Essays on Development Economics

Scott Weiner

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2021

Abstract

Essays on Development Economics

Scott Weiner

This dissertation consists of three essays, each covering very distinct topics under the broad umbrella of Development Economics, each set in a different region of the developing world (Latin America, Sub-Saharan Africa, and South Asia). The one element that loosely ties them together is that they each seek to add, in a small way, to our understanding of factors that contribute to, and in some cases may entrap people in, poverty: factors such as (lack of) geographic mobility, hunger, and disease.

In the first chapter, I use the natural experiment of military conscription in Argentina, which randomly assigned not only military service, but also the location of service, to study the effect of this temporary displacement on long-run migration rates. I then use a rich source of administrative earnings and employment data to investigate the labor-market implications of conscription and, in particular, displacement. I find that conscription on the whole caused a small increase in the likelihood of appearing in the formal labor force, and a small increase in earnings particularly for those who were assigned to serve in the Navy. Assignment to military service outside of one's province of origin increased the likelihood of living outside the province of origin by 2.5 percent, and while the net effects of this displacement on earnings and employment are imprecisely estimated, the evidence suggests that there are modest long-term benefits of conscription in Argentina that are not fully attributable to displacement.

In the second chapter, I investigate the effects of Ramadan on calorie consumption and labor supply among Muslim households in rural Malawi. Across four rounds of household survey data,

I find no evidence of a decrease in calorie consumption during Ramadan on average. I do, however, find evidence that working-age people reduce their weekly work by about three hours, or nearly 20 percent, on average. This finding on calories shows substantial variation across the different rounds of data. The evidence presented calls into question the hypothesis that consumption during Ramadan should fall more dramatically when the holiday overlaps with the harvest (when baseline consumption levels are relatively high compared to the rest of the year), compared to when Ramadan falls near the annual hunger season (when baseline consumption levels tend to be much lower). I discuss potential implications of this variation for our understanding of seasonal consumption patterns.

The third and final chapter, which is authored jointly with Kaivan Munshi and Nancy Luke, discusses a randomized intervention conducted in rural South India aimed at improving rates of treatment completion for tuberculosis. Tuberculosis (TB), despite being a highly treatable disease, kills well over 1 million people every year, with 95 percent of cases and deaths appearing in developing countries. India bears the largest TB burden of any country, with more than 25 percent of the world's total yearly cases. A key factor for successful management of TB is ensuring that patients complete the full six-month (or more) treatment regimen: missing even a few doses of the prescribed medications increases the likelihood of relapse and development of a drug-resistant strain of TB, which is much more difficult and costly to treat effectively. We conduct an intervention allowing patients to select a community member to serve as a Directly Observed Treatment (DOT) provider to help ensure compliance with the full treatment regimen. Although patients assigned a Community DOT provider report significantly more frequent visits and higher rates of satisfaction compared to our control group, we do not find any significant improvement in treatment outcomes among those assigned this intervention. We explore several potential explanations for this finding and suggest potential avenues for future research.

Table of Contents

List of Tables

List of Figures

Acknowledgements

There are so many people and institutions that I want to acknowledge that I feel compelled to break it down into categories:

1. **Chapter 1:** This project was a sprawling undertaking with a vast array of people who helped in different capacities along the way. Jorge Mangonnet and Julia Rubio were instrumental in helping me develop contacts in Argentina. My understanding of the history and context of conscription in Argentina were greatly enhanced through conversations with Rut Diamint, Jorge Battaglino, Ernesto Schargrodsky, Lourdes Puente, Osvaldo Tosco, Nadia Kreizer, Máximo Pérez, Jorge Antelo, Guillermo Casal, Carlos Pérez Aquino, Diego Sombra, Eduardo Gaundo, Gustavo Isaac, and Eduardo Stafforini. Pablo Panero and the staff at the Army's Office of Recruitment were very generous with their time and willingness to share information. Manuel Puente, David Alfaro Serrano, Pablo Warnes, and Facundo Galván also provided helpful insights about working in Argentina. My co-advisors, Suresh Naidu and Eric Verhoogen, provided invaluable insight in framing my research questions, guidance for conducting research in the field, encouragement to seek out the data that could provide interesting economic insights, and feedback on the various aspects of the research as it progressed. The paper also benefited from useful comments from and discussions with Rodrigo Soares, Miguel Urquiola, Jack Willis, François Gerard, Michael Best, Doug Almond, Jonas Hjort, Chris Cotton, Ashna Arora, Nandita Krishnaswamy, Anurag Singh, Lorenzo Lagos, Yue Yu, Amanda Awadey, Yogita Shamdasani, Kiki Pop-Eleches, Sun Kyoung Lee, Wojciech Kopczuk, Varanya Chaubeym and various participants in the Applied Micro and Development Colloquia at Columbia, as well as attendees of the APPAM Student Conference in April 2018. I gratefully acknowledge funding support from Columbia's Institute for Latin

American Studies (ILAS) Travel Grant, and from the Student Research Grant Program of Columbia's Center for Development Economics and Policy (CDEP). This project would not have been possible if not for a number of people and institutions that facilitated my access to the various data sources I utilize in this paper. I thank the National Institute of Statistics and Censuses (INDEC) of Argentina for providing a portion of the underlying data making this research possible (via the IPUMS-International collection). I thank Ernesto Schargrodsky and Martín Rossi for sharing supplementary data related to conscription assignments and cutoffs from their 2011 paper. I am deeply indebted to Ernesto Calvo for generously sharing several years' worth of voter registration data. I thank Andrés Drenik for alerting me to the existence of the SIPA earnings data in the Argentine Ministry of Labor and for connecting me to his contacts there. I am very grateful to Danilo Trupkin, Bernardo Díaz de Asterloa, and Juan Zabala Suárez for their role in granting me access to these earnings data. I am particularly indebted to Guillermo Cruces for his advocacy on my behalf, without which it is unlikely that I ever would have been granted access to the SIPA data. Finally, I am extremely grateful to the dedicated civil servants at the Ministry of Labor who generously took time out of their day to assist me in accessing the SIPA data, both in person and remotely: Lucía Tumini, J. Sebastián Rotondo, and in particular M. Victoria Castillo Videla, who generously helped me to run countless analyses remotely over the course of several months, and on multiple occasions made sure I had the results I needed promptly in the face of a pressing deadline.

2. **Chapter 2:** This paper benefited from discussions with Supreet Kaur, Eric Verhoogen, Suresh Naidu, Kaivan Munshi, and Doug Almond, and from feedback from participants at the Development Colloquium at Columbia. Even more so, it benefited from the discussion at my defense with my dissertation committee – Eric Verhoogen, Suresh Naidu, Rodrigo Soares, Miguel Urquiola, and Jack Willis – who helped guide me toward a more scaled back, and thus more realistic and achievable, framework for the paper. Olivier Ecker, Rachel Gilbert, and Kate Schneider generously shared time, data, programs, and other valuable in-

formation with me that were immensely helpful in the process of constructing measures of calories from the Malawi IHS data. This paper makes use of data from Malawi's IHS2 (2004–2005), IHS3 (2010–2011), IHS4 (2016--2017), and IHS5 (2019--2020), generously made available to researchers by the National Statistics Office of Malawi and the World Bank.

3. **Chapter 3:** The data used in this paper are from the South India Community Health Study (SICHS), directed by Kaivan Munshi, Yale University, Nancy Luke, Pennsylvania State University, Soumya Swaminathan, WHO, and Shantidani Minz, Christian Medical College, Vellore, India. SICHS was funded by the Eunice Kennedy Shriver National Institute for Child Health and Human Development (01 HD058831-01), the Keyes Fund through the Faculty of Economics at the University of Cambridge, and the Population and Training Center at Brown University. We are grateful to all of our collaborators at the National Institute for Research in Tuberculosis in Chennai and the Christian Medical College in Vellore, Tamil Nadu.

4. **General acknowledgements:** I'd like to thank all of my committee members (Eric Verhoogen, Suresh Naidu, Rodrigo Soares, Miguel Urquiola, and Jack Willis) for their patience, understanding, and flexibility as I slowly made way through this process. They gave me second, third, and fourth chances – whether or not I deserved them – to prove that I could ultimately pull this together. I also want to thank Navin Kartik, Amy Devine, and the Columbia Dissertation Office for their help and accommodation over the past year.

   I am grateful to Dr. Karimi Mailutha, Dr. Shirley Matthews, Dr. Alan Kouzmanoff, Jen Zobel Bieber, Dr. Christine Li, Barbara Nordin, Deborah Malatesta Crowley, Dr. Mohamed Elsamra, and Dr. Jason Greif for their various efforts to help guide me through various stages of the Ph.D. process.

   I am, of course, deeply grateful and indebted to my advisors, Eric Verhoogen and Suresh Naidu, who along with providing helpful substantive advice and mentorship throughout my time as a Ph.D. student, went above and beyond the call of duty to help me get to the finish line, continuing to meet with and advise me even after I entered the purgatory of no longer being an enrolled student, and even after my defense. Both maintained remarkable patience with me and displayed remarkably limited levels of anger and annoyance in our interactions, even when circumstances would rightly have called for some level of annoyance.

I am also eternally grateful to Kaivan Munshi, who took a chance on me as an undergrad, inspired me to continue pursue research in Development, helped teach me to think carefully as an economist, was instrumental in getting me to Columbia in the first place, and even took another chance on me towards the end of the Ph.D. program.

I want to thank Chris Cotton and Shreya Mathur for their friendship and support, particularly over the last year and a half.

My grandmother Lois has waited, and waited, and waited for this dissertation to be finished, and through it all has never given up on me. I promise I'll call more regularly now that it's submitted.

I am incredibly thankful to my in-laws, Monica and Rick Segal, for graciously taking me into their home (rent-free, no less) for most of the duration of the pandemic, providing food, childcare, support, entertainment, and companionship over the course of a very difficult year. Monica in particular is essentially superhuman, and I am in awe every day with how much she takes on and how little seems to faze her.

My parents have put up with me at my worst, believed in me when they probably shouldn't have, provided consistent guidance and support as circumstances have shifted in unanticipated ways, talked me off several proverbial ledges, and helped me push through various obstacles and stumbling blocks to reach this point. I could not have done it without them.

Above all, I am endlessly, unspeakably, unimaginably grateful to my wife Caroline, who has been a constant source of support dating back long before we both embarked on our doctoral journeys, and has taken on a disproportionate share of... well, everything, for far longer than was originally agreed to. On top of this, she has, despite a lack of training in economics, been an invaluable sounding board for my research, and her unparalleled gifts as a writer have helped me push through the most stubborn cases of writer's block. I'm not sure what I possibly could have done in my life to deserve a partner of her caliber, but in the immortal words of Alexander Hamilton (by way of Lin-Manuel Miranda), she is truly the "best of wives and best of women."

All errors throughout this document are my own.

Dedication

For Julia, my light, my joy. Thank you for being exactly who you are.

Chapter 1: Mobilizing Forces: Military Conscription as a Driver of Mobility in Argentina

1.1 Introduction

A broad literature in economics documents the presence of unexploited potential returns to migration within both developing and developed countries. That is to say, workers have the potential to increase their earnings simply by moving to a different location within the same country, but often are not taking advantage of this opportunity. Indeed, a striking proportion of people live quite close to their place of birth. In Argentina, for example, as of the Census of 2001, about 80 percent of native-born citizens lived in their province of birth (constructed from IPUMS data, Minnesota Population Center 2019). In the US, the corresponding figure was about 69 percent as of 2009 (Molloy, Smith, and Wozniak 2011). While such figures do not necessarily imply a market failure or aggregate welfare loss, they strongly suggest that workers are not choosing their location of residence solely on the basis of where they would receive the highest returns to their work and skills. This could mean that households are more vulnerable to shocks to the local labor market than would be implied by a standard spatial equilibrium model in which workers are perfectly mobile.

One possible explanation for the high proportion of people living near their place of birth is that migration might be more difficult for people with no experience living outside of their original hometown. In such cases, people might find it particularly difficult to imagine finding a new place to live, a new job, new social connections, and the like. A corollary to this is that if these same people were to be induced to live outside of their hometown even temporarily, it might make it easier for them to migrate later on in life. In other words, moving once, or being temporarily displaced, might make it easier to relocate in the future. This, in turn, could make it easier to pursue new economic opportunities in a different part of the country, or to leave in the face of a local economic downturn, such as the departure of a major local employer. In keeping with this idea, Bryan, Chowdhury, and Mobarak (2014) offer groundbreaking experimental evidence that a one-time incentive to migrate temporarily for seasonal work encouraged workers to migrate temporarily not just once, but also again the following year without any further incentive. A question that this study could not speak to is whether those same workers might, in the long run, be more likely to migrate permanently in pursuit of economic opportunity.

This paper is the first, to my knowledge, to specifically address the question of whether a temporary relocation can encourage higher migration rates for individuals in the long run, and to investigate what implications this might have for labor market outcomes. Specifically, I exploit random variation in both the assignment of military conscription in Argentina and, for a subset of conscripts, in the assignment of the location of service. Estimates from my preferred specifications indicate that while conscription in general did not have an appreciable effect on long-term migration, *displacement*, that is, being assigned to military service outside of one's home province, increased the likelihood of being observed in recent years living outside of that province by approximately 2.5 percent. The same specifications suggest that displacement was associated with a small and not statistically significant increase in earnings of about 0.3 percent, though this might mask a larger return accrued to those actually induced to migrate. Separately, I find that conscription on the whole caused a small increase in the likelihood of appearing in the formal sector, and a small increase in earnings particularly for those who were assigned to serve in the Navy.

Several features of the conscription system in Argentina are useful for the purposes of this paper:

First, conscription was assigned through a random lottery, which determined not only whether or not young men[1] were assigned to serve, but also the branch of the military in which they would serve – either the Army, Air Force, or Navy. This randomization permits a causal interpretation for estimated effects of conscription on various outcomes of interest.

[1]Only men were subject to conscription.

Second, this assignment was based on objective and observable characteristics: the year of birth and the last 3 digits of an individual's National ID number, hereafter referred to as the DNI (*Documento Nacional de Identidad*). This meant that even if an individual were to defer service for a number of years, his assignment was fixed by the initial drawing and would not change – thus, the assignment was not easily susceptible to manipulation or strategic behavior. Even the wealthy and well-connected generally found it difficult, though not impossible, to avoid military service if assigned. The fact that these criteria determined the conscription assignment also make it possible to observe the intended assignment using current administrative records, and thus to study the long-term effects of the program without directly observing whether an individual actually completed his assigned service.[2] See Section 1.2.1 for more details on the history and implementation of the conscription system.

Third, the fact that branch assignments were randomized also means that for the subset of the population living close to an Army base, but far from an Air Force or Navy base, there was random assignment to complete military service close to or far from home. Though I am not able to directly observe where (or if) an individual completed his service, typically, conscripts were sent to the nearest available base corresponding to their assigned branch. Henceforth, I will refer to the assignment of an individual to a branch of the military that did not have a corresponding base in his home province as a (temporary) displacement.[3]

A fourth feature that makes this an attractive research setting is the relative lack of confounding factors. Consider, for example, the conscription setting that has been studied most extensively, starting with Angrist (1990): the case of the Vietnam Draft in the US. Though there was nearly random assignment of draft status, deferrals for students allowed many wealthier children to avoid the draft, and among those who were drafted, many were sent to an unfamiliar place, fought in a violent and traumatic conflict, and upon returning home were at least supposed to have had access to various veterans' benefits, including college education and healthcare coverage. It would be

[2]Of course, this also means that I am restricted to Intent-to-Treat estimates, which might understate the effects of conscription on those who actually completed the service.

[3]Some people may have been required to serve far from home even if there had been a base nearby. This may cause me to underestimate the effects of displacement by classifying some leavers as stayers.

challenging for any study to separate out these different factors in understanding the long-term effects of the draft assignment. In the Argentine case, with the exception of the brief Malvinas (Falklands) War with the UK, conscripts served in peacetime. As the service was structured to be a basic obligation of citizenship, no special benefits were established for ex-conscripts (again, with the exception of those who fought in the Malvinas). In the popular imagination, a colloquial nickname for the mandatory military service program, *colimba*, is reputed to derive from a description of daily life as a conscript: *correr, limpiar, barrer* – run, clean, sweep (Ablard 2017).[4] In other words, the 1–2 years spent as a conscript were widely felt to be tedious, a waste of time, and not necessarily an opportunity to learn new skills, but also generally less traumatic than being sent into combat. That said, I do not rule out the possibility that conscripts may have gained valuable skills or social connections, nor that they may have been exposed to some level of violence even in peacetime. Rather, I argue that in an analysis of conscription in Argentina, it is reasonable to attribute any effects uncovered directly to some aspect of the experience of conscription, rather than to external or secondary factors such as benefits or programs for veterans.

A final attractive feature of this setting is the availability of rich sources of administrative data. I was able to access several years of voter rolls, which should capture nearly the entire universe of adult citizens of Argentina (as long as they are still alive and in the country), along with over 20 years worth of monthly, employer-employee matched records of formal sector earnings collected by the Ministry of Labor. Because both data sources include the year of birth and DNI, I am able to link them to each other and to each individual's conscription assignment to build a detailed picture of the long-term economic effects of this natural experiment.

As noted above, this paper uses the random assignment of individuals to a military branch without a base nearby as a metric for temporary displacement. Of course, the differential effect of being assigned to one branch of the military versus another cannot necessarily be attributed solely to displacement, even if this is one difference between these assignments. Different branches might provide different training, experience, levels of exposure to violence, and access to social

[4]The etymology is not entirely clear.

networks and employment opportunities. In order to address this concern, I take a difference-in-differences based approach. The simplest version of this approach is as follows: because all provinces but one housed at least one Army base, I take the effect of being assigned to the Air Force (Navy) compared to the Army for people from provinces *with* an Air Force (Navy) base, and then compare this difference to the analogous difference for people from provinces *without* an Air Force (Navy) base. For this latter group (i.e. provinces without an Air Force or Navy base), assignment to a branch for which there is no base in the province is a displacement, but in all other respects the difference between the experience of being assigned to the Air Force (Navy) compared to the Army would ideally be comparable.

Indeed, this is the identifying assumption underlying this analysis: that the difference between the experience of serving in the Air Force or Navy compared to serving in the Army (in terms of training, exposure to violence, etc.) is orthogonal to the presence of an Air Force/Navy base in a province. In other words, if an Air Force or Navy base were to be built in a province that previously did not have one, the assertion is that the effects of service in the Air Force/Navy *compared to the Army* would fall into line with the effects observed in other provinces that already had Air Force/Navy bases present.[5] This assumption would be violated if in provinces lacking Air Force/Navy bases, the distribution of skills pre-conscription or of opportunities available afterwards is such that workers in these provinces would have benefited differentially from the training received in the Air Force/Navy versus training received in the Army. If, however, this identifying assumption holds, then the difference-in-differences estimates as described can be fully attributable to the effect of displacement. It is important to note that this does *not* require the locations of bases to be as-if-randomly assigned, nor do provinces with and without Air Force/Navy bases need, in principle, to be similar on observable measures.[6] It also does not need to be the case that the effects of conscription are uniform across people from different provinces. As long as heterogeneity in the effects of conscription itself can be effectively captured by controlling for differences in the effect

[5]This is equivalent to the assumption of parallel trends.

