masters) as well as non-degree (including post-doctoral), training for individuals from the LMIC site is generally the focus of the program. In the earlier days of the program, many of the longer term trainees came from institutions in LMICs other than the main site described in the grant.

FIC specifies in the grant announcements that the long-term trainees should be given incentives to go back to their country of origin. A survey carried out by FIC in 2002 found that a return rate of over
80% at that time.\footnote{https://www.fic.nih.gov/News/GlobalHealthMatters/july-august-2012/Pages/hiv-aids-aitrp-program-anniversary.aspx last accessed 10-8-19} This return home is not necessarily to take up a research position, or to the LMIC site involved in the program, moreover FIC prides itself on graduating trainees assuming high level positions in government and multilateral organizations. Incentives to return home include ‘sandwich training’, strategic selection of trainees, re-entry funding, visa restrictions and formal return agreements and contracts with their training institution.

The returning trainees studied in this paper are those African scientists who participated in long-term FIC AITRP supported training at American institutions between 1988 and 2014 inclusive. The FIC AITRP program was one of the first programs around the world to engage researchers from Africa in systematic training in frontier research, and has boasted as contributing towards enormous achievements in terms of research and public health outcomes, HIV and otherwise, in African countries. As just one example, FIC AITRP trainees and collaborators were responsible for the landmark 2011 study HPTN 052\footnote{https://www.annalsofglobalhealth.org/articles/10.5334/aogh.2432/} which revealed the personal and public health benefits of early treatment and led to treatment guidelines on treatment as prevention. At the time, Executive Director of UNAIDS, Michel Sidibé, described the results of HPTN 052 as a ‘breakthrough’ and ‘a serious game changer [that]... will drive the prevention revolution forward.’\footnote{https://www.unaids.org/en/resources/presscentre/pressreleaseandstatementarchive/2011/may/20110512pstrialresults}

## 4 Data and Sample Characteristics

This section provides a detailed description of the process through which the data used in the econometric analysis are assembled. I describe (1) the sample of returning African scientists trained in the United States; (2) the sample of non-migrant African scientists affected by these returns, and the set corresponding control scientists to which they will be compared; (3) outcome variables used in the study. I also present relevant descriptive statistics.

### Sample of Returning African Scientists Trained in the United States

Names of African trainees participating in the FIC AITRP program are gathered directly from the records of American institutions involved. The 20 (out of 29 total universities and 2 training hospitals identified receiving AITRP grants) American universities involved in the AITRP with African trainees are contacted, and 14 of these universities provide data on their long-term African trainees. The universities provide the names, home country, degree (if any) and year(s) of training for their long-term
African trainees between 1988 and 2014. The names of the trainees are then matched with publication data, if any, using the Elsevier Scopus database, and information on institution of return for each trainee is gathered based on publication affiliations post graduation. Resumes of trainees are gathered using a combination of internet searches and direct email correspondence to ascertain the returnee’s role in the institution of return, as well as institutions of non-publishing returnees.

The sample of trainees contains 421 African researchers who took part in long-term training in the United States over the full time period (1988-2014). Trainees are from a range of African countries, with each American university hosting trainees from a variety of countries. I identify those trainees who returned home to Africa following graduation using affiliation information from publication data and information on graduation year. The trainee is considered to have returned home if they published following the year of graduation from the training program and their affiliation was in their home country. Out of 284 returnees with a publication record with affiliation details following graduation, 242 trainees (85% of all those continuing to publish, or 57% of all trainees) in the sample returned home, while 34 of them remained in a developed country (mostly the United States), and 8 moved to African countries other than their home country. Because some returnees are affiliated with more than one institution in their home country after training, I identify 316 unique return events across institutions in 15 African countries between 1988 and 2014 (Figure 3; Table 1). As the timing of the return home cannot be ascertained precisely from publication records I use the year of graduation from the American program as a proxy year for return events.

Sample of Treated, Non-Migrant African Scientists

I measure the impact of exposure to American-trained scientists returning to African institutions on local scientists. Therefore I focus on scientists working in the institutions in Africa at the time of return of the FIC AITRP trainee. Those scientists affiliated with the institution that the FIC AITRP trainee moves (back) to (publishing within the 3 years before the FIC AITRP trainee graduates, with over 75% of publications in those 3 years affiliated with that African institution), and working in HIV related research (i.e. published at least one HIV related publication in the 3 years prior) are considered treated by the return event of the American-trained scientist.

In order to identify those scientists treated by the return event of American-trained scientists I use publication data in the Elsevier Scopus database to generate a sample of scientists and associated publication history affiliated with each African institution between 1988 and 2014.

The use of publication data in studies of this type (namely, in generating a plausible set of scientists

---

6Some of the universities state that their records of trainees in the 1990s is poor - and so it is possible that the sample is biased towards more recent time periods.
in a particular location associated with their full publication record) comes with two major challenges. First, generating a full scientist level publication record is complicated by the fact that scientists may have common names (for example, Smith J), therefore it is difficult to determine which Smith J published which paper, or a single scientist may go by more than one version of a name. Second, understanding where scientists are located given that an affiliation in a publication may not accurately represent the full-time location of a scientist, and that the scientist needs to publish in order for the researcher to see their affiliation — which for the African scientists is not always the case in each year. Fortunately, the first issue is resolved using the Elsevier Scopus publication database which provides an author identifier for each author in every publication contained in the database. The author identifier is developed using an algorithm that incorporates scientist name, coauthors and topic type and allows for scientists to change affiliations across publications. The second issue is resolved using a rule of thumb — if a scientist classifies her affiliation as being in a certain country in over 75% of her publications over a defined, multi year period, she is considered as being based in that institution in that time period. The first year of treatment is considered as the treatment in the main analysis. A single scientist can be treated by multiple returnees coming from multiple American institutions in the same year.

Carrying out this procedure gives 1,740 scientists treated at some point during their career by a return event of a FIC AITRP trainee. As the first year of treatment is the main event considered, several of the returnees drop out of the sample (those who were the second or third trainee to return to the same institution within the career of a treated non-migrant), leaving 112 returning FIC AITRP trainees considered in the main analysis. This gives a median of 11 scientists (mean 19) treated by the return of a single FIC AITRP trainee (Figure 4). I match the list of treated non-migrant African scientists with their full publication record and generate variables of publication rate, collaboration outcomes and content of the research. The complete list of references used in each scientist’s publications are also gathered from Elsevier Scopus database.\(^7\)

**Sample of Control Scientists**

In order to identify the effect of return of an American-trained scientist I could examine changes in non-migrant African scientist’s output after the return event, relative to before the return. However, because the return event is correlated mechanically with career age, and calendar year, the specifications must include life cycle and period effects (Levin and Stephan 1991). The control group that pins down the counterfactual age and calendar time effects are those scientists who never experience the return of a sample FIC AITRP trainee in their institution.

\(^7\)Publication data in the Elsevier Scopus database is again use to generate a sample of scientists

using code developed by Rose and Kitchin (2019)
affiliated with institutions in countries that were at some point involved in the FIC AITRP (to ensure that scientists included are from countries that are similarly connected with the United States and equally economically and politically stable). The control scientists are culled from this universe of African scientists who are affiliated with institutions that did not receive a returning FIC AITRP trainee in the time period of the scientist’s career. The control scientists are chosen using a coarsened exact matching procedure so that their involvement with American-based scientists (from American FIC AITRP training institutions and any other American institutions), type of research (HIV or otherwise), and publication rate matches that of the treated scientists at the counterfactual date of treatment (Appendix A provides more details on construction of the control group). Combining the treated and control samples allows me to estimate the effect of the return of an American-trained scientist in a difference-in-differences framework.

In addition to quantitative data on publication outcomes, I conducted 16 interviews with NIH FIC staff, American-based scientists and administrative staff involved in the FIC AITRP grants, as well as with African trainees and other scientists in institutions receiving returnees. These interviews were carried out on the telephone, or in person where possible. They ranged from 30 minutes to 2 hours, and the primary purpose of the interviews is to illuminate mechanisms of impact of the FIC AITRP and returning trainees.

4.1 Measurement

I use a variety of different measures all of which are generated using publication records of the non-migrant African scientists. First, I generate a number of publication count measures to identify changes in publication rate of scientists. Second, I measure collaboration patterns. Finally, I measure knowledge flows between returning trainees and their American-based networks using a variety of approaches.

Measurements can be divided into more general measurements (such as publication counts and collaborations with any American-based scientist), and measurements that incorporate the returnee or the returnee’s American training institution (such as rate of collaboration with the returning scientist or returnee’s American training institution). As the treated scientists are matched one to one with control scientists in the CEM matching procedure based on pre-treatment (or counterfactual year of treatment) variables, a counterfactual year of return, returnee and returnee’s American training institution is given to each control scientist. The control scientists are then assigned to their matched treated scientist’s returnee and associated American training institution, and measurements based on the specific (counterfactual) returnee can be generated for both treated and control scientists. For those treated and control scientists who experience more than one (counterfactual) returnee in a given year, outcomes are
measured for each returnee, and associated American training institution, and the maximum value is taken.

