Migration and Local Problem-Solving*

Caroline Fry[1] and Gauri Subramani[2]

[1]Shidler College of Business, University of Hawai'i at Manoa

[2]College of Business, Lehigh University

August 28, 2024

Abstract

Migrants play an important role in transferring knowledge and resources across organizations and borders, which in turn affects the rate of innovation of those who remain at home. However, research on how migrants influence the *direction* of innovation of their non-migrant peers is limited: How does exposure to a migrant affect non-migrants' research agendas? In the context of South-South migration of female scientists, we compare the disease orientation of non-migrants' research output before and after the migration of a peer against a control group of non-migrants whose peers applied for, but did not receive, a fellowship for foreign study. We document that non-migrants publish more after their peers migrate, and their relative rate of publication in diseases that are prevalent in both home and host countries increases. Our results suggests that migrants can broker access to knowledge and collaborations that can increase their home country peers' focus on shared local problems. This study underscores how bridging networks of knowledge workers can enhance and direct innovation on pressing problems.

*Both authors contributed equally. This research was supported by the NBER Small Grants Science of Science funding. We are grateful to Akanksha Roy, Muzammil Khan, Sanchita Shrivastava, and Ben Subramani for research assistance.

1 Introduction

Migrants play an important role in the exchange of knowledge and resources between organizations and across borders (Nanda and Khanna, 2010; Agrawal et al., 2011a; Borjas and Doran, 2012; Ganguli, 2015; Wang, 2015; Choudhury and Kim, 2019; Hernandez and Kulchina, 2020; Balachandran and Hernandez, 2021). A growing body of literature highlights the positive impact of migrants, both those who stay in their host location and those who return back home, on innovation outcomes for individuals remaining at home. These outcomes include increased access to knowledge and international networks (Kahn and MacGarvie, 2016b; Fry and Furman, 2023) as well as enhanced productivity (Kerr, 2008; Choudhury, 2016; Fry, 2023a). Although these studies enhance our understanding of how skilled migration affects innovation rates in home regions, the influence of migration on the *direction* of innovation at home remains unclear. Specifically, how does second-hand exposure to other contexts, or migrant brokerage, shape non-migrants' research agendas?

In this paper, we aim to understand how migrant brokerage influences the research agendas of non-migrants remaining at home by examining the extent to which exposure to a migrant affects their focus on local problems. For instance, scientists in countries where malaria is endemic may be more or less likely to research malaria after a key individual from their institution migrates to another country. Evaluating how migrants affect their home-country peers' research can inform our understanding of what drives the direction of innovation more broadly. We know from prior work that research and innovation do not address the issues impacting all regions and populations equally (Acemoglu and Zilibotti, 2001; Moscona and Sastry, 2022; Vakili and McGahan, 2016; Koning et al., 2021; Einiö et al., 2019; Koning et al., 2021). For instance, there is limited research on diseases affecting low-income countries (Vakili and McGahan, 2016; Kremer and Glennerster, 2004). Since local scientists and innovators are particularly well-suited to tackle issues specific to their regions and communities, it is essential to better understand the factors that shape their propensity to study these local challenges.

On the one hand, upon being exposed to new environments, migrants themselves might shift their focus away from locally relevant issues and towards problems they encounter in their new locations (Fry, 2023d). Non-migrants in their home locations could be influenced by the migrant's research agenda or by the new opportunities that the migrant facilitates. If migrants transfer knowledge, connections, and resources that are less relevant to local issues, non-migrants may also be less likely to focus on locally-relevant problems.

On the other hand, the migration of an individual may sometimes lead those who remain at home to concentrate more on local issues. It could be that migrants broker access to knowledge, connections, and resources that can be leveraged to study problems that are specific to their home locations. We argue that exposure to a migrant who can link previously less well-connected parts of the network that face *shared problems* is particularly likely to increase non-migrants' emphasis on the set of shared, locally-relevant issues. Migrants traveling to locations with shared challenges can enable the formation of collaborations, the recombination of knowledge and the generation of new solutions to local problems and highlight opportunities that inspire non-migrants to tackle local challenges. For instance, scientists from developing countries might discover alternative solutions or valuable opportunities when they visit other nations in the Global South facing similar problems, such as diseases like malaria.

We conduct our analysis using data on the migration of female scientists who apply for a fellowship administered by the Organization for Women in Science for the Developing World (OSWD). This fellowship program provides opportunities for female scientists in developing countries to study in another developing country. To assess the influence of migration of an OWSD fellow on the research agendas of non-migrant home country researchers, we construct a panel dataset of non-migrants' publication outcomes. We focus on non-migrants at the OWSD fellowship applicants' home institutions and in the same scientific fields, limiting to fellows and non-migrants who publish in disease related

topics. Combining publication data with data on country-specific disease prevalence, we analyze changes in the probability that a non-migrant publishes on diseases that are more prevalent in their home (and/or the migrant's host) country before and after the migration event. We compare the publication output of non-migrants whose peers received an OWSD fellowship with those of non-migrants whose peers applied for but did not receive the fellowship. This design allows us to control for career, field, disease, and temporal changes in publication output.

We find that non-migrants publish more after the migration event, as compared to non-migrants from institutions with an unsuccessful fellowship applicant. This increase in publications is primarily driven by additional publications related to neglected tropical diseases (NTDs).[1] Additionally, treated non-migrants tend to increase their research focus on diseases that are prevalent in their home countries following the migration of a peer scientist, particularly those diseases that are also prevalent in the host country. Conversely, they experience a smaller increase in research productivity in diseases prevalent in the home country but not prevalent in the host country, and they do not increase their research on diseases specific to the host country that are not also prevalent in the home country. We also document that this effect on overlapping home and host issues is mainly driven by publications on neglected tropical diseases, indicating that migration and representation on a global stage plays a crucial role in advancing much needed research on specific topics.

We investigate possible mechanisms for these effects and do not find evidence that direct collaboration with the migrating fellow is the primary driver of non-migrants' increased focus on local research topics. Instead, we find suggestive evidence that the migrating fellow acts as a bridge to host country knowledge and collaborators, and that this access to foreign knowledge, connections, and resources partially drives the observed

[1]Neglected tropical diseases(NTDs) are localized problems primarily present in low-income and tropical areas, affecting more than 1 billion people worldwide (https://www.who.int/news-room/questions-and-answers/item/neglected-tropical-diseases last acceessed 8.23.24).

relative increase in research on local problems.

Our findings contribute to the existing literature on the role of migration and brokerage in innovative outcomes. While an extensive prior literature has documented the important role of migrants in global knowledge diffusion and network formation (Saxenian, 2002, 2005; Agrawal et al., 2006; Saxenian, 2007; Kerr, 2008; Oettl and Agrawal, 2008; Agrawal et al., 2011b; Choudhury, 2016; Kahn and MacGarvie, 2016b; Kerr, 2018; Choudhury and Kim, 2019; Fry and Furman, 2023), we highlight a novel way by which migrants influence innovation that happens in their home environment: shaping home country peers' research *direction*. Of note, we highlight the important role of shared problems across contexts, a finding that has implications for theory on migrant brokerage and network dynamics, as well as for policy and practice supporting the migration of knowledge workers.

Our results also speak to the broader literature on the determinants of the direction of innovation. We identify a mechanism by which research on specific topics can be fostered. Researchers have provided micro-level evidence on the role of organizational structure (Balachandran and Eklund, 2024), access to collaborators (Catalini, 2018; Catalini et al., 2020), funding availability (Myers, 2020), tools and data (Nagaraj et al., 2020; Nagaraj, 2022; Nagaraj and Tranchero, 2023; Furman and Teodoridis, 2020), and exposure to problems (Koning et al., 2021; Moscona and Sastry, 2022; Truffa and Wong, 2023; Fry, 2023b; Andrews and Smith, 2023) in shaping the direction of innovation. We show that an individual migrating from a developing country can orient the direction of non-migrant peers' research agendas towards home country problems, particularly those shared with the host country. More broadly, our findings highlight how generating ties within distributed networks of knowledge workers, especially those with shared challenges, can enhance innovation on pressing problems.

This research has significant implications for managers, organizations, entrepreneurs, and policymakers seeking to enhance and direct innovation. Organizations, broadly de-

fined, particularly those aiming to make progress on specific topics, can effectively leverage migrant hires and outbound talent to provide a strategic advantage in innovation. By integrating individuals with diverse international experiences and perspectives, it is possible to stimulate creativity, spur cross-country collaborations, and develop innovative, global solutions. Further, this research suggests that in order to direct innovative outcomes, particular attention be paid to facilitating ties across parts of the network with shared problems: internal or external to the organization. For policymakers, the insights from this research can provide evidence of the value of opportunities for international research exchange and inform the development of frameworks to harness the potential of networks built by mobile scientists and innovators.

2 Migration, Brokerage, and the Direction of Innovation

2.1 Migrants and Brokerage

Migrants play a key role in both transferring knowledge and connections from their home regions to their new locations (Borjas and Doran, 2012; Ganguli, 2015; Wang, 2015; Choudhury and Kim, 2019; Fry, 2023a; Fry and Ganguli, 2023). Whether they remain in their host location or return home, migrants can act as a bridge between individuals in their home and host regions. This brokerage effect has been well-established in scientific and innovative fields; for example, migrants can facilitate the formation of new collaborations and boost the patenting and publication productivity of their non-migrating peers (Choudhury, 2016; Fry and Furman, 2023; Fry, 2023a; Xie and Freeman, 2023).

While we know from previous work that network connections brokered by migrants can affect the volume of non-migrants' innovative output, the effect of these network connections on the *direction* of innovation conducted by non-migrants is not well-understood. Specifically, there is no clear evidence about how and when exposure to a migrating peer affects the local focus of individuals' research. Given that problems faced by low-income

countries, for example, are understudied (Kremer and Glennerster, 2004; Vakili and McGahan, 2016; Moscona and Sastry, 2022), if migration away from these countries affects the level of research conducted on local topics by non-migrating peers, this could have important implications for the direction of innovation more broadly.

2.2 The Effect of Migration on Non-Migrants' Research Agendas

It could be that exposure to a migrant leads non-migrants to learn about and subsequently focus more on non-local problems. Prior research has documented that researchers and innovators tend to focus on issues that they are exposed to (Einiö et al., 2019; Koning et al., 2021; Moscona and Sastry, 2022; Fry, 2023c; Truffa and Wong, 2023), and two recent studies document that researchers are more likely to study problems that are local to their host locations after migration (Fry, 2023d). Therefore, migrant researchers themselves might shift their focus away from issues local to their home regions and towards problems they are exposed to and encounter in their new locations. Non-migrating peers may be influenced directly by the migrant's research agenda, particularly if the migrant collaborates with the non-migrants, provides them with mentorship and insight into new topics.

Irrespective of their own research agendas, migrants could also indirectly facilitate access to international opportunities–knowledge, connections, and resources– which, given the geographic variation in problems around the world, could be less likely to be related to the non-migrant's local environment. Insofar as a migrant facilitates access to knowledge, connections, and resources that are less oriented toward local problems, non-migrants may take advantage of these opportunities and shift their research away from local topics.

Alternately, it could be that following the migration of a peer, non-migrants increase their focus on local problems. Personal relevance can influence the choice of topics studied by scientists and innovators (Einiö et al., 2019; Koning et al., 2021), and innovators are

more likely to focus on topics on which they have a unique advantage (Fry, 2023c), including those with significance in their home locations. Additionally, it is costly to make changes in research orientation (Myers, 2020), so researchers studying local topics may be unlikely to shift their agendas even after exposure to new ideas. Therefore, non-migrant peers may actually focus *more* on local problems if migrants directly or indirectly broker access to knowledge, connections, and resources from previously less well connected parts of the network that can be leveraged to study important home-country topics. This is particularly likely if there is overlap between local problems in the home country and problems in the host context.

2.3 The Power of Shared Problems

In contexts where there are shared problems in the home and host countries, non-migrant peers can gain access to knowledge, collaborators, and resources that are relevant to their local environments. This, in turn, can enhance their research on home-country topics, particularly on the set of shared problems, and increase their investments in this work in multiple ways.

First, non-migrants can collaborate with researchers in the host location who face similar challenges. By leveraging their overlapping motivations and expertise, they can focus on identifying common problems to address together. In fact, previous research has shown that when barriers to collaboration are reduced, researchers often align their agendas (as measured by the work they cite) more closely (Catalini, 2018).