[6]This is true of any difference-in-differences analysis: treatment and control groups do not need to be statistically balanced at baseline as long as there are parallel trends.

of assignment to the Army, then it will be possible to isolate a pure displacement effect, separated from the effects of service in the Air Force or Navy.

This paper offers two main contributions to the literature on the economics of migration: it identifies a potential barrier to migration that has received little attention to this point, and studies the long-run effects of lowering this barrier. This follows in the long tradition in development economics grappling with the puzzle of large, persistent wage gaps within a country, with wages generally low in rural sectors and higher in urban areas, and the related question of why such wage gaps are not closed via workers migrating (see, for example: Lewis 1954; Harris and Todaro 1970; Young 2013; Bryan et al. 2014; Munshi and Rosenzweig 2016; Bazzi et al. 2016; Morten and Oliveira 2018; Lagakos et al. 2018; Bryan and Morten 2019; Lagakos et al. 2020. For a recent overview of this literature, see Lagakos 2020). In the canonical spatial equilibrium model of Rosen (1979) and Roback (1982), any differences across locations in quality of life due to real wages or amenities present an arbitrage opportunity for workers that can be exploited through migration (assumed to be costless); as such, in equilibrium, at least at the margins, workers will be indifferent between their chosen location and the next-best one, and thus will move away in the face of any local market downturn, causing rents to fall and real wages to rise in the negatively-shocked region, and essentially smoothing the shock across the entire economy. Starting from this framework, Topel (1986) was first (or among the first) to note that introducing costly migration to this model implies that less mobile workers will face larger negative earnings effects in the face of a negative shock, and many recent papers (for example: Yagan 2014; Bartik 2018; Zabek 2019) have documented that workers do not fully adjust to local downturns and thus do face negative labor market effects. These results highlight the importance of gaining a better understanding of the barriers to migration.

A number of studies have documented evidence of various barriers to migration in both developing and developed economies, from high rents due to zoning regulations (Hsieh and Moretti 2019), to informal insurance networks (Munshi and Rosenzweig 2016), to basic road quality and transportation costs (Morten and Oliveira 2018). The idea that a lack of experience living outside of one's hometown might itself be a barrier to migrating has intuitive appeal: learning by doing is a feature of all kinds of human activities, it is easy to imagine that moving to a new place could be one such activity.[7] It is also backed by a number of pieces of suggestive evidence: Angrist and Chen (2011) find that Vietnam conscripts are more likely to live outside of state of birth than non-conscripts; Cadena and Kovak (2016), Schündeln (2013), and Basso and Peri (2020) all document how people who have immigrated internationally tend to migrate within their new home country at much higher rates than native-born citizens, thus helping to equilibrate local labor markets; Malamud and Wozniak (2012) provide quasi-experimental evidence that going to college (which at least for some people means temporarily relocating) causes people to have a higher propensity to migrate.[8] Most notably, Bryan, Chowdhury, and Mobarak (2014) show that experimentally inducing people to migrate seasonally from rural Bangladesh to the capital city makes them more likely to temporarily migrate again the following year.[9] But in general, even if we see that people who have experience moving also have a greater propensity to migrate again, it is difficult to establish a causal relationship; as with any study of migration, one has to contend with the fact that migrants are in most cases self-selecting, and thus that the association between moving experience and future mobility could be driven by some underlying preference or higher suitability for migration among the migrant group compared to the stayers. This paper is the first, to my knowledge, to directly investigate the causal relationship between temporary displacement and subsequent increases in mobility.

This study also contributes to the literature that uses various types of random experiments or quasi-random relocation shocks (such as natural disasters) to study the economic effects of migration (examples include Bauer et al. 2013; Bryan et al. 2014; Chetty et al. 2016; Chyn 2018;

[7] Another plausible mechanism is that strong social ties bind people to their hometown (cf. Zabek 2019), and that leaving even temporarily helps to proverbially "cut the cord" and make these ties less constraining.

[8] It should be noted, however, that in this last paper, the authors do not find evidence that the increased likelihood of migration is associated with the likelihood of going to college out of state. Nonetheless, even going to an in-state college often means establishing a temporary residence outside one's family home, so it could still be that this experience away from home is at least part of what encourages higher migration rates.

[9] Indeed, this study was in large part inspired by this finding. What the Bryan et al. (2014) experiment could not answer – at least not yet – is whether their "treated" group was in the long run also more likely to migrate *permanently*: to do so would require tracking down and following up with participants over several decades.

Deryugina et al. 2018; Nakamura et al. 2019; Sarvimäki et al. 2019; Becker et al. 2020). The specific setting of this paper offers two key innovations to the existing body of research. First, in contrast to most migration-inducing shocks, peacetime conscription is temporary and does not generally entail a catastrophic, life-altering shock (e.g. a war, volcanic eruption, or hurricane) that makes it impossible to return home. This type of shock, while quite useful as a natural experiment, is often quite permanent and could potentially come with a large (usually negative) wealth shock and psychological distress, thus making it more difficult to attribute outcomes specifically to the relocation, and calling into question the implication that similar effects would be observed for the same people if they had simply decided to migrate on their own. Second, related to the previous point, because I observe both stayers and leavers *among* conscripts, along with an exempted control group, I can directly compare the long term outcomes of conscripts who (likely) stayed close to home and those who were sent far away. If the conscription experience is otherwise similar for stayers and leavers (though this is not a given), it allows me to isolate the effect of displacement from other aspects of the experience, thus potentially offering a higher degree of external validity in the measured effects of relocation.

Finally, this paper also contributes to the vast literature looking at the effects of military service on a wide range of outcomes. This includes numerous studies of the effects of military service on labor market outcomes (Berger and Hirsch 1983; Angrist 1990; Imbens and van der Klaauw 1995; Buonanno 2006; Paloyo 2010; Angrist and Chen 2011; Angrist et al. 2011; Galiani et al. 2011; Grenet et al. 2011; Bauer et al. 2012; Card and Cardoso 2012), along with studies looking at effects on educational attainment (Cipollone and Rosolia 2007; Keller, Wagener, and Poutvaara 2010); health (Bedard and Deschênes 2006; Angrist, Chen, and Frandsen 2010; Autor, Duggan, and Lyle 2011); alcohol and tobacco use (Goldberg et al. 1991; Eisenberg and Rowe 2009); mortality (Hearst, Newman, and Hulley 1986; Conley and Heerwig 2012); and crime and violent behavior (Yager, Laufer, and Gallops 1984; Beckerman and Fontana 1989; Rohlfs 2010; Galiani, Rossi, and Schargrodsky 2011; Lindo and Stoecker 2014; Gibbons and Rossi 2020).

This study is also among the first to investigate the effects of conscription in a developing country, with the main other example of which I am aware being the study by Galiani, Rossi, and Schargrodsky (2011) of the effects of conscription in Argentina on crime, which serves as the starting point for this paper.[10] Given that conscription is currently more prevalent in developing countries than richer ones, might have different effects on people in the context of a developing country, and could even play a role in the process of economic development and structural transformation, there is good reason to investigate the labor market effects of conscription itself, even beyond the lens of displacement and migration.

1.2 Background

1.2.1 Conscription in Argentina[11]

A Brief History of Conscription in Argentina

Argentina's system of conscription, or *Servicio Militar Obligatorio* (Mandatory Military Service), was first enacted in 1901, and remained in place until 1995. When it was first introduced, conscription was presented more as a nation-building project than one of national defense or war-readiness. At that time, unlike in the European countries on which the conscription model was based, there was little concern about war with neighboring countries. There was, however, substantial concern among elites about domestic unrest. A very large fraction of the country's population was foreign-born or first-generation native-born, and elites saw these new arrivals as "unruly and unassimilated" (Ablard 2017), with little attachment to their new country or sense of Argentine national identity. With the influx of European immigrants also came new ideas and social movements seen as dangerous to the state; in particular, anarchism and syndicalism were popular among the large Italian immigrant community at the turn of the century.

Conscription offered a way to "Argentinize" the (male) populace by, at least in theory, train-

[10] More recently, Ertola Navajas et al. (2020) and Gibbons and Rossi (2020) have also studied long-term effects of conscription in Argentina.

[11] Much of the discussion contained herein draws from the descriptions in Galiani, Rossi, and Schargrodsky (2011) and Ablard (2017), supplemented by my own discussions with Rut Diamint and Jorge Battaglino at Universidad Torcuato di Tella, and with several officials in the Ministry of Defense of Argentina.

ing a generation of obedient, loyal soldiers with a strong sense of Argentine identity and pride. Conscripted forces could also be used to suppress domestic unrest (particularly labor and radical activists), help establish a stronger state presence in remote regions, and violently subdue and ultimately assimilate the indigenous population. Particularly with reforms introduced in 1911, military service was presented as an essential component of citizenship, intimately tied to the introduction of universal secret and obligatory male suffrage: the reform created a system to identify and register all available conscripts, and used that list to build up voter rolls. Identity documents issued in conjunction with this reform were then used to monitor and document compliance with both military service requirements and obligatory voting. By the 1950s, conscription was firmly established as a rite of passage for Argentine men.

For the most part, conscripts completed their military service in peacetime, with the important exception of the 1982 Malvinas (Falklands) War against the UK. However, this does not necessarily imply that military service was always peaceful. Along with the aforementioned role of subduing domestic unrest and indigenous peoples, the military was heavily involved in Argentine politics throughout the 20th century, with frequent coups (some aided by conscripted soldiers) overthrowing democratic governments and periods of rule by military regimes.[12] The last military dictatorship was particularly notorious for human rights abuses, including kidnappings, torture, forced disappearances, and extrajudicial executions. The extent to which conscripts were involved in these abuses is controversial. Though no conscripts have been formally implicated in any of the legal cases surrounding human rights violations, conscripts were involved in skirmishes with left-wing guerrilla forces, and it is possible that some were involved in other violent activities. With the return to democracy in 1983, the military's prestige and influence in public life greatly diminished, leading to a rapid reduction in the fraction of men called to perform military service. Public outcry surrounding the death in 1994 of a young conscript, Omar Carrasco, at the hands of military superiors, led to the suspension of mandatory military service and the transformation of

[12]See Potash (1969, 1980, 1996) for a comprehensive account of the military's role in Argentine politics throughout the century. Notably, this three-volume account makes almost no mention of conscripted soldiers, focusing almost entirely on the actions and decisions of enlisted military officers.

the military to an all-volunteer force.[13]

The Conscription Process

During the era in which mandatory military service was in effect, the process followed a largely stable timeline:

The first step was issuing DNIs to young men (and ultimately women as well) in the year that they would turn 18. For people born in or after 1968, DNIs were issued at birth.[14] While it was required and expected that all citizens would receive their IDs at the appointed time, it is possible that some people received them on a delay. In theory this could allow individuals to try to manipulate their conscription assignment by delaying their DNI assignment, but there is no evidence, to my knowledge, of this happening, certainly not at any appreciable scale.

Next, on May 31 of the year before a given cohort would be called to serve, the *sorteo,* or draft lottery, took place in the headquarters of the National Lottery in Buenos Aires, and was broadcast live over the radio across the country. Prior to 1976, this took place in the year that the cohort would turn 20 years old, so that they would begin their service at age 21. Due to a legal change passed in the late 1960s, later cohorts (those born in 1958 or later) were sorted in the year in which they would turn 18, and began service at age 19.[15] The lottery was conducted as follows: 1000 balls numbered from 1–1000 were placed in a lottery wheel to be drawn at random, with each ball representing a sort order for conscription. All men in the given cohort who shared the same last three digits of the DNI would then receive the same sort assignment. Thus, starting with 000 and going through to 999, one ball would be drawn for every possible 3-digit combination of DNI endings. The final results of this sort process would be published in newspapers across the country the following day; see Figure 1.1 for an example of how the sort results would be recorded.[16] A

[13]This case is cause for further caution about the level of violence to which conscripts were subjected during their service. Ablard (2017) notes that newspapers "frequently reported on the physical abuse of soldiers at the hands of their superior officers," which suggests that this case was not an outlier.

[14]People born before 1968 still had to wait until they were approaching age 18 to receive their documents, even as younger children were receiving them almost immediately at birth.

[15]This meant that men born in 1956 and 1957 were automatically exempted from military service.

[16]The sort results given in this figure ultimately never went into effect, as mandatory military service was suspended in 1995, a few months after this lottery was completed.

lower sort number meant one was more likely to be exempted from military service, but the exact cutoffs would be determined at a later date.

Figure 1.1: Example of *sorteo* result, 1976 cohort

DNI (last 3 digits)	Draft sort order
000	923
001	343
002	627
⋮	⋮
998	276
999	974

In the months after the *sorteo*, all men in the sorted cohort would receive a medical examination to determine whether or not they were physically fit for service. Importantly, this process was completed after the lottery, but before the draft cutoffs were announced, so people would not know with certainty at the time of this examination whether or not they would be expected to serve. In theory, this should mean that it was difficult to selectively avoid service based on their assignment. In practice, however, while it was impossible to know where the exact cutoffs would fall, it was generally feasible to guess with a reasonable degree of accuracy, especially if one's assigned number was very low or very high. As such, people assigned a high sort number might have more incentive to exert effort to manipulate the results of their medical examination, perhaps by bribing or otherwise persuading a doctor to declare them unfit for service. Indeed, Galiani, Rossi, and Schargrodsky (2011) show some evidence of such manipulation, in that in certain cohorts, the rates of actually completing service are somewhat lower among people with higher sort numbers compared to those with lower numbers who were also ultimately assigned to service.[17] However, they

[17]It is worth noting here that among the cohorts for which they have this information, the replication data from Galiani et al. (2011) show that the fraction of people assigned to serve who ultimately completed military service was consistently lower than the fraction deemed medically fit to do so, suggesting that some individuals found ways to

also find a dramatic jump in completing service around the cutoff points, suggesting that there was in fact a high degree of compliance with the assigned service.[18] Anecdotally, even fairly well-off and well-connected people found it difficult to get out of military service completely, and at least some considered the service a duty of citizenship and did not make an effort to avoid it. Still, if there was in fact some self-sorting out of the conscription assignment, while it does not affect the validity of the Intent-to-Treat estimates that I produce, it may somewhat complicate their interpretation, as these measures would encompass the net effect of both being assigned to military service and exerting greater effort to avoid conscription.[19]

Several months later, three cutoffs were announced determining conscription assignments. Sort numbers less than or equal to the first, lowest cutoff number would be exempt from any military service. Above this first cutoff and up to the second, conscripts were assigned to the Army, which took by far the largest share of conscripts. Above the second cutoff and up to the final cutoff, conscripts were assigned to the Air Force. Sort numbers above the highest cutoff were assigned to the Navy.[20]In most years, these cutoffs were assigned uniformly across the country. But for certain cohorts, there was geographic variation in the assignments. Men born in 1955 and 1965 faced different cutoffs depending on which of the 5 Army Corps they fell under, based on where they lived. The variation is more substantial among men born in and after 1966, as cutoffs were for these cohorts were determined separately for each of the 29 Military Districts. Legally, men assigned to serve in the military could automatically request a deferral for up to 2 years without an

avoid military service other than obtaining a medical exemption. Specifically, I calculate from their data that 70.1 percent of men assigned to serve in the military from the 1958-1962 cohorts did so, while 8.6 percent of assignees received a medical exemption. This suggests that the remaining 21.3 percent likely found a different way to avoid military service (unless there is some substantial mismeasurement in one or more of these variables).

[18]All of this suggests that it might be a good setting for a Regression Discontinuity analysis, as it is reasonable to assume that the effort function for getting a medical or other exemption is continuous around the cutoff, given that the precise cutoff is not known in advance. Galiani et al. (2011) also provide evidence of a lack of manipulation in the neighborhood of these cutoffs. This could potentially be a useful extension to the analysis I present.

[19]The use of medical exams as a screening mechanism might also mean there was positive selection into military service, as poorer people may have been less healthy, and thus less likely to pass the medical exam (cf. Angrist and Krueger 1994). However, unlike the possible manipulation of medical results just discussed, there is no reason to expect this effect to vary by the assigned sort number, so it should not introduce any further bias or complication to ITT estimates.

[20]For certain cohorts or subsets of cohorts, one or more of the branches were excluded from the assignments, such that in effect there were only one or two cutoffs. In other cases, the entire cohort was assigned to military service, meaning that the first cutoff was effectively 0, as the sort assignments ranged from 1 to 1000.

excuse, and could defer for up to 10 years in order to complete secondary or university education. A few other categories of people were also granted automatic exemptions, such as clergymen or sole providers for dependent parents or children. In addition, Congress would, from time to time, pass amnesty bills, allowing people who had avoided military service for a certain period of time without a formal exemption to be formally excused. Together, all of these factors may help explain why a nontrivial fraction of the male population was able to avoid military service despite having been assigned to it. Anecdotally, people who did comply with their service assignment tended not use the available deferrals, preferring to serve with other people the same age and to get military service over with as quickly as possible.

Finally, in the following year, often in March, conscripts would begin their term of military service. For the first 3 months, they would undergo basic training, and would be required to live full time in the barracks. After this, they would receive their permanent posting. Conscripts in the Army and Air Force were expected to serve for one year total, though occasionally they would be retained for a few additional months. Conscripts in the Navy were required to serve for 2 years. Once they had completed their required service, this was noted in their ID document. This certification of having completed (or been exempted from) mandatory service was – at least on paper – legally required in order to be employed in a formal sector job, and to get access to some social services.[21]

Life as a Conscript

In order to better understand the "treatment" to which conscripts in this natural experiment were subjected, it will be useful to offer a brief discussion of the actual experience of conscription.

In the popular imagination, conscription was largely considered a wasted year (or two), especially for men from middle or higher class backgrounds. The tasks assigned to conscripts were generally menial, and the stipends paid were in most years extremely small.[22] Unlike the oft-

[21] It is not clear how strictly this requirement was enforced.

[22] I do not have data on the stipends issued from year to year, but my understanding is that while stipends were actually fairly generous for a stretch of years mid-century, enough that poorer conscripts could actually send some money home to their families, in time inflation and budget cuts dramatically reduced stipends to the point where they

studied former conscripts in the US, no special GI-Bill-type benefits were conferred to conscripts in Argentina over those who were exempted from conscription due to a low lottery number; conscription was structured as a duty for all male citizens, and thus not something that merited any special compensation. The one exception to this was for veterans of the Malvinas (Falklands) War against the UK: people who actually fought in this war are in theory entitled to a suite of benefits. Otherwise, any effects of conscription in Argentina on labor market outcomes are directly attributable to some aspect of the experience of military service, rather than to any health or educational benefits offered afterwards.