**Rate of Publication**

Measures corresponding to the rate of publication include the number of publications each year a scientist is an author on, and an additional measure weighting each publication by its journal impact factor — a measure of the frequency with which the average article in a journal has been cited in a particular year. Key word searches of the title, abstract and keywords in each publication in a given year for a sample scientist gives publication outcomes in a given research area. The probability that at least one of these publications contains keywords associated with HIV related research is measured.

**Collaboration Rates**

Collaboration patterns of non-migrant African scientists is measured using co-authoring patterns in publication data. Using author written parsing script I extract coauthor names and affiliations from the scientist’s publication record to generate collaboration counts (both absolute and dummy indicator) across a variety of groups.

Measures of collaboration rates with any American-based scientist, or any scientist affiliated with an American training institution involved in the FIC AITRP are generated. In addition to collaboration rates with these two groups of scientists, more specific measures based on collaboration rates with the returning trainee and scientists from their American training institution are generated.

**Knowledge Flows**

I measure knowledge flows between the sample scientist and the returning trainee, the returnee’s American training institution and any American-based scientist in two main ways. First, following a long line of research, I use publication to publication citations to measure knowledge flows among scientists (Jaffe et al 1993; Jaffe and Trajtenberg 1999). Taking the non-migrant scientist’s full list of references used in their publications, I measure the extent to which they cite publications authored by the returning trainee, the returning trainee’s American institution, as well as any American-based scientist\(^8\) in each year before and after the return event.

Second, knowledge flows are measured through measures of intellectual proximity of publication output. This exploits key-words used to describe the content of life sciences, the Medical Subject

---

\(^8\)Just publications authored by returnee and American scientists in the 5 years prior to the return event are considered in order to account for the possibility that the publication themselves may be influenced by the return (as per Azoulay et al 2012).
Heading (MeSH) terms. This controlled vocabulary indexes articles for PubMed and the terms are assigned by profession librarians, ensuring consistency across life sciences. Because of the large number of scientists who do not publish in a given year, the MeSH based measures are aggregated at the level of the sample scientist before and after the return event.

I measure intellectual proximity between the publication output of sample scientists and that of the returning trainee, as well as the returning trainee’s American training institution FIC grant principal investigator (PI) through the construction of a continuous measure of intellectual distance on the basis of MeSH terms. Following the method in Boudreau et al (2016), for an individual scientist in the sample, the total MeSH terms across their publications in the five years before and after the return event are used to generate two vectors (pre and post return) of MeSH terms with a corresponding probability of use per year. A vector of MeSH terms with associated probability is also generated for the publications of the returning trainee publications during their fellowship, as well as the publications of the returning trainee’s American training institution FIC AITRP PI publications in the five years prior to the return year. This latter vector is matched with the pre and post vectors of the sample scientists as per the treatment year and returning trainee or returnee’s American training institution FIC AITRP PI. A measurement of intellectual distance between the sample scientists and the associated returning trainee or returnee’s American training institution FIC AITRP PI is calculated as the cosine between the vectors before and after the return event and expressed as a percentile across all sample values. The value of 1% reflects the least (or closest) and 100% the greatest distance. Therefore, a negative coefficient reflects a movement towards the returning trainee or the returnee’s American training institution FIC AITRP PI.

Measurements for control scientists are made in a symmetric way, using the counterfactual return event date, returning trainee and American training institution.

4.2 Descriptive Statistics

I identify 1,740 distinct scientists who are affiliated with an African institution at the time of the return of a FIC AITRP trainee. The matching procedure identifies a control scientist for 1,657 (95%) of the treated scientists, treated by 112 unique FIC AITRP trainees.

FIC AITRP Trainees

Figure 5 illustrates the publication trends of the 288 FIC AITRP trainees who have a publication record following graduation. The differential trends of trainees known to return home versus those remaining in developed countries shown in the figure is most likely due to a selection effect resulting
in different types of trainees not returning home. One important point to note from the figure (panel d) is that the returning trainees continue their collaborative relationship with their American training institution following graduation.

Descriptive statistics for the sample of 112 FIC AITRP trainees used in the main analysis is provided in Table 2. The average returnee graduated from their fellowship in 2004. 65% of returnees returned to institutions with broader institutional programs, and almost 80% returned to institutions in which there had been previous returnees from the FIC AITRP program. In the five years following return home, most of the returning scientists publish some research (with an average of 5.5 publications in the five years after return), particularly in HIV. 63% of returnees continue to co-author with scientists from their American training institution, and 75% co-author publications with scientists from the institution they return to.

Non-Migrant African Scientists

The descriptive statistics in Table 3 pertain to the set of $2 \times 1,657 = 3,314$ matched treated and control scientists. The covariates of interest are measured prior to the return of the trainee (or counterfactual). A few of the covariates are balanced between treated and control scientists by virtue of the CEM procedure — for instance, the career age at the time of return. However, the observed balance in other statistics, such as the five year stock of number of American coauthored publications, number of any FIC AITRP American institution co-authored publications, and number of HIV publications at time of (counterfactual) return is not guaranteed.

While publication outcomes are well matched at baseline, there are differences between the mean likelihood of treated and control scientists to co-author with the returnee, and the returnee’s training institution. This is consistent with the view that moves are not random, and that there are some people in destination locations who a mover is already connected to. Because these baseline differences make it difficult to ascertain co-authoring behavior in the absence of the return event, those non-migrant scientists who have co-authored with the trainee or the trainee’s American training institution are removed from the analysis in robustness checks.

The estimation sample includes observations 5 years before and after the return event (or counterfactual). The result is a balanced panel dataset with 36,454 scientist-year observations.

---

9FIC AITRP grants pre-specify at least one institution in Africa in which the United States based grant holder carries out a variety of shorter term programs in.
5 Results

5.1 Econometric Framework

In order to identify the effect of the return of an American-trained scientist, I compare a non-migrant African scientist’s outcomes after the return of the American-trained scientist to their institution relative to before, using a scientist fixed effect specification. The estimating equation (equation 1) relates non-migrant African scientist $i$’s outcomes in year $t$ to the return of a FIC AITRP trainee in their affiliated institution.

$$E \left[ y_{it} | X_{it} \right] = \exp \left[ \beta_0 + \beta_1 \text{AFTER\_RETURN}_{it} + \beta_2 \text{AFTER\_RETURN}_{it} \times \text{RETURN\_INSTITUTION}_i + f(\text{AGE}_{it}) + \delta_t + \gamma_i \right]$$

(1)

Where $y$ is the outcome measure, AFTER\_RETURN denotes an indicator variable that switches to one the year after the FIC AITRP trainee returns to the home country, RETURN\_INSTITUTION is an indicator variable that switches to one if the sample scientist is treated. $f(\text{AGE})$ corresponds to a flexible function of the non-migrant African scientist’s career age, with calendar year fixed effects and non-migrant African scientist fixed effects. Standard errors are clustered at the level of the institution.

The majority of the dependent variables of interest are skewed and non-negative (Figure 6 illustrates the distribution of publications the year of the return (or counterfactual)). Due to the large number of zero’s in the dataset, most of the specifications are estimated in two ways; first using an inverse hyperbolic sine ordinary least square estimate. And second, outcomes are converted into dummy outcomes (given a 1 if they happen at all in a given year), and a second set of estimates use a linear probability model.

5.2 The Impact of a Returnee on Non-Migrant African Scientist Performance

Table 4 presents the core results estimating the specification presented in equation 1. These provide strong support for the expectations of the paper, that return home of an American-trained scientist results in an increase in performance of non-migrant scientists as well as a relative increase in performance in the field of the returning scientist. Overall, I find that non-migrant scientists increase their rate of publications following the return home of a FIC AITRP trainee, as indicated by the estimates for
AFTER RETURN × RETURN INSTITUTION being positive and statistically significant (column 1, 2, 3, 7). I find a sizeable and significant 6% increase in the annual number of publications of a non-migrant scientist following the return of an American-trained scientist, as compared with a scientist not subject to the return of an American-trained scientist (column 1). To verify that this isn’t driven by an increase of publications in low quality journals, column 3 measures the change in publications weighted by their journal impact factor. The significant increase for the treated group as compared to the control group suggests that scientists increase both quantity and quality of publications following the return event. The increases in rate of publication are solely due to increases in HIV related research (column 5, 8). Column 8 shows that treated scientists are 3.5 percentage points more likely to publish in HIV related research following the return event. With the average probability of publishing in HIV related research of 0.32, this post return increase is around a 10% increase on the mean. However, the non-migrants do not experience a significant increase in the rate of first authored publications (column 4), which raises questions on the role of the African scientists on projects, and the possibility that they are a ‘long-arm’ of the American labs, a concern of which I return to in the discussion.