Beyond collaboration, non-migrants who learn about research on shared problems conducted in other settings might find opportunities to leverage these findings in their own work. They might either apply these insights directly or combine them with their existing research to develop novel solutions to pressing issues in their home contexts. The concept of recombining ideas to spur innovation, as discussed by Weitzman (1998) and Fleming (2001), illustrates how such cross-border exchanges can lead to breakthroughs

that might not have been possible within the confines of a single research environment. In line with this idea, existing research has found that exposure to new ideas leads to their use in novel ways. In the context of migration, Choudhury and Kim (2019) highlight how migrants transfer knowledge across borders and engage in the recombination of knowledge from both their home and host countries. The authors find that migrants from India and China to the United States engage in patenting at their American employers that leverages herbal knowledge from their home countries.

Moreover, exposure to diverse perspectives on shared problems can stimulate creative thinking, encouraging non-migrant researchers to approach these problems in new and innovative ways (Garfield et al., 2001; Fink et al., 2012). This is likely to be particularly true if migrants are connecting parts of the network with key similarities. For example, "small world" networks, or those in which actors are both connected to each other and connected by persons who know each other well, can stimulate creativity and improve performance (Uzzi and Spiro, 2005).

Finally, the experience of witnessing other researchers tackling similar problems, or the opportunity to collaborate with them, can serve as a significant source of motivation for non-migrant researchers. We know from prior work that exposure to new information can stimulate motivation and a desire to learn more (Ditta et al., 2020). In our context, this motivation might stem from a reinforced sense of the importance of their work or the recognition that others are also invested in solving similar issues.

In the remainder of this section, we introduce a setting in which it is likely that migrant brokers can facilitate connections between previously less well connected parts of the network with important, shared problems: South-South migration of scientists.

2.4 South-South Migration of Scientists

There exists significant variation in the prevalence of problems (and relevant solutions) around the world. For example, neglected tropical diseases such as malaria, schis-

tosomiasis, and African sleeping sickness are predominantly found in low-income countries. This is true not only for diseases; Moscona and Sastry (2022) show that the distribution of crops and pests has significant variation worldwide. Further, prior research has documented differential rates of research into the problems and populations prevalent in the Global South in a variety of contexts, including diseases (Vakili and McGahan, 2016), crops and pests (Moscona and Sastry, 2022), psychological research (Apicella et al., 2020), and topics related to management research (Nagaraj and Yao, 2023).

Researchers in the Global South are uniquely positioned to study local issues, as they are directly exposed to these challenges and have a comparative advantage in addressing them (Fry, 2023b). Their proximity to these grand challenges often fuels their motivation. However, in low-income countries, support for research is low, and these researchers could face constraints that inhibit access to knowledge, collaborators and resources which limit their propensity to study local problems.

One way to overcome these limitations is to access international knowledge, collaborators and resources via migrant brokers (Kahn and MacGarvie, 2016a; Fry, 2023a). In particular, for scientists who migrate within the Global South, it is likely that they are moving to and from a set of countries with greater overlap in local problems than, for example, migrants from the Global South to the Global North. Considering the above, access to international collaborators and relevant knowledge and resources may be particularly valuable for scientists exposed to a South-South migrant and are likely to increase their propensities to study local problems.

This leads us to our first hypothesis:

Hypothesis 1: Peers at the home institutions of scientists who migrate within the Global South are more likely to focus their research on local problems.

Given the discussion above, we posit that this effect will be more salient when the home and host countries have overlapping sets of problems. Thus, our next hypothesis follows:

Hypothesis 2: Peers at the home institutions of scientists who migrate within the Global South are more likely to focus their research on local problems if the problem is shared with the migrant's host country.

3 Setting, Data and Empirical Strategy

One of the main barriers to studying how exposure to a migrant affects research direction is that there are issues related to the endogeneity of network connections and the difficulty of identifying a counterfactual network (Manski, 1993; Jackson and Wolinsky, 1996). Further, since migration of researchers and innovators is often selective and primarily involves the global elite in developed countries who travel within already well-connected networks (Azoulay et al., 2022), it is challenging to identify settings in which there is observable South-South migration of scientists.

To resolve these challenges, we explore the impact of the migration of female scientists who move between countries in the Global South and are supported by a PhD fellowship operated by the Organization for Women in Science for the Developing World (OWSD). We compare changes in research output of non-migrating peers of a successful OWSD PhD fellow to changes in research output of non-migrating peers of an unsuccessful OWSD PhD fellowship applicant. The peers of the unsuccessful applicant allow us to identify counterfactual outcomes in the absence of the OWSD fellowship.

3.1 Organization for Women in Science for the Developing World

OWSD provides grants to female scientists to study abroad during their PhDs. Female scientists in developing countries apply to OWSD for funding to support their PhD studies after having obtained approval from their proposed host institution in another country in the Global South. Fellows can apply for a full-time fellowship, which allows them to complete their PhD degrees at a host institution for up to four years, or they

can pursue a “sandwich" fellowship which enables them to study for a shorter period in another country and then return to their home institutions to complete their degrees.

Fellows are selected based on the feasibility of completing their proposed research project, the match between the proposal and the potential host institution, and the quality of the connection that the applicant has made with the host and other relevant partner organizations. OWSD also aims to achieve diversity each year in terms of the fellows' locations and scientific fields. Importantly for our study, there is no evidence that OWSD selects fellows based on the alignment of their peers' research interests and the prevalence of these topics in their home or proposed host countries.

Unfortunately, the unsuccessful OWSD applicants in our sample are unlikely to have opportunities to serve as migrant brokers between their peers and others outside of their institutions. In the absence of the OWSD fellowship, access to migration opportunities, and international networks is limited for female scientists in the Global South. OWSD applicants are generally “young women who may never have left their home town before, let alone their country" (Quadrio-Curzio et al., 2020). Indeed, our conversations with OWSD administrators confirmed that the majority of unsuccessful applicants fail to visit their proposed host institutions, providing us with a reasonable control group to which we can compare the impact of migrant brokerage.

3.2 Sample construction

3.2.1 OWSD applicants

We begin with a sample of OWSD PhD fellows who applied to the program from 1996 to 2016 and agreed to have their data used for research. We have data on 315 applicants in total, which includes data on 134 successful applicants and 181 unsuccessful applicants.[2] We have information on their home institution, proposed host institution,

[2]There are a total of 2,952 applicants, out of which which 654 were awarded fellowships. Our sample of 315 applicants is representative of all applicants in terms of country of origin and field of study.

year of fellowship, and field of study. The home institution is where they are based at the time of applying for the OWSD fellowship. We match the names, institutions and field of study of the fellows with their full publication record, if any, using the Elsevier Scopus database.

3.2.2 Non-migrant scientists

As we are primarily interested in the effects of being *exposed* to a migrating fellow as opposed to the effects on the fellow directly, we generate a sample of applicants' non-migrating peers.

To identify non-migrant peers, we select individuals from the Elsevier Scopus publication database who are affiliated with applicants' home institutions in the three years prior to fellowship application and who publish in the same broad field (such as agricultural research or health research).[3] We extract each non-migrant scientist's entire publication history from the Elsevier Scopus publication database to generate annual variables on research output as described below. We retain publication records of sample scientists for the five years prior and up to ten years following the fellowship application year.[4]

Measuring the extent to which researchers focus on local problems is challenging. We do so by concentrating our publications and measures on diseases (with more details in the next section) and leverage geographic variations in disease burden worldwide to create clear indicators of both the local nature of issues and the focus of a scientist's research agenda. Given this focus, we restrict our sample to the set of OWSD applicants and peers who ever publish in *any* disease. This leaves us with 3,667 treated non-migrant scien-

[3]Many scientists in our sample are exposed to more than one OWSD applicant. We treat each peer $\times$ applicant pair as a separate observation, and cluster standard errors at the level of the peer. Appendix Table A16 shows our results limiting only to the set of peers treated by a single applicant.

[4]For ancillary analysis, we impute the gender of each peer scientist (that is, peers of the OWSD applicant) in the sample. We use a commercial database obtained from the company Ethnic Technologies that identifies gender using first names. However, we only have data on first names for 45% (8,542) of the individuals in our data because the remainder of our individuals do not have a first name listed on their publications. For the sample of people for whom we do have first names, we can identify gender unambiguously for 35% (6,887 people). Our gender-identified sample of peers is 27% female and 73% male.

tists (those whose peers received an OWSD fellowship) and 5,256 control scientists (those whose peers applied for but did not receive an OWSD fellowship) who ever publish in diseases.[5] This corresponds to 198 OWSD applicants. Just over 59% of the non-migrant disease sample is from Nigeria, with other common home countries being South Africa, Zimbabwe, Cameroon, Bangladesh, Ethiopia, and Kenya. Around 40% of the sample identified South African host institutions, with other common host institutions situated in China, Malaysia, India, Pakistan, Kenya, and the Philippines.[6]

3.3 Data and measurement

3.3.1 Disease publications and prevalence

In addition to annual counts of total publications, we develop several measures of research direction. For all peer scientists' publications, we identify whether each publication is related to diseases that are prevalent in the peer and applicant's home country and/or the applicant's proposed host country.[7]

To measure the disease focus of each publication, we use a key word search to identify whether a publication covers a specific disease topic out of all 156 diseases[8] included in the Global Burden of Disease database (i.e., for each disease we identify whether it is mentioned in the title or abstract of the publication). 30% of publications in our sample match to at least one disease.

We then identify the country-specific prevalence of each of our 156 diseases using

[5] In total (not limiting to those who ever publish on diseases), we have 7,831 treated non-migrant scientists and 11,639 control scientists.

[6] Appendix Table 1 lists the set of home and proposed host countries for non-migrating peers.

[7] As a robustness check, we validate our main results using data on crop and pest prevalence. This is discussed in greater detail in Appendix B. In a manner identical to the disease identification, we use key word searches of the title and abstract of the publications to identify whether each publication covers a crop or pest topic. We next merge the set of all sample publications to data on country-specific crop/pest prevalence in using the Centre for Agriculture and Bioscience International's (CABI) Crop Protection Compendium (CPC). Results of these analyses are presented in Appendix Table 14.

[8] We exclude diseases that are not endemic to physical environments; eg, "falls", "drowning", "fire, heat, and hot substances".

the Global Burden of Disease database. The majority of our analysis utilizes the disease prevalence measure of the percent of 'Disability-Adjusted Life Years' in a country in the year prior to the publication year (to allow for publication lag). We consider a disease to be of 'high prevalence' in a country if its prevalence is greater than the median disease prevalence as calculated from our full set of diseases and countries.[9] There is significant variation in highly-prevalent diseases across home and host countries. For some home countries, high-prevalence diseases include liver cancer and diabetes, whilst for others they include malaria and schistosomiasis.[10]

We leverage our data on diseases featured in publication abstracts and titles to generate several key measures that capture the local nature of research. In our primary analysis, we measure individual-level annual publication counts to evaluate the number of articles a non-migrant publishes[11] on diseases that are of high prevalence in:

- The home country
- The host country[12]
- Both the home and host countries
- The home country but *not* the host country
- The host country but *not* the home country

Figure 1 provides a graphical representation of the five different sets of publications we explore in this paper.

[9] Appendix Table 2 shows our results using alternative measures of disease prevalence.

[10] We are particularly interested in research on Neglected Tropic Diseases (NTDs); thus, in additional analysis, we narrow our set of diseases to specifically examine migrant peers' proclivities to publish on NTDs.

[11] Around 8% of our sample publications cover more than one disease. In this instance, if there is at least one disease covered in the publication in the named measurement category, the publication takes the value of 1 for that measure. If there is more than one high prevalence disease covered in the named category, the publication takes the value of the count of diseases covered. These measures therefore reflect the number of disease publications per person per year.

[12] For migrant peers, the host country is where their peer who applies for the fellowship identifies as the country they will visit.

3.3.2 Collaborations

To explore any mechanisms driving our observed effects, we evaluate whether the migrant facilitates access to knowledge, connections, and resources directly via collaborating with non-migrant peers or indirectly via brokering international collaborations. We generate variables that reflect the rate at which non-migrant peers collaborate with both the OWSD applicant and with individuals in the applicant's host country. Namely, we evaluate the author lists of each publication in our sample and identify if they include: 1) the OWSD applicant, 2) anyone in the host institution, and 3) anyone in the host country.

3.3.3 Knowledge recombination

We also explore the extent to which the migrant facilitates access to new knowledge from the host institution. We collect data from Elsevier Scopus on all host institutions' publications in applicants' areas of study. We then extract the keywords associated with each publication and institution to construct a list of keywords that describe the research conducted in each host institution in each year.[13] For example, for a specific host institution that had a fellow visiting for a PhD in biology, we collect data on all publications in biology and generate a list of all of the keywords indexing these articles by year of publication. We count the proportion of host institution keywords that the non-migrating peer uses in their own research before and after the migration event. Through this analysis, we are able to evaluate if, after the migration event, non-migrating peers' research more closely resembles the work is done at the host institution.[14]

[13] Keywords are indexed by Elsevier Scopus.