Based solely on the popular conception, it may seem reasonable to imagine that conscription really was little more than a temporary displacement and wasted 1–2 years of potential labor market experience. However, this is not a consensus view. Supporters of mandatory military service (including people more recently arguing for a return to the program) characterized it as a way to learn discipline and to mature; it was also billed as an on-ramp to the formal labor market and upward social mobility, and for some the first opportunity to access basic medical care, sanitation, and sometimes even basic education and literacy training (Ablard 2017).[23] It is also plausible that conscription offered an opportunity to form new social connections that ultimately may have been economically advantageous, at least for some, and it may also be that some conscripts learned practical skills that were transferable to the civilian labor market, though I do not have evidence specifically validating this. Finally, as was mentioned in the previous sub-section, completing one's conscription assignment offered a "passport" to the formal labor market in a literal sense: if one's identification documents did not indicate completion of required military service, he would (in theory) be ineligible for formal sector employment.

On the other hand, conscription may have also had direct negative impacts on recruits. Despite the characterization of Argentina's conscription program as a "peacetime" draft, there are many reports of conscripts being exposed to some level of violence, even beyond those who were actually

would cover little more than the purchase of cigarettes.

[23]Several current and former military professionals independently mentioned to me in conversation that conscription for many rural young men offered their first encounter with a toothbrush.

sent to fight in the Malvinas War. Reportedly, conscripts routinely faced harassment or hazing at the hands of military officers and strict punishments for minor offenses. Occasionally, this harassment was even lethal, as in the case of the killing of conscript Omar Carrasco that ultimately led to the suspension of mandatory military service in Argentina. Conscripts were also at times involved with quelling domestic unrest, supporting coups, and engaging in skirmishes with guerrillas, militants, or even rival military factions. It is unclear if conscripts were involved in or witness to the various human rights abuses of the Dirty War and the National Reorganization Process under the military junta of the 1970s and '80s, but there is no direct evidence (to my knowledge) of involvement of conscripts, and in fact in rare cases conscripts were victims of such abuses, or were even "disappeared".

All of this is to say that conscription on its own, at least for some recruits, was not merely a temporary displacement. However, displacement was part of the conscription experience for many recruits. Even those who were assigned to serve quite close to home were not allowed to leave the for their first 3 months of service. In fact, the earlier years of the program when the focus was more on nation-building and assimilation of immigrants' sons, active efforts were made to integrate people from different parts of the country, so even if their was a base near their home, they were likely to be assigned elsewhere. Over time, as budgets became more strictly constrained, conscripts started to be more frequently sent to the nearest available base.[24] Given that after the 3-month basic training period, conscripts were allowed to leave the base on weekends, there is good reason to think that it might have made a difference whether or not one was stationed reasonably close to home, as many conscripts would go home on weekends if it was feasible to do so.

The details of this displacement are central to the main research question of this paper, i.e. the effect of temporary displacement on long-term migration decisions. Given that I use the absence of a base corresponding to the assigned military branch in a given province as an indicator of "displacement," it is important to note that in a number of ways, conscripts whom I characterize as "stayers" (i.e. not displaced) may also have faced some level of displacement, which may cause

[24]I do not have information about specific deployments of troops or how these decisions were made, nor do I have information on the timing of this shift towards keeping conscripts closer to home.

me to underestimate the true displacement effect. As noted, conscripts may not have been assigned to the nearest available base. But even if the nearest available base was in the same province, and even if one was assigned to this base, Argentina's provinces are quite large geographically, so that base might be quite far from home. Conversely, there could be cases where there is no base for one's assigned branch in one's own province, but the nearest base is in a neighboring province and is quite close to the provincial border, such that this assignment does *not* actually require venturing particularly far from home.[25]

1.2.2 Conscription as Development Policy

While many countries have, like Argentina, moved away from universal or random conscription in favor of volunteer-based forces, many of the countries that have retained conscription-based systems are developing or middle-income countries.[26] Yet very few studies of military conscription have looked at its effects in a developing country. The primary exception, to my knowledge, is the study of the effects of conscription on crime in Argentina by Galiani, Rossi, and Schargrodsky (2011), to which this paper is, in part, a follow-up.[27]

Since the effects of conscription in developing countries have received little attention in the literature, it is reasonable to ask whether these effects might be different from those observed in developed countries. In particular, one might reasonably wonder whether military conscription could help facilitate the process of structural transformation that historically has been crucial to countries' economic development, perhaps helping to incorporate people from rural, underdeveloped areas into formal labor markets, or making them more willing and able to move into more productive urban areas and/or out of less productive agricultural work.

Certainly, Argentine military propagandists and proponents of conscription embraced this line

[25] All of this suggests that it could be useful to develop a more refined measure of displacement. It would certainly be useful to have some individual-level data on deployments, even for a subset of the population, but I was not able to access any such data.

[26] Examples include Thailand, Iran, Brazil, Colombia, Egypt, Russia, and Turkey. See CIA (2019) for a complete list of military service obligations by country.

[27] As has been noted, Ertola Navajas et al. (2020) and Gibbons and Rossi (2020) also look at the effects of conscription in Argentina, but neither of these papers focuses on migration or labor market outcomes.

of thinking. Politicians and military officers offered panegyrics to the "habits of order and discipline" that conscripts would learn, the skills that men of the lower classes could bring back to teach their families, and the benefits of medical attention and basic education that conscripts might receive for the first time (Ablard 2017). Conscription was characterized as essentially "a social service directed to the physical and moral uplift of poor Argentine men," "a passport to formal sector employment," and a source of skills that would allow men to earn a living and support a family (Ablard 2017). Indeed, I do find some empirical evidence that conscription increased rates of formal labor force participation, and perhaps even earnings.

Yet basic neoclassical economic theory, as articulated at least as early as Oi (1967), would suggest that conscription is highly inefficient: if it really provided such a great benefit, people would presumably be willing to volunteer for the service. Setting wages and benefits sufficient to attract people to the military, if the labor market functions well, would be a more efficient way of attracting the people who would be most interested and capable to join the service, minimizes deadweight loss for people taken out of the labor market or higher education. Mandatory conscription, on the other hand, in this view functions essentially as a tax, requiring conscripts to sacrifice valuable labor market experience in exchange for generally below-market wages. This idea has encouraged the move to smaller, better paid, and highly professionalized fighting forces in many countries, including Argentina.

Of course, especially in a developing country, labor markets might not function very well, such that perhaps mandatory service could actually offer a long-term benefit to at least some people who would not otherwise volunteer for it.[28] Interestingly, even in relatively affluent countries, the evidence on the long-term labor market effects of conscription is not as clear-cut as the standard theory would suggest. While some studies find a negative effect of conscription on earnings (Angrist (1990) in the US for Vietnam era conscripts; Imbens and van der Klaauw (1995) in the Netherlands; Buonanno (2006) in the UK), others find no significant effect (Bauer et al. (2012) and

[28]It's also worth noting that governments have budget constraints, so if leaders perceive a need to build a large military force, it might be impractical to offer a sufficiently high wage to attract the desired number of people, in which case making service mandatory, while still allocatively inefficient, might be difficult to avoid.

Paloyo (2010) in West Germany; Angrist and Chen (2011) and Angrist, Chen, and Song (2011) in follow-up studies of Vietnam conscripts in the US;[29] Grenet et al. (2011) in the UK; Albrecht et al. (1999) in Sweden). Other studies have even found positive effects of conscription, at least for some segments of the population: Card and Cardoso (2012) find a significant 4–5% increase in earnings for conscripts with six or fewer years of formal education in Portugal (and no significant effect on more educated groups); Berger and Hirsch (1983) document a slight earnings premium for Vietnam-era veterans in the US without a high school diploma, along with a slight penalty for more educated veterans, though these results may not fully account for the non-random selection bias in veteran status. While it is difficult to draw broadly generalizable conclusions from these varied results, there is at least a plausible case to be made that conscription in some settings might benefit at least a segment of the population, particularly less educated or lower-skilled workers.

There is also some direct evidence that military service might increase long-term geographic mobility. For example, Angrist and Chen (2011) find that being drafted increases the likelihood of living outside of one's birth state by about 3 percentage points. Relatedly, Pingle (2007) notes that much of the recently observed decline in the yearly rate of interstate migration within the US can be attributed to the decrease in the share of the population that is active-duty military, as active-duty military members tend to move much more often than civilians.[30] Malamud and Wozniak (2012) also find that the increased likelihood of going to college in order to *avoid* the Vietnam draft caused people to be more likely to migrate, but interestingly they find that this effect is not necessarily associated with going to an out-of-state college.

Despite the fact that primary focus of this paper is on the impact of displacement due to conscription on long-term mobility, a positive result to this effect does *not* imply that the most efficient way of improving labor market mobility in pursuit of an economic development strategy would be to draft everyone into the military. As discussed in this subsection, the effects of military con-

[29] This finding does not necessarily contradict the earlier one: they continue to find an initial decrease in earnings for veterans compared to non-veterans, but over time the effect of lost experience fades, and is ultimately canceled out by an increase in educational attainment. It is also worth noting that even Angrist (1990) does not find a negative effect of conscription on non-White veterans.

[30] Of course, being reassigned within the military is not the same as moving voluntarily.

scription are highly context-specific and potentially quite heterogeneous, and it is reasonable to imagine that there could be more direct ways of promoting similar outcomes of mobility and/or formal labor market integration without mandatory military service. A compulsory program of any sort is also quite heavy-handed: given that the long-term benefits may only accrue to certain segments of the population (e.g. workers with low levels of education), a voluntary program might be more sensible.

Finally, even if it is the case that conscription could help to facilitate the process of structural transformation and state-building in a developing country, a policymaker would need to weigh this potential benefit against legitimate concerns about increased violence and militarization among conscripts, including the potential risk to stable, democratic government that could also (arguably – see for example Acemoglu et al. 2019) be important for economic development. Indeed, recent evidence from Argentina supports the idea that former conscripts have a more militaristic world-view (Ertola Navajas et al. 2020), and the experience in Argentina itself with human rights abuses and economic mismanagement under military rule provide ample reason for caution.

1.3 Data Sources and Construction

This paper relies primarily on two sources of individual-level data: complete voter registration rolls from several recent elections, and monthly employer-employee matched earnings data from Argentina's Social Security system, the *Sistema Integrado Previsional Argentino* (SIPA).

1.3.1 Voter Rolls (2011-2017)

This dataset contains the name; DNI; sex; birth year; street address; and the province, department, and electoral circuit of current residence for all registered voters.[31] I primarily use the rolls corresponding to the 2017 election cycle for the main analysis, but I have available complete rolls corresponding to 2011, 2013, 2015, as well as partial rolls from 2003, for supplementary analysis. These records, in principal, are meant to capture the full universe of Argentine citizens living in

[31] Birth years are excluded for some female voters, including all female residents of the City of Buenos Aires, and a small percentage of those living elsewhere.

the country at the time of each election, as voting is compulsory for all Argentine citizens and voter registration is automatic. Of course, in practice, the voter rolls may not be perfectly accurate. People may not update their address immediately upon moving, which could cause me to understate rates of internal migration. Individuals' sex, birth year, or even DNI might be recorded incorrectly, which could result in incorrect measurement of conscription status and could make it more difficult to match to employment records (discussed below). There may be some native citizens who are simply missing from the voter rolls, though this should be a very small fraction of the population. Even accurate records may not perfectly reflect the universe of people who were subject to conscription: citizens who have died or migrated out of the country *should* be missing from the rolls,[32] and some people who *are* on the rolls may be naturalized citizens who were never eligible for conscription.[33] Despite these concerns, this data source should suffice to capture a nearly complete universe of potential ex-conscripts and potential prime-age workers. This in turn allows me to observe whether an individual ever appears in the formal labor force during the period for which I have data, and to consider this as an outcome variable of interest.

In addition, I use the voter roll data to impute the following variables, which will be crucial for my analysis:

Province of origin, i.e. the province in which an individual resided at the time that their DNI was issued.[34] There is no (documented) function explicitly linking DNIs to the province in which they were originally issued, nor is this information included in the voter roll data. However, two key facts make it possible to use the voter roll data to generate a reasonable guess as to where many individuals were living when their DNIs were first issued. The first fact is that DNIs were not assigned at random. Rather, in order to ensure that each number would uniquely identify an

[32]This could introduce survivorship bias.

[33]It is my understanding that a certain range of DNIs is reserved for non-native citizens, so I exclude from analysis DNI ranges that are sparsely populated and numbers higher than 50,000,000.

[34]Note that this is not necessarily the province of birth, but for most cohorts it represents the province of residence at the time when they received their conscription assignment, as DNIs were assigned shortly before this. However, due to a reform to the system for DNI assignments, cohorts born in and after 1968 were assigned a DNI immediately at birth. For these cohorts, the imputed province of origin will represent the province of birth, which could potentially mean that conscription variables are assigned incorrectly in the analysis due to the variation in cutoffs across provinces for cohorts born in and after 1965.

individual, the documents were produced in a centralized source in the capital and distributed from there to offices in each province. In most cases, the documents were printed in numerical order, and sent to provincial offices in packets containing a range of consecutive numbers.[35] Thus, in most cases, two individuals with numerically close DNIs were likely living in the same province at the time when they first received their documents, unless the numbers fall on opposite sides of a cutoff point between two ranges assigned to different provinces. The second key fact is the observation that motivated this paper: that a large majority of people live in their province of birth. This, together with the details of the process through which DNIs were assigned, suggests that in looking at any range of DNIs together with the current province of residence, we should expect to see a particular province disproportionately represented up to some cutoff, after which point a different province should be disproportionately represented. By identifying clusters of provinces of current residence in adjacent or near-adjacent DNIs, I derive a guess as to the province of origin for all individuals in a certain numerical range, and thus also classify individuals currently living in a different province as migrants. A description of the details of this procedure can be found in Appendix A.1.2.

Recently lived or currently living outside province of origin. This is the primary migration outcome that I use for analysis, essentially an indicator variable for migration out of the province of origin. It excludes individuals for whom the province of origin could not be imputed, and takes a value of 1 for those currently living (according to the 2017 voter roll) in a different province than the province of origin. Of course, some people may have moved out of their home province for some time and later moved back. In an effort to catch at least some of these cases and classify them as migrants,[36] I turn to an identifier variable included in the 2017 voter roll. This variable begins with a single letter that corresponds to a province's postal code, followed by a long series of numbers. I interpret this province postal code as a marker of a recent residence, corresponding

[35] I was able to confirm the details of this process by speaking to Dr. Diego Sombra at the National Registry of Persons (RENAPER), the agency that issues identification documents.

[36] My ideal outcome measure would be an indicator for "ever migrated", but this is infeasible as it would require me to be able to observe each individual's lifetime history of residences.

to the location approximately 10 years prior.[37] This then allows me to classify people who have "recently" lived outside of their province of origin as migrants, even if their current residence is in the province of origin. The variable takes a value of 0 if both the "recent" and current residences are the same as the province of origin, or if the "recent" residence is missing but the current residence is the same as the province of origin.

1.3.2 SIPA Earnings Data (1995-2016)

This is dataset provides employer-employee matched monthly earnings for all formally employed workers in Argentina from 1995 to 2016. It also includes some basic demographic information about individuals such as sex and year of birth, along with details about the employer such as firm size, industry, and the age of the firm. Certain variables are available for a subset of years, including the type of employment contract, whether the employee is in a management position, and whether the firm is public sector. Notably, this excludes people working in the informal sector, as well as self-employed people. It also excludes people working "*in gray*," essentially employees hired as contract workers in order to skirt various labor regulations. The dataset also does not include individuals' occupations, education levels, or hours worked in each month, it also does not distinguish between regular pay and bonus pay.

From this dataset, I generate the following outcomes:

Participation in the formal labor force. Formal work is generally more stable and higher paying than informal or contract work, so the simple fact of whether (and how often) someone appears in the formal workforce is a useful economic indicator. Using the voter roll data as the universe of potential workers, for each year of the SIPA data I select the men who would have been prime age workers (25–54 years old) at the end of that year, I then use the DNI to match these to the SIPA

[37] I have not been able to verify my interpretation of this variable, however the evidence that it represents a recent residence is fairly strong: it is highly correlated with the current residence variable, and I have asked Argentine colleagues to provide some examples of people who had migrated to Buenos Aires from a different province, and recent migrants seemed to be coded with their former province, while people who had migrated many years prior had a code corresponding to their current residence. My best guess is that the province code corresponds to the place of residence at the time that these records were digitized in their current format. There are a few possible methods through which I might be able to test my interpretation of this variable more rigorously, but I have not been able to do this at the time of this writing.

data.[38] Those who never appear in the formal employment data while prime age receive a value of 0, otherwise they are assigned a value of 1. Workers who are never prime age between 1995 and 2016 (i.e. those born in 1940 and earlier or 1992 and later) are excluded from all analysis of labor market outcomes.[39]

*Average lifetime earnin*gs. My primary measure of earnings is based on the average amount earned per prime-age working month (i.e. excluding months with little to no earnings reported). Calculating this measure presents a few challenges, including: identifying whether a person should be considered fully working in a given month; identifying whether pay that is substantially higher or lower than other months represents a data entry error (for example, a missing or extra 0), or if it should be included in the average; and averaging out earnings over several years when there is no consensus standard for price levels in Argentine data for much of the 2000s and 2010s.[40] For information about the time series of price levels that I constructed, see the following subsection (1.3.3). I use those price levels to convert nominal wages to real wages, using January 2005 as a base. In an effort to make sure the average I construct is reasonably representative of a worker's typical monthly earnings, I identify and exclude months from the average where the values seem excessively high or low. Specifically, I deem income for a given month to be excessively high if it is the largest amount the worker earned from the given firm in the given year, and if the natural log of the ratio of that month's earnings divided by the worker's *second*-highest monthly earnings from the same firm that year is greater than 1.6 (roughly 5$\times$ higher).[41] I similarly deem all earnings below the 0.5th percentile of (prime-age male) worker earnings for the month in question to be excessively low to include in the worker's average. Note that this measure excludes

[38]For details on the process of matching across these datasets, including statistics on the successful match rate, see Appendix A.1.1.

[39]The last cohort to have any members conscripted into the military was born in 1975, so anyone born after this is in any case excluded from analysis due to collinearity of the birth year fixed effect and conscription status.

[40]The SIPA dataset unfortunately does not include hours worked, but in general I would not want to penalize the average earnings calculation by including months in which the worker put in substantially fewer hours than usual. I also would generally not want to include, for example, large retirement bonuses in the average (these are often seem to be orders of magnitude larger than the typical month's work).