I explore the dynamics of these effects in Figure 7. I estimate a specification in which the treatment effect is interacted with a set of indicator variables corresponding to a particular year relative to the trainee return. One point worth noting from these figures is that effects do not appear to be transitory, and although the results are noisy due to the small sample, there does not appear to be a pre-trend. The rest of the paper explores the mechanisms by which a return migrant impacts non-migrant performance.

5.3 Return Migrant as a Core/Periphery Bridge

Figure 8 illustrates initial support that the return migrant provides a core/periphery bridge between institutions in Africa and the United States. Depicting the institutional collaborative network both before the FIC AITRP program begins (panel a), and after (panel b), the figure shows that following the program, treated institutions (lighter gray circles) become more central to the full network of African and American institutions, as compared to the control institutions (white circles). The figure also illustrates that all of the African institutions in the sample become more connected over the full time period. This fact further necessitates the use of control scientists in the sample to account for this trend. Subsequent evidence that the return migrant is providing a core/periphery bridge is explored through unpacking each of the proposed mechanisms by which second-hand brokerage can operate.

Borrowing Knowledge

I first test whether knowledge flows from the returnee’s network in the United States increase after
the returnee’s arrival. This would be suggestive of a bridge being formed, and the actor associated with
the broker borrowing access to the knowledge that the broker has access to.

I find that treated scientists tend to cite publications authored by scientists from the returnee’s
training institution (or counterfactual) more following the return event (Table 5 column 2, 3, 4). With
an increased probability of citing the American training institution of the returning scientist of 2.2
percentage points, this gives an economically significant 29% increase over the mean. I also measure
any change in overlap of the research agenda of treated scientists with that of the returnee’s training
institution PI (or counterfactual). As illustrated by the negative coefficient in Table 5 column 5 which
represents a lower percentile in key word overlap, treated scientists are much more likely to move their
research agenda closer to that of the returnee’s American training institution PI than control scientists.
Having defined the outcome variable in terms of percentiles, we can interpret the coefficient as indicating
that treated non-migrants move around two percentiles towards the American PI following the return
event, compared to control scientists. A similar magnitude is found for those publications not co-
authored with scientists from the American training institution (Table 5 column 6), suggesting that the
non-migrants are learning about the research in the United States, independent of their collaborations.

Borrowing Network Position

Second, I assess whether association with a broker allows actors to borrow the network position of
the broker through measuring non-migrant African scientist’s collaboration rates with American-based
scientists that the returnee is connected to.

Figure 9 depicts the collaboration rates of treated scientists with various different groups of American
scientists. Panel B illustrates that the treated scientists are more likely to coauthor a publication with an
American-based scientist affiliated with a FIC AITRP training institution after the return event. Table
6 provides the regression counterpart to Figure 9, and columns 1, 2 and 3 illustrate that non-migrant
scientists are more likely to collaborate with scientists from the United States, in particular those from
the training institution of the returnee (or counterfactual). Treated scientists are 33% more likely to
publish with the American training institution following the return event (column 4). These are mostly
new collaborators for these non-migrants (Table 6 column 5, 6). This provides supporting evidence that
under certain conditions, actors associated with a broker assume the network position of the broker.

As discussed in Section 2, one would expect non-migrants with fewer connections to the core prior to
the return event to experience greater improvements to their performance if the return migrant allows
the non-migrant to borrow their network position through some kind of sponsorship. Tables 7 explores
heterogeneity in the effect of the returnee through separating the treated and control scientists into three
groups based on their network prior to the return: 1. those who publish with the returnee’s American
training institution prior to the return, 2. those who publish with OECD based scientists in over 75% of their publications prior to the return, and 3. the remainder. The same difference-in-difference regression is run on the three groups separately. As seen in columns (3) and (6), the greatest impact of the returnee is felt by those scientists less well connected with OECD scientists prior to the return. This also serves as a robustness check as the movement of the returning scientist to an institute is likely to be endogenous to the scientists in the institute who were already within the same close network. The fact that the greatest effect is not experienced by those non-migrants with a collaborative history with the returnee’s American training institution (columns 1 and 4) is comforting.

Another piece of supporting evidence that the impact of a return event is at least in part due to borrowing the returnee’s network position is the differential impact of a returnee according to their role back home. One the one hand, if the non-migrants benefit predominantly from borrowing the knowledge of the returnee, it might be expected that the benefits are greater when the returnee is an active scientist. On the other hand, if the non-migrants benefit predominantly from borrowing the network position of the returnee, it might be expected that the benefits are greater when the returnee has more of an administrative, or outward facing role in their institution. Interviews suggest that the latter is true, and the following quotation from an interview with an American-based PI involved in the FIC AITRP confirms how the role of a returnee can influence their impact:

“One of my trainees was chairman of the School of Medicine…. He had a credible skill set, but he was not able to put his skill set to use because he was more administrative. He called on us. The next thing we knew we were doing in-country training at his behest….. He was able to nurture a mini [American institution] back home. He was able to do that because of his position.”

Results in Table 8 are consistent with the qualitative evidence. For a sub-sample of returnees for whom full career information was obtained, cross-tabs of the post-pre difference in treated non-migrant outcomes are calculated according to the role that the returnee assumes on returning home. Table 8 shows that there is a larger positive change for non-migrants who have returnees who are occupying administrative positions on their return home. Although the sample of returnees is extremely small, and thus any findings must be taken with a grain of salt, this suggests that those in an administrative position are able to exert greater positive spillovers onto the non-migrants in their institution.
5.4 Alternate Explanations

I consider two alternative explanations that could be driving the observed effect. Namely, team work with the returnee, and knowledge flows from the returnee.

Team Work Benefits

Science is increasingly carried out in teams (Wuchty et al 2007). Prior work finds that the co-location of scientists results in increased collaboration rates, and more correlated research trajectories (Catalini 2018). On this basis, and the significant frictions associated with collaborating with scientists in developing countries, the return home of a scientist should result in increasing rates of collaboration between the returnee and the non-migrants. This increased collaboration could increase the rate of publication outcomes, particularly of those in the field of the trainee, due to improvement in the skills within the team. However, theory relating to complementary skills suggests that the formation of teams between a specialist trained abroad, and a generalist trained in the home country may be challenging (Jones 2008). Furthermore - assortative matching theory suggests that the combined output is that of the least productive member of the team (Jones et al 2008; Ahmadpoor and Jones 2018), dis-incentivising the returnee to collaborate with home country scientists. Consistent with Catalini 2018, Figure 10 illustrates that treated scientists are much more likely to co-author with the returnee following their return. However, very few people actually collaborate with the returning trainee, just 66 treated scientists (less than 4%). And this is mostly people who collaborated with FIC AITRP trainees before the event. I consider if it is these people driving the result in Table 9. Splitting the sample into two groups: those who have pre-return characteristics that are correlated with collaborating with the returnee (columns 1 and 3) and those who don’t (columns 2 and 4). If benefits from team work with the returnee are driving the main result, I would expect that those with the characteristics correlated with collaborating with the returnee are also the ones who benefit the most from the arrival of the returnee in their institution. However, table 9 shows a different story. Those scientists who are less likely to co-author with the returnee (or counterfactual) experience the greatest positive impact from the return event. Furthermore, the results are robust to removing publications coauthored with the returnee (Table 4, column 2). I therefore do not think that team work benefits are driving the observed results.

Knowledge Flows from the Returnee

Economic geography has long documented a relationship between physical proximity and knowledge transfer (Jaffe et al 1993). Mobile scientists carry knowledge with them, and knowledge flows in the form of citations to a mover’s pre-move publications are found to increase in the destination following a move (Azoulay et al 2012; Ganguli 2015). If this is new knowledge, this could improve publication outcomes
of non-migrants. I measure changes in citation rates to the returning trainee’s (or counterfactual) pre-graduation publications as well as any movement towards the research of the returning trainee. If benefits from knowing about the returnee’s research are driving the main result, this may cast doubt on the hypothesis that non-migrant benefit from the formation of a core/periphery bridge, as it could just be due to new knowledge coming into the institution, irrespective of the involvement of the core. However, table 10 illustrates the citation rates to the returnee’s publications do not increase for treated scientists following the return (column 1), and that research agenda of the treated scientists does not move towards that of the trainee’s following their return (column 2).

5.5 Robustness and Sensitivity Checks

The main threat to identification in this study is the possibility that the treated institutions are getting better, and particularly that they are becoming more internationally connected, just prior or at the same time as the return home of the Fogarty trainees. A few tests help to understand if this is driving the observed results. First, I re-match the 1,740 treated scientists with the same individual level pre-return covariates, and this time overlay pre-return institution level covariates of institution size, productivity and collaboration rates with OECD country institutions. This results in a smaller sample of treated and control scientists (due to the difficulty in finding a similar scientist in a similar institution at the same time) of 2,780 scientists (matches are found for 80% of the treated scientists). I run the main specification regressions on this sample of individual and institution-level matched scientists in Table 11. The results are very similar on this matched sample, providing support that the effect is not driven by selection of the returnees to better performing, or better connected, institutions.