[14] We limit to pre-migration event host institution keywords to avoid any contamination that may occur because of the migration event itself (that is, if the migrating peer introduces home country ideas or topics to the host institution).

3.4 Empirical strategy

In order to identify the effect of the migration event on non-migrant scientists' research agendas, we leverage a differences-in-differences design. We compare publication patterns on specific disease topics before a peer's move versus after, relative to the change in non-migrant-exposed scientists' outcomes, using a scientist fixed effect specification. Our preferred analysis uses data at the person-year level and our outcome variables measure the annual count of publications in each category of interest (publications on home topics, publications on host topics, publications on topics prevalent in both home and host contexts, and publications on topics prevalent only in home or host contexts). The estimating equation (1) relates a non-migrant scientist i's publications in year t to the timing of the fellow's migration event. The first year of treatment is the year after fellowship application (to allow for publication lag and migration time). We estimate the following:

$$E\left[y_{it}|X_{it}\right] = exp\Big[\beta_0 + \beta_1(After_Fellowship \times Successful_Applicant)_{it} + \beta_2(After_Fellowship)_t + \gamma(publications)_{it} + \sigma_t + \tau_i + f(AGE_{it})\Big] \quad (1)$$

Where $\gamma(publications)_{it}$ is a control for an individual's annual count of publications, σ_t are year fixed effects, τ_i are individual fixed effects, and $f(AGE_{it})$ is a flexible function for scientist's career age. Standard errors are clustered at the person level.

While our primary results are based on a person $\times$ year-level dataset, we also conduct analyses at the person $\times$ year $\times$ disease-level; that is, we construct a dataset that identifies the number of publications an individual has on each of the 156 diseases in our sample in each year. In these specifications, we include person $\times$ disease and disease $\times$ year fixed effects, as well as person $\times$ year fixed effects. This allows us to account for unobserved propensities of individuals to publish in particular diseases overall, variation in disease research output trends over time, and person-level general research output trends over

time.

Our main identifying assumption in our baseline analysis is that in the absence of the OWSD fellowship, the research trajectory of the treated scientists, or non-migrants exposed to a migrating female fellow, would have following the same trend as that of the control scientists, or non-migrants exposed to female scientists who were unsuccessful in their fellowship application. We validate this assumption in two main ways. First, we are evaluating peers' publication patterns, not fellows' own publications; we anticipate that an applicant's fellowship outcome is unlikely to be based on their peers' research trajectories. Second, to test whether there were already differences in publication patterns between our treatment and control groups in the pre-treatment periods, we estimate event study specifications. Figure 3 presents these results for publications on home-country diseases. The figure demonstrates that there is no significant pre-trend; treated and control peers publish on diseases present in their home countries at equivalent rates prior to the migration event.[15]

4 Results

4.1 Descriptive Statistics

In Table 1, we provide details on the treated and control samples at the year of application. Columns 1 and 2 show averages and standard errors in brackets for our control and treatment groups, respectively. Column 3 shows the raw difference in means between treated and control groups, and column 4 shows the t-statistic for significance of

[15]Treatment occurs at various times in our setting; not all fellows are treated in the same year. When treatment timing varies, our estimation approach using difference-in-differences estimation compares already-treated units and later-treated units (a "forbidden comparison" (Goodman-Bacon, 2021)), which can negatively weight units that are treated earlier and lead to inaccurate estimates. As a result, we also evaluate our results using differences-in-differences estimators that avoid comparing already-treated units to later-treated units. We identify that the estimators developed by Borusyak et al. (2022) and Gardner (2022) are best suited for our setting. Appendix B provides further detail about our empirical approach. We conduct our main analysis using these estimators and Figure B1 shows pre-trend figures using these results. Our results are very similar to what we find in our Ordinary Least Squares analyses.

that difference.[16] As we see in this table, there is balance of key productivity statistics, career, gender and research direction at the time of treatment. Even so, we note that any differences between treated and control scientists in terms of absolute values are not of particular concern to us empirically; for our empirical approach, we rely on the assumption that the trends in these rates are relevant, and we control for this by including individual-level fixed effects in our regressions.

Across the sample of treated and control scientists, on average, sample scientists publish more on home country diseases than host country diseases, and these papers on home-country topics make up a large proportion (on average, 42%) of their overall publications pre-treatment.[17] Collaborations with outside researchers are fairly rare, and non-migrants produce just 0.09-0.11 publications with host country researchers in the five years prior to treatment.

4.2 Main Finding

4.2.1 Changes in disease focused research

We start by examining the effect of exposure to a migrant on non-migrants' overall productivity. Table 2 presents these results.[18] We show that non-migrants whose peers received an OWSD fellowship publish more research after their exposure to a migrating fellow (column 1). Non-migrants' annual publication outputs are 6.3% higher following their peers' initiation of a fellowship abroad.

After controlling for this overall increase in annual publications, we see that treated non-migrants increase their relative publication output on disease topics (column 2) whilst decreasing the relative focus on non-disease topics (column 3), and the majority of this in-

[16] The t-test compares the control and treatment group means and assumes a null hypothesis that the two means are equal. If the t-statistic is greater than 1.96, we can reject the null hypothesis that there is no difference between means at a 5% level of significance.

[17] As in our analysis, our count of total disease publications measures the number of diseases covered in publications by the focal individual.

[18] Note that the coefficients for $After_Fellowship$ and $Successful_Applicant$ are absorbed by fixed effects for years since treatment, and individual level fixed effects.

crease in disease research is in neglected tropical diseases (NTDs) (column 4) which are diseases predominantly affecting communities in the Global South. In the remainder of the paper, we explore the relative focus on different categories of diseases, controlling for overall annual publications throughout which allows us to measure the relative increase in a topic focus accounting for changes in productivity.

4.2.2 Changes in research oriented towards local problems

Our central question is: do non-migrants publish more on diseases that are prevalent in their home countries after the migration of a peer from their institution? Table 3 shows results for publication counts on diseases that are high prevalence in home (and host) countries. Figure 1 provides a graphical representation of the five different sets of publications we explore in this table. We start by evaluating the change in the count of publications on high-prevalence home country diseases. We show the raw data representing the changes in publication output on locally prevalent diseases plotted in Figure 2 and the corresponding regression results in column 1 of Table 3. Consistent with our first hypothesis, we see that non-migrants are more likely to publish on diseases that are high prevalence in their home countries after the migration event (column 1). We also observe that they are more likely to publish on diseases that are high prevalence in their migrant peers' host countries (column 2). As predicted by our second hypothesis, these effects are mostly driven by publications on diseases that are high prevalence in *both* home and host countries (column 3).

Unpacking these measures reveals that non-migrants also increase their publications on diseases with high prevalence in their home countries and *not* the host country, even if the increase is modest (column 4). However, there is no corresponding increase in research on diseases prevalent only in the host country and not in the home country (column 5). This suggests that non-migrants are intensifying their research focus on local diseases without expanding their work on diseases not prevalent in their home countries.

Figures 2 and 3 illustrate the temporal nature of these changes in research focus. The figures both reveal that non-migrants make a relatively quick change in research focus towards home country diseases following the migration of a peer (about two years after the migration event, we see a noticeable change), and this effect is sustained (although perhaps with slightly less intensity) until 10 years after the migration event.

We consider how some of our measures may contribute to the results. Appendix Table A2 shows our results using alternative measures of disease prevalence. We also consider different ways of thinking about topic overlap between countries; Appendix Table A3 looks at migrants who move within the same continent and between countries that have a high proportion of shared diseases. We see that even using these alternate measures, the propensity at which non-migrants increase their relative focus on local, and particularly on shared, problems goes up.

In Appendix Table A4, we show that this increased focus on local topics is exclusively driven by non-migrants' increased publications in high prevalence neglected tropical diseases (NTDs). In fact, there is no effect on publications on non-NTD topics. Regarding the significance of the diseases studied, we demonstrate in Appendix Table A5 that the diseases researched by non-migrants have a high prevalence across several low-income countries.[19] These diseases represent significant grand challenges to which non-migrants in the Global South are directly exposed, giving them a unique comparative advantage in studying these issues.[20]

[19] In Appendix Table A5 we identify whether the diseases they research are global diseases (i.e. diseases that are responsible for more than half of the world's Disability Adjusted Life Years (DALYS)), and are high prevalence across all high-income countries or low-income countries. For example, malaria is a disease that predominantly impacts low-income countries, and has high global prevalence overall, but HIV/AIDS is responsible for significant DALYS in high-income countries and is therefore considered a 'high-income' and a 'low-income' disease in the analysis.

[20] In Appendix Table A6, we also provide evidence that the change in home and host country disease focus is driven by those with prior publications in home country problems, suggesting that it is driven by those who have a pre-existing interest and a comparative advantage in studying these problems.

4.3 Potential Mechanisms

We investigate several potential mechanisms that could be driving our observed effects. Namely, migrants might collaborate with researchers at home and directly transfer new knowledge or resources that support research into local problems. Or, they could indirectly facilitate access to knowledge, resources, and collaborators relevant to local problems.

4.3.1 Direct effect

First, we consider that migrants could directly influence their home country peers' research trajectories either by collaborating with them, advising them, or acting as role models who research local topics on an international stage. In Table 4, we start by looking at whether successful fellows themselves orient their research trajectories towards home country topics (columns 1 and 2). Interestingly, we see that treated fellows are not significantly more likely to research home country diseases after they migrate.

We evaluate our peer effects separately for non-migrants based on whether or not their applicant peer publishes in home country diseases.[21] Column 3 shows these results for peers of applicants who never publish on home country diseases, and column 4 shows results for peers of applicants who do publish on home country diseases. In both of these samples, we see a similar magnitude of increase in the non-migrating peer's propensity to publish on home country diseases, which further suggests that collaboration with the applicant peer, or specific role model or advising channels, are not likely to be driving the publication outcomes we observe. Note also that collaboration with a fellow is extremely unlikely; just 0.6% of sample scientists in this sample ever have a publication with a fellow. In addition to a low rate of collaboration with the fellow themselves, we see that our sample of non-migrant scientists are not more likely to publish with fellowship recipients

[21] We identify whether fellowship applicants publish in any home country disease at any point in the 5 years before or 10 years after applying for the fellowship.

following the fellowship (column 5). In column 6, we drop non-migrants who publish with the fellow from our sample and our results remain robust.

Together, we fail to find strong evidence that indicates that non-migrants' increased research investment in local diseases is driven by their migrant peers directly. We therefore examine indirect pathways by which non-migrants' research agendas may be affected.

4.3.2 Indirect effect

Another way in which exposure to a migrant can affect a home country peer is if they facilitate connections between their host country and home country peers. As shown in Fry and Furman (2023), fellowships of this nature can increase the rate at which home country peers collaborate with the migrant's new peers in the host country. Given this, we examine the coauthorship patterns with host country researchers to understand the extent to which access to collaborators could be driving the changes we see in non-migrants' research agendas.

Our results are presented in Table 5. In columns 1 and 2, and consistent with prior research in this area, we see that after being exposed to a migrating peer, non-migrants are more likely to coauthor on any subject with a researcher from the host institution and host country as compared to the control non-migrants. To further investigate this collaboration mechanism, we examine the content of publications coauthored by non-migrants and host country scientists. We evaluate whether these publications are on high prevalence home country diseases (column 3), non-home country diseases (column 4), shared home and host diseases (column 5), and diseases that are prevalent only in the home country and not host country (column 6). We see that coauthored publications focus predominantly on diseases that are prevalent in *both* home and host countries, and that there is no effect on research on diseases that are present in one environment but not the other. These results indicate that the migration event creates opportunities for home

and host peers to invest in areas of common interest and relevance.

Lastly, migrants can have an impact by facilitating the flow of knowledge between countries. Several studies have documented that migrants play this role (Wang, 2015; Kahn and MacGarvie, 2016b). It is plausible that an increase in access to new knowledge, particularly pertaining to problems shared between the home and host countries, could influence a non-migrants' research orientation.

Specifically, we are interested in evaluating if after treatment, non-migrating peers' research begins to more closely resemble the work being done in the host institution. We investigate this by testing the non-migrants change in use of key words associated with publications authored by individuals in the host institution before and after the migration event. Table 6 shows these results. In columns 1-4, we see that treated non-migrants use a higher proportion of host institution keywords in all of their publications after the migration event. Then, we focus specifically on publications on home country diseases (column 5) and both home and host country diseases (column 6). We see that the increase in use of host country keywords is both greater in magnitude and more significant in these contexts. That is, non-migrating peers are more likely to use host institution keywords in their publications on both local and shared problems after exposure to a migrating peer.[22]

Together, this analysis provides suggestive evidence that non-migrants' increased investment in local research following the migration of a peer from their institution is driven at least in part by gaining access to connections and knowledge in the context of shared problems.