[41]I make one exception to this, which is if the maximum earning amount that would otherwise be excluded is roughly at or below the 4th percentile of worker-firm earnings for the given month: in this case, I assume that the other months probably do not represent a full working month, and exclude them.

months in which a worker does not appear in the formal sector, so average earnings of workers who are unemployed for long stretches or earn much of their income seasonally would be overstated. On the other hand, if a worker is employed part-time in the formal sector and supplements his income with informal work, I would underestimate his true monthly earnings. And of course, if a worker never appears in the formal labor force during the period covered by the SIPA data, or only appears outside of his "prime" working years, he will be excluded from analysis on this outcome entirely. For regressions on this outcome, I use the natural-log of this measure, so that the regression coefficients can be interpreted as a percent increase/decrease in earnings.

1.3.3 Additional Data

The following data are also necessary for my analysis:

Conscription assignments. These provide the correspondence between the last three digits of the DNI and the lottery sort order for each cohort from 1927 to 1976, as well as the cutoff points for exemption and for each branch. A small number of cases will be excluded from analysis because of internal inconsistencies in reported cutoffs, e.g. the highest number assigned to the Army exceeds the lowest number sorted to the Air Force. In cohorts born after 1965, some individuals from provinces that are divided into multiple military districts fall into a range of draft lottery numbers where their assigned branch would depend on the district; these ambiguous cases are also excluded from most analyses.[42]

Military base locations. Taken from the *Libro Blanco de la Defensa* (Ministerio de Defensa, Argentina 2015), I compiled a list of the provinces that currently have Army, Air Force, and Navy bases.[43] Any changes in the locations of bases either during the period when conscription was in place or afterwards are thus not reflected in the analysis that follows. This could well mean that the presence or absence of a base in a given province is recorded incorrectly, especially for earlier

[42]For specifications that do not look at specific branch effects, some of these individuals might be included if their draft number is high enough that they were surely not exempt, even if it's unclear exactly which branch they would have served in.

[43]In some cases it was unclear whether a location on the map was actually a base where large numbers of people would be posted; in these ambiguous cases, I assumed they were in fact bases, which might mean that I classified some "leavers" incorrectly as "stayers".

cohorts. It's also true that provinces are, in general, very large geographically, and in many cases very sparsely populated, such that having a base in your province would *not* necessarily imply that you completed your service close to home.

Price levels. There is broad acknowledgment that starting in 2007, the administration of President Nestor Kirchener exerted political pressure on the national statistics agency, INDEC, to manipulate its price level calculations in order to make inflation seem lower than it truly was (see Cavallo 2013; Cavallo, Cruces, and Perez-Truglia 2016; Cavallo and Rigobon 2016). This practice continued throughout the administration of his successor (and wife), Cristina Fernández de Kirchener, such that there are now no official national[44] price level statistics available between 2007 and mid-2016. Because inflation was very high over this period, to use nominal earnings to construct lifetime earnings measures would severely exaggerate earnings levels of younger people and those who appeared more often in the formal labor force in later years. As such, I combined price level data from several different sources (including Secretaría de Modernización 2018, which includes price level statistics compiled by provincial governments and the City of Buenos Aires for much of the period of data manipulation, as well as the replication data provided by Cavallo et al. 2016). Wherever possible (and reliable), I used INDEC-produced statistics for Greater Buenos Aires (as these are the most consistently available over time). For the period when these are unavailable and/or unreliable, I use data produced by a company called (confusingly) Buenos Aires City (and downloaded from the replication data for Cavallo et al. 2016), which was run by the former head of INDEC who was removed in favor of an appointee more amenable to data manipulation. Where this source is also unavailable, I use the index produced by the government of the City of Buenos Aires.[45]

[44] In practice, "national" price level statistics often in reality are based on Greater Buenos Aires price levels.

[45] Specific price levels used for each month, along with the source for each, are available from the author upon request.

1.4 Empirical Strategy

1.4.1 Effects of Conscription

To begin the analysis, I estimate the effects of the conscription on various outcomes of interest, using the basic specification:

$$Y_{icp} = \alpha + \beta\, Draft_{icp} + \delta_c \times \theta_p + \varepsilon_{icp}, \tag{1.1}$$

where Y_{icp} denotes the outcome of interest measured for individual i in birth-year cohort c from place-of-origin p, $Draft_{icp}$ is an indicator (or set of indicators) for conscription assignment status, δ_c is a full set of cohort fixed effects, and θ_p is a full set of province-of-origin[46] fixed effects. Note that conscription status can refer simply to whether an individual was assigned to any military service or not, or to a set of indicators for the specific branch of the military to which he was assigned. Because all of these are determined by random assignment, this specification yields a consistent estimate for the causal impact of one's conscription assignment on the given outcome. However, because I do not observe whether or not any individual actually completed his assigned service, all analyses will produce Intent-to-Treat estimates; my estimates may understate the true treatment effects. Birth-year cohort fixed effects are necessary for any specification as the underlying randomization is conducted at the cohort level: without these fixed effects, because earlier cohorts tended to have higher percentages assigned to military service, the coefficient on conscription would in part reflect age effects. Though the lottery assigning draft numbers was always conducted nationally at the cohort level, for some cohorts – mostly those born in 1965 or later – cutoffs for assignment to each military branch exhibited relatively high variability across different provinces, and in some cases even within a single province.[47] This makes it necessary, for

[46] In principle, it could be helpful to define the place of origin at a sub-province level in order to, for example, examine heterogeneous effects on people from rural versus urban areas, or to derive a more precise measure of the distance to the nearest military base. It might be possible to do this as an extension to the imputation procedures described in Appendix Section A.1.2.

[47] In cases where province-cohorts are divided across different military "zones" (either the 5 Army Corps or the 29 military districts), I exclude from analysis anyone whose draft number makes his conscription status ambiguous.

the affected cohorts, to interact the cohort fixed effects with the military "zone," i.e. the geographic level at which the cutoffs varied. Again, failing to do so for these cohorts would result in a bias in the estimated effects of conscription toward the average outcomes of people from provinces with more conscripts. In order to maintain a consistent set of controls all cohorts, I interact *all* cohort fixed effects with province-of-origin fixed effects, even when cutoffs for a given cohort are uniform across the country.

1.4.2 Displacement Effect: Simple Example

While the above specification is econometrically well-identified, it does not allow me to distinguish the *displacement* effect from any other possible effects of conscription. In order to demonstrate how I attempt to isolate the displacement effect from other effects of conscription, it will be useful to consider a slightly simplified example. To begin, consider the ideal experiment we would want to run. Suppose that we had a system in which conscription status was randomly assigned, and conscripts were also randomly assigned to complete their service close to or far away from home. Then we could run the very simple regression:

$$Y_i = \tilde{\alpha} + \tilde{\beta} Draft_i + \tilde{\gamma} Draft_i \times Far_i + \tilde{\varepsilon}_i$$

and the coefficient $\tilde{\gamma}$ would give an estimate of the effect of displacement (being sent far from home), net of the more general effect of conscription, on any outcome of interest. Of course, this idealized set up is not exactly what took place in reality. Even if there may have been an element of randomness in the assignment of locations for conscripts in Argentina, I do not observe the actual locations in which individuals completed their service. There is, however, a fairly close analogue that I *do* observe. Consider the case of an individual living in hypothetical Province A, which contains an Army base but no Air Force base. [48] If we restrict our attention to just this province, and exclude (for now) people who were assigned to the Navy, we can run the following regression:

[48] In all of the discussion that follows in this subsection, "Navy" could substitute for "Air Force" without any substantive change in the interpretation.

$$Y_{ic} = \tilde{\alpha}^A + \tilde{\beta}^A AnyDraft_{ic} + \tilde{\gamma}^A AirForce_{ic} + \tilde{\delta}_c^A + \tilde{\varepsilon}_{ic}^A, \tag{1.2}$$

where $AnyDraft$ is an indicator for assignment to any military branch – in this case, either the Army or the Air Force, and $\tilde{\delta}_c^A$ is a complete set of cohort fixed effects. Because the branch assignments were designated through a random lottery drawing (within each cohort c), this specification allows us to estimate a fully-identified causal impact on outcome Y for people from Province A of being assigned to the Air Force as compared the Army, represented by the coefficient $\tilde{\gamma}^A$. Importantly, for reasons of cost and ease of coordination, conscripts were generally stationed as close as possible to their home. Thus, for conscripts from Province A, the assignment to serve in the Army or the Air Force determined whether they would have to complete their service far from home, or have a reasonably high chance of staying nearby.[49] If the experience of serving in the Air Force was identical to that of serving in the Army, then the only difference between these assignments would be attributable to the temporary displacement. In this case, we would have a close approximation of our idealized experiment. Of course, one might reasonably be concerned that the experience of being assigned to the Air Force would *not* have been identical to that of the Army. Fortunately, I do not need to make such a strong assumption, as I can run the same regression (1.2) from above for individuals from hypothetical Province B, which has both an Air Force and an Army base. This will produce an estimate of the differential effect of being assigned to the Air Force versus the Army when there is no displacement. These two regressions can be pooled into a single Difference-in-Differences specification as follows:

$$\begin{aligned} y_{it} = {} & \tilde{\alpha}^B + \tilde{\alpha}^D ProvA_p + \tilde{\beta}^B Draft_{ipc} + \tilde{\beta}^D Draft_{ipc} \times ProvA_p \\ & + \tilde{\gamma}^B AirForce_{ipc} + \tilde{\gamma}^D AirForce_{ipc} \times ProvA_p + \tilde{\delta}_c^A + \tilde{\delta}_c \times ProvA_p + \tilde{\varepsilon}_{ipc} \end{aligned}$$

[49]It is worth emphasizing that it need not be the case that everyone who served in the Army was assigned to the base closest to their home. Even if many people are not sent to the closest Army base, if there is no Air Force base in the province, we can be sure that serving in the Air Force required relocating, compared with some unknown but likely fairly high probability of staying local when serving in the Army. This may bias estimates of displacement effects toward zero, meaning that they would represent a lower bound (in absolute value) for the true effects.

where $\tilde{\gamma}^D$ estimates the difference between the $Air\ Force - Army$ effect for Province A (which lacks an Air Force base) as compared to Province B (which has one). If the only difference for people in Province A versus Province B of the experience of serving in the Air Force (as compared to the Army) is that people from Province A have to leave the province while people from Province B do not, then $\tilde{\gamma}^D$ can be characterized as an estimate of the effect of temporary displacement.

1.4.3 Displacement: Main Specification

My preferred empirical specification differs only slightly from the simplified version described above; the primary difference is that I consider both Air Force and Navy service together. While nearly every province has an Army base, only some have Air Force or Navy bases. Because Air Force and Navy bases appear in different but partially overlapping sets of provinces, a few additional interaction terms are needed in order to properly calculate the displacement effect. To be precise, I estimate the following regression equation:[50]

$$\begin{aligned}
Y_{iczp} = {} & \beta_1 AnyDraft_{iczp} + \beta_2^A AirForce_{iczp} + \beta_2^N Navy_{iczp} \\
& + \beta_3^A AnyDraft_{iczp} \times NoBaseAirOnly_p + \beta_3^N AnyDraft_{iczp} \times NoBaseNavyOnly_p \\
& + \beta_3^E AnyDraft_{iczp} \times NoBaseEither_p \\
& + \beta_4^A AirForce_{iczp} \times NoBaseNavyOnly_p + \beta_4^N Navy_{iczp} \times NoBaseAirOnly_p \\
& + \beta_5^A AirForce_{iczp} \times NoAirBase_p + \beta_5^A Navy_{iczp} \times NoNavyBase_p \\
& + \delta_c \times \theta_p + \varepsilon_{iczp},
\end{aligned}$$

or, somewhat more succinctly:

[50] Because the only province that has a Navy base but no Air Force base is the very small and somewhat idiosyncratic province of Tierra del Fuego, it is actually excluded from specifications that separate the $AirForce \times NoBase$ and $Navy \times NoBase$ effects, so all interaction terms with $NoBaseAirOnly$ are dropped.

$$
\begin{aligned}
Y_{iczp} = {} & \beta_1 AnyDraft_{iczp} + \beta_2 AirNav_{iczp} \\
& + \beta_3 AnyDraft_{iczp} \times NoBase_p^C + \beta_4 AirNav_{iczp} \times NoBase_p^C \\
& + \beta_5 AirNav_{iczp} \times NoBase_p + \alpha_c \times \theta_p + \varepsilon_{iczp}
\end{aligned}
\tag{1.3}
$$

In equation 1.3, $AnyDraft_{iczp}$ represents and indicator for assignment to *any* of the three military branches; $AirNav_{iczp}$ represents two separate indicator variables, $AirForce_{iczp}$ and $Navy_{iczp}$;[51] δ_c represents a complete set of cohort fixed effects; and θ_p represents a complete set of province-of-origin fixed effects. The term $NoBase_p^C$ serves as an indicator that there is no base corresponding to a *different* (or "complementary") branch than the one with which the term is interacted. For example, for a person assigned to the Air Force while his native province has no Navy base, the term $AirNav_{iczp} \times NoBase_p^C$ would correspond to $AirForce_{iczp} \times NoNavyBase_p$, and would take a value of 1 regardless of whether or not the province has an Air Force base. The term $AnyDraft_{iczp} \times NoBase_p^C$ represents two regression terms simultaneously: $AnyDraft_{iczp} \times NoAirBase_p$ and $AnyDraft_{iczp} \times NoNavyBase_p$. While we are not necessarily interested in the coefficients on these $NoBase_p^C$ interaction terms themselves, they need to be included in order to correctly pool together the $AirForce_{iczp} \times NoAirBase_p$ and $Navy \times NoNavyBase_p$ Diff-in-Diff specifications into a single regression that properly accounts for heterogeneity in effects across provinces.

The main coefficient of interest is β_5, which gives an estimate of the effect of being assigned to a branch of the military with no base in one's native province, i.e. assignment to the Air Force from a province with no Air Force base, and/or assignment to the Navy from a province with no Navy base. The identifying assumption required for a causal interpretation of the β_5 coefficient is akin to the standard difference-in-differences assumption of parallel trends. Specifically, we need to assume that the effects of the experience of serving in the Navy or Air Force – that is,

[51] It could alternatively represent a composite variable for Air Force *or* Navy if we wanted to lump these effects together.

of all aspects of that experience *other than* the distance from home, that are distinct from the experience of serving in the Army – are orthogonal to one's proximity to the nearest Navy or Air Force Base. Note that this does *not* require that the locations of Navy or Air Force bases be "as-if" randomly assigned, nor that provinces with a Navy or Air Force base be similar on observable characteristics to those without one. It also does not require that the effects of conscription be uniform across different provinces. What it does require is that if there is heterogeneity in the effects of conscription between provinces with and without Air Force or Navy bases, that this heterogeneity can be captured by controlling for differential effects of serving in the Army. Put differently, the assumption is that if we were to build an Air Force (Navy) base in a province that previously did not have one, the difference in the effect of serving in the Air Force (Navy) versus the Army for people in that province would not systematically differ from the observed effect on people in provinces that had already had an Air Force (Navy) base.

It is not obvious that this identifying assumption is a reasonable one, but it is at least to some extent testable. One possible test would be perform separate regressions for each province and/or each cohort of various outcomes on $AnyDraft_{iczp}$, $AirForce_{iczp}$, and $Navy_{iczp}$, such that the coefficients on Air Force and Navy reflect the difference between serving in these branches versus the Army. If those differences are reasonably stable within a province over time, and especially if they are reasonably similar to estimates for other provinces that also have (or lack) an Air Force or Navy base, this would suggest that it is sensible to think that the Air Force/Navy branch-specific effects (separate from any displacement effects) are comparable across provinces regardless of the presence or absence of the corresponding base. Another potential test would be to interact various province-level statistics (e.g. province HDI or GDP per capita, out- or in-migration rates, average education levels, etc.) with the regressors from the main specification. If this does not change coefficient estimates substantially, and these additional terms do not have too much explanatory power, it suggests that we may not need to be overly concerned about confounding province characteristics.

Finally, perhaps the best possible way of ensuring that displacement effects are not being driven by province-level characteristics would be to identify cases where bases were closed down or new bases were opened. Such a scenario would allow us to observe how the Air Force/Navy vs. Army effect changes within a province as people from that province assigned to the Air Force/Navy go from being "leavers" in one cohort to "stayers" in the next (or vice versa). Essentially, this would be a Triple-Difference analysis, comparing the Air Force/Navy vs. Army effect (first difference) between provinces with and without an Air Force/Navy base (second difference) before versus after a base is built or decommissioned (third difference). Unfortunately, at the time of writing, I do not have the necessary information on base openings or closures to perform this analysis.

One final concern to note about my analysis is that there are several potential sources of attenuation bias, particularly when it comes to the effect of displacement, but also for the more straightforward conscription and military branch effects. First, because I do not observe compliance with conscription assignments, all estimates are of "Intent-to-Treat" effects rather than (generally preferable) LATE estimates. Second, there is reason to be concerned about incorrect imputation of treatment assignment variables: if there is any typo in the DNI or year of birth, conscription treatment variables become completely incorrect, as we will assign the wrong draft lottery number, and potentially also the wrong applicable cutoffs. Third, because of the way I impute province of origin (see Section 1.3.1), it is very likely that some percentage of these values is incorrect. This, in turn, might lead me to characterize an individual as being in a "No Base" province for his assigned branch when in fact his province does contain a base for that branch, or vice versa: I could be identifying some subset of "stayers" and "leavers" incorrectly. Errors in the imputed province could also bias estimates of simple conscription effects for cohorts in which the assignment cutoffs vary depending on the location by causing me to misstate the conscription assignment. All of these could attenuate my estimates of both conscription and displacement effects. A fourth set of potential sources of measurement error applies specifically to the measurement of the displacement effect, but not conscription branch effects: the metric I use for displacement is imprecise and prone to a number of potential errors. The locations of the bases themselves may have changed over time, and may have even been recorded incorrectly. I classify conscripts as “leavers” or “stayers” based on the absence or presence of a base corresponding to their branch assignment in their home province, but on top of the possibility that my province imputations are flawed, provinces themselves are quite large in terms of area, and some bases may be farther from population centers than others, so even if the base to which an individual was assigned was in his home province, he may have had to travel quite far to get there, and thus may have experienced a very similar displacement to his peers coming from outside of the province. Conversely, a person living quite close to the border of another province may be classified as a “leaver” while having in fact served quite close to home. Finally, we might be concerned that even people deployed quite close to home spent most of their time on and around the base rather than at home, so even though this would likely be an easier adjustment for people in such a scenario compared to others who were sent hundred of miles away, this experience, which I essentially use as a my experimental control group, may still have had some features of a temporary displacement.