Table 12 provides additional evidence to ascertain the robustness of the results. First I verify that the effect is not driven by a few returnees affecting a large number of non-migrants. Column (2) provides the estimate without returnees who affect large numbers of non-migrants. The finding is robust and actually greater without these returnees. To provide further evidence that the effect is driven by the returnee and not other institutional factors, I verify that the effect is sensitive to the qualification of the returnee, by removing those returnees who received a Ph.D during their studies in the United States in column (3). As expected, the coefficient is smaller than the baseline which is in line with expectations that those returnees with more experience in the United States, and thus more embedded in the network, exert a greater spillover. Column (4) includes country time trends, and column (5) institution time trends, to remove concerns that the effect is driven by improvements in country level, or institution level capacity that coincides with the timing of the return. The inclusion of the time trends doesn’t change the coefficient of interest by much, which is reassuring, but it does increase the standard errors (which is not surprising as it is a demanding specification) which leaves an insignificant
finding. To verify that the control sample is not contaminated by the treatment as well — biasing the result, column (6) removes from the sample those scientists that are ever collaborators with the treated scientists. The large increase in the coefficient following the removal of this group of contaminated control scientists suggests that their inclusion biases the result downwards, and so the main result is a conservative estimate. Finally, in columns (7) and (8), I conduct simulation studies to validate the quasi-experiment exploited in the paper. In column (7) I keep just the pre-event data, and generate a placebo return year two years prior to the actual return year. I run the baseline specification with this placebo return year and find a precisely estimated zero effect. This is reassuring that the returnees are not returning to institutions that are improving in the years before the return. In column (8), I keep the control sample only and generate placebo return years for control scientists, where dates are drawn at random from the empirical distribution of return events for the actual returnees. I replicate the main specification but limit the sample to the set of 1,657 control scientists. The effect of return is again reassuringly precisely estimated at zero.

Attrition of trainees arising from the use of just trainees with a publication record post-graduation creates two potential concerns. First, the results could suffer from selection bias. Trainees without publication records following program participation, or those not returning home, differ systematically from those that return home and continue to publish. Although the data on pre-return characteristics are limited, trainees who return and publish are more likely to have studied for a PhD in the United States and to have published prior to their graduation. They are also more likely to be from a country with a greater level of scientific capacity. There are no significant differences in the period of the fellowship or the US institution that they attended. If these differences are indicative of differences in a trainee’s potential impact on non-migrants, the results would be biased and should be interpreted as conditional on trainees returning home and continuing to actively publish. I run regressions with interactions of pre-return trainee covariates to assess the hypothesis that pre-return characteristics of trainees affects the magnitude of the impact on non-migrants (Appendix B: Table B1). There is no discernible difference in impact according to the returnee pre-return characteristics. One point to note, however, is that although the main result is relatively stable to inclusion of covariates of the returnee publication record during their fellowship and PhD degree status, the inclusion of a dummy variable reflecting scientific advancement of the home country reduces the main coefficient. Although inconclusive due to the noisy nature of the estimates, this does suggest that the findings in this paper are more relevant for low income countries with relatively more advanced scientific capacity, which is an interesting avenue for future research. Second, those without a publication record may move to institutions in which non-migrants in the control group are working. This is a threat to identification because the control group may be affected by the treatment, although the implications depend on how they impact non-migrants. Unfortunately, this is not testable, but a lack of research productivity, and alumni surveys finding that many take up senior positions in government or non-profits (which are not the same institutions as the
control group) suggests that their impact may be minimal.

6 Discussion

This paper offers a new perspective on the consequences of return migration of high-skilled workers from developed to developing countries, exploring how returnees can provide a bridge in the network affecting the performance of developing country non-migrants. Through an examination of the impact of the return home of African scientists after training in the United States on non-migrant scientists working in the institutions they return to, the results show that the publication rates of African non-migrant scientists increase following the return event. Furthermore, this increase in publications is mostly in the field of study of the returning scientist. The relationship is contingent on a lack of prior connectivity of the non-migrant.

These findings shed light on the phenomenon of association with a broker, a potentially critical but under-recognized mechanism that shapes the performance of outsiders. Although extant research has long explored the impact of networks on performance, particularly in knowledge production, it has generally considered actors within a network as a function of their ties. Failure to account for the notion that broader features of a network — including the structure (in particular where there are insiders and outsiders - or core/periphery), as well as indirect ties — can also affect performance would lead to an incomplete understanding of how individuals affect the performance of others.

To examine the effects of association with a broker within a core/periphery network, I introduce the concept of a core/periphery bridge. With many network exhibiting a core/periphery structure, the impact of an individual bridging the core and periphery is less well understood. Beyond documenting the performance implications of association with a core/periphery bridge for periphery actors, I also provide supporting evidence of the existence of two concurrent mechanisms driving the impact: (1) borrowing the knowledge that the broker has access to, and (2) borrowing the network position of the broker.

The findings inform the topical debate on immigration and mobility, in particular the impact of migration of high-skilled individuals in and out of developing countries. To date, much research on this topic has focused on the mobile individual themselves. The clarification of conditions under which non-migrants benefit from return migration of others, and when they can borrow the knowledge and network position of a return migrant — or broker — provides an opportunity to analyse spillovers from migration. The extent to which second-hand brokers can adopt their own brokerage position in the absence of the return migrant, the codified nature of knowledge being transferred, as well as the motives and incentives
of the return migrant to both share knowledge and connections should be important considerations of future research and program design aimed to promote sharing or borrowing of a returnee’s access.

The magnitude of the results, as well as the finding that the impact is greatest for more peripheral scientists, are consistent with research on the removal of frictions to accessing inputs for scientific production. Ding et al (2010), for example, studied the impact of the arrival of information technology, in the form of BITNET and Domain Name System, in an American institution on scientists’ productivity and collaboration patterns. They found that the arrival of the internet had a positive impact on publication and collaboration rates, particularly for female or scientists in lower tier institutions. That improvements in scientists’ access to inputs particularly affects those traditionally more marginalized suggests a need for policy interventions targeting access to collaborations, resources and knowledge for ‘outsiders’ in an innovation system.

This study has four major limitations. The main challenge, as with many network studies is assessing whether the shock affected the control scientists as well. Given that the control and treated scientists are loosely in the same network (global science) it is plausible to think that there may be ripple effects. If it affects them positively — I provide an underestimate of the main effects. Limitations in the data also prevent me from controlling for the return home of non-FIC AITRP scientists. If other returnees are arriving in the institutions at the same time this could be problematic for the estimate. Second, the FIC AITRP studied is specifically designed to engage developing country sites. Thus the question of whether these results are generalizable to other forms of training programs is unclear. Third, the study is limited to returning HIV researchers in Africa in a time when HIV research, particularly that done in or on African populations, was very topical. Whether these findings are generalizable to other fields and countries is also unclear. Fourth, I use publication records as a proxy for performance. Whether this is a true reflection of performance is unclear. Particularly concerning is the fact that first (and last) authored publications do not significantly increase following return events. This raises questions on whether capacity is actually improving — or whether the relationship with American-based scientists is one in which African scientists carry out low-skilled field work tasks to deliver samples to the United States for analysis. Future work using more reflective measures of actual scientific capacity is necessary.

Despite the above mentioned limitations, the findings have important implications for the valuation of international training programs and policies on bringing people home, decisions on engagement with international scientists and for developing country science more generally. The majority of impact evaluations and cost benefit calculations from international training programs of developing country high-skilled workers just pertain to the individual trained. The benefits to non-migrants estimated in this study suggest that re-conceptualizing the unit of analysis of such estimates is important to understand ‘bang for the buck’ of programs sending individuals abroad, and the relative merits of incentivizing the
return home of high-skilled nationals.

A back of the envelope calculation finds that one returnee contributes around 4 additional publications of non-migrants in the five years after returning. With the returnee publishing an average of 5 publications in the five years after return, this spillover effect is 80% of the effect of the returnee themselves on the institution’s innovation output. Given that the average cost of one within sample Fogarty trainee is around USD $144,000 (in today’s USD), this suggests that the cost of one African publication, including the returnee’s publications as well as the spillover, is approximately USD $16,000. While this estimate does not incorporate re-direction of research funding dollars to the treated institutions, and so is not an absolute measure of cost per African publication, it is interesting to note that this is a fraction of estimates of the cost of other programs designed to encourage publication output. As two examples, Myers (2018) estimates that the cost to the NIH of a single publication coming out of their R01 grant program is between USD $344,000 and $665,000, and Jacob and Lefgren (2011) estimate the cost to the NIH of one publication is $1.7 million. It is important to note as well that my estimate is likely to be an under-estimate of the spillover effect. The sample of scientists treated by the returnee in this study is very narrowly defined. I only consider those already publishing at the time of the returnee, thus this estimate doesn’t include future students of the returning trainee.