4.4 Alternative Explanations

To ensure that the increase we observe in non-migrants peers' relative focus on problems shared by the home and the host country is not driven by alternative explanations,

[22]Note that for this analysis it is difficult to explore the use of keywords in non-home country publications given the relative dearth of those publications and our use of a single post-treatment measure which drastically reduces the sample size available for analysis.

we conduct several robustness tests.

First, to account for the possibility that there are person-disease trends or disease-specific trends in research that might be driving our results, we conduct our analysis at the person × year × disease level in Table 7. Note that this analysis includes scientist × disease fixed effects as well as disease × year fixed-effects and scientist × year fixed effects. The scientist × disease fixed effects remove time-invariant differences in the extent to which scientists research on particular diseases. Disease × year fixed effects remove the possibility that any trends observed are driven by general changes in the level of research into given diseases over the observed time period. Scientist × year fixed effects account for any changes in productivity of scientists over years which could generate changes in levels of research in given diseases. As we see in Table 7, our results are robust to this alternative data structure. In fact, the coefficient on the diseases prevalent at home and not in the host country is close to zero using this specification, suggesting that the small effect picked up in the baseline specification was due to one of these alternative channels (such as an increase in research more generally into those diseases).

Second, we explore the possibility that there are overlapping features of home and host countries other than shared problems that are driving our results. For example, are publications on shared topics increasing because home and host country researchers speak the same language or are more geographically proximate? In Appendix Table A7, we look at shared topic publication outcomes, splitting our sample of non-migrant fellows with and without shared language and on the same and different continents. We also interact shared continent and shared language with our post × treated variable to control for the effect of shared continent and shared language on publications in shared diseases. In all specifications, the effect on shared topic publications remains positive and significant, implying that home and host similarity on dimensions other than disease prevalence is not driving the observed effect.

Third, we evaluate the possibility that the observed effect is driven by OWSD fel-

lows who are traveling from institutions with a greater focus on shared home and host diseases. In other words: are fellows traveling to countries with similar problems more likely to be from institutions with a focus on those problems, and is that driving our effects? We test this in a few ways in Appendix Table A8. First, we drop the minority of non-migrants who had coauthored publications with anyone in the host country prior to the migration event, and second, we split the sample of non-migrants into those in an institution with a high focus on home and host diseases, and the remainder. We fail to find any evidence that institutions with prior relationships with the host country or that non-migrants in institutions with high home and host disease focus are driving the observed effects. Further, we run the full analysis excluding migrants who travel to South Africa, the most common host country in our sample in Appendix Table A9. Our results remain robust.

Fourth, we compare peers of OWSD fellows who had “sandwich" fellowships for just to those who had longer, full-time PhD fellowships. Of our sample, 36% apply for full-time fellowships. Presumably, the sandwich fellows had less time in which to build relationships and connect their home and host country peers. In Table A10, we see that our results are meaningfully smaller for sandwich fellows. This suggests that the intensity of the connection to the host institution does matter for the outcomes we find. While not the focus of the paper, in a reduced sample of non-migrants for which we could obtain gender data, we show that the female non-migrants are more likely to increase their local research focus following the migration event (Appendix Table A11). Given the potential for homophily, this is likely to be the group of non-migrants most affected by the departure of a female fellow from their institution, thus providing further evidence that the observed effects are less likely to be driven by institution-specific characteristics, and seem to be driven by the migration event itself.

Fifth, it could be that non-migrating peers are more likely to study problems that are of high prevalence globally (for example, HIV) and we are inaccurately characterizing

these as shared home and host problems as opposed to problems that are shared across a variety of countries because of their global prevalence. In addition to dropping NTDs with higher global prevalence and attention (malaria, HIV and TB) in Appendix Table A4, in columns 2-4 of Table A12, we limit only to diseases with high global prevalence. Namely, we want to test whether non-migrants are orienting their efforts towards diseases that are high prevalence in general, not necessarily specific to the home and host country. We see that non-migrants are actually no more likely to focus on high global prevalence diseases that are also high prevalence in home not host countries, host only, and diseases that are not high prevalence in the home country, after the migration event. This result suggests that we are observing an increase in focus towards diseases specifically prevalent in the home and host countries, as opposed to towards diseases that are important globally.

We also test whether our effects are mechanical by randomly assigning treated peers to a placebo host country. Appendix Table A13 presents these results. The results of this counterfactual exercise are directionally similar but meaningfully smaller (2% of the size) of the effects we find in our main analysis.[23].

Lastly, we are proposing that the effects of migrant brokerage are indirect, but what if the non-migrant also migrates and that is driving our observed effects. We evaluate this concern in Appendix Table A14 by excluding migrating non-migrants from the sample (i.e. any non-migrant who is affiliated with another institution in another country at any point in the post-fellowship period). The results remain similar using this alternative sample.

Overall, these analyses show that even accounting for additional institution and disease-level observable characteristics, the effects we find on non-migrating peers' research trajectories are robust.

[23]Of note, many of the placebo host countries disease prevalence overlaps with actual home countries and so in column 2 we drop those observations, and find a null effect of migration to this placebo host on home and host disease research focus

4.5 Crop and Pest Results

To complement our analysis of disease publications and to check whether the results we find are unique to the disease setting, we evaluate publication patterns on crops and pests.[24] This analysis is less granular as our data on crop and pest prevalence is binary and time invariant; there is a dummy variable that takes the value of 1 if a specific crop or pest has appeared a country, 0 otherwise, but we do not know when this measure was collected or how prevalent in a continuous manner the crop or pest is.

We conduct our baseline analysis with this data in an identical manner to that conducted on the disease outcomes and explore if there are changes in the frequency with which agriculture research focused non-migrants publish on local or global crops and pests. In Appendix Table A15, we see that non-migrants are more likely to publish on crops and pests that are present in their home countries and on those that are present in their peer's host countries after the migration event. However, this effect is not significant when we focus only on crops and pests that are present in home or host countries (columns 4 and 5). These findings mirror our disease results.

5 Discussion

In this paper, we explore the extent to which the migration of a female scientist from a developing country influences the research agendas of their non-migrating peers. Specifically, we ask whether the non-migrant scientists at the home institution change in their propensity to research more local problems in the context of disease research. We document that following the migration of a peer, home scientists are more productive. We find that this effect is driven by an increase in non-migrants' propensities to publish papers on locally-prevalent diseases. This appears to be driven specifically by diseases shared by the home and host countries. We explore mechanisms and find suggestive evidence

[24]This analysis is discussed in greater detail in Appendix B.

that an increase in collaborations with host country scientists and an increase in access to knowledge from the host institution drive our observed changes in research trajectory.

We contribute to two main literatures. First, we contribute to a growing literature on the role of migration in organizational innovation outcomes (Saxenian, 2002, 2005, 2007; Agrawal et al., 2006; Kerr, 2008; Oettl and Agrawal, 2008; Agrawal et al., 2011b; Kerr, 2018; Kahn and MacGarvie, 2016b; Fry and Furman, 2023; Choudhury and Kim, 2019; Choudhury, 2016). We build on Choudhury and Kim (2019) who find that ethnic migrant inventors bring knowledge from their home regions to firms they move to in the U.S., and this knowledge is both used directly and recombined in patents received by their U.S. employers. Our results extend this work by demonstrating that by facilitating the flow of knowledge, connections and resources back to their home regions —especially from countries facing similar issues— migrants can shape the research orientation of their peers back at home. By highlighting the important role of shared problems in network outcomes, we propose that the impact of migrant brokerage depends on the characteristics of the host locations.

Second, our results contribute to literature on the direction of innovation. Prior literature has explored drivers such as funding availability, exposure to problems, access to tools and data and access to collaborators (Kremer and Glennerster, 2004; Kyle and McGahan, 2012; Catalini, 2018; Catalini et al., 2020; Furman and Teodoridis, 2020; Myers, 2020; Nagaraj et al., 2020; Nagaraj, 2022; Nagaraj and Tranchero, 2023; Koning et al., 2021; Moscona and Sastry, 2022; Andrews and Smith, 2023; Fry, 2023b; Truffa and Wong, 2023). We build on this work and provide evidence on a novel mechanism: migration of a peer. This is particularly relevant for those based in the Global South, with limited access to opportunities, knowledge and collaborators, and for whom local problems represent particularly salient and under-researched challenges.

This study has implications for organizational-level decision making and is particularly relevant for managers, organizations, entrepreneurs, and policy makers seeking to

foster innovation. Our findings highlight the potential for migrants to indirectly affect the trajectory of their peers' innovative work through acting as brokers across borders. These results underscore how even the temporary migration of knowledge workers can enhance and direct innovation on pressing problems.

This paper is not without limitations. First, we are not able to establish the optimal allocation of attention towards research topics. Future work should seek to understand the macro-effects of this change in research focus and its impact on health and economic outcomes. Second, we measure effects on incumbent scientists which could give us a lower bound of these estimates. Future research should explore impacts on future generations in home institutions. Third, we use publication data which has limitations in that we are unable to observe individual contribution to team publications, effort exerted, and scientific work that does not result in publication outcomes captured in international databases. Lastly, we examine a short-term migration program in which the majority of migrants either remain affiliated with their home institution or return home following full-time migration. Future research should seek to understand variation in outcomes based on migrant time abroad and the nature of the migration event itself.

That said, we believe that there are important implications for both theory and practice related to capturing benefits from outbound migration and improving and directing innovative capacity. More broadly, our work suggest that building ties across distributed networks, and particularly across parts of the network with shared problems, is one way that firms, organizations and nations can advance research and build their competitive advantage.

References

Acemoglu, D. and Zilibotti, F. (2001). Productivity Differences*. *The Quarterly Journal of Economics*, 116(2):563–606.

Agrawal, A., Cockburn, I., and McHale, J. (2006). Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography*, 6(5):571–591.

Agrawal, A., Kapur, D., McHale, J., and Oettl, A. (2011a). Brain drain or brain bank? the impact of skilled emigration on poor-country innovation. *Journal of Urban Economics*, 69(1):43–55.

Agrawal, A., Kapur, D., McHale, J., and Oettl, A. (2011b). Brain drain or brain bank? The impact of skilled emigration on poor-country innovation. *Journal of Urban Economics*, 69(1):43–55.

Andrews, M. and Smith, A. (2023). Do local conditions determine the direction of science? evidence from land grant colleges. *Evidence from Land Grant Colleges (September 28, 2023)*.

Apicella, C., Norenzayan, A., and Henrich, J. (2020). Beyond weird: A review of the last decade and a look ahead to the global laboratory of the future.

Azoulay, P., Jones, B. F., Kim, J. D., and Miranda, J. (2022). Immigration and Entrepreneurship in the United States. *American Economic Review: Insights*, 4(1):71–88.

Balachandran, S. and Eklund, J. (2024). The Impact of Partner Organizational Structure on Innovation. *Administrative Science Quarterly*, 69(1):80–130. Publisher: SAGE Publications Inc.

Balachandran, S. and Hernandez, E. (2021). Mi casa es tu casa: immigrant entrepreneurs as pathways to foreign venture capital investments. *Strategic Management Journal*, 42(11):2047–2083.

Borjas, G. J. and Doran, K. B. (2012). The Collapse of the Soviet Union and the Productivity of American Mathematicians*. *The Quarterly Journal of Economics*, 127(3):1143–1203.

Borusyak, K., Jaravel, X., and Spiess, J. (2022). Revisiting Event Study Designs: Robust and Efficient Estimation. Technical Report 2108.12419, arXiv.org. Publication Title: Papers.

Catalini, C. (2018). Microgeography and the direction of inventive activity. *Management Science*, 64(9):4348–4364.

Catalini, C., Fons-Rosen, C., and Gaulé, P. (2020). How do travel costs shape collaboration? *Management Science*, 66(8):3340–3360.

Choudhury, P. (2016). 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(3):585–610.

Choudhury, P. and Kim, D. Y. (2019). The ethnic migrant inventor effect: Codification and recombination of knowledge across borders. *Strategic Management Journal*, 40(2):203–229. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/smj.2977.

Ditta, A. S., Strickland-Hughes, C. M., Cheung, C., and Wu, R. (2020). Exposure to information increases motivation to learn more. *Learning and Motivation*, 72:101668.

Einiö, E., Feng, J., and Jaravel, X. (2019). Social Push and the Direction of Innovation. SSRN Scholarly Paper ID 3383703, Social Science Research Network, Rochester, NY.

Fink, A., Koschutnig, K., Benedek, M., Reishofer, G., Ischebeck, A., Weiss, E. M., and Ebner, F. (2012). Stimulating creativity via the exposure to other people's ideas. *Human Brain Mapping*, 33(11):2603–2610. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.21387.