1.5 Results

Table 1.1 shows some basic summary statistics for the men included in the data I have available. There are a few points worth highlighting from this table. First, note that the percentage of people assigned to serve in the military is generally decreasing over time. This highlights the importance of controlling for birth cohort fixed effects, otherwise effects of older age would be incorrectly attributed to being assigned to complete military service. Second, notice that the percentage of the cohort for whom the conscription assignment is ambiguous jumps for cohorts born in and after 1965: this is because these cohorts were assigned to different branches by based on the military district in which they resided. Some of the more populous provinces are broken up into multiple military districts, resulting in cases where because I only observe the province of origin and not the military district, I cannot ascertain the applicable branch assignment. For a similar reason, my estimate of the percentage of the cohort assigned to military service (Column 4) begins to diverge dramatically from the estimate in the data from Galiani et al. (2011) (Column 7): because they do not observe the place of origin for these cohorts, they can only ascertain conscription assignments for those whose draft numbers were so high or so low that their assignment status would hold regardless of where they lived.[52] Finally, it is worth comparing the numbers of people in each cohort in my data (Column 1) to the number given by Galiani et al. (2011) (Column 6). For earlier cohorts, my data tend to undershoot the cohort counts, which is likely largely attributable to mortality, as my data come from 2017 voter rolls rather than contemporaneous counts. This could mean that I exclude some people from my analysis who appear in the earlier years of earnings data, due to my inability to match those people to the recent voter rolls. For younger cohorts, the counts seem to align reasonably well, with my counts actually exceeding those in the Galiani et al. (2011) data in some cases. This could perhaps mean that I am incorrectly including some people who were naturalized relatively recently and were thus not subject to the conscription lottery.

[52]This could also bias estimates of conscription effects substantially for these cohorts if I did not include province (or military district) fixed effects.

Table 1.1: Summary Statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Cohort	# Men (1000s)	Age Served	Age rec'd DNI	% Assigned to Serve	% Assgn. Unclear	#Men in GRS data	% Assigned to Serve	% Actually Served
1941	87	21	18	100	2.1	161	100	50.9
1942	93	"	"	96.0	2.1	165	95.9	50.0
1943	101	"	"	88.9	4.8	170	88.8	50.4
1944	110	"	"	77.1	5.4	175	77.1	48.0
1945	119	"	"	73.7	0	180	73.8	44.6
1946	123	"	"	79.1	0	185	79.1	43.5
1947	131	"	"	72.1	0	190	72.1	40.6
1948	141	"	"	70.7	0	196	70.6	39.1
1949	148	"	"	78.6	0	201	78.6	39.8
1950	158	"	"	76.0	0	202	78.6	42.9
1951	165	"	"	87.1	0	204	87.1	36.0
1952	169	"	"	87.5	0	205	87.6	41.1
1953	175	"	"	85.6	0	206	85.6	43.7
1954	180	"	"	93.2	0	208	93.1	46.4
1955	184	"	"	98.2	0.2	209	97.6	46.1
1956	193	N/A	"	0	0	211	0	–
1957	197	N/A	"	0	0	212	0	–
1958	197	19	"	82.8	0	217	82.6	45.0
1959	201	"	"	68.1	0	219	68.1	46.3
1960	205	"	"	66.1	0	220	66.0	45.6
1961	210	"	"	65.0	0	221	65.1	45.9
1962	218	"	"	68.1	0	225	68.1	46.1
1963	218	"	"	65.1	0	228	65.1	43.5
1964	221	"	"	60.0	0	231	60.1	39.2
1965	223	"	"	59.8	1.5	234	57.1	38.2
1966	221	"	"	31.3	15.2	238	10.5	20.4
1967	226	"	"	23.9	17.7	241	2.5	15.6
1968	227	"	0	34.8	15.3	245	11.7	16.7
1969	239	"	"	39.4	17.2	248	20.1	17.0
1970	250	"	"	44.7	17.0	252	24.8	18.6
1971	264	"	"	19.3	26.3	255	6.1	11.6
1972	265	"	"	9.3	13.2	259	0.9	6.0
1973	269	"	"	16.6	25.3	265	4.1	7.6
1974	281	"	"	18.0	19.7	271	8.1	7.4
1975	297	"	"	19.4	20.4	277	0.1	8.3
1976	303	N/A	"	0	0	283	0	–

This table shows statistics for cohorts born between 1941 and 1976. Earlier cohorts are excluded as workers from these cohorts would not be considered "prime-age" by 1995, the earliest year for which I have earnings data. Column (1) lists the total number of men (in 1000s) in the cohort who appear in the 2017 voter rolls. Compare this to the number from the Galiani et al. (2011) replication data given in Column (6). Column (2) indicates the age at which the cohort served in the military, Column (3) gives the age at which the cohort received their DNI paperwork (0 indicates "at birth"). Column (4) gives the percentage of men from Column (1) who were asigned to serve in the military, and (5) gives the percentage for whom the military branch assignment was ambiguous. Column (7) gives the percentage assigned to service in the military according to the Galiani et al. (2011) replication data. Column (8) gives the percentage that actually served in the military, also from the Galiani et al. (2011) replication data.

1.5.1 Conscription Effects

To begin the main analysis, I investigate the overall effects of conscription and of being drafted into each of the three branches of the military: the Army, Air Force, and Navy. While this has not been the main focus of this paper, it provides a set of well-identified results to serve as a benchmark for the rest of the analysis to follow. Table 1.2 shows the overall effects of being conscripted into the military, as well as the separate effects of being drafted into each branch, on long-term migration, formal labor force participation, and earnings.[53] In all cases, these effects represent the difference compared to men who were not assigned to serve in the military, after controlling for province-cohort fixed effects.[54]

Table 1.2: Effects of Conscription

	Live outside province of origin		Appear in formal labor force		ln(Avg. earnings)	
	(1)	(2)	(3)	(4)	(5)	(6)
Draft (any)	.056		.128***		.0015*	
	(.041)		(.044)		(.0009)	
Army		.014		.092*		.0014
		(.044)		(.048)		(.0010)
Air Force		.081		.120		.0012
		(.083)		(.093)		(.0019)
Navy		.287***		.202***		.0032**
		(.068)		(.076)		(.0016)
Prov×BirthYr FEs	Y	Y	Y	Y	Y	Y
Observations	5,387,016	5,383,226	5,387,016	5,383,226	3,721,971	3,674,018
Mean(Y)	21.82	21.87	69.49	68.65	7.06	7.06

Robust standard errors in parentheses. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. Province-cohorts uniformly exempted or assigned to the same branch are excluded. Variation in the number of observations from Columnns (1) to (2) (and similarly is, in (3) vs. (4) and (5) vs. (6)) is due to exclusion in Column (1) of province-cohorts uniformly drafted that are included in Column (2) because they were drafted into different branches, and due to inclusion of people in (1) who were known to have been assigned to military service, but for whom the assigned branch is ambiguous such that they are excluded from (2).

[53]See Section 1.3 for definitions of each of these variables.

[54]Note that I use heteroskedasticity-robust standard errors but I do not cluster standard errors at the province-cohort level. This is because *within* each province-cohort, the actual randomization is done on the individual level. More precisely, it is done based on the last 3 digits of the DNI, but there is no reason to think that people sharing the same DNI last 3 digits would be more likely to share any other pre-treatment traits compared to any other randomly selected individual of the same province-cohort, and thus no compelling reason to cluster standard errors at this level.

The first thing to note about these results is that the effect sizes are generally quite small, even those that are highly statistically significant. For example, the estimated effect of being assigned to the Navy on living outside one's province of origin carries a p-value < .001, while the magnitude of about 0.3 percentage points indicates a likelihood of migration about 1.3 percent higher than the population mean of 21.9 percent. It bears repeating that there are at least two main reasons why the effect sizes I find here are likely to be underestimates: (1) all estimates are Intent-to-Treat, and thus will be biased towards zero to the extent that people were able to avoid completing military service (or joined the service voluntarily despite not being assigned to it); (2) there are various potential sources of measurement error in the assignment variables, the province-of-origin variables, the outcome measures, and even potentially the ID and birth year variables (as discussed in Section 1.3), which may attenuate my estimates.

Conscription overall does not have a significant effect on migration, but does significantly increase the likelihood of appearing in the formal labor force by 0.13 percentage points (about 0.2 percent higher than the population mean), and causes a not-quite-significant increase in average prime-age earnings by about 0.15 percent (.0015 log-points). Together, this suggests that it is unlikely that conscription is solely a temporary displacement, even if many felt the experience to be a waste of time. That is to say, there would appear to be some effects of conscription that are driven by some mechanism outside of migration alone: perhaps some benefits of training, social connections, or even basic education or health services offered to conscripts – though I cannot distinguish between these mechanisms. This stands somewhat in contrast to the finding by Galiani, Rossi, and Schargrodsky (2011) that conscription has negative labor market consequences, though they similarly find very small and often not significant effects, and the differences could be attributable to the different timing at which the outcomes were measured, or to the somewhat indirect way they constructed variables of formal labor force participation, unemployment, and earnings based on occupational categories. That the experience of conscription in Argentina had effects beyond simple displacement is consistent with the findings in that paper that conscription is associated with an increase in crime rates, and also with more recent work (Ertola Navajas et al.

2020; Gibbons and Rossi 2020) suggesting that it affected long-term personality traits, beliefs, and even rates of intimate partner abuse and violence.

When broken down to assigned branches, it appears that most of the effects are being driven by service in the Navy. Given that the Navy is the branch with the fewest number of bases, thus requiring relocation for the greatest fraction of draftees, this is consistent with (though not necessarily evidence for) the hypothesis that the migration effects, and at least some of the labor-market effects, might be driven by the initial displacement. However, it is worth reiterating that the Navy required a longer period of service of two years (instead of one), and so might have had a deeper effect due to having more time to impart particular skills, more time to forge social connections, and even perhaps due to the displacement itself lasting longer. For consistency and comparability across outcomes, I have only included in the sample men who would have been of prime working age (defined as 25–54) during at least one year of the time period for which I have wage data, i.e. from 1995–2016. This means that cohorts born before 1941 are excluded from the analysis.[55] My findings on migration, the only outcome that I can measure for earlier cohorts, remain qualitatively quite similar when these cohorts are included.

1.5.2 Displacement Effects

Next, I turn to the primary question of interest for this paper, in which I attempt to isolate the long-term effect of being sent far from home. Table 1.3 shows the results from my preferred specifications on the same three outcomes considered in Section 1.5.1.

[55]No one born after 1975 was eligible for conscription, so these cohorts are necessarily excluded from the analysis.

Table 1.3: Effects of temporary displacement

	Live outside province of origin		Appear in formal labor force		ln(Avg. earnings)	
	(1)	(2)	(3)	(4)	(5)	(6)
Draft (any)	.080	.082	.127*	.129*	.0010	.0010
	(.066)	(.066)	(.075)	(.075)	(.0016)	(.0016)
Draft×NoBaseNavyOnly	−.056	−.062	−.062	−.070	−.0006	−.0008
	(.100)	(.100)	(.114)	(.114)	(.0024)	(.0024)
Draft×NoBaseAirOnly	−2.57	–	.136	–	−.0676*	–
	(2.05)		(1.55)		(.0362)	
Draft×NoBaseEither	−.149	−.146	−.018	−.012	.0024	.0025
	(.114)	(.114)	(.117)	(.117)	(.0024)	(.0024)
Air Force	−.101	−.040	−.040	.039	−.0035	−.0018
	(.107)	(.119)	(.122)	(.138)	(.0026)	(.0031)
Navy	.011	−.024	.176*	.130	.0004	−.0005
	(.087)	(.092)	(.101)	(.107)	(.0022)	(.0023)
Air×NoBaseNavyOnly	.161	.105	.261	.188	.0076	.0061
	(.169)	(.176)	(.198)	(.207)	(.0042)	(.0044)
Navy×NoBaseAirOnly	−.298	–	2.82	–	.0011	–
	(3.32)		(2.98)		(.0649)	
Air/Navy×NoBase	**.485***		**−.131**		**.0027**	
	(.116)		(.131)		(.0027)	
Air×NoBaseEither		**.307**		**−.374**		**−.0022**
		(.214)		(.231)		(.0048)
Navy×NoNavyBase		**.552***		**−.043**		**.0044**
		(.130)		(.148)		(.0031)
Prov×BirthYr FEs	Y	Y	Y	Y	Y	Y
Observations	5,328,605	5,326,367	5,328,605	5,326,367	3,633,357	3,631,483
Mean(Y)	21.83	21.83	68.58	68.58	7.06	7.06

Robust standard errors in parentheses. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. Province-cohorts uniformly exempted or assigned to the same branch are excluded. Columns (2), (4), and (6) exclude the only province that has a Navy base but no Air Force base, which is the extremely small and rather idiosyncratic province of Tierra del Fuego. Results are nearly identical for the coefficients of interest when this province is included.

Consistent with my hypothesis, the conscription effects on migration appear to be concentrated among those who are assigned to a branch with no base in their home province. Specifically, none of the military branches have a significant effect on migration for people living in provinces that contain a base corresponding to that branch. In contrast, being assigned to serve in the Air Force or Navy when there is no corresponding base in the province of origin increases the long-term likelihood of living outside of that province by almost 0.5 percentage points, a 2.2 percent increase over the population average. This result is highly statistically significant ($p < .001$). Looking at the

Navy and Air Force effects separately (in Column 2), we see that the above effect is largely being driven by the enrollment in the Navy. Being assigned to the Air Force when there is no Air Force base in one's home province does not appear to be significantly different from being assigned to the Air Force when there is a base in the province. However, I do find a positive estimate and in terms of magnitude, the effect is a bit more than half the size of the (highly significant) effect of being assigned to the Navy in a province with no Navy base, and is an order of magnitude larger than the (slightly *negative*) baseline Air Force effect for men from provinces that did have an Air Force base.

All of this suggests that being temporarily relocated via assignment to a branch that necessitated completing one's military service outside of his home province is associated with a significantly higher likelihood of living outside that home province later in life. In other words, a temporary relocation as a young adult does seem to help lower the barriers to leaving one's native area later in life. This naturally lends itself to the question of whether the temporary displacement has discernible implications for one's long-term labor market outcomes and economic wellbeing. Columns (3)–(6) of Table 1.3 show the net effects of temporary displacement on formal labor force participation and log average formal monthly earnings. It is important to note that these are the *overall* effects of displacement, rather than the effects for those actually induced to eventually migrate, because a large fraction of temporarily displaced conscripts *do not* end up migrating out of their home province. As such, the fact that we do not see significant gains in formal labor force participation or earnings as a result of the displacement does not necessarily indicate that there are no unexploited returns to migration. Rather, because the fraction of people induced to migrate is quite small in absolute terms, it could well be that the returns to these workers in terms of employment and wages are being masked by the null, or even potentially negative, effects on people who returned to their native province after conscription and remained there (or avoided serving altogether).

It is also worth noting that we continue to see certain effects of conscription on labor market outcomes in these results that appear to be unrelated to the presence or absence of a nearby base.

For example, we see that the draft overall has a small positive effect on formal labor force participation, and that being sent to the Navy has an additional positive effect on top of this. We also saw in Table 1.2 that at least the Navy has a small positive earnings effect. Because we see little evidence of a migration effect for conscripts with a base nearby, it is unlikely that these facts could be explained by the displacement effect of living on a base, even if that base was close to home. Rather, this suggests that there may have been some level of labor-market value in the conscription experience, perhaps due to some aspect of the training, development of "non-cognitive" skills like discipline or grit, or valuable social connections.

Finally, it is worth noting that while never significant in the above results, for certain outcomes the effect of being assigned to the Air Force when there no Navy base nearby is comparable or even larger than the Air Force effect when there is no Air Force base nearby. This is of concern because it may indicate a violation of the identifying assumption discussed in the previous section: that the Air Force − Army effect in provinces with both an Air Force and Army base can serve as a reasonable control for the non-displacement aspects of the Air Force assignment in provinces that do not have an Air Force base. If there are appreciable differences in the value of being assigned to the Air Force that are potentially correlated with the presence of an Air Force base, then we might not be able to ascribe a causal interpretation to the Air Force × No Base coefficient, measuring specifically the displacement effect on a given outcome.

Of course, the ability to ascribe a causal interpretation to the estimates of the displacement effects depends on the validity of the underlying identifying assumptions. Because I cannot directly prove that these assumptions hold (and indeed, there is at least some potentially contradictory evidence), I present an alternate set of specifications in Tables 1.4 and 1.5, in which I focus on provinces that do have an Army base, but lack either an Air Force base, a Navy base, or both. To understand the advantage of this approach, consider a province that does have an Army base, but no Navy base. For men from this province, the random assignment to the Army versus the Navy is, in effect, an assignment to perform military service either (likely) fairly close to home, or relatively far away with (near) certainty, respectively.[56] Because this assignment is fully random, there is little question as to the validity of the causal interpretation of estimates from these specifications. The concern with this set of specifications, however, is that the difference between serving in the Navy (or Air Force) versus the Army may not simply be the distance traveled to reach the base. Different branches may offer different training, skills, and a different network of peers and officers, and importantly the standard service term for the Navy was two years as opposed to one year for the Air Force and Army. Further, several provinces are quite large geographically, so even reporting to the closest base in such a province might mean traveling quite far from home; conversely, for some provinces, the closest Air Force or Navy base might have not have been inside of the province, but might have still been just across provincial boundaries and thus quite close to home for at least of subset of residents.

Table 1.4: Migration response: Army (base nearby) vs. Air Force/Navy (no base nearby)

	Live outside province of origin			
	No Air Base (excl. Navy) (1)	No Navy Base (excl. Air Force) (2)	Neither Navy nor Air Base (3)	Missing Air and /or Navy Base (4)
Draft (any)	-.101	-.010	-.095	-.015
	(.095)	(.060)	(.095)	(.060)
Air Force (no base)	.298*		.302*	.249
	(.179)		(.179)	(.177)
Navy (no base)		.528***	.755***	.530***
		(.092)	(.148)	(.092)
Prov×BirthYr FEs	Y	Y	Y	Y
Observations	1,286,061	2,945,748	1,389,109	3,015,221
Mean(Y)	26.73	22.42	26.76	22.52

Robust standard errors in parentheses.$^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$
Province-cohorts uniformly exempted or assigned to the same branch are excluded.

With all this said, Table 1.4 shows results for long-term migration that are fairly consistent with the estimates presented in the preferred specifications, and consistent with the idea that being

[56] It is certain in the sense that if men assigned to the Navy do indeed serve in the Navy, there is no way to do so in their home province. This does not account for people getting out of the service or switching to a different branch; it could also be that some of these people are assigned to a base that is still fairly close to home despite being in a different province.

assigned to live outside of one's home province could make it substantially more likely that a person moves again to a new province later in life. It is worth noting that assignment to the Air Force does not seem to have quite as strong an effect as assignment to the Navy, which might suggest that living outside of one's home province for a longer time period causes a larger increase in the propensity to migrate, though I cannot rule out the possibility that the difference is due to other aspects of the experience of being assigned to either of these branches.[57] However, it is notable that the coefficients on both the Air Force and Navy are very similar to those estimated in Table 1.3, meaning that using the effects of serving in the Air Force/Navy (versus the Army) when there *is* a base nearby as a control does not change the estimated effects substantially.[58] This, in turn, suggests that it might be reasonable to attribute the migration effects of conscription largely to displacement, bolstering the claim of a causal relationship.