Not only is there a plethora of programs, initiatives and strategies around the world targeting global training or experience for individuals, employees and firms, but the findings of this study are general and applicable to a range of settings. First, I anticipate that alternative programs and policies creating core/periphery bridges under similar conditions would result in similar outcomes. Second, the findings are relevant across a variety of settings that exhibit core/periphery structures. As just one example, hiring strategies of firms, particularly of entrepreneurial firms, should consider how to best leverage bridges with the central network.

This paper raises more questions than it answers. Is the same phenomenon observed if the foreign trained remains abroad? A body of research on the benefits of a skilled diaspora to developing countries has documented the role these individuals play in knowledge flows, remittances and trade. But an understanding of the dynamics in terms of the network is less well understood. Are bridges to the center more important for periphery actors operating in industries or settings that have a dense center? And what happens in settings with more tacit knowledge, or uncertainty? Future work should seek to explore these questions and further our understanding through leveraging the concept of a core/periphery bridge.
References


Borgatti, S.P., Everett, M.G. 2000 Models of Core/Periphery Structures Social Networks 21(4): 375-395


Burt, R.S. 2002 Bridge Decay Social Networks 24(4): 333-363


Burt, R.S., 2010 Neighbor Networks: Competitive Advantage Local and Personal Oxford University Press

Catalini, C., 2018 Microgeography and the Direction of Inventive Activity Management Science 64(9): 4348-4364

Choudhury, P., 2015 Return Migration and Geography of Innovation in MNEs: A Natural Experiment of Knowledge Production by Local Workers Reporting to Return Migrants Journal of Economic Geography 16: 585-610


Ding et al., 2010 The Impact of Information Technology on Academic Scientists’ Productivity and Collaboration Patterns Management Science 56(9): 1439-1461


Fafchamps, M., Leij, M.J., Goyal, S., 2006 Scientific Networks and Co-authorship Working Paper


Gibson, J., McKenzie, D., 2014 Scientific Mobility and Knowledge Networks in High Emigration Countries: Evidence from the Pacific Research Policy 43(9): 1486-95

Granovetter, M.S., 1973 The Strength of Weak Ties The American Journal of Sociology 78(6): 1360-1380


Hwang, K., 2008 International Collaboration in Multilayered Center-Periphery in the Globalization of Science and Technology Science, Technology and Human Values 33(1): 101-133


Jacob, B.A., Lefgren, L., 2011 The Impact of Research Grant Funding on Scientific Productivity Journal of Public Economics 95(9-10): 1168-1177


Kahn, S., MacGarvie, M., 2016b Do Return Requirements Increase International Knowledge Diffusion? Evidence from the Fulbright Program Research Policy 45(6): 1304-1322


Mohnen, M., 2016 Stars and Brokers: Peer Effects Among Medical Scientists Working Paper

Myers, K., 2018 Managing the Rate and Redirection of Science: Evidence from NIH Data Working Paper


Zelnio, R., 2012 *Identifying the Global Core-Periphery Structure of Science* Scientometrics 91(2): 601-615


Figure 3: Number of FIC AITRP Trainee Returnees to African Countries

Note: Aggregate counts of sample FIC AITRP trainees known to return back to Africa between 1989-2014.
Figure 4: Histogram of Number of Non-Migrants Each Returnee Impacts

Note: I compute the number of treated sample non-migrants affected by each FIC AITRP trainee, and the frequency of returnees impact those number of non-migrants is given on the y-axis.
Figure 5: Publication Trends for FIC AITRP trainees

(a) Mean number of publications

(b) Mean number of HIV publications

(c) Mean number of publications with American-based coauthors

(d) Mean number of publications with American training institution

Note: Publication trends for the 284 FIC AITRP trainees who have a publication record following graduation (242 who return home, 34 who remain in a developed country, 8 who move to another African country) are plotted for the five years before and after graduation.
Figure 6: Histogram of Publication Rate at the Time of Return

Note: I compute the number of publications in the year of return authored by the 3,314 sample treated and control scientist.
Figure 7: Impact of FIC AITRP Trainee Return on Non-Migrant African Scientists’ Publication Outcomes

(a) Number of publications
(b) Any publication
(c) Any HIV publication
(d) Number of HIV publications

Note: The solid blue dots in the above plots correspond to coefficient estimates in panels (a) and (d) stemming from conditional (scientist) fixed effects ordinary least squares specifications in which inverse hyperbolic sine outcomes are regressed onto year effects, article age effects, as well as 10 interaction terms between treatment status and the number of years before/after the return of a trainee (the indicator variable for treatment status interacted with the year of return is omitted). And coefficient estimates stemming from conditional (scientist) fixed effects Linear Probability Model specifications in panel (b) and (c) in which publication dummy variables are regressed onto year effects, article age effects, as well as 10 interaction terms between treatment status and the number of years before/after the return of a trainee (the indicator variable for treatment status interacted with the year of return is omitted). All specifications also include a full set of lead and lag terms common to both the treated and control articles to fully account for transitory trends in citations around the time of the return. The 95% confidence interval robust standard errors clustered around the institution is plotted with light blue bars.
Figure 8: Pre and Post-Program Collaboration Network of Within Sample African Institutions, as well as American Institutions Involved in FIC AITRP

(a) Pre-Program Network (1985-1989)  
(b) Post-Program Network (2015-2019)

Note: The collaboration network of the publications in medical related subjects of 152 African institutions that are included in the sample (and publish in 1985-1989), as well as the 14 sample American institutions involved in the FIC AITRP program, excluding publications authored by FIC AITRP African trainees, is plotted for the pre-program period (1985-1989) in panel (a) and post-program period (2014-2019) in panel (b). American institutions are represented by gray circles, African institutions that are treated between 1988 and 2014 are gray circles, and African institutions never treated between 1988 and 2014 are white circles. The lines between circles represent a collaborative link. The size of the circles in the network are adjusted according to the log of the number of publications they produce in the time frame.
Figure 9: Impact of FIC AITRP Trainee Return on Non-Migrant African Scientists’ Collaborations

(a) Any publication with United States based coauthors
(b) Any publication with American FIC AITRP institution coauthors
(c) Any new United States based coauthor
(d) Any new American FIC AITRP institution coauthor

Note: The solid blue dots in the above plots correspond to coefficient estimates stemming from conditional (scientist) fixed effects Linear Probability Model specifications in which coauthoring dummy variables are regressed onto year effects, article age effects, as well as 10 interaction terms between treatment status and the number of years before/after the return of a trainee (the indicator variable for treatment status interacted with the year of return is omitted). The specifications also include a full set of lead and lag terms common to both the treated and control articles to fully account for transitory trends in citations around the time of the return. The 95% confidence interval robust standard errors, clustered around the institution is plotted with light blue bars.
Note: The average probability of a scientist in the treated sample publishing with the returning FIC AITRP trainee in each year before and after the return (or counterfactual) is plotted.
Table 1: FIC AITRP Returnees by Country and American Training Institution

<table>
<thead>
<tr>
<th>Country</th>
<th>Number of Returnees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kenya</td>
<td>73</td>
</tr>
<tr>
<td>Uganda</td>
<td>60</td>
</tr>
<tr>
<td>Zambia</td>
<td>35</td>
</tr>
<tr>
<td>Tanzania</td>
<td>20</td>
</tr>
<tr>
<td>Botswana</td>
<td>19</td>
</tr>
<tr>
<td>Malawi</td>
<td>9</td>
</tr>
<tr>
<td>Senegal</td>
<td>6</td>
</tr>
<tr>
<td>Nigeria</td>
<td>5</td>
</tr>
<tr>
<td>Mozambique</td>
<td>4</td>
</tr>
<tr>
<td>Rwanda</td>
<td>3</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>3</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>2</td>
</tr>
<tr>
<td>Central African Republic</td>
<td>1</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>1</td>
</tr>
<tr>
<td>Lesotho</td>
<td>1</td>
</tr>
</tbody>
</table>

American Training Institution

<table>
<thead>
<tr>
<th>University</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Washington</td>
<td>84</td>
</tr>
<tr>
<td>Case Western Reserve University</td>
<td>40</td>
</tr>
<tr>
<td>Harvard School of Public Health</td>
<td>26</td>
</tr>
<tr>
<td>Johns Hopkins University</td>
<td>21</td>
</tr>
<tr>
<td>Vanderbilt University</td>
<td>16</td>
</tr>
<tr>
<td>University of Alabama at Birmingham</td>
<td>15</td>
</tr>
<tr>
<td>Baylor College of Medicine</td>
<td>8</td>
</tr>
<tr>
<td>Dartmouth College</td>
<td>7</td>
</tr>
<tr>
<td>Duke University</td>
<td>7</td>
</tr>
<tr>
<td>Brown University</td>
<td>5</td>
</tr>
<tr>
<td>University of Nebraska, Lincoln</td>
<td>4</td>
</tr>
<tr>
<td>Emory University</td>
<td>3</td>
</tr>
<tr>
<td>State University of New York at Buffalo</td>
<td>3</td>
</tr>
<tr>
<td>University of Maryland Baltimore</td>
<td>3</td>
</tr>
</tbody>
</table>