Fleming, L. (2001). Recombinant Uncertainty in Technological Search. *Management Science*, 47(1):117–132. Publisher: INFORMS.

Fry, C. and Furman, J. L. (2023). Migration and Global Network Formation: Evidence from Female Scientists in Developing Countries. *Organization Science*. Publisher: INFORMS.

Fry, C. and Ganguli, I. (2023). Return on Returns: Building Scientific Capacity in AIDS Endemic Countries.

Fry, C. V. (2023a). Bridging the Gap: Evidence from the Return Migration of African Scientists. *Organization Science*, 34(1):404–432. Publisher: INFORMS.

Fry, C. V. (2023b). Crisis and the trajectory of science: Evidence from the 2014 ebola outbreak. *Review of Economics and Statistics*, 105(4):1028–1038.

Fry, C. V. (2023c). Crisis and the Trajectory of Science: Evidence from the 2014 Ebola Outbreak. *The Review of Economics and Statistics*, 105(4):1028–1038.

Fry, C. V. (2023d). Local researchers, local problems. *Local Problems (June 14, 2023)*.

Furman, J. L. and Teodoridis, F. (2020). Automation, Research Technology, and Researchers' Trajectories: Evidence from Computer Science and Electrical Engineering. *Organization Science*, 31(2):330–354. Publisher: INFORMS.

Ganguli, I. (2015). Immigration and Ideas: What Did Russian Scientists "Bring" to the United States? *Journal of Labor Economics*, 33(S1):S257–S288. Publisher: The University of Chicago Press.

Gardner, J. (2022). Two-stage differences in differences. arXiv:2207.05943 [econ].

Garfield, M. J., Taylor, N. J., Dennis, A. R., and Satzinger, J. W. (2001). Research Report: Modifying Paradigms—Individual Differences, Creativity Techniques, and Exposure to Ideas in Group Idea Generation. *Information Systems Research*, 12(3):322–333. Publisher: INFORMS.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

Hernandez, E. and Kulchina, E. (2020). Immigrants and foreign firm performance. *Organization Science*, 31(4):797–820.

Jackson, M. O. and Wolinsky, A. (1996). A Strategic Model of Social and Economic Networks. *Journal of Economic Theory*, 71(1):44–74.

Kahn, S. and MacGarvie, M. (2016a). Do return requirements increase international knowledge diffusion? evidence from the fulbright program. *Research Policy*, 45(6):1304–1322.

Kahn, S. and MacGarvie, M. J. (2016b). How Important Is U.S. Location for Research in Science? *The Review of Economics and Statistics*, 98(2):397–414.

Kerr, W. R. (2008). Ethnic Scientific Communities and International Technology Diffusion. *The Review of Economics and Statistics*, 90(3):518–537.

Kerr, W. R. (2018). *The Gift of Global Talent: How Migration Shapes Business, Economy & Society*. Stanford University Press, Stanford.

Koning, R., Samila, S., and Ferguson, J.-P. (2021). Who do we invent for? Patents by women focus more on women's health, but few women get to invent. *Science*, 372(6548):1345–1348. Publisher: American Association for the Advancement of Science Section: Report.

Kremer, M. and Glennerster, R. (2004). *Strong medicine: creating incentives for pharmaceutical research on neglected diseases*. Princeton University Press.

Kyle, M. K. and McGahan, A. M. (2012). Investments in pharmaceuticals before and after trips. *Review of Economics and Statistics*, 94(4):1157–1172.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3):531–542. Publisher: [Oxford University Press, Review of Economic Studies, Ltd.].

Moscona, J. and Sastry, K. (2022). Inappropriate Technology: Evidence from Global Agriculture.

Myers, K. (2020). The Elasticity of Science. *American Economic Journal: Applied Economics*, 12(4):103–134.

Nagaraj, A. (2022). The Private Impact of Public Data: Landsat Satellite Maps Increased Gold Discoveries and Encouraged Entry. *Management Science*, 68(1):564–582. Publisher: INFORMS.

Nagaraj, A., Shears, E., and de Vaan, M. (2020). Improving data access democratizes and diversifies science. *Proceedings of the National Academy of Sciences*, 117(38):23490–23498. Publisher: Proceedings of the National Academy of Sciences.

Nagaraj, A. and Tranchero, M. (2023). How Does Data Access Shape Science? Evidence from the Impact of U.S. Census's Research Data Centers on Economics Research.

Nagaraj, A. and Yao, H. (2023). Geographic bias in management scholarship: Data-driven estimates and trends.

Nanda, R. and Khanna, T. (2010). Diasporas and domestic entrepreneurs: Evidence from the indian software industry. *Journal of Economics & Management Strategy*, 19(4):991–1012.

Oettl, A. and Agrawal, A. (2008). International labor mobility and knowledge flow externalities. *Journal of International Business Studies*, 39(8):1242–1260. Publisher: Palgrave Macmillan.

Quadrio-Curzio, A., Blowers, T., and Thomson, J. (2020). Women, science and development: The leading role of OWSD. *Economia Politica*, 37(1):1–12.

Saxenian, A. (2002). Silicon Valley's New Immigrant High-Growth Entrepreneurs. *Economic Development Quarterly*, 16(1):20–31. Publisher: SAGE Publications Inc.

Saxenian, A. (2005). From Brain Drain to Brain Circulation: Transnational Communities and Regional Upgrading in India and China. *Studies in Comparative International Development*, 40(2):35–61.

Saxenian, A. (2007). *The New Argonauts: Regional Advantage in a Global Economy*. Harvard University Press, Cambridge, Mass., first thus edition edition.

Truffa, F. and Wong, A. (2023). Undergraduate Gender Diversity and Direction of Scientific Research.

Uzzi, B. and Spiro, J. (2005). Collaboration and Creativity: The Small World Problem. *American Journal of Sociology*, 111(2):447–504. Publisher: The University of Chicago Press.

Vakili, K. and McGahan, A. M. (2016). Health Care's Grand Challenge: Stimulating Basic Science on Diseases That Primarily Afflict the Poor. *The Academy of Management Journal*, 59(6):1917–1939. Publisher: Academy of Management.

Wang, D. (2015). Activating Cross-border Brokerage: Interorganizational Knowledge Transfer through Skilled Return Migration. *Administrative Science Quarterly*, 60(1):133–176.

Weitzman, M. L. (1998). Recombinant Growth*. *The Quarterly Journal of Economics*, 113(2):331–360.

Xie, Q. and Freeman, R. B. (2023). Creating and Connecting US and China Science: Chinese Diaspora and Returnee Researchers.

Figures and Tables

Figure 1: Home and host country overlap

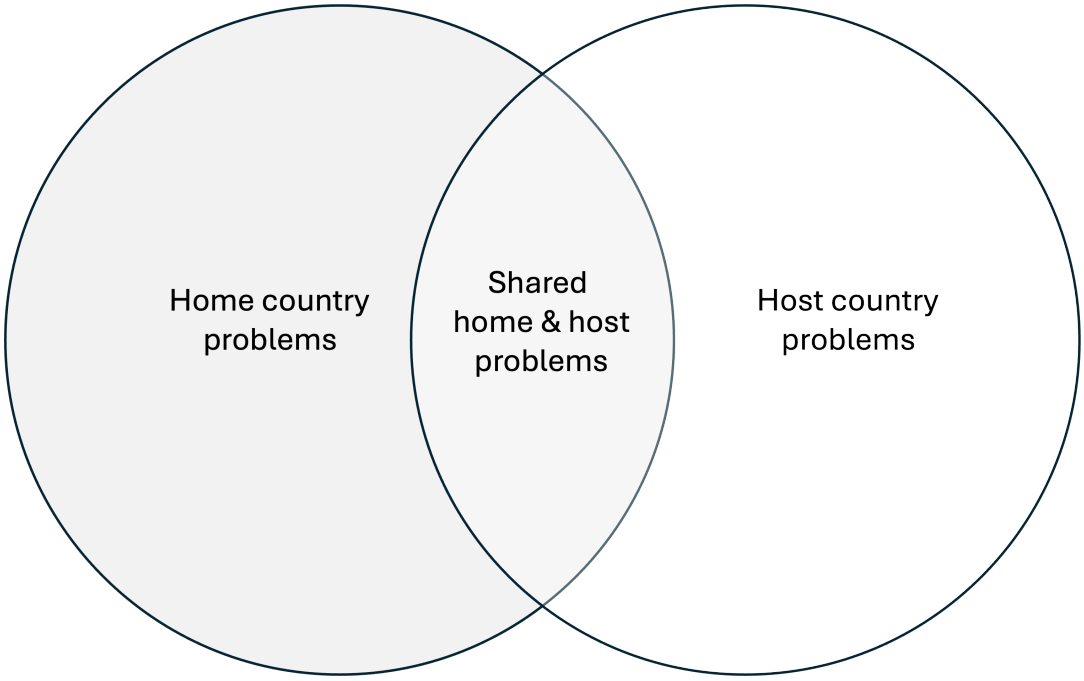
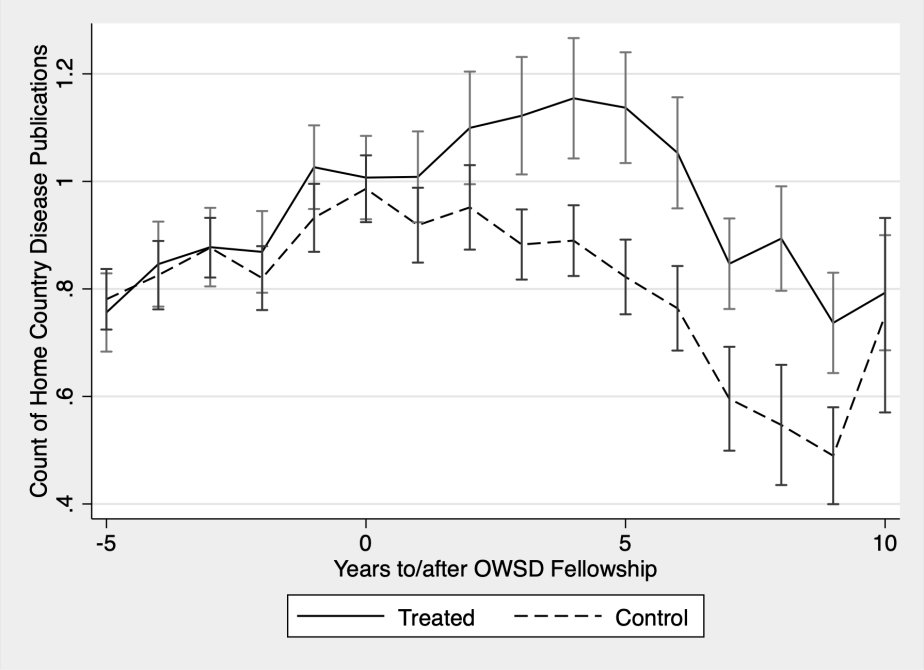
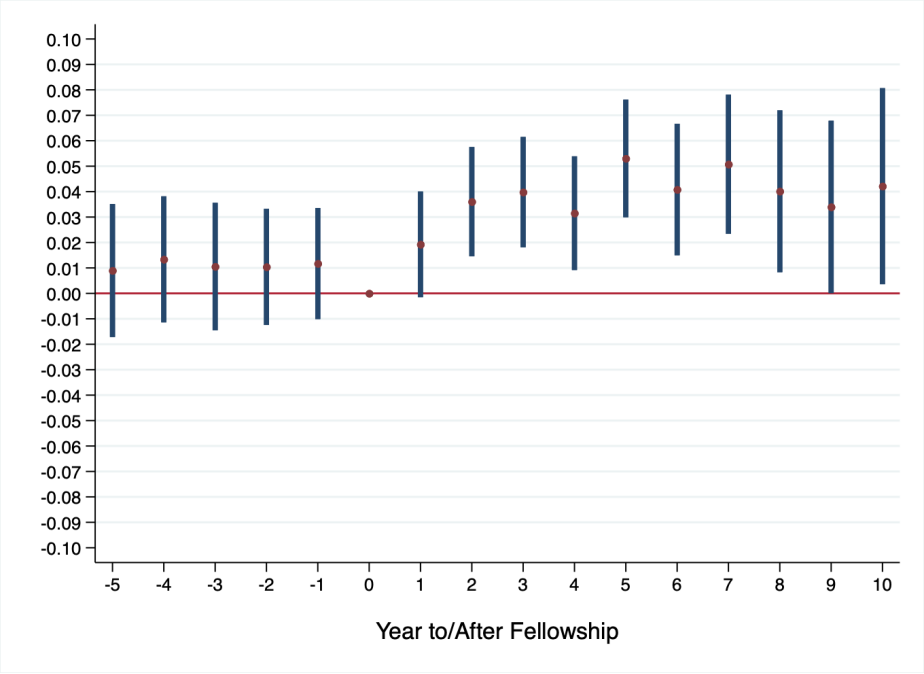


Figure 2: Changes in Home Country Disease Publication Rates



Notes: This figure plots raw trends of average publication rates on home-country diseases (per non-migrant peer scientist) for the five years before and ten years after the fellowship.