[57]It could also be that I have incorrectly identified the locations of Air Force bases, though it seems less likely that I have incorrectly recorded a province as missing a base, and more likely that in some provinces I incorrectly identified non-base Air Force properties that did not actually house troops as bases. Of course, given that I used records giving current base locations as my data source and that the prestige, budget share, and most importantly number of personnel dedicated to the military have fallen precipitously since the return to democracy in 1983, it is quite possible that several bases have been decommissioned that would in fact have held conscripts, and as such that men I've categorized as "displaced" were in fact able to serve quite close to home.

[58]With the exception of Column (3), which shows a larger effect for assignment to the Navy.

Table 1.5: Labor market effects: Army (base nearby) vs. Air Force/Navy (no base nearby)

	No Air Base (excl. Navy)	No Navy Base (excl. Air Force)	Neither Navy nor Air Base	Missing Air and /or Navy Base
	Appear in Formal Labor Force			
	(1)	(2)	(3)	(4)
Draft (any)	.094	.089	.082	.094
	(.092)	(.063)	(.091)	(.063)
Air Force (no base)	-.281		-.296	-.321*
	(.186)		(.186)	(.184)
Navy (no base)		.087	.343**	.091
		(.103)	(.151)	(.103)
Prov×BirthYr FEs	Y	Y	Y	Y
Observations	1,286,061	2,945,748	1,389,109	3,015,221
Mean(Y)	71.39	69.40	70.97	69.28
	ln(Average Earnings)			
	(1)	(2)	(3)	(4)
Draft (any)	.0037**	.0017	.0038**	.0017
	(.0018)	(.0013)	(.0018)	(.0013)
Air Force (no base)	-.0042		-.0043	-.0034
	(.0037)		(.0037)	(.0036)
Navy (no base)		.0039*	.0019	.0039*
		(.0021)	(.0030)	(.0021)
Prov×BirthYr FEs	Y	Y	Y	Y
Observations	913,262	2,031,954	980,636	2,076,254

Robust standard errors in parentheses.$^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$
Province-cohorts uniformly exempted or assigned to the same branch are excluded.

Table 1.5 shows analogous specifications for formal labor force participation and log-earnings. Unlike the corresponding migration results, no obvious pattern jumps out in these results. Depending on the specification, we see some positive earnings effects from assignment to the Army or Navy, and possible positive effects on formal labor force participation from assignment to the Navy, but these effects are not consistent across specifications, and do not seem to align in any clear way with displacement. Indeed, all the coefficients on the effect of assignment to the Air Force are *negative* (though generally not significant) for both formal labor force participation and earnings.

While this does seem to imply that the effects of conscription were not *restricted* to effects from displacement – that is to say that the labor market effects of conscription seem to derive at least in part from some mechanism other than displacement – it does *not* mean that increased migration rates failed to translate to labor market gains. Because many workers were *not* induced to migrate by the initial displacement, all we can say is that the average net effect of the displacement was insufficient to improve overall earnings or employment, but this does not preclude the possibility that there were in fact positive effects on those who were induced to migrate, these might just be undetectable due to null (or even possibly negative) effects on those whose migration decisions were unaffected.

1.6 Conclusion

When we look at the high proportions of people who live relatively close to their place of birth, it is easy to imagine that this might be simply because people simply prefer the familiarity of their native place, where they are most likely to have family ties, a broader social network, and a large amount of useful local knowledge. However, this paper, while not contradicting this hypothesis, provides evidence that there is more to it than this. Specifically, I find evidence that temporarily displacing someone from their native area makes them about 2–3% more likely to move away again in the long run.[59] If it were the case that people had full information about the potential benefits of migrating elsewhere, and simply chose not to do so because these benefits were outweighed by the disutility of living outside of their native place, we would not expect those preferences – and the resulting decisions about migration – to be changed by spending a short period living outside of that place. Yet I find that random conscription assignment in Argentina to a branch of the military that did not have a base nearby was associated with an increase in the long-term likelihood of living outside one's native province, and that this effect did not apply to people assigned to the same branch of the military but for whom there was a corresponding base nearby.

However, in contrast to the hypothesis that this conscription program constituted a pure waste

[59]The true effect is likely larger, for various reasons discussed throughout the paper.

of time, or that any long-term labor market benefits would be confined to those induced by displacement to migrate at higher rates, I find evidence of modest labor market benefits to conscripts that do not seem to be restricted to those who served far from home. Specifically, I find a 0.1 percentage-point increase in formal labor force participation rates that appears to apply to draftees across the board (though this effect is not significant when restricting attention to provinces with bases missing), a small additional increase of about 0.2 percentage points for being assigned to the Navy, and a small increase in earnings of about 0.3–0.5 percent from assignment to the Navy irrespective of place of origin.[60] None of these effects is restricted to conscripts who were sent far from home, suggesting that they are probably attributable to some other aspect of military service.[61] Moreover, I do not see any significant difference in the effects on these outcomes among the "displaced" conscripts when compared with their non-displaced counterparts.

This is not to say that there are no unexploited returns to migration in this setting. It's important to keep in mind that the increase in migration rates among displaced conscripts is fairly small, such that even a substantial earnings and/or employment benefit accrued to those induced to migrate may be disguised by the null (or even negative) effects on the people who returned home after military service and stayed there. Still, it seems clear that the labor market benefits of conscription in Argentina, though they may be small, are not solely attributable to displacement. Further research would be required to try to identify the specific mechanism for these effects, and to understand whether there are any conclusions that we might be able to generalize to other developing countries. We cannot say from this study whether these effects could be generated through a similar training program, perhaps unconnected to the military, and perhaps even voluntary: there is not enough information to know whether the "mandatory" and "military" aspects of mandatory military service were necessary to achieve the observed labor market effects, and whether this might be different in a developing country compared to a developed one. What we *can* say is that given the small magnitude of observed effects, it's quite probable that the costs of this program, in terms of lost

[60] These effects are also likely somewhat understated because I do not observe whether individuals actually completed their assigned military service.

[61] It's unlikely that these effects are attributable to the minor displacement of living close to home but in barracks, as I do not find any migration effects among the conscripts that I classify as more likely to have served close to home.

civilian labor market experience, the fiscal cost of administering the program, and the less tangible costs in terms of crime, violence, and attitudes toward politics and democracy, could easily have outweighed the modest benefits. There are likely many more cost-effective and humane way of achieving comparable or better outcomes.

A number of migration-related questions remain open for further research. The most pressing question is whether there was in fact an economic benefit to those who were induced to migrate in the long-term by the initial temporary displacement: did these people earn more, were they better able to adjust via migration to local labor-market shocks, were they made more flexible in any other way in terms of being able to switch between jobs or industries more easily? Is there an alternative intervention that would be *more* effective in terms of encouraging future migration, or less authoritarian than compulsory relocation? All of these questions remain unanswered. Also unanswered is *who were* the people induced to migrate – did they tend to be higher or lower skilled workers, is it really the case that this program helped to incorporate people from poorer, remote areas into more productive sectors? Lastly, given the potentially high costs and the unclear benefits of inducing higher migration rates, it is reasonable to ask whether it would be better to implement place-based policies that target workers in areas facing a localized downturn, rather than trying to encourage these workers to migrate in search of opportunities elsewhere.

Chapter 2: Does Ramadan Affect Daily Caloric Intake? Evidence from Rural Malawi

2.1 Introduction

Ramadan is the holiest month on the Islamic calendar, a time for prayer, giving to charity, gathering with friends and family, and, perhaps most notably, refraining from food and drink during daylight hours. A number of papers in economics have taken advantage of (plausibly exogenous) year-to-year variation in the timing of Ramadan, as well the differences in the length of the fast due to latitude and the time of year, to study the effects of Ramadan on various outcomes, including fetal health (and related later-life consequences) (Almond and Mazumder 2011; van Ewijk 2011; Majid 2015), work output (Campante and Yanagizawa-Drott 2015; Hu and Wang 2019; Schofield 2020), and even traffic accidents (Gulek 2021). Most economic studies either rely on indirect evidence of this, or accept that the mechanism underlying the effects they observe is something of a black box.[1]

This paper considers the setting of rural Malawi, and looks at whether it is indeed true that caloric intake consistently falls for Muslim households during Ramadan. I find that there is not consistent evidence of such an effect in this setting. Nevertheless, I do find evidence of a decrease in labor supply, which calls into question whether the effects of Ramadan that have been observed in other settings, particularly effects on work and productivity, are necessarily attributable to a reduction in calories consumed. I also find some potentially interesting variation in this result across different rounds of data. Specifically, the variation I observe is consistent with the idea that the effect of Ramadan, at least in this context, might depend upon the timing of the holiday relative

[1] To my knowledge, the only economics papers that try to directly estimate the effect on caloric intake are Schofield (2020), and its predecessor paper (Schofield 2014).

to the agricultural cycle. While it is important to note that the evidence for this interaction is far too limited to draw any firm conclusions at present, if the pattern I observe were found to apply more broadly, it would have interesting implications for the broader understanding of Ramadan observance and of seasonality in consumption.

To study this question, I use data from Malawi's Integrated Household Survey (IHS), Rounds 2–5 (NSO 2005b; 2012; 2018; 2021). In each round of this survey, a module is administered asking about a large number of different food items that household members may have consumed. For any item that consumed within the last week, respondents also indicate the total quantity consumed by all household members, allowing me to construct an estimate of total calories consumed by the household. Under the assumption that households interviewed during Ramadan are not systematically different from those contacted at other times of year, I use a Difference-in-Differences approach to look at whether calorie consumption per person in Muslim households is significantly different from consumption in non-Muslim households during Ramadan. Comparing this difference to households observed at other times of year then provides an estimate of the effects of Ramadan on caloric intake in Muslim households. I use an analogous methodology to study whether Ramadan affects the number of hours worked per week by Muslim working-age adults.

The primary finding in this analysis is that Ramadan has no significant effect on average caloric intake for Muslim households. Of course, the absence of evidence for such an effect does not necessarily imply that no such effect exists. However, a 95% confidence interval gives us a lower bound on this effect of a reduction by approximately 120 calories per person per day, or about a 5% reduction in calories, which would be a much smaller effect than has been identified in other research (e.g. Schofield 2020). Nonetheless, I also find a decrease of almost 20% in hours worked, a reduction of about 3 hours per week for working adults.

It is possible that the decrease in hours worked is attributable to something other than a reduction in calories. However, I cannot rule out the possibility that a decrease in caloric intake could be a contributing factor. Because my measure of calories consumed is given at the household level rather than the individual level, and because children, the elderly, and the ill would not be expected to fast during Ramadan, it is possible that I am underestimating the true effect on caloric intake among healthy, working-age adult Muslims. It is also plausible that temporary hunger due to fasting during daylight hours could make it more difficult to do work, even if the calorie deficit is negated by the post-fast meal (this would be consistent with the findings in Hu and Wang 2019, for example).

One other possibility to consider is that the effect of Ramadan could be different from year to year, depending on where it falls in the agricultural cycle, especially in a country like Malawi where there is pronounced seasonal variation in food consumption. In this light, it is interesting to note that I do find a significant decrease in consumption among Muslims during Ramadan in one round of data, in which Ramadan falls just before the onset of the annual hunger season. Interestingly, I also find some (limited) evidence of a potential increase in calorie consumption during Ramadan in two more recent rounds of survey data in which Ramadan overlapped with the main harvest season, which is a time of relative abundance for most households. This is somewhat puzzling from a theoretical perspective: if baseline consumption levels are higher during the harvest season, we would generally expect marginal utility of consumption to be lower, and thus more willingness to decrease total consumption during Ramadan when it overlaps with the harvest. Conversely, when Ramadan falls closer to the hunger season, we would expect baseline household consumption to be quite low already, such that households would be less willing to reduce consumption any further. If anything, we seem to see the opposite effect: consumption during Ramadan falls by significantly more when it Ramadan occurs shortly before the hunger season compared to when it overlaps with the harvest.

We should be quite cautious in trying to explain or interpret differences between rounds – there is no evidence of a causal relationship, and with only four rounds of survey data to draw from, these differences could be purely idiosyncratic or coincidental and not indicative of any broader pattern that would be consistent over time. However, it is at least worth noting the possibility of a link between the timing of Ramadan and the agricultural season. This is particularly true in the context of Malawi, where, as in many other poor countries and regions, food insecurity is a recurring seasonal phenomenon, and this seasonal dimension has important implications for food prices, labor markets, the efficacy of informal insurance mechanisms, and the risk of famine (see Devereux et al. 2008; Chirwa et al. 2012). However, if we were able to establish evidence for such a pattern, it would be consistent with the hypothesis that when food is abundant, post-fast meals are larger, and might even outweigh the effects of the fast, while when food stocks are running low, these meals become more modest and do not cancel out the effects of the fast. It could also potentially speak to one reason we see such dramatic seasonal variation year after year in consumption in places like Malawi: if households increase consumption during Ramadan when it falls during the harvest, and only use it as an opportunity to cut back on consumption – and possibly to save a bit more food for the hunger season – when Ramadan falls shortly before the hunger season, it could mean that they are simply not taking the annual hunger season into consideration when deciding on their food consumption until it is in the very near future, and thus more salient in their minds. If this sort of salience effect plays an important role in economic decision making, it could help explain the persistence of seasonal hunger and other similar phenomena, and could also help policymakers design tools to address these sorts of concerns. Of course, more research would be needed to draw any conclusions as to this specific hypothesis.

This paper contributes to several strands of economic research. First and foremost, it contributes to the literature studying the effects of Ramadan on various social and economic outcomes. A number of papers have looked at the effects of Ramadan on overall productivity (Campante and Yanagizawa-Drott 2015; Hu and Wang 2019; Schofield 2020), subjective wellbeing (Campante and Yanagizawa-Drott 2015), fetal and long-term health outcomes (Almond and Mazumder 2011; van Ewijk 2011; Almond et al. 2014; Majid 2015; Lee et al. 2020), traffic accidents (Gulek 2021), and more (Haruvy et al. 2018; Hodler et al. 2020; Shalihin et al. 2020; Wang et al. 2020). It also contributes to a broader literature on the relationship between religion and economic and social outcomes (McCleary and Barro 2006; Hoverd and Sibley 2013; Benjamin et al. 2016; Iyer 2016; Kuran 2018; D'Haene et al. 2019). This is, to my knowledge, the first paper to directly study the impact of Ramadan on calorie intake in a poor, Sub-Saharan African country such as Malawi.

Second, this paper contributes to the vast literature using survey data, in particular from the World Bank's Living Standards Measurement Studies (LSMS), to measure caloric intake at the household level. In particular, it relates to several papers using Malawi's IHS data (which is part of the LSMS project) to study questions of food security and access to nutrition in that country (Ecker and Qaim 2011; Chirwa et al. 2012; Headey and Ecker 2013; Pauw et al. 2014; Verduzco-Gallo et al. 2014; Beck et al. 2015; Gilbert et al. 2019; Schneider 2021). One major challenge in collecting and analyzing item-by-item food consumption data is describing quantities consumed (Smith et al. 2014 discusses this challenge, along with many others). Forcing households to report quantities in standard units such as liters or kilograms is likely to result in very noisy estimates and strain on the respondent. On the other hand, allowing respondents to provide broad, non-standard units, as the Malawi IHS does, creates difficulties for the data analyst hoping to translate "heap" or "plate" to grams, in order to measure quantities of calories or other vital nutrients; several researchers have pointed out inconsistencies in the default conversion factors provided with the Malawi data, and have proposed procedures to address these concerns (Ecker and Qaim 2011; Verduzco-Gallo et al. 2014; Gilbert et al. 2019). I develop several refinements to the procedure for constructing kilogram and calorie metrics from the IHS data; this is the first paper to my knowledge to construct calorie measures for the recently-released IHS5 data, and the only one to my knowledge to pool data from all four rounds from IHS2 (2004–2005) to IHS5 (2019–2020). I will explain why I think these refinements offer a meaningful improvement to the quality of the calorie measure, and briefly discuss potential implications for the collection of similar data in the future.

Lastly, as mentioned, this paper makes a small contribution to the broad literature on seasonal hunger and food insecurity. This is a phenomenon that is widely documented in various regions throughout the world where people depend largely on rainfed agriculture (Dercon and Krishnan 2000; Devereux et al. 2008; Khandker and Mahmud 2012; Devereux et al. 2012; Bryan et al. 2014; Basu and Wong 2015). It is also poorly understood from an economic perspective, in that a standard economic model of consumption decisions would suggest that households would respond to a repeated seasonal decline in the availability of food by adjusting consumption levels in order to save in other seasons. This paper offers some preliminary suggestive evidence as to what may be the mechanisms underlying this phenomenon.

2.2 Background

2.2.1 Ramadan

Fasting during daylight hours each day of the month of Ramadan is one of the central tenets (or Five Pillars) of the Islamic faith: all adult Muslims are expected to do so, with exceptions for the ill, the elderly, etc. (Almond and Mazumder 2011). The Islamic calendar follows a lunar cycle, and each calendar year is about 11 days shorter than a year on the Gregorian calendar. This means that the timing of Ramadan changes from year to year, and cycles through each season over the course of about 33 years.

A growing body of economic literature has used the changing timing of Ramadan and length of the fast day as a source of exogenous variation to study the effects of Ramadan observance on a number of outcomes. Almond and Mazumder (2011) were to my knowledge the first in the economics literature to use Ramadan timing, and variation in the length of the fast, as a source of exogenous variation. They find that *in utero* exposure to Ramadan during the first month of pregnancy increased the likelihood of disability as an adult by around 20%. They also find that the longest observed Ramadan fast (in terms of the length of the fast day) overlapping with the first month of pregnancy decreased the male-to-female sex ratio at birth by more than 6 percentage points, and that the overlap of a longer Ramadan fast day with pregnancy at any stage decreased birth weights by about 18 grams on average. Importantly, these results are consistent with the medical literature on effects of skipping meals during pregnancy, and do not require a reduction in calories over the course of the day. van Ewijk (2011), Almond et al. (2014), and Majid (2015) build on these findings, showing that *in utero* exposure to Ramadan has persistent adverse effects throughout the life-cycle: lower reading and math scores among children; decreases in hours worked among adults; and poorer health in old age, including increases in heart problems and Type

2 diabetes.