Note: This table provides details on the sample of 242 scientists who are trained in the United States in long-term training programs supported by the FIC AITRP and return home following their graduation (graduating between 1989 and 2014).
Table 2: Summary Statistics for Within Sample FIC AITRP Trainee Returnees (N=112)

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>median</th>
<th>std. dev.</th>
<th>min.</th>
<th>max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year of graduation</td>
<td>2004</td>
<td>2006</td>
<td>7.03</td>
<td>1989</td>
<td>2014</td>
</tr>
<tr>
<td>Ph.D degree</td>
<td>0.098</td>
<td>0</td>
<td>0.30</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Masters degree</td>
<td>0.51</td>
<td>1</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Already published before fellowship</td>
<td>0.39</td>
<td>0</td>
<td>0.49</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Career age at fellowship if already published</td>
<td>3.53</td>
<td>2</td>
<td>4.20</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>Lag between graduation and publication in home country</td>
<td>3.95</td>
<td>3</td>
<td>3.79</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>Return to institution with broad institution program</td>
<td>0.65</td>
<td>1</td>
<td>0.48</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Career number of publications</td>
<td>29</td>
<td>13</td>
<td>40.03</td>
<td>1</td>
<td>194</td>
</tr>
<tr>
<td>Post graduation number of publications</td>
<td>5.56</td>
<td>3</td>
<td>5.84</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>Career number of HIV publications</td>
<td>22.27</td>
<td>7.5</td>
<td>34.13</td>
<td>0</td>
<td>161</td>
</tr>
<tr>
<td>Post graduation number of HIV publications</td>
<td>4.54</td>
<td>2</td>
<td>5.26</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>Publish with U.S. coauthors post graduation</td>
<td>0.76</td>
<td>1</td>
<td>0.43</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Post graduation number of publications with U.S. coauthors</td>
<td>4</td>
<td>2</td>
<td>5.25</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>Publish with any U.S. training institution coauthors post graduation</td>
<td>0.71</td>
<td>1</td>
<td>0.46</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Post graduation number of publications with any U.S. training institution coauthors</td>
<td>3.48</td>
<td>2</td>
<td>4.79</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>Publish with own U.S. training institution coauthors post graduation</td>
<td>0.63</td>
<td>1</td>
<td>0.49</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Post graduation number of publications with own U.S. training institution coauthors</td>
<td>3.02</td>
<td>1</td>
<td>4.67</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>Publish with return institution coauthors post graduation</td>
<td>0.75</td>
<td>1</td>
<td>0.43</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Post graduation number of publications with return institution coauthors</td>
<td>3.75</td>
<td>2</td>
<td>4.85</td>
<td>0</td>
<td>23</td>
</tr>
</tbody>
</table>

Note: The sample consists of 112 scientists who are trained in the United States in long-term training programs supported by the FIC AITRP and return home following their graduation (graduating between 1989 and 2014). These scientists are the first to return to an institution during the career of a sample non-migrant scientist. Post graduation publications are those published in the five years following the graduation date.
Table 3: Statistics for Non-Migrant African Study Scientists the Year of the Return of a Trainee

<table>
<thead>
<tr>
<th></th>
<th>Control Scientists (N = 1,657)</th>
<th>Treated Scientists (N = 1,657)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>median</td>
</tr>
<tr>
<td>Career age</td>
<td>9.98</td>
<td>8</td>
</tr>
<tr>
<td>Number of publications</td>
<td>6.84</td>
<td>5</td>
</tr>
<tr>
<td>Last author publications</td>
<td>1.38</td>
<td>0</td>
</tr>
<tr>
<td>Number of JIF weighted publications</td>
<td>6.02</td>
<td>3.03</td>
</tr>
<tr>
<td>Number HIV publications</td>
<td>2.54</td>
<td>1</td>
</tr>
<tr>
<td>Number publications with returning trainee</td>
<td>0.0012</td>
<td>0</td>
</tr>
<tr>
<td>Number publications with U.S. coauthors</td>
<td>1.47</td>
<td>0</td>
</tr>
<tr>
<td>Number publications with any U.S. training institution</td>
<td>0.57</td>
<td>0</td>
</tr>
<tr>
<td>Number publications with returnee’s U.S. training institution</td>
<td>0.37</td>
<td>0</td>
</tr>
</tbody>
</table>

Note: This study sample consists of 3,314 African scientists who were actively publishing in HIV related research at the time of the return (or counterfactual return) of a FIC AITRP trainee. All variables are measured using scientist level data gathered from the Elsevier Scopus database, and measurements are made for the five years before the return, unless stated otherwise stated.
<table>
<thead>
<tr>
<th>(1) num pubs without returnee</th>
<th>(2) num pubs returnee</th>
<th>(3) JIF weighted pubs</th>
<th>(4) num first authored pubs</th>
<th>(5) num HIV pubs</th>
<th>(6) non HIV pubs</th>
<th>(7) any pub</th>
<th>(8) any HIV pub</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFTER RETURN × RETURN INSTITUTION</td>
<td>0.062*</td>
<td>0.061*</td>
<td>0.062*</td>
<td>0.013</td>
<td>0.059**</td>
<td>0.022</td>
<td>0.023</td>
</tr>
<tr>
<td>Number of scientists × year observations</td>
<td>36,454</td>
<td>36,454</td>
<td>36,454</td>
<td>36,454</td>
<td>36,454</td>
<td>36,454</td>
<td>36,454</td>
</tr>
<tr>
<td>Number of institutions</td>
<td>440</td>
<td>440</td>
<td>440</td>
<td>440</td>
<td>440</td>
<td>440</td>
<td>440</td>
</tr>
</tbody>
</table>

[a] Estimates in columns (1)-(6) stem from fixed effects ordinary least square specifications with dependent variables being inverse hyperbolic sine of counts of outcomes per scientist per year. Estimates in columns (7) and (8) stem from fixed effects linear probability model specifications with dependent variables being dummy outcomes of the event occurring. All models incorporate a full suite of calendar year, career age and scientist fixed effects.

[b] Heteroskedastic robust standard errors, clustered at the institution, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.
Table 5: Knowledge Flows from Returnee’s International Network to Non-Migrant African Scientist

<table>
<thead>
<tr>
<th>(1) cite any U.S.</th>
<th>(2) cite any U.S. training institution</th>
<th>(3) cite specific U.S. training institution</th>
<th>(4) cite specific U.S. training institution PI</th>
<th>(5) percentile distance from U.S. training institution PI</th>
<th>(6) percentile distance from U.S. training institution PI of non-coauthored pubs</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFTER RETURN × RETURN INSTITUTION</td>
<td>0.039** (0.018)</td>
<td>0.034** (0.016)</td>
<td>0.022** (0.0087)</td>
<td>0.0069* (0.0039)</td>
<td>-1.93** (0.95)</td>
</tr>
<tr>
<td>Mean of the DV</td>
<td>0.45</td>
<td>0.34</td>
<td>0.075</td>
<td>0.020</td>
<td></td>
</tr>
<tr>
<td>Number of scientists</td>
<td>3,314</td>
<td>3,314</td>
<td>3,314</td>
<td>3,314</td>
<td>2,267</td>
</tr>
<tr>
<td>Number of scientists × year observations</td>
<td>36,454</td>
<td>36,454</td>
<td>36,454</td>
<td>36,454</td>
<td>2,267</td>
</tr>
<tr>
<td>Number of institutions</td>
<td>440</td>
<td>440</td>
<td>440</td>
<td>440</td>
<td>320</td>
</tr>
</tbody>
</table>

[a] Estimates stem from fixed effects ordinary least square specifications with dependent variables being the change in a dummy variable if they cite the U.S. based researcher in the five years before and after the return event (or counterfactual) in column (1) (2) (3) and (4), controlling for overall scientist’s number of citations in that year, and the change in percentile intellectual distance between the returnee’s U.S. training institution FIC AITRP PI research output (or counterfactual) and African scientists, before and after the return year (or counterfactual) in column (5), and excluding the sample scientist’s coauthored publications with the U.S. training institution scientists in column (6). Models in columns (1) - (4) incorporate a full suite of calendar year, career age and scientist fixed effects. Models in columns (5) and (6) incorporate career age at treatment and treatment year fixed effects.