Figure 3: Differences in publication rate into home country diseases between treated and untreated non-migrant peer scientists



Notes: This figure shows coefficient estimates stemming from conditional (scientist) fixed-effects OLS specifications. The inverse hyperbolic sine outcomes are regressed onto year effects and scientist age effects, and include interaction terms between treatment status and the number of years before/after the OWSD fellowship, and a control variable representing the annual overall publication count for each scientist. The 90% confidence interval robust standard errors clustered around the scientist is plotted with solid bars.

Table 1: Summary statistics for non-migrant peer scientists (at year of application)

	(1) Control Sample	(2) Treatment Sample	(3) Difference Between Groups	(4) T-Stat
Female	0.30	0.25	0.05	3.19
	(0.46)	(0.43)		
Career Age	10.40	9.15	1.25	8.33
	(7.24)	(6.74)		
Total publications	11.92	11.06	0.86	2.79
	(14.88)	(14.05)		
Total disease publications	5.12	5.12	-0.00	-0.01
	(10.18)	(10.83)		
Total home disease publications	4.59	4.60	-0.01	-0.02
	(9.43)	(9.97)		
Total host disease publications	3.36	3.38	-0.02	-0.10
	(7.79)	(8.70)		
Total home and host disease publications	3.25	3.29	-0.04	-0.24
	(7.57)	(8.50)		
Total home only disease publications	1.34	1.31	0.04	0.44
	(4.21)	(3.57)		
Total host only disease publications	0.11	0.09	0.02	1.55
	(0.58)	(0.77)		
Total count of collaborative publications with host country researchers	0.09	0.11	-0.03	-2.20
	(0.54)	(0.59)		
Observations	5256	3667	8923	

Notes: This table depicts summary statistics for our treatment and control samples. All productivity statistics are generated for the 5 years prior to treatment. Columns 1 and 2 show averages and standard errors for our control and treatment groups, respectively. Column 3 shows the raw difference in means between control and treatment groups, and column 4 shows the t-value resulting from a t-test for the treatment and control samples. The t-test compares the control and treatment group means and assumes a null hypothesis that the two means are equal. If the t-statistic is greater than 1.96, we can reject the null hypothesis that there is no difference between means at a 5% level of significance.

Table 2: Changes in overall publication output of non-migrating peers after OWSD fellowship

	Number of Publications	Number of Disease Publications	Number of Non-Disease Publications	Number of NTD Disease Publications	Number of Non-NTD Disease Publications
	(1)	(2)	(3)	(4)	(5)
After Fellowship × Successful Applicant	0.0634***	0.0223***	-0.0125**	0.0212***	0.0058
	(0.015)	(0.007)	(0.006)	(0.006)	(0.006)
Total Observations	107576	107576	107576	107576	107576
Mean of Dep. Variable	2.2171	0.7643	1.4528	0.4317	0.3326
Author FE	X	X	X	X	X
Year FE	X	X	X	X	X
Career Age FE	X	X	X	X	X
Time Since Grant FE	X	X	X	X	X
Annual Pubs		X	X	X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts. [b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Table 3: Changes in disease topic publications of non-migrating peers after OWSD fellowship

	Number of Home Country Publications	Number of Host Country Publications	Number of Home and Host Country Publications	Number of Home not Host Country Publications	Number of Host not Home Country Publications
	(1)	(2)	(3)	(4)	(5)
After Fellowship × Successful Applicant	0.0285***	0.0217***	0.0210***	0.0150***	-0.0004
	(0.008)	(0.007)	(0.007)	(0.005)	(0.002)
Total Observations	107576	107576	107576	107576	107576
Mean of Dep. Variable	0.9111	0.6835	0.6636	0.2474	0.0198
Author FE	X	X	X	X	X
Year FE	X	X	X	X	X
Career Age FE	X	X	X	X	X
Time Since Grant FE	X	X	X	X	X
Annual Pubs	X	X	X	X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts.[b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Table 4: Impact of fellow's research focus on non-migrating peer's research trajectory

	Number of Home Country Publications (fellow sample)		Number of Home Country Publications (non-migrant peer sample fellow not home country disease researcher)	Number of Home Country Publications (non-migrant peer sample fellow home country disease researcher)	Number of publications with fellow (non-migrant peer sample)	Number of Home Country Publications (non-migrant peer sample excluding those who ever coauthor with fellow)
	(1)	(2)	(3)	(4)	(5)	(6)
After Fellowship × Successful Applicant	0.1051	0.0086	0.0307***	0.0318***	0.0004	0.0289***
	(0.108)	(0.080)	(0.011)	(0.012)	(0.000)	(0.008)
Total Observations	240	240	69606	37970	107576	107178
Mean of Dep. Variable	0.4344	0.4344	0.9379	0.8620	0.0005	0.9064
Author FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Career Age FE	X	X	X	X	X	X
Time Since Grant FE	X	X	X	X	X	X
Annual Pubs		X	X	X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts. [b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Table 5: Changes in host country collaborations of non-migrating peers after OWSD fellowship

	Number of Publications Coauthored with Host Institution	Number of Publications Coauthored with Host Country	Number of Home Country Publications Coauthored with Host Country	Number of Non-Home Country Publications Coauthored with Host Country	Number of Home and Host Country Publications Coauthored with Host Country	Number of Home not Host Country Publications Coauthored with Host Country
	(1)	(2)	(3)	(4)	(5)	(6)
After Fellowship × Successful Applicant	0.0034**	0.0119***	0.0117***	0.0022	0.0116***	0.0004
	(0.001)	(0.004)	(0.003)	(0.003)	(0.003)	(0.000)
Total Observations	107576	107576	107576	107576	107576	107576
Mean of Dep. Variable	0.0058	0.0311	0.0307	0.0298	0.0307	0.0307
Author FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Career Age FE	X	X	X	X	X	X
Time Since Grant FE	X	X	X	X	X	X
Annual Pubs	X	X	X	X	X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts. [b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Table 6: Changes in use of host institution keywords of non-migrating peers after OWSD fellowship

	Proportion of Host Institution Keywords Used				Proportion of Host Institution Keywords Used in Home Country Disease Publications	Proportion of Host Institution Keywords Used in Home and Host Country Disease Publications
	(1)	(2)	(3)	(4)	(5)	(6)
After Fellowship × Successful Applicant	0.0062**	0.0061**	0.0055*	0.0056*	0.0068**	0.0066**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Total Observations	12254	12254	12254	12254	11872	11588
Mean of Dep. Variable	0.2779	0.2779	0.2779	0.2779	0.2759	0.2771
Author FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Career Age FE	X	X	X	X	X	X
Annual Pubs		X		X	X	X
Annual Keywords			X	X	X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed counts of the number of host affiliation keywords in publications pre-migration event in the migrant's field of study that are used by non-migrating peers before and after the migration event, divided by the overall number of keywords in the non-migrating peers publications before and after the migration event. There is one observation per sample scientist (pre and post migration event) which combines all keywords used in publications in these two time periods. The sample includes only non-migrating peers for whom host institution publications could be obtained. [b] Heteroskedastic robust standard errors are given in parentheses.

Table 7: Changes in disease topic publications of non-migrating peers after OWSD fellowship at the peer, year disease level

	Number of publications				
	(1)	(2)	(3)	(4)	(5)
After Fellowship × Successful Applicant × High Home Prevalence Disease	0.0011***				
	(0.000)				
After Fellowship × Successful Applicant × High Host Prevalence Disease		0.0008***			
		(0.000)			
After Fellowship × Successful Applicant × High Prevalence Home and Host Disease			0.0010***		
			(0.000)		
After Fellowship × Successful Applicant × High Prevalence Home Not Host Disease				0.0003	
				(0.000)	
After Fellowship × Successful Applicant × High Prevalence Host Not Home Disease					-0.0005**
					(0.000)
Total Observations	16781856	16781856	16781856	16781856	16781856
Mean of Dep. Variable	0.0065	0.0065	0.0065	0.0065	0.0065
Author Year FE	X	X	X	X	X
Disease Year FE	X	X	X	X	X
Author Disease FE	X	X	X	X	X
Career Age FE	X	X	X	X	X
Time Since Grant FE	X	X	X	X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts of publications in a given disease of the specific type (e.g. high home prevalence) in a given year for the focal scientist. [b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Online Appendices for "Migration and Local Problem-Solving"

Appendix A: Appendix Figures and Tables

Table A1: Home and Host Countries

Home Country	Home Count	Percent	Host Country	Host Count	Award Granted	No Award
Bangladesh	214	2.4	Argentina	97	97	0
Benin	112	1.26	Benin	175	175	0
Cameroon	1,373	15.39	Botswana	26	10	16
Ethiopia	309	3.46	Brazil	95	44	51
Ghana	129	1.45	China	1977	959	1,018
Kenya	434	4.86	Egypt	33	33	0
Madagascar	31	0.35	Ghana	117	58	59
Malawi	3	0.03	India	652	523	129
Nepal	7	0.08	Kenya	172	93	79
Nigeria	5,279	59.16	Malaysia	828	781	47
Rwanda	1	0.01	Morocco	2	2	0
Senegal	29	0.33	Nigeria	67	59	8
South Africa	404	4.53	Pakistan	522	495	27
Sudan	82	0.92	Senegal	71	0	71
Tanzania	57	0.64	Singapore	9	9	0
Uganda	121	1.36	South Africa	3839	1,785	2,054
Yemen	6	0.07	Tanzania	6	1	5
Zambia	9	0.1	Thailand	17	0	17
Zimbabwe	321	3.6	The Philippines	128	44	84
Eswatini	2	0.02	Uganda	90	88	2
Total	8,923	100		8923	5,256	3,667

Notes: This table shows the frequency of OWSD applicants and peers from our set of home and host countries. The Home Count column captures the number of peer $\times$ applicant observations from each home country while Host Count captures the number of peer $\times$ applicant observations associated with each home country. The Award Granted column shows how many of these observations are affiliated with an applicant who received the fellowship, while No Award counts the number of observations affiliated with an unsuccessful applicant.

Table A2: Alternative High Home Prevalence Measures

	Number of Publications on Diseases Above Median Prevalence	Number of Publications on Diseases With Any Prevalence	Number of Publications on Diseases with Above 1 Percent Prevalence	Number of Publications on Diseases with Above 0.1 Percent Prevalence
	(1)	(2)	(3)	(4)
After Fellowship × Successful Applicant	0.0285***	0.0233***	0.0176**	0.0283***
	(0.008)	(0.008)	(0.007)	(0.008)
Total Observations	107576	107576	107576	107576
Mean of Dep. Variable	0.9111	1.0070	0.5166	0.8986
Author FE	X	X	X	X
Year FE	X	X	X	X
Career Age FE	X	X	X	X
Time Since Grant FE	X	X	X	X
Annual Pubs	X	X	X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts. [b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Table A3: Alternative High Home Problem Overlap Measures

	Number of Publications	Number of Home Country Publications	Number of Home and Host Country Publications	Number of Publications	Number of Home Country Publications	Number of Home and Host Country Publications
	(1)	(2)	(3)	(4)	(5)	(6)
After Fellowship × Successful Applicant	0.0723***	0.0118	0.0059	0.1274**	-0.0270	-0.0258
	(0.020)	(0.011)	(0.011)	(0.051)	(0.025)	(0.025)
After Fellowship × Successful Applicant × Home Host Same Continent	0.0083	0.0264*	0.0285*			
	(0.029)	(0.015)	(0.015)			
After Fellowship × Successful Applicant × Home Host High Disease Overlap				-0.0688	0.0617**	0.0518**
				(0.054)	(0.027)	(0.026)
Total Observations	107576	107576	107576	107576	107576	107576
Mean of Dep. Variable	2.2171	0.9111	0.3513	2.2171	0.9111	0.3513
Author FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Career Age FE	X	X	X	X	X	X
Time Since Grant FE	X	X	X	X	X	X
Annual Pubs		X	X		X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts. [b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Table A4: Changes in NTD disease topic publications of non-migrating peers after OWSD fellowship