Campante and Yanagizawa-Drott (2015) use variation in the length of the Ramadan fast based on a country's latitude and the dates of Ramadan to show that the holiday slows GDP growth in Muslim countries, but improves measures of subjective wellbeing. Hu and Wang (2019) study hourly productivity among salespeople working at a cosmetics outlet, a job requiring little physical activity but a relatively high level of cognitive effort, and find that productivity starts to wain only in the last two hours of the fast, by which point Muslim workers would have been fasting for 12 hours, and that these workers quickly recover to full productivity after sunset when they are able to eat and drink again. Schofield (2020) looks at agricultural productivity, and finds that Ramadan has no significant effect on farm work hours, but a significant decline in output of about 1 percent, consistent with a fall in productivity for Muslim workers of approximately 20–40 percent. She also finds suggestive evidence that this decline is attributable to reduced calorie intake, particularly because productivity does not recover immediately after Ramadan, consistent with the time needed for the body to recover from caloric deficits accumulated over a month of decreased food intake. Other recent papers have studied effects of Ramadan on traffic accidents Gulek (2021), prosocial behavior (measured via a Dictator Game) Haruvy et al. (2018), and public support for and incidences of terrorism Hodler et al. 2020. These latter two papers demonstrate some important aspects of Ramadan observance outside of fasting. In Haruvy et al. (2018), people were more generous during Ramadan when they were abstaining from food (but not after eating their evening meal), whereas generally people become less generous when abstaining from food, suggesting that religious fasting may serve as a reminder to be more charitable to those in need. Along similar lines, Hodler et al. 2020 find that in years and regions when the Ramadan fast was longer due to longer daylight hours, public support for terrorism declined in polling, and an hour increase in the length of the fast day corresponded to a 2–3 percentage point drop in terrorist attacks during that year, particularly in the sorts of attack that require some level of public support or complicity (such as providing shelter or an escape route) – suggesting that the more intensive fast may perhaps have promoted greater religious introspection and rejection of violence.

A common theme throughout these papers is that it is difficult to assess the precise mechanism driving the observed effects – whether it is lack of food, dehydration, or some other social or behavioral aspect of Ramadan observance. Of the papers mentioned in the preceding discussion, the only one that attempts to directly measure the effect of Ramadan on daily calorie intake is Schofield (2020), who estimates a decrease in caloric intake among Muslims in India of approximately 600 calories per day during Ramadan.

Other papers in the medical literature (Karaağaoğlu and Yücecan 2000; Toda and Morimoto 2004; Ziaee et al. 2006) have attempted to study the direct health effects of Ramadan fasting. They each find evidence of weight loss during Ramadan, and also report some minor adverse effects such as irritability or loss of interest in work. This consistent finding of weight loss would generally imply that calorie expenditures are exceeding calorie intake, and thus that there is likely a sustained reduction in caloric intake. However, Sadeghirad et al. (2012) find in a meta-analysis of studies of health effects of Ramadan substantial heterogeneity across different settings. They do find consistent, significant evidence of weight loss, averaging to about a 1.25 kg reduction, over the course of the month; this seems to be true in most regions that they consider, and it also seems to be the case that this weight is re-gained on average by 2–6 weeks after the end of Ramadan. However, paradoxically, they do not find an alignment between weight loss and calorie reduction. In studies conducted in the Middle East and East Asia, they find calorie intake reductions of about 150–200 calories per person per day on average. On the other hand, in studies conducted in North Africa, they document an *increase* on average of over 250 calories consumed per day, despite the evidence of weight *loss* in this same set of countries at a similar scale to the weight loss observed in the Middle East and East Asia. Clearly, there is still much to be learned about the effects of Ramadan in different settings.

This paper looks at the relationship between the Ramadan fast and caloric intake, as well as labor supply, in the southern African country of Malawi. In principle, it seems intuitive that the Ramadan fast would generally cause a reduction in calorie consumption, though the evidence for this is mixed, as discussed above. In reality, people might eat more or better food during

Ramadan, especially during the post-sundown *Iftar* meal, and poorer Muslims may benefit from shared meals and charity, as wealthier people in the Islamic faith are expected to give charitable contributions during Ramadan. I also consider ways in which the timing of Ramadan might interact with agricultural seasons. For example, during the yearly hunger season (see Section 2.2.3), it would seem logical that households who would be skipping meals even in the absence of the fast would be unlikely to reduce consumption further for the Ramadan fast. Conversely, during the peak season, we might in theory expect a larger reduction in consumption, which in turn might generate some quantity of savings to carry over for the rest of the year. On the other hand, if post-fast meals are quite large and are attended by the entire family, including those (such as children and the elderly) who are not expected to fast, we might even see an increase in calorie intake for the household overall during Ramadan. All of this suggests that it could be fruitful to study empirically how overall consumption behavior changes during the Ramadan fast. The primary question is whether the Ramadan fast leads to a reduction in food consumption overall during that month. Secondary to this question is whether/how that effect varies depending on timing of Ramadan relative to the agricultural cycle.

2.2.2 Challenges in measuring calorie consumption

Household calorie intake has long been recognized as a vital indicator of welfare and a necessary component in the calculation of poverty statistics. This is particularly true in developing countries, where food expenditures tend to make up the large majority of most households' expenditures Subramanian and Deaton (1996); Smith et al. (2014); Eli and Li (2020). There are a number of challenges when it comes to collecting and analyzing data on food consumption. In general, it is considered best practice to measure households' food intake by asking them to recall quantities consumed of an extensive list of food items. However, such surveys vary widely in their design and are prone to various sources of measurement error Gibson et al. (2014); Smith et al. (2014). As Beck et al. (2015) note, even when using survey data that have already been collected, methodology can make a large difference in measures of great importance to policymakers: they estimate a decrease in poverty rates of 3.4–8.4 percentage points in Malawi (depending on methodological choices) from the 2004–2005 agricultural cycle to the 2010–2011 cycle, in contrast to the disappointing official estimate of just a 1.8 percentage point decrease.

It is important to acknowledge that calories consumed are not a perfect indicator of household welfare. For one, there is the widely documented phenomenon of the Engel curve: as households get richer, the share of their budget devoted to food expenditure tends to decline. Further, wealthier households tend to substitute cheaper sources of calories to more preferred ones, which tends to result in an increase in the amount spent per calorie Subramanian and Deaton (1996). Looking exclusively at calories can also be misleading, as individuals consuming adequate levels of calories might still face significant deficiencies in vital micronutrients Ecker and Qaim (2011); Headey and Ecker (2013). [2]

Whereas caloric intake tends to be highly correlated with other measures of household welfare, particularly in developing countries, looking at the evolution of per capita calorie consumption over time can also generate misleading comparisons. Deaton and Drèze (2009) offer a stark example of this in India, in which they show that, despite rapid economic growth from the 1980s to early 2000s, per capita calorie consumption actually declined. They speculate that it might be possible to account for this at least in part through increased access to labor saving technologies and declining levels of physical activity. Eli and Li (2020) follow up on this; constructing a measure of energy expenditure, they find that totally energy expenditure decreases only minimally, largely because the proportion of children in the population falls dramatically over this period. They suggest, following evidence from Duh and Spears (2017), that improvements in sanitation and the disease environment might account for the apparent discrepancy between falling calorie consumption and improved metrics of health and nutrition. Specifically, if fewer calories are being wasted in the form of diarrhea, this might apply a reduction in total household caloric requirements. Deaton (1997) also describes complications when trying to account for different caloric needs of different household

[2] Because of this, it is generally important to consider measures of dietary diversity, and ideally also specific nutrient intake, to get a more complete picture of household nutritional status. This is beyond the current scope of this paper.

members. Of particular note, it is, for obvious reasons, impossible to use household level data to determine intra-household food allocation. This might present a challenge for measurement of the effect of Ramadan, as not every household member will generally be required to observe the fast.

Smith et al. (2014) assess surveys used to measure caloric intake across 100 different countries in order to identify best practices for data collection, and the extent to which surveys adhere to these best practices. Their results are generally underwhelming. Many of the surveys included fail to meet some of the most important criteria that they describe. Fortuitously for this paper, Malawi's IHS food intake modules stand out in this regard, largely conforming to best practices. In particular, Malawi's questionnaires collect information about a wide variety of specific food items, their provenance (own production, purchase, or gift), use a one-week recall (which is thought to be a reasonable timeframe for most households), account for seasonality in the survey design by spacing out interviews relatively evenly across the year, and make some effort to account for food consumed outside of the home.[3]

Even so, potentially important sources of measurement error remain. One of the most vexing problems in working with the Malawi data is one for which Smith et al. (2014) offer no clear-cut solution: the problem of unit measurements.[4] Malawian households are given flexibility to report quantities of consumption of each item from a wide variety of non-standard units. Converting these units to a standard metric on the back end poses a significant challenge. Various researchers (Ecker and Qaim 2011; Verduzco-Gallo et al. 2014; Beck et al. 2015; Gilbert et al. 2019) have noted that the unit conversions provided by Malawi's National Statistics Office (NSO) are incomplete and in many cases appear incorrect or internally inconsistent. Each have used different methodologies to try to address this concern. I describe the methodology that I employ in Section 2.3.2.

[3]Perhaps even more impressive is that these practices were implemented even in the earlier rounds of the survey, well before the Smith et al. (2014) paper was released.

[4]Every method presented for collecting food quantities comes with significant drawbacks, so Smith et al. (2014) do not take a firm position on which methodology is best for collecting food quantities. They do, however, recommend use of demonstration methods, in which respondents use photos or other reference points to help clarify quantities consumed. In recent rounds, the IHS has provided a photo guide for reference purposes for certain food items and units; there might be room to expand upon this practice.

2.2.3 Seasonal hunger: the "father of famine"

A number of considerations demonstrate the importance of focusing attention on the seasonal nature of hunger for many of the world's poorest people. First, seasonality itself can both cause and exacerbate hunger, particularly in many of the poorest developing countries. Devereux, Vaitla, and Hauenstein Swan (2008), in a discussion of the 2002 famine in Malawi that resulted in tens of thousands of hunger-related deaths, argue that part of the reason that the famine was so severe was that households' coping resources were largely depleted during the hunger seasons of several preceding years. In this context, they refer to seasonal hunger as the "Father of Famine," arguing that it is impossible to understand and prevent famine without first tackling annual hunger seasons. With much of the developing world still heavily dependent upon rainfed subsistence agriculture, it is perhaps not surprising that seasonality plays a larger role in such places than in more developed countries. Another possible risk of ignoring seasonality is that it may lead to a severe underestimate of rates of poverty and food insecurity. Khandker and Mahmud (2012), in their extensive study of seasonal hunger, suggest that "most of the world's acute hunger and undernutrition occur in the annual hunger season." Further, they argue that many households that would usually be classified as non-poor on average actually do slip below the poverty line or face food deficiencies during the hunger season. This is consistent with other literature suggesting that volatility often causes as much or more hardship for the world's poor than low average income itself (see for example, Collins et al. 2010). Seasonal hunger might also affect other aspects of economic life in developing countries. For one thing, it is well-documented in the literature that many people in developing countries depend on informal insurance arrangements to smooth economic shocks (see for example, Townsend 1994, and Munshi and Rosenzweig 2016). When there is an aggregate shock that affects the entire insurance network simultaneously, as might be the case in a hunger season, these sorts of arrangements are limited in their effectiveness. There are also several channels through which seasonal hunger could generate a poverty trap, pushing people to take actions that, while perhaps necessary for short-run survival, may be harmful in the long run. Examples of this include working on others' land for a low wage (c.f. Jayachandran 2006) at the expense of planting or harvesting on one's own land; selling off productive assets at a price far below their long-term value (c.f. Rosenzweig and Wolpin 1993); consuming crops before fully ripe, thus sacrificing much of the nutritional value; and selling crops during the peak harvest season when prices are lowest rather than saving them for later in the year when prices will usually be substantially higher (Devereux, Vaitla, and Hauenstein Swan 2008). Hunger could also have a direct effect on future earnings by decreasing labor productivity at harvest, precisely when this labor is needed to generate the next year's income. Finally, seasonal hunger might generate an intergenerational poverty trap, as childhood malnourishment, even if temporary, can have permanent effects on cognitive development, especially when coupled with other diseases to which children become more vulnerable when malnourished. Parents might also decide to take children out of school during hunger seasons due to lack of money (Devereux, Vaitla, and Hauenstein Swan 2008). Another reason to focus on seasonal hunger is that it is a widespread phenomenon, documented in many parts of the developing world. Bryan, Chowdhury, and Mobarak (2014) and Khandker and Mahmud (2012) study the "monga" season in Bangladesh, particularly in the northern region of Rangpur. Basu and Wong (2015) conduct an intervention in West Timor, Indonesia, in order to try to assist people in saving to dull the impact of the hunger season. Dercon and Krishnan (2000) visit households in Ethiopia repeatedly over a short period of time and find that many slip into and out of poverty over the course of a year. Devereux, Vaitla, and Hauenstein Swan (2008) document evidence of hunger seasons in Niger, northern Ghana, Namibia, and Malawi. In fact, though hunger seasons are primarily a feature of poor, rural areas, there is evidence of a similar phenomenon of cyclical hunger even in the US, driven by the monthly receipt of food stamps (Shapiro 2005; Hastings and Washington 2010).

Lastly, the existence of seasonality in consumption is to some extent a puzzle for standard economic theory of consumption and saving. Consider the canonical Consumption Euler Equation:

$$u'(c_t) = \delta(1+r)\mathbb{E}\left[u'(c_{t+1})\right].$$

Under the common simplifying assumption[5] that $\delta = \frac{1}{1+r}$, where δ is the subjective time-

[5]This rules out the idea of individuals steadily increasing or decreasing consumption over the course of their lives

discount rate and r is the interest rate, this suggests that people should be roughly trying to smooth their consumption today to match their expected consumption tomorrow.[6] Any seasonality in consumption is difficult to explain with this model; if every year, around the same time, there is a season in which income falls, this would surely not take most people by surprise. As such, this should be incorporated into expectations, and individuals should save (or borrow) to avoid seasonal hunger. If people are observed regularly failing to smooth this predictable cyclical drop in income and consumption, there are only two possible explanations: either (1) people face some form of binding constraints in both their ability to borrow and to save, such that their consumption is closely tied to their current income, or (2) something in this model is fundamentally flawed.[7]

Brune et al. (2016) conduct a randomized experiment in an attempt to investigate this question of why people in Malawi fail to save when grain prices are low in order to be able to weather or even take advantage of high prices later in the year. They randomly offer some farmers access to a standard savings account, and another group a combination of this savings account and a "commitment" savings account, for which they have to designate a date before which they will not be allowed to withdraw funds. High take up for the "commitment" account would provide suggestive evidence that people's failure to save up for the hunger season is due to a problem with self-control, rather than simply a lack of access to savings or a persistent lack of foresight. Interestingly, they find higher take up of the offer that includes the commitment option than the standard savings account, but they find that people place little money in this commitment account. They suggest that perhaps the constraint to saving is a social one: that having extra money stored up subjects one to pressure to help other family and community members with this money.

In this paper, I analyze the effects of Ramadan on calorie consumption at different points of the agricultural cycle. My results, as I discuss in Section 2.4, provide suggestive evidence that people do *not* seem to use the Ramadan fast as a way to shift consumption from the seasons of relative

[6]Technically, people are predicted to smooth expected marginal utility of consumption. If their utility function for consumption is such that $u'''(c) > 0$, they might choose to build up precautionary savings today, consuming less today than tomorrow on average, in order to insure themselves against a potentially severe negative future shock.

[7]It is worth noting that these possibilities are not mutually exclusive; it could well be that both of these factors play a role in explaining seasonal hunger.

abundance (when the marginal utility of consumption should be lower) to other parts of the year. Indeed, somewhat paradoxically, I do find evidence of a decrease in consumption during Ramadan in a year when it falls shortly before the annual hunger season, despite the fact that we would expect many households to be running low on food at this point, and thus to have little leeway for a reduction in food intake from this benchmark. As this finding is based on only four rounds of cross-sectional survey data, I cannot feign any certainty that this result would replicate across other years, and I have no direct evidence for any particular explanation or mechanism underlying my results. However, I can rule out the idea that (1) savings are significantly higher when Ramadan falls closer to the peak season , and that (2) the absolute reduction in calories due to Ramadan would be weakly decreasing as the holiday falls closer to the hunger season. This could be consistent with the idea that people do not begin trying to plan and save for the hunger season until it is imminent – perhaps because of short-sightedness in planning and decision-making, or perhaps because of constraints that make it difficult or very costly to start saving earlier in the year (e.g. food spoilage, which could act as a negative and decreasing interest rate on savings). It could also be consistent with the idea that post-fast meals are simply larger affairs during the peak season, while closer to the hunger season households throughout the community lack the resources to put together a large communal meal. These are all interesting possibilities that merit further investigation, but it is beyond the scope of this paper to distinguish between different mechanisms that might account for the observed relationship between the timing of Ramadan within the agricultural cycle and its effects on consumption among Muslim households.

2.2.4 The setting: Malawi

Chirwa, Dorward, and Vigneri (2012)[8] document that over 80 percent of the population of Malawi is dependent on agriculture, and that this is largely rain-fed, with less than 5 percent of cultivated land being irrigated. This makes households highly susceptible to agricultural seasons. It is also a very poor country, with 56% of the rural population below the poverty line (and indeed,

[8]The remainder of this description of Malawi's agricultural cycle comes largely from this same work.

as has been discussed, this might underestimate the percentage who fall below the poverty line at some point each year), and 57% reporting inadequate food consumption. The primary staple crop is maize, and the pattern of seasonal hunger in the country closely tracks the crop cycle of maize. Maize is generally harvested between April and June. Prices are lowest during this time as crops flood the market. The proportion of households falling below the poverty line is also lowest at this time, consistent with the idea of poverty being a partly seasonal phenomenon. Most households begin to run out of their own food stocks around August or September, and maize prices jump at this point. According to survey data from 2004–2005,[9] 81% of households who grew food crops during this agricultural season had exhausted their stocks by December. Food prices continue to climb substantially from December through March, which is generally the hunger season. Many use casual labor, or *ganyu*, as a source of income during this time; large numbers of people looking for work at the same time tends to drive down wages. So the hunger season is characterized by the triple-burden of minimal food reserves, high market food prices, and low wages. On top of this, doing *ganyu* work on others' land takes away from households' ability to tend to their own land. Finally, early maize crops start to come in in March. People will boil maize in its less ripe "green" form in order to be able to eat it immediately. While this reduces the total caloric content of the crop, and thus potentially reduces the total availability of calories for the rest of the year, it enables households to put an end to the hunger season and to be reasonably well-nourished in time to harvest the bulk of their crop.

One feature of Malawi's demographics that makes it an interesting setting to study the impact of Ramadan is that Muslims are a relatively small minority population in the country – approximately 14 percent based on my calculations from the IHS5 data (NSO2021). This proportion is large enough that we should expect to encounter a reasonably large sample of Muslim households in any nationally-representative survey. But importantly, it is also small enough that it seems reasonable to think that general equilibrium effects of Ramadan or other major Muslim holidays on prices for non-Muslims should be limited. Specifically, we might be concerned that a decrease in demand

[9]Specifically, the IHS2 dataset (NSO (National Statistics Office) 2005b), which will be one of my main data sources.

for food among Muslims during Ramadan could lower food prices paid by non-Muslims, thereby increasing caloric intake among non-Muslims and exaggerating the estimated effect of Ramadan on Muslims' calorie consumption. This would be a greater threat to identification in a country with a larger proportion of Muslim households.

2.3 Data and Empirical Strategy

2.3.1 Data

The data I use for this paper come from the Malawi IHS (Integrated Household Survey), Rounds 2–5 (NSO 2005b; 2012; 2018; 2021), unless otherwise noted. I include rural households only, and for consistency across rounds also exclude households from the small island district of Likoma, which was until recently excluded from the survey. I construct the following key variables for analysis:

Interview date: In certain instances, there is conflicting or inconsistent information about the interview date in the data provided. Because the purpose of this paper was to look at how consumption changes at Ramadan, accounting for seasonal consumption patterns, knowing the precise date was fairly important. To this end, in order to verify that the interview dates were recorded correctly, I grouped households by village, knowing that teams of interviewers would generally go to a village together for a few days at a time. In cases where the interview dates were unclear or inconsistent with neighboring households, I used the interview dates of the neighboring households as well as information about the interviewers in order to impute the correct interview date.

Ramadan dates: The month of Ramadan begins at different times across the world, according to different local standards. The most common standard is the sighting of the new crescent moon, indicating that the following day will be the first day of the new month. Similarly, the end of Ramadan is also usually determined based on the sighting of the following new crescent moon, with the caveat that the month will always be no fewer than 29 days and no more than 30 days. Wherever possible, I used reports from Malawian news outlets or statements from local religious leaders to determine the start and end dates of Ramadan. Where this information was unavailable, I consulted the Islamic Crescent Observation Project or ICOP (International Astronomical Center 2021), a crescent-sighting tracking website, which accumulates local reports of new moon sightings as well as astronomical maps of when moon sightings would have been expected in different parts of the world. Where I could not find information specific to Malawi, I relied on references from nearby countries, namely Zambia, Tanzania, and South Africa, which typically follow the same calendar; I also cross-referenced the astronomical maps on this site to make sure a crescent-sighting would have been possible on the designated date. Finally, I generated a variable indicating whether or not an interview overlapped with Ramadan. Because the interview is backward-looking, I classified it as occurring during Ramadan if it took place at least two days after the first day of Ramadan, so that at least three days of the previous week would have been fast days. If the interview took place after the final day of the fast, I did not classify it as taking place during Ramadan, even if several days of the prior week would have been fast days, because the end of Ramadan is marked with feasts and celebrations that might obscure the effects of fasting.

Muslim household: The religion of each adult household member was given separately, so designating a household as "Muslim" or "non-Muslim" was not straightforward in cases where different household members practice different religions. I identified a household as Muslim if either (1) at least half of the household members related to the household head (i.e. excluding friends, servants, etc.) were Muslim, or (2) if the household head was Muslim and at least one of his or her family members in the household was also Muslim.

Household size/adult equivalent: I included in the measure of household size all household members who were staying in the household at least one day during the week leading up to the interview.[10] Then, in order to adjust for the lower caloric requirements of children, I used information from the NSO (2005a) to construct a Malawi-specific discount factor by age group for children under 16 years of age.

Average hours worked in the last week (household): In constructing this metric, I summed

[10] In cases when half or more of the listed household members would not be considered "present" by this definition, I used the full size of the household. In most cases, this seemed to produce more reasonable numbers of calories per person per day.

hours worked for each individual across different categories (household agriculture, work in non-agricultural household business, work for a wage/salary, casual day labor, work at apprenticeship). I made a few adjustments to these measures, including truncating any hours reported above 84 hours per week total, and only including the maximum value of "working in household business" and "running household business" if both values were 10 or more hours, as these often seemed to overlap. Once this measure was constructed, I took an average of hours within each household, excluding (a) household members who were not of "working age",[11] (b) those who did not work at all, and (c) servants or lodgers included in the household roster.

In the following subsection, I explain the methodology that I use to construct my primary outcome measure of interest: calorie intake.

2.3.2 Measuring calories consumed

I constructed total calorie[12] counts for the household for the week using the following procedure:[13]

1. *Perform consistency checks on the default unit conversions provided with the data download.* For each food item, I compared the provided unit conversions (a) across the 3 regions; (b) across different sub-units of the same unit type (for example, comparing a "small" pail to a "medium" or "large" pail); and (c) across different units that would generally have similar conversions (for example, a "pail" and a "basin"). In all cases I corrected conversions that seemed unlikely or implausible with a number that seemed reasonable and more in line with similar observations.

[11] I defined "working age" quite liberally to include anyone between 16 and 69 years old, in order to exclude as few households with working adults as possible while calculating labor supply mainly among those who would be expected to fully participate in the labor force when and if possible.

[12] Throughout this paper, I refer to "calories" under the typical colloquial definition, which in the scientific literature would generally be referred to as kilocalories or (capitalized) "Calories."

[13] Despite being widely regarded for their detailed treatment of theoretical and practical considerations for analyzing consumption data and living standards, neither Deaton (1997) nor Deaton and Zaidi (2002) give any notable guidance for the most vexing issues I encountered in working with the Malawian food survey data: converting units to standard metrics, converting standard quantities to calories, and identifying and cleaning errors, inconsistencies, or outliers in large and complex food consumption datasets.

2. *Cross-check the provided calorie conversions* for the most recent round of data (Round 5) against those from Round 3, and fix those that seem implausible or vary excessively from other metrics. Where these metrics were in conflict or seemed implausible, I checked them against (1) the MAFOODS database (MAFOODS 2019) and (2) the USDA database (Haytowitz et al. 2019). In general, I kept the most recent provided calorie conversions unless there was a large difference and/or a compelling reason to think that one of the other sources was more accurate.

3. *Average across regions*. Although there may be good reason to think that there are systematic differences in the quantity received from a plate in one region versus another, I chose to use the average factor across regions, because differentiating by region seemed to add unnecessary noise that might have biased the calorie measure. It is also true that interviewers in all regions had the same photo guides on which they were supposed to base the unit selections; units were meant to represent a similar amount across all households.

4. *Impute missing unit types using provided units*. In many cases, I observed that certain units had a consistent relationship to others. For example, a small pail generally had a conversion factor that was quite close to the factor given for a five-liter bucket. So in cases where only one of these units was provided, I added the other unit based on this observation in order to be able to convert more observations.

5. *Match up "other" units to recognized units where possible*. In many cases, interviewers typed in custom units that exactly or nearly matched existing units. Where possible, I matched these typed-in units to a corresponding standard unit.

6. *Convert to kilograms using the exact sub-unit*. Match unit code directly to the constructed list of unit-to-kg conversions for each food item.

7. *If unmatched, try alternate sub-units and/or removing sub-unit*. For example, if the specified unit was "4B" for "medium pail", and no conversion is given for "4B", I attempt to convert

using “4A” (“small pail”) or just “4” (“pail”).

8. *If still unmatched, assign a default unit.* For each food type, I specified a unit that seemed like a plausible default. For example, for items that usually come as pieces like mangos or cucumbers, I used “Piece” as a default. For smaller grains or flours, I used “Cup.” For items that generally come in bunches, I used “Heap” or “Plate,” depending on what conversions were available or were possible to plausibly construct.

9. *Complete the final conversion.* Multiply the specified quantity by the kg-per-unit conversion, then by the calorie-per-kg and edible-portion-per-kg factors.[14] This provides a sum for the total calories consumed by all household members over the last week.

10. *Divide by adjusted household size* to produce a measure of calories per adult-equivalent for the week, then divide by 7 to convert to a measure of average calories per day.

11. *Address outliers.* In all rounds of data, there are a number of households for which the raw metrics would indicate tens of thousands of calories consumed per household member per day. Because extreme outliers could significantly affect my regression analyses, I chose to censor households with calories consumed per adult-equivalent below the 2nd percentile or above the 95th percentile for each round of data. This, including the asymmetry of the cutoff points, is in line with what other researchers have done when using the Malawi IHS data.[15] An alternative to dropping these households from the analysis of calories would be to Winsorize these outliers, for example, assigning any measure below the 2nd percentile with 2nd

[14] I also implemented a few checks at this point in the process to make sure the quantity being converted seemed relatively reasonable before assigning a calorie value.

[15] Ecker and Qaim (2011), for example, in data generously shared with the author, excluded all households with calculated calories per person below 500 or above 5000, which matches roughly to my percentile cutoffs. The asymmetry in the cutoffs is also reflective of the heavily right-tailed nature of the raw distribution: the minimal level of calories I can observe is 0, whereas there is no maximum; further, very low quantities are more likely to be accurate, especially in a country with high levels of poverty and food insecurity, whereas reports of very high levels of caloric intake (in excess of 6000 calories per day for an entire week, for example) are highly implausible and much more likely to be attributable to measurement error. I would thus argue that the measurement error itself is likely to be highly asymmetric and right-tailed, though it is difficult to prove this directly. Gilbert et al. (2019) and Verduzco-Gallo et al. (2014) drop households reporting more than 8,000 or fewer than 200 calories per person per day, which seemed perhaps excessively permissive.

percentile value, and those above the 95th percentile with the 95th percentile variable. While in many circumstances, this would be preferable to excluding such cases, in this case there is good reason to think that households with very large calculated levels of caloric intake are more likely to have some food quantity recorded incorrectly, and are not necessarily more likely to fall near the top of the calorie consumption distribution. Because of this, I opted to exclude these households rather than to potentially introduce a substantial source of noise into my analysis.

This procedure is largely in line with what other researchers working with these data have done in constructing a measure of calories consumed. However, I added a number of consistency checks throughout the procedure, including cross-checks and corrections of the provided unit conversions, in an effort to try to construct an internally consistent and generally reasonable metric. I also introduced default measures, which allows me to avoid having to convert every possible unit provided without counting everything unmatched as zeros. Both of these innovations should help generate a somewhat more accurate measure of calories consumed. The primary difference between my approach and the one employed by Verduzco-Gallo et al. (2014) is that I rely on my cleaned version of NSO's provided unit conversions where possible, complemented by a set of default units, whereas Verduzco-Gallo et al. (2014) preference their own construction of "implicit" unit conversion factors, based on a comparison of the median price per unit to the price paid per kilogram, wherever feasible. While my methodology is admittedly more ad-hoc, it allows me to be somewhat more assured that the quantities imputed are based more directly on the quantities that respondents were asked to select from, and alleviates the potential concern that using consumption data to construct conversions for the same set of data could amplify pre-existing biases or measurement error.[16]

These concerns also speak to potential survey design issues. Allowing for a vast array of units to choose from (including an "Other" option) is theoretically appealing as a way to facilitate recording household consumption exactly as reported with minimal distortion from the interviewer.

[16]Having said this, calculating implicit conversions could be a useful complement to the methodology I present; it could provide a useful consistency check, especially in cases where I resort to default units, and could also help identify systematic biases in calculated price per kg due to inaccurate unit conversions.

For this reason, Smith et al. (2014) neither favor nor disfavor this approach compared to potential alternatives, stating that more evidence is needed to determine what approach provides the most accurate results. However, if "pail," "heap," and "bunch" can have such vastly different interpretations from one household to the next, it is unclear how much value we truly derive from allowing for such a large set of options in terms of increased accuracy, especially if there is additional noise coming from the unit-to-kg conversions themselves, and if many units have no conversion specified at all. Given this, it might make more sense to limit the number of potential units, possibly sacrificing some ability to reflect a respondent's exact words, but perhaps increasing the likelihood of recording the actual quantity consumed with a reasonable degree of accuracy and consistency across households. Alternatively, interviewers could be trained to try to direct respondents to more readily convertible units, while still allowing them to choose from a broader menu when necessary. In addition, as Smith et al. (2014) advocate, it could be productive to expand the use of demonstration methods, such as photo examples, that help ensure that the respondent can give the interviewer a clear sense of the total quantity consumed of any particular good.

2.3.3 Identification strategy

The regression specification for the main analysis of this paper is as follows:

$$\begin{aligned} Y_{idt} = {} & \beta^{M} Muslim_{idt} + \beta^{R} Ramadan_{t} + \beta^{MR} Muslim_{idt} \times Ramadan_{t} \\ & + \gamma_{d} \left[\times \tau_{t} Round_{t} \right] + \mu_{t} \times \rho_{d} Region_{d} + \varepsilon_{idt}, \end{aligned} \tag{2.1}$$

where outcomes Y are observed for household i in district d at time t.

This specification sets up a difference-in-differences analysis in which I analyze differences in outcomes – namely calorie intake and labor supply – between Muslims and non-Muslims during Ramadan and in all other parts of the year. In this framework, β^{MR} represents the difference-in-differences in outcomes, and is the main coefficient of interest. Because Muslim households are disproportionately concentrated in a relatively small number of districts, I include district fixed effects (γ_d) in all specifications, and allow these to vary from round to round of the survey data by interacting them with τ_t, a set of round fixed effects.[17] In addition, in order to control for potentially large seasonal differences in consumption, I include monthly fixed effects (μ_t) interacted with region fixed effects (ρ_d) to account for differences in timing of the agricultural seasons in different parts of the country (specifically, North, Centre, or South). The identifying assumption, as with any difference-in-differences analysis, is that of parallel trends. Specifically, this assumption means that the comparison of Muslim households to non-Muslim households, outside of Ramadan, serves as a valid counterfactual for the expected difference between Muslim and non-Muslim households observed during Ramadan, if Ramadan did not take place.

On the one hand, the assumption of parallel trends seems quite reasonable, as households were selected randomly and interviews were scheduled to ensure that, within each district, interviews were spread out evenly across the months of the year. Nonetheless, there are a number of potential caveats to bear in mind. For one, surveys are cross-sectional, so each household was only interviewed once. Thus, we might see differences in various measures that are simply due to heterogeneity between households and between different villages, even within the same district. Put differently, there could be unobserved omitted variables that make households differ from one another and potentially affect our observations of Ramadan. For example, if a very wealthy household was visited during Ramadan, and much poorer households were visited at other times of the year, this could bias our estimates of the effect of Ramadan. Related to this concern, because Ramadan takes place for only one month out of the year, and because Muslims make up a relatively small proportion of the population in Malawi, the number of Muslim households observed during Ramadan for any given year will be quite small, giving us relatively low statistical power to detect differences between these households and those observed at other times of the year.

Another potential concern is that changes to consumption behavior during Ramadan affect consumption decisions for the rest of the year. If consumption decreases (increases) during Ramadan,

[17] Of course, this interaction with round fixed effects is omitted in analyses that include only one round of data.

households on a fixed budget may have more (less) money available to spend on food in the rest of the year. This could exaggerate the observed effects of Ramadan by increasing the difference from the rest of the year.[18]

A final threat to identification comes from potential measurement error. In particular, if Muslims are more likely to report consumption of certain types or quantities of foods during Ramadan, then any inaccuracies in the conversion factors used for these foods might bias our measure of the effects of Ramadan on caloric intake. Similarly, if other particular groups have a tendency to consume certain food items, and these are measured inaccurately, this could also bias our results by skewing the measures of caloric intake that we rely upon for comparison. Results might also be sensitive to different ways of addressing outliers in consumption measures, and excluding outliers could introduce selection bias. Finally, as is generally the case when there is measurement error, it could introduce attenuation bias wherein noise in the data biases our coefficient estimates towards zero and makes it more difficult to detect the true effect that we wish to identify.

2.3.4 Seasonality

Appropriately controlling for seasonality is vital to any analysis of consumption data in Malawi. Considering the dramatic seasonal fluctuations in food prices, along with the large percentage of households whose consumption follows a pronounced seasonal trajectory, to not account for the time of year in which a household was interviewed would introduce significant bias into consumption measures.

In thinking about how to appropriately construct measures that will control for seasonality in consumption, it is important to note that there are regional differences in the timing of seasons in different parts of the country, and that there is some variation from one year to the next as to the actual timing of the harvest, as well as the other agricultural seasons (dry season and rainy season). To account for this, my preferred specifications all include Month×Year×Region fixed effects.

One potential downside of controlling for Month×Year effects (e.g. March 2004, April 2004,

[18] This is not necessarily a problem. It would affect the interpretation of the coefficient on $Muslim \times Ramadan$, but the coefficient would still accurately reflect the difference between Ramadan and the rest of the year.

etc.), as opposed to simple Month-of-Year effects (e.g. March, April, etc.), is that in allowing for different Month effects in each year, we may capture some meaningful inter-household variation as part of a seasonal effect, when in fact it would be correctly attributed to the specific composition of households.[19] However, given that in some rounds of survey data, I see sizable effects of Ramadan on the overall population, even after including Month×Year fixed effects, these are likely seasonal effects that have not been fully accounted for with these controls, so I am not very concerned about these seasonal controls being excessively aggressive. On the other hand, we might consider controlling for a narrower timeframe than months. As an alternative, I repeated the analysis using Half-Month fixed effects (results not shown).[20] This seemed to have very little effect on the overall results.

2.4 Results

Table 2.1 shows summary statistics for calories consumed per age-adjusted household member per day for each round of survey data. A few things are worth pointing out. First, note that the distribution of calorie consumption in Round 2 is substantially higher than in other 3 rounds. For a number of reasons, this is unlikely to be a reflection of truly higher levels of caloric intake in the country during this year compared to the latter years. Indeed, it directly contradicts the finding of increases in per capita calorie consumption from Round 2 to Round 3 documented in Verduzco-Gallo et al. (2014). More likely, this is due to differences in the specific implementation of the food consumption module in this round, or even perhaps in the design of the household roster and counts of household members. Fortunately, I can rule out the possibility that this anomaly is attributable to an issue in my calorie conversion methodology, as I find this same discrepancy between Round 2 and the other rounds when I construct an analogous calorie metric using the household level total calorie counts produced by other researchers (Ecker and Qaim (2011); International Food Policy

[19]To illustrate this concern, consider an extreme example in which we introduced fixed effects for every interview date. While this would very specifically capture differences in the timing of seasons from year to year, it would generate very imprecise estimates of these effects and likely wash out many meaningful differences between households observed on different dates.

[20]These results are available from the author upon request.

Research Institute (IFPRI) (2020)).[21]

Table 2.1: Daily Calories per Adult-Equivalent by Round

	Round 2 (04-05)		Round 3 (10-11)		Round 4 (16-17)		Round 5 (19-20)	
	All HH	Incl. HH	All HH	Incl. HH	All HH	Incl. HH	All HH	Incl. HH
HH Mean Calories	2,706.42	2,687.42	2,106.49	2,049.84	2,093.05	2,001.22	2,207.89	2,148.10
	(10.47)	(11.09)	(10.64)	(10.96)	(9.69)	(10.19)	(11.03)	(12.04)
HH Median	2,539.56	2,522.56	1,979.83	1,914.88	1,970.29	1,862.61	2098.20	2028.13
HH 10th %ile	1,485.21	1,463.14	1,070.57	1,036.97	1,083.21	1,042.35	1,108.25	1,074.81
HH 90th %ile	4,219.26	4,206.27	3,499.25	3,418.68	3,370.18	3,223.91	3,713.38	3,611.14
Observations	10,491	9,143	11,412	9,368	11,577	9,431	10,633	8,687

Means and standard errors (in parentheses) for calories are calculated using sampling weights; median and percentile measures are