[b] Heteroskedastic robust standard errors, clustered at the institution, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.

c] The sample sizes vary in the last two columns as not every scientist in the sample publishes both before and after the return event, and thus is excluded from the analysis (particularly with a pubmed id which is necessary to use the MeSH terms). Just 70% of publications in the full sample have a pubmed id associated with them.
Table 6: Impact of FIC AITRP Trainee Return on Non-Migrant African Scientists’ Collaborations

<table>
<thead>
<tr>
<th></th>
<th>(1) collaborate with U.S. coauthor</th>
<th>(2) collaborate with any U.S. training institution coauthor</th>
<th>(3) collaborate with specific U.S. training institution coauthor</th>
<th>(4) collaborate with specific U.S. training institution PI</th>
<th>(5) any new U.S. coauthors</th>
<th>(6) any new specific U.S. training institution coauthors</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFTER RETURN × RETURN INSTITUTION</td>
<td>0.032*** (0.012)</td>
<td>0.024** (0.0097)</td>
<td>0.0087** (0.0043)</td>
<td>-0.00012 (0.0016)</td>
<td>0.032*** (0.010)</td>
<td>0.0061* (0.0035)</td>
</tr>
<tr>
<td>Mean of the DV</td>
<td>0.22</td>
<td>0.11</td>
<td>0.026</td>
<td>0.0062</td>
<td>0.20</td>
<td>0.092</td>
</tr>
<tr>
<td>Number of scientists</td>
<td>3,314</td>
<td>3,314</td>
<td>3,314</td>
<td>3,314</td>
<td>3,314</td>
<td>3,314</td>
</tr>
<tr>
<td>Number of scientists × year observations</td>
<td>36,454</td>
<td>36,454</td>
<td>36,454</td>
<td>36,454</td>
<td>36,454</td>
<td>36,454</td>
</tr>
<tr>
<td>Number of institutions</td>
<td>440</td>
<td>440</td>
<td>440</td>
<td>440</td>
<td>440</td>
<td>440</td>
</tr>
</tbody>
</table>

[a] Estimates stem from fixed effects linear probability model specifications with dependent variables being dummy outcomes of the event occurring. All models incorporate a full suite of calendar year, career age and scientist fixed effects.
[b] Heteroskedastic robust standard errors, clustered at the institution, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.
Table 7: Breakdown of Publication Outcomes by Non-Migrant African Scientist Network Characteristics

<table>
<thead>
<tr>
<th></th>
<th>(1) num pubs</th>
<th>(2) HIV pub</th>
<th>(3) HIV pub</th>
<th>(4) num pubs</th>
<th>(5) HIV pub</th>
<th>(6) HIV pub</th>
</tr>
</thead>
<tbody>
<tr>
<td>connected to training</td>
<td>0.033</td>
<td>0.018</td>
<td>0.041**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>institution</td>
<td>(0.038)</td>
<td>(0.066)</td>
<td>(0.017)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>connected to OECD</td>
<td>0.0001</td>
<td>0.012</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>not connected</td>
<td>0.076*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>connected to training</td>
<td>2.50</td>
<td>0.55</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>institution</td>
<td>1.62</td>
<td>0.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>not connected</td>
<td>1.39</td>
<td>0.27</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mean of the DV

| Number of scientists   | 195          | 195         | 195         | 1128         | 1128        | 1128        |
| number of scientists x | 2,145        | 2,145       | 2,145       | 12,408       | 12,408      | 12,408      |
| year observations      | 1,991        | 21,901      | 1,991       | 21,901       |             |             |
| Number of institutions | 64           | 64          | 64          | 246          | 246         | 246         |
| Number of returnees    | 63           | 63          | 63          | 90           | 90          | 90          |

[a] The sample of scientists is split into three groups: those who have published with the returning trainee’s U.S. training institution before the return in columns (1) (4), those who have published with OECD collaborators in over 75% of their publications in the 5 years before the return year (or counterfactual) (but never the training institution) - columns (2) (5), and those who have less than 75% of their publications with an OECD collaborator in the 5 years before the event in columns (3) (6).

[b] Estimates stem from fixed effects ordinary least square specifications with dependent variables being inverse hyperbolic sine of counts of outcomes per scientist per year in columns (1)-(3), and fixed effects linear probability model specifications with dummy outcomes in columns (4)-(6). All models incorporate a full suite of calendar year, career age and scientist fixed effects.

[c] Heteroskedastic robust standard errors, clustered at the institution, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.
Table 8: Difference in Non-Migrant Scientist Change in Publication Rate by Returnee Role

<table>
<thead>
<tr>
<th>Returnee Role</th>
<th>Administrative</th>
<th>Technical/Teaching</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average difference</td>
<td>0.70 (8.10)</td>
<td>-0.63 (8.09)</td>
</tr>
<tr>
<td>between post-return</td>
<td></td>
<td></td>
</tr>
<tr>
<td>and pre-return number</td>
<td></td>
<td></td>
</tr>
<tr>
<td>of publications (sd)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of scientists</td>
<td>30</td>
<td>493</td>
</tr>
<tr>
<td>Number returnees</td>
<td>4</td>
<td>29</td>
</tr>
</tbody>
</table>

Note: The sample of treated scientists for whom full resume information on the returnee was found was split into two groups: those who have a returnee taking up an administrative position in column (1) and those who have a returnee taking up a technical or teaching position in column (2). The difference between the total number of pre and post return publications is given.
Table 9: Breakdown of Outcomes by Non-Migrant African Scientist Likelihood to Coauthor with Returnee

<table>
<thead>
<tr>
<th></th>
<th>(1) num pubs</th>
<th>(2) HIV pub</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>high probability</td>
<td>low probability</td>
</tr>
<tr>
<td>AFTER RETURN × RETURN INSTITUTION</td>
<td>0.015 (0.11)</td>
<td>0.064* (0.034)</td>
</tr>
<tr>
<td>Mean of the DV</td>
<td>3.06</td>
<td>1.40</td>
</tr>
</tbody>
</table>

Number of scientists | 261 | 3,053 | 261 | 3,053 |
Number of scientists × year observations | 2,871 | 33,583 | 2,871 | 33,583 |
Number of institutions | 83 | 431 | 83 | 431 |
Number of returnees | 73 | 112 | 73 | 112 |

[a] A predicted probability of coauthoring with the returning trainee is generated by assigning linear predictions from a fitted logit model of pre-return scientist characteristics (collaboration and publication record) on the probability of collaborating with the returning trainee. The sample of scientists is split into two groups: those who have a high predicted probability to coauthor with the returning trainee (or counterfactual) in columns (1) (3), those who have a low predicted probability to coauthor with the returning trainee (2) (4).

[b] Estimates stem from fixed effects ordinary least square specifications with dependent variables being inverse hyperbolic sine of counts of outcomes per scientist per year in columns (1) (2). Estimates stem from fixed effects linear probability model specifications with dependent variables being dummy outcomes of the event occurring in columns (3) (4). All models incorporate a full suite of calendar year, career age and scientist fixed effects.

[c] Heteroskedastic robust standard errors, clustered at the institution, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.
Table 10: Knowledge Flows to Returnee Publications

<table>
<thead>
<tr>
<th></th>
<th>(1) citations to returnee publications</th>
<th>(2) percentile distance from returnee publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFTER RETURN × RETURN INSTITUTION</td>
<td>-0.00034 (0.0019)</td>
<td>0.71 (0.55)</td>
</tr>
<tr>
<td>Mean of the DV</td>
<td>0.0028</td>
<td></td>
</tr>
<tr>
<td>Number of scientists</td>
<td>3,314</td>
<td>868</td>
</tr>
<tr>
<td>Number of scientists × year observations</td>
<td>36,454</td>
<td>868</td>
</tr>
<tr>
<td>Number of institutions</td>
<td>440</td>
<td>199</td>
</tr>
</tbody>
</table>

[a] Estimates stem from fixed effects linear probability models with dependent variables being the change in dummy outcomes if the scientist cites the returning trainee (or counterfactual) pre-return publications in years before or after return (or counterfactual) in column (1), controlling for overall number of citations of sample scientist, and being the change in percentile intellectual distance between the returnee’s publications (or counterfactual), before and after the return year (or counterfactual) in column (2). All models incorporate career age at treatment and treatment year fixed effects.

[b] Heteroskedastic robust standard errors, clustered at the institution, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.
Table 11: Main Specification Using Reduced Sample with Individual and Institution Re-Return Characteristics Match

<table>
<thead>
<tr>
<th></th>
<th>(1) num pubs</th>
<th>(2) any HIV pub</th>
<th>(3) collaborate with any U.S. training institution coauthor</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFTER RETURN × RETURN INSTITUTION</td>
<td>0.062*</td>
<td>0.030</td>
<td>0.018**</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.019)</td>
<td>(0.0082)</td>
</tr>
<tr>
<td>Mean of the DV</td>
<td>1.50</td>
<td>0.32</td>
<td>0.10</td>
</tr>
<tr>
<td>Number of scientists</td>
<td>2,780</td>
<td>2,780</td>
<td>2,780</td>
</tr>
<tr>
<td>Number of scientists × year observations</td>
<td>30,580</td>
<td>30,580</td>
<td>30,580</td>
</tr>
<tr>
<td>Number of institutions</td>
<td>233</td>
<td>233</td>
<td>233</td>
</tr>
</tbody>
</table>

[a] Estimates in columns (1)-(2) stem from fixed effects ordinary least square specifications with dependent variables being inverse hyperbolic sine of counts of outcomes per scientist per year. Estimates in columns (3) stem from fixed effects linear probability model specifications with dependent variables being dummy outcomes of the event occurring. All models incorporate a full suite of calendar year, career age and scientist fixed effects.

[b] Heteroskedastic robust standard errors, clustered at the institution, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.
Table 12: Sensitivity Checks

<table>
<thead>
<tr>
<th>Dept Var: mm pubs</th>
<th>(1) benchmark specification</th>
<th>(2) without ‘gregarious’ returnees</th>
<th>(3) without Ph.D returnees</th>
<th>(4) with country time trends</th>
<th>(5) with institution time trends</th>
<th>(6) with contaminated controls</th>
<th>(7) placebo test 1</th>
<th>(8) placebo test 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFTER RETURN × RETURN INSTITUTION</td>
<td>0.062*</td>
<td>0.079**</td>
<td>0.054*</td>
<td>0.043</td>
<td>0.074</td>
<td>0.19***</td>
<td>-0.011</td>
<td>-0.0088</td>
</tr>
<tr>
<td>(0.034)</td>
<td>(0.031)</td>
<td>(0.032)</td>
<td>(0.034)</td>
<td>(0.045)</td>
<td>(0.035)</td>
<td>(0.026)</td>
<td>(0.033)</td>
<td></td>
</tr>
<tr>
<td>Mean of the DV</td>
<td>1.54</td>
<td>1.52</td>
<td>1.54</td>
<td>1.54</td>
<td>1.54</td>
<td>1.40</td>
<td>1.43</td>
<td>1.52</td>
</tr>
<tr>
<td>Number of scientists</td>
<td>3,314</td>
<td>2,856</td>
<td>3,314</td>
<td>3,314</td>
<td>3,314</td>
<td>2,419</td>
<td>3,314</td>
<td>1,657</td>
</tr>
<tr>
<td>Number of scientists × year observations</td>
<td>36,454</td>
<td>31,416</td>
<td>36,454</td>
<td>36,454</td>
<td>36,454</td>
<td>26,609</td>
<td>19,884</td>
<td>18,227</td>
</tr>
<tr>
<td>Number of institutions</td>
<td>440</td>
<td>411</td>
<td>428</td>
<td>440</td>
<td>440</td>
<td>314</td>
<td>440</td>
<td>391</td>
</tr>
</tbody>
</table>

[a] Estimates stem from fixed effects ordinary least square specifications with dependent variables being inverse hyperbolic sine of counts of publications per scientist per year. All models incorporate a full suite of calendar year, career age and scientist fixed effects.

[b] Heteroskedastic robust standard errors, clustered at the institution, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.
Appendix A: Construction of the Control Group

I detail the procedure implemented to identify the control scientists that help to account for life-cycle and secular trends in the difference-in-differences specification. Publication outcomes might be subject to life-cycle patterns, with outcomes reflecting the trends of the age of the scientist. Also - scientific productivity, particularly in Africa, is rapidly changing over time. Therefore relying on scientists treated earlier or later as an implicit control group may not fully capture these time-varying omitted variables.

To address this concern, I create a sample of control scientists to account for time varying variables in the difference-in-differences specification. Specifically I identify a control scientist who is ‘similar’ to each treated scientist and assign to them their matched treated scientist’s counterfactual return event (returning scientist / return year / American training institution). The control scientists are selected from a universe of possible scientists who are based in FIC AITRP countries and affiliated with institutions that never receive a FIC AITRP returnee in their career lifetime.

The universe of possible control scientists is generated using affiliation data from Elsevier Scopus publication database with inclusion criteria such that the scientist must have published at least three times in their lifetime and at least once as first or last author (to remove technicians or incidental publishers). The institution of each scientist is determined as being the institution in which they are affiliated with in a given time period in over 75% of their publications (to avoid visiting or honorary appointments).

The list of covariates used to identify ‘similar’ control scientists for each treated scientist such that the following conditions are met:

1. treated scientists exhibit no differential output trends relative to control collaborators up to the time of return (or counterfactual);
2. treated scientists exhibit no differential trends in terms of international, particularly American, collaborations relative to control collaborators up to the time of return (or counterfactual);
3. treated scientists exhibit no differential trends in terms of their field of study relative to control collaborators up to the time of return (or counterfactual);
4. the distribution of career age at the time of return (or counterfactual) are for similar treated and control scientists.

Coarsened exact matching. To meet these goals, I implement the nonparametric ‘coarsened exact
matching’ (CEM) procedure (Iacus, King and Porro 2011) to identify a control scientist for each treated scientist. The first step is to select a set of covariates on which to guarantee balance, and the second is to create a large number of (coarse) strata that covers the entire support of the joint distribution of the covariates in the previous step. In a third step, each observation is allocated to a stratum and for each treated observation, a control is selected from the same stratum. If there are multiple choices possible, one is selected randomly. If the treated observation is unmatched it is removed from the sample. In this implementation, control scientists are recycled, and so a small number serve as the control for several treated observations (which is accounted for in the specification through the use of scientist specific identifiers by which to cluster standard errors).

**Implementation** I identify controls based on the following set of covariates (t denotes year of return): career age at t, a dummy for any HIV publication in the four years before t, a dummy for any publications with United States based coauthors in the four years before t, a dummy for any publications with coauthors at American institutions involved in FIC AITRP program in the four years before t, and dummy for any publications with United States based coauthors in years t-1, t-2, t-3 and t-4.

I implement the CEM procedure year by year, with replacement. Specifically, in year t, I

1. eliminate from the set of potential controls all those who begin their publication record after year t-1;
2. create the strata using the variables described above;
3. identify within each strata a control for each treated unit, randomly selecting one if there are more than one match;
4. assign the control a counterfactual returnee/year of return/returnee American training institution based on the matched treated returnee;
5. repeat these steps for year t + 1

I match 1,657 of 1,740 treated scientists (95%). In the sample of 1,657 treated + 1,657 controls = 3,314 scientists, there is no evidence of preexisting trends in output (figure A.1).
Figure A1: Publication Trends for Treated and Control Scientists

(a) Mean number of publications

(b) Any HIV publication

(c) Any publication with American-based coauthors

(d) Any new American-based coauthors
Appendix B: Additional Results
Table B1: Sensitivity Checks - Returnee Pre-Return Characteristics

<table>
<thead>
<tr>
<th>Dept Var: num pubs</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFTER RETURN ×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RETURN INSTITUTION</td>
<td>0.062**</td>
<td>0.054</td>
<td>0.054*</td>
<td>0.0040</td>
<td>-0.013</td>
</tr>
<tr>
<td></td>
<td>(0.034)</td>
<td>(0.033)</td>
<td>(0.032)</td>
<td>(0.0045)</td>
<td>(0.037)</td>
</tr>
<tr>
<td>AFTER RETURN ×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RETURN INSTITUTION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>× returnee productivity during fellowship</td>
<td>0.007</td>
<td></td>
<td></td>
<td>-0.0009</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFTER RETURN ×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RETURN INSTITUTION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>× returnee PhD</td>
<td>0.13</td>
<td></td>
<td></td>
<td>0.19*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.082)</td>
<td></td>
<td></td>
<td>(0.10)</td>
<td></td>
</tr>
<tr>
<td>AFTER RETURN ×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RETURN INSTITUTION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>× returnee from more advanced country</td>
<td>0.075</td>
<td></td>
<td></td>
<td>0.084</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
<td></td>
<td></td>
<td>(0.058)</td>
<td></td>
</tr>
</tbody>
</table>

Number of scientists | 3,314     | 3,314     | 3,314     | 3,314     | 3,314     |
Number of scientists × year observations | 36,454 | 36,454 | 36,454 | 36,454 | 36,454
Number of institutions | 440 | 440 | 440 | 440 | 440

[a] Estimates stem from fixed effects ordinary least square specifications with dependent variables being inverse hyperbolic sine of counts of publications per scientist per year. All models incorporate a full suite of calendar year, career age and scientist fixed effects, and terms of the interaction of post-event and covariates are included.

[b] Heteroskedastic robust standard errors, clustered at the institution, are given in parentheses. *, **, *** denote statistical significance at p-values of 0.1, 0.05 and 0.01 respectively.