	Number of High Prevalence Home Country NTD Disease Publications	Number of Any Prevalence Home Country NTD Disease Publications	Number of Publications on High Prevalence Home Country Non-NTD Diseases	Number of High Prevalence Home Country NTD Disease Publications Excl. HIV and TB	Number of High Prevalence Home Country NTD Disease Publications Excl. malaria
	(1)	(2)	(3)	(4)	(5)
After Fellowship × Successful Applicant	0.0280***	0.0251***	0.0089	0.0197***	0.0210***
	(0.007)	(0.007)	(0.007)	(0.005)	(0.006)
Total Observations	107576	107576	107576	107576	107576
Mean of Dep. Variable	0.4537	0.4881	0.4573	0.2421	0.2801
Author FE	X	X	X	X	X
Year FE	X	X	X	X	X
Career Age FE	X	X	X	X	X
Time Since Grant FE	X	X	X	X	X
Annual Pubs	X	X	X	X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts.[b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Table A5: Global importance of non-migrating peer disease publications

	Number of Publications on High Prevalence Global Diseases	Number of Publications on High Prevalence High-Income Country Diseases	Number of Publications on High Prevalence Low-income Country Disease
	(1)	(2)	(3)
After Fellowship × Successful Applicant	0.0192**	0.0045	0.0245***
	(0.008)	(0.007)	(0.008)
Total Observations	107576	107576	107576
Mean of Dep. Variable	0.8548	0.5278	0.8987
Author FE	X	X	X
Year FE	X	X	X
Career Age FE	X	X	X
Time Since Grant FE	X	X	X
Annual Pubs	X	X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts. [b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Table A6: Effect of non-migrating peer's prior research focus on changes in research trajectory

	Number of Publications (no prior home pubs)	Number of Home Country Publications (no prior home pubs)	Number of Home and Host Country Publications (no prior home pubs)	Number of Publications (prior home pubs)	Number of Home Country Publications (prior home pubs)	Number of Home and Host Country Publications (prior home pubs)
	(1)	(2)	(3)	(4)	(5)	(6)
After Fellowship × Successful Applicant	0.0453*	0.0142	0.0034	0.0754***	0.0293***	0.0248***
	(0.026)	(0.009)	(0.008)	(0.017)	(0.009)	(0.009)
Total Observations	32182	32182	32182	75394	75394	75394
Mean of Dep. Variable	1.6823	0.1678	0.1161	2.4454	1.2283	0.8973
Author FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Career Age FE	X	X	X	X	X	X
Time Since Grant FE	X	X	X	X	X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts. [b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Table A7: Peer, year disease level, control for same continent, language and travel time

	Number of publications	Number of publications (not shared continent)	Number of publications (shared continent)	Number of publications	Number of publications (not shared language)	Number of publications (shared language)	Number of publications
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
After Fellowship × Successful Applicant × High Prevalence Home and Host Disease	0.0011**	0.0012**	0.0012***	0.0017***	0.0017***	0.0005*	0.0020***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)
After Fellowship × Successful Applicant × High Prevalence Home and Host Disease × Home/Host Shared Continent	0.0002						-0.0012***
	(0.001)						(0.000)
After Fellowship × Successful Applicant × High Prevalence Home and Host Disease × Home/Host Shared Language				-0.0012**			-0.0003
				(0.001)			(0.001)
Total Observations	16781856	7818408	8963448	16781856	5084040	11697816	16781856
Mean of Dep. Variable	0.0065	0.0074	0.0057	0.0065	0.0088	0.0055	0.0065
Author Year FE	X	X	X	X	X	X	X
Disease Year FE	X	X	X	X	X	X	X
Author Disease FE	X	X	X	X	X	X	X
Career Age FE	X	X	X	X	X	X	X
Time Since Grant FE	X	X	X	X	X	X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts of publications in a given disease of the specific type (e.g. high home prevalence) in a given year for the focal scientist. [b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Table A8: Accounting for Existing Home and Host Similarity

	Number of Home Country Publications (excl prior host country connections)	Number of Home and Host Country Publications (excl prior host country connections)	Number of Home Country Publications (low institution prior home and host pubs)	Number of Home and Host Country Publications (low institution prior home and host pubs)	Number of Home Country Publications (high institution prior home and host pubs)	Number of Home and Host Country Publications (high institution prior home and host pubs)
	(1)	(2)	(3)	(4)	(5)	(6)
After Fellowship × Successful Applicant	0.0287***	0.0211***	0.0376***	0.0395***	0.0227**	-0.0023
	(0.008)	(0.007)	(0.012)	(0.011)	(0.010)	(0.010)
Total Observations	101686	101686	53563	53563	54013	54013
Mean of Dep. Variable	0.8587	0.6147	0.6533	0.4842	1.1667	0.8416
Author FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Career Age FE	X	X	X	X	X	X
Time Since Grant FE	X	X	X	X	X	X
Annual Pubs	X	X	X	X	X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts. [b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Table A9: Robust to Excluding South African Host Countries

	Number of Home Country Publications	Number of Host Country Publications	Number of Home and Host Country Publications	Number of Home not Host Country Publications	Number of Host not Home Country Publications
	(1)	(2)	(3)	(4)	(5)
After Fellowship × Successful Applicant	0.0252**	0.0243**	0.0212**	0.0190**	0.0031
	(0.011)	(0.010)	(0.010)	(0.008)	(0.003)
Total Observations	59743	59743	59743	59743	59743
Mean of Dep. Variable	0.9544	0.7632	0.7314	0.2229	0.0317
Author FE	X	X	X	X	X
Year FE	X	X	X	X	X
Career Age FE	X	X	X	X	X
Time Since Grant FE	X	X	X	X	X
Annual Pubs	X	X	X	X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts. [b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Table A10: Sandwich fellows versus non-sandwich fellows

	Number of Publications (not sandwich fellowship)	Number of Home Country Publications (not sandwich fellowship)	Number of Home and Host Country Publications (not sandwich fellowship)	Number of Publications (sandwich fellowship)	Number of Home Country Publications (sandwich fellowship)	Number of Home and Host Country Publications (sandwich fellowship)
	(1)	(2)	(3)	(4)	(5)	(6)
After Fellowship × Successful Applicant	0.0922***	0.0422***	0.0299**	0.0537***	0.0162*	0.0102
	(0.028)	(0.015)	(0.014)	(0.018)	(0.010)	(0.009)
Total Observations	42740	42740	42740	64836	64836	64836
Mean of Dep. Variable	2.1810	0.8789	0.3336	2.2409	0.9323	0.3629
Author FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Career Age FE	X	X	X	X	X	X
Time Since Grant FE	X	X	X	X	X	X
Annual Pubs		X	X		X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts. [b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Table A11: Effect of non-migrating peer's gender on changes in research trajectory

	Number of Publications (female non-migrant)	Number of Home Country Publications (female non-migrant)	Number of Home and Host Country Publications (female non-migrant)	Number of Publications (male non-migrant)	Number of Home Country Publications (male non-migrant)	Number of Home and Host Country Publications (male non-migrant)
	(1)	(2)	(3)	(4)	(5)	(6)
After Fellowship × Successful Applicant	0.0510	0.0703***	0.0708***	0.0514*	0.0244	0.0165
	(0.041)	(0.026)	(0.027)	(0.029)	(0.016)	(0.015)
Total Observations	10534	10534	10534	28255	28255	28255
Mean of Dep. Variable	1.9713	0.9684	0.7033	2.5306	1.0041	0.7418
Author FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Career Age FE	X	X	X	X	X	X
Time Since Grant FE	X	X	X	X	X	X
Annual Pubs		X	X		X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts. [b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Table A12: Just High Prevalence Diseases

	Number of publications				
	(1)	(2)	(3)	(4)	(5)
After Fellowship × Successful Applicant × High Prevalence Home and Host Disease	0.0010***				
	(0.000)				
After Fellowship × Successful Applicant × High Prevalence Home and Host Disease (high global prevalence disease)		0.0009***			
		(0.000)			
After Fellowship × Successful Applicant × High Prevalence Home not Host Disease (high global prevalence disease)			0.0004		
			(0.000)		
After Fellowship × Successful Applicant × High Prevalence Host not Home Disease (high global prevalence disease)				-0.0005*	
				(0.000)	
After Fellowship × Successful Applicant × High Prevalence Not Home Disease (high global prevalence disease)					-0.0008***
					(0.000)
Total Observations	16781856	16781856	16781856	16781856	16781856
Mean of Dep. Variable	0.0065	0.0065	0.0065	0.0065	0.0065
Author Year FE	X	X	X	X	X
Disease Year FE	X	X	X	X	X
Author Disease FE	X	X	X	X	X
Career Age FE	X	X	X	X	X
Time Since Grant FE	X	X	X	X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts. High global prevalence diseases are those with above median global prevalence. [b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Table A13: Placebo Host Country

	Number of publications			
	(1)	(2)	(3)	(4)
After Fellowship × Successful Applicant × High Prevalence Home and Host Disease	0.0006**	-0.0002		
	(0.000)	(0.001)		
After Fellowship × Successful Applicant × High Prevalence Home not Host Disease			0.0009***	
			(0.000)	
After Fellowship × Successful Applicant × High Prevalence Host not Home				-0.0004***
				(0.000)
Total Observations	16781856	2295540	16781856	16781856
Mean of Dep. Variable	0.0065	0.0061	0.0065	0.0065
Author Year FE	X	X	X	X
Disease Year FE	X	X	X	X
Author Disease FE	X	X	X	X
Career Age FE	X	X	X	X
Time Since Grant FE	X	X	X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts. Each sample scientist is randomly assigned a placebo host country and home and host diseases are identified using the placebo host. In column 2 we drop person, disease, year observations for whom the placebo host disease prevalence is identical to the actual host disease prevalence. [b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Table A14: Excluding Migrating 'Non-Migrants'

	Number of Home Country Publications	Number of Host Country Publications	Number of Home and Host Country Publications	Number of Home not Host Country Publications	Number of Host not Home Country Publications
	(1)	(2)	(3)	(4)	(5)
After Fellowship × Successful Applicant	0.0220**	0.0147*	0.0141*	0.0113*	-0.0006
	(0.009)	(0.008)	(0.008)	(0.006)	(0.002)
Total Observations	81791	81791	81791	81791	81791
Mean of Dep. Variable	0.8643	0.6567	0.6362	0.2281	0.0205
Author FE	X	X	X	X	X
Year FE	X	X	X	X	X
Career Age FE	X	X	X	X	X
Time Since Grant FE	X	X	X	X	X
Annual Pubs	X	X	X	X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts. [b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Table A15: Changes in crop and pest topic publications of non-migrating peers after OWSD fellowship

	Number of Home Country Publications	Number of Host Country Publications	Number of Home and Host Country Publications	Number of Home not Host Country Publications	Number of Host not Home Country Publications
	(1)	(2)	(3)	(4)	(5)
After Fellowship × Successful Applicant	0.0169**	0.0137**	0.0125**	0.0062	0.0012
	(0.007)	(0.007)	(0.006)	(0.004)	(0.004)
Total Observations	135979	135979	135979	135979	135979
Mean of Dep. Variable	0.3960	0.3821	0.2889	0.1071	0.0932
Author FE	X	X	X	X	X
Year FE	X	X	X	X	X
Career Age FE	X	X	X	X	X
Time Since Grant FE	X	X	X	X	X
Annual Pubs	X	X	X	X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts. [b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Table A16: Robust to Excluding Non-Migrants Exposed to Multiple Applicants

	Single Exposure	Two Exposures	Three Exposures	More than Three Exposures	More than Five Exposures
	(1)	(2)	(3)	(4)	(5)
DV: Number of home country publications					
After Fellowship × Successful Applicant	0.0319**	0.0335*	0.0069	0.0233*	0.0305**
	(0.013)	(0.018)	(0.023)	(0.013)	(0.014)
Total Observations	39366	20124	10362	37723	23008
Mean of Dep. Variable	0.8795	0.7701	0.8169	1.0451	1.1009
Author FE	X	X	X	X	X
Year FE	X	X	X	X	X
Career Age FE	X	X	X	X	X
Time Since Grant FE	X	X	X	X	X
Annual Pubs	X	X	X	X	X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: [a] Estimates stem from fixed-effects OLS regression specifications in which dependent variables are inverse hyperbolic sine transformed annual counts. [b] Heteroskedastic robust standard errors, clustered at the level of the scientist, are given in parentheses.

Appendix B: Alternative Measures and Empirical Approaches

0.1 Alternative Measures: Crop and Pest Analysis

We validate our main results using data on crop and pest prevalence. In a manner identical to the disease identification, we use key word searches of the title and abstract of the publications to identify whether each publication covers a crop or pest topic. We next merge the set of all sample publications to data on country-specific crop/pest prevalence in using the Centre for Agriculture and Bioscience International's (CABI) Crop Protection Compendium (CPC). This database provides comprehensive information on crops, pests, and their interactions, and provides detailed information on the geographic distribution of all crops and pests globally. CABI provides a prevalence measure that takes the value of 1 if a crop or pest is present in a country and 0 otherwise, and so we extract that measure for the non-migrant's home/host country. In total, we have a sample of 7,551 crops and pests, 585 of which are published on by at least one individual in our sample. 22.63% of publications match to at least one of these crops or pests.

Because our focus is on identifying research productivity in local/global diseases and crops/pests, we constrict our sample to the set of peers who we can plausibly expect to publish in these topics. We limit our sample to peers who ever publish on *any* crop/pest (we describe the construction of our disease and crop/pest sample in the following section). This leaves us with 7,831 treated non-migrant scientists (those whose peers received an OWSD fellowship) and 11,639 control scientists (those whose peers applied for but did not receive an OWSD fellowship), of which 4,164 treated scientists and 6,383 control scientists ever publish in crops. This corresponds to 211 OWSD applicants, with a median OWSD applicant treating 35 stayers in their home institution. Our main analysis is conducted at the applicant $\times$ peer level.

We leverage our data on publication matches to diseases and crops/pests to generate several key measures of local/global nature of the research. In our empirical analyses, we measure individual-level annual publication counts on our diseases and organisms of interest. In our primary analysis, we measure individual-level annual publication counts to evaluate the number of articles an individual publishes on crops/pest that are present in:

- The home country
- The host country[1]
- Both the home and host countries
- The home country but *not* the host country
- The host country but *not* the home country

[1]For migrant peers, the host country is where their peer who applies for the fellowship identifies as the country they will visit.

We first focus on the full set of crops and pests that are present in our home and host countries. Table A14 shows these results. We see that stayers are more likely to publish on crops and pests that are present in their home countries and on those that are present in their peer's host countries. However, this effect is not significant when we focus only on crops and pests that are present in home or host countries (columns 6 and 7). Treated researchers are thus increase their focus on organisms that appear in *both* their home and their peers' host environments, but not those that only appear either at home or in the host country.

0.2 Alternative DiD estimation

The basic setup of differences-in-differences analysis relies on comparing treated and untreated units with the assumption that their outcomes would have been similar in the absence of treatment. The standard approach to this analysis is a two-way fixed effects regression, which regresses the outcome of interest on a unit's treatment status and includes unit and time fixed effects. However, when treatment timing varies, this estimation approach compares already-treated units and later-treated units (a "forbidden comparison" (Goodman-Bacon, 2021)), which can negatively weight units that are treated earlier and lead to inaccurate estimates.

There are several novel estimation approaches that seek to address this issue and estimate causal parameters of interest in contexts with variation in treatment timing (de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Borusyak et al., 2022; Gardner, 2022). These estimators vary in how they construct comparison groups and what assumptions they impose on parallel trends. The estimator proposed by Callaway and Sant'Anna (2021) does not require strongly balanced data (that is, for all units to be observed in all periods), but when estimating each group-time average treatment effect (ATTGT), only observations that are balanced within a specific 2x2 (two-period and two-group) are used for the estimator. Thus, when groups are not observed in all periods, estimation is conceptually more complicated. The estimator proposed by Sun and Abraham (2021) requires a strongly balanced panel, which we do not have.

The imputation-based estimators developed in Borusyak et al. (2022) and Gardner (2022) fit a TWFE regression using observations only for units and time periods that are not-yet-treated. They then infer the never-treated potential outcome for each treated unit using the predicted value from this regression. This provides an estimate of the treatment effect for each treated unit, and these individual-level estimates are aggregated to create estimates of the average treatment effect on treated units (ATT) (Roth et al., 2023). Both estimators return identical point estimates but differ in their standard errors (Butts and Gardner, 2022). These approaches accommodate panels that are not strongly balanced and allow for event-study

estimation; thus, these methods are best suited for our setting. Figure B1 below shows our pretend results for home country publications using both Boryusak and Gardner estimation.

0.3 Parallel trend violations

A key assumption with differences-in-differences methods is the parallel trends assumption, which assumes that the average outcome for treated and untreated samples will evolve along parallel paths. Instead of requiring that parallel trends hold, Rambachan and Roth (2023) propose a robust inference approach. This approach measures the sensitivity of pre-trends by examining how results change under different assumptions about potential violations of parallel trends.

To confirm that our results are robust to potential differences in pre-trends, we adopt this inference approach and identify by how much the slope of pre-trends can change without leading us to observe a null treatment effect. Figure B2 shows the results of this sensitivity analysis. We see that the critical value of M is 0.3, which means we can reject a null effect unless we are willing to allow for the linear extrapolation across consecutive periods to be off by more than 0.7 percentage points.

References

Borusyak, K., Jaravel, X., and Spiess, J. (2022). Revisiting Event Study Designs: Robust and Efficient Estimation. Technical Report 2108.12419, arXiv.org. Publication Title: Papers.

Butts, K. and Gardner, J. (2022). did2s: Two-Stage Difference-in-Differences. *The R Journal*, 14(3):162–173.

Callaway, B. and Sant'Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230.

de Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review*, 110(9):2964–2996.

Gardner, J. (2022). Two-stage differences in differences. arXiv:2207.05943 [econ].

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2):254–277.

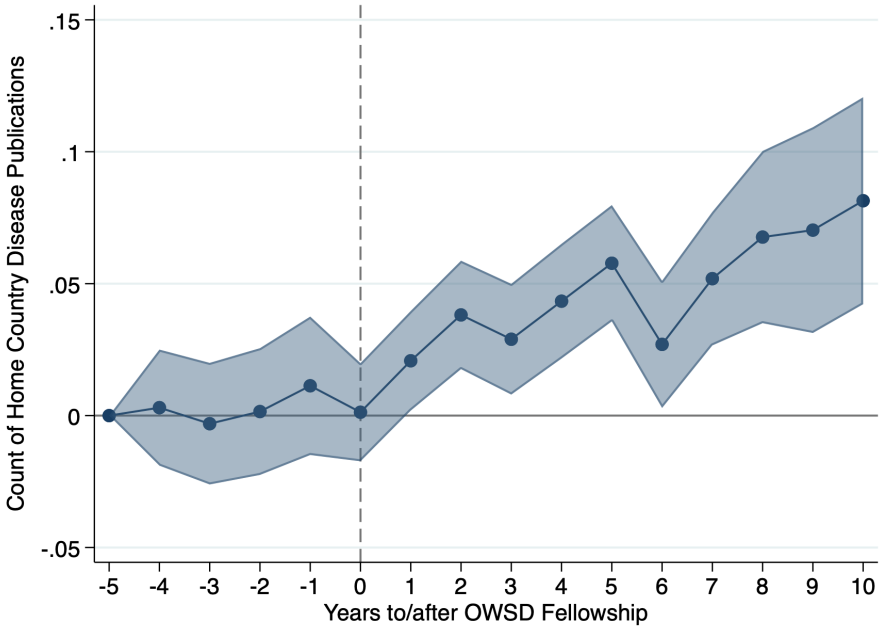
Rambachan, A. and Roth, J. (2023). A More Credible Approach to Parallel Trends. *The Review of Economic Studies*, page rdad018.

Roth, J., Sant'Anna, P. H., Bilinski, A., and Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*, 235(2):2218–2244.

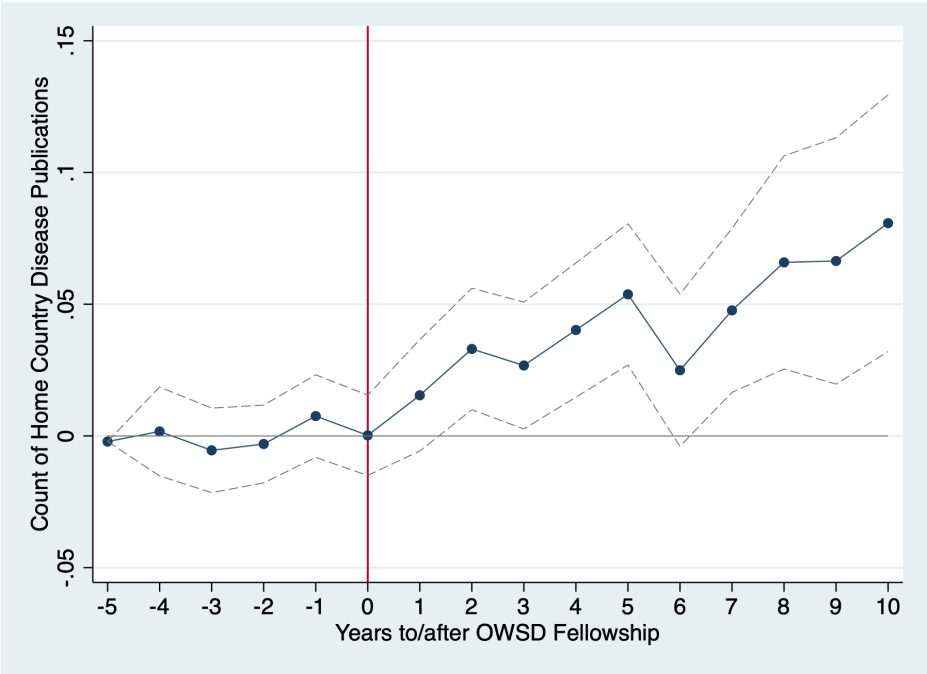
Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2):175–199.

Figure B1: Alternative DiD Estimation Pretrends

Panel A: Boryusak Estimation

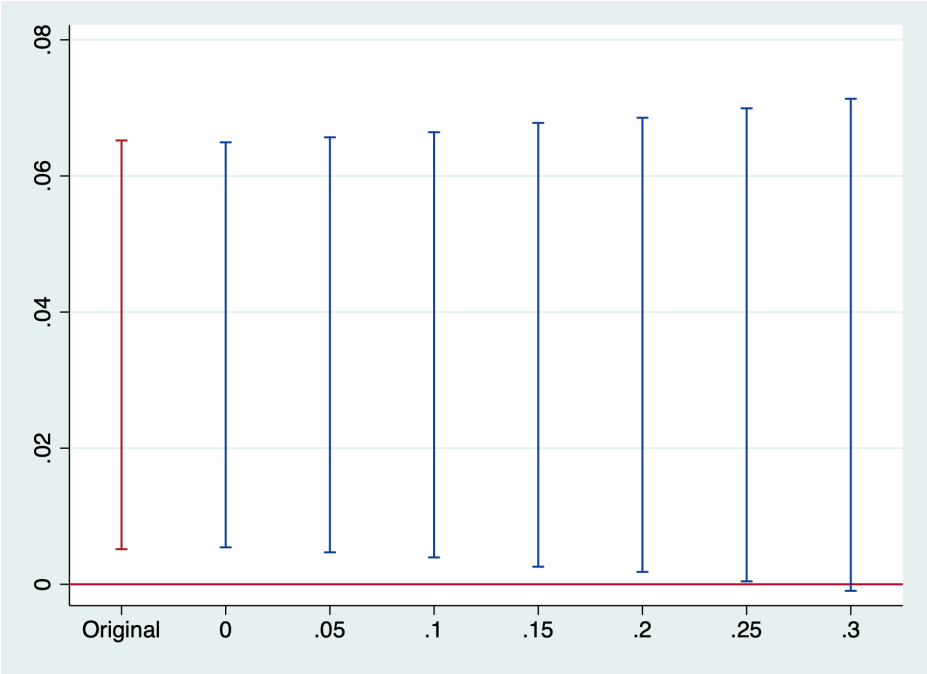


Panel B: Gardner Estimation



Notes: These figures shows coefficient estimates stemming from conditional (scientist) fixed-effects OLS specifications using the Boryusak estimation (Panel A) and Gardner estimation (Panel B) methods. The inverse hyperbolic sine outcomes are regressed onto year effects and scientist age effects, and include interaction terms between treatment status and the number of years before/after the OWSD fellowship, and a control variable representing the annual overall publication count for each scientist. The 95% confidence interval robust standard errors clustered around the scientist is plotted with solid bars.

Figure B2: Sensitivity Analyses (Rambachan and Roth, 2023)



Notes: This figure shows the results of sensitivity analysis based on Rambachan and Roth (2023). The original estimate (in red) is our OLS estimate. This approach formalizes the intuition that pre-trends are informative about violations of parallel trends by allowing the slope of the pre-trends to change by no more than some constant 'M' across consecutive periods. M = 0 corresponds with allowing only for linear violations of parallel trends, and larger values of M allow for larger deviations from linearity. We see that the critical value for a significant effect is M = 0.25. That is, we can reject a null effect (the treatment effect in a particular period is zero) unless we are willing to allow for the linear extrapolation across consecutive periods to be off by more than 0.25 percentage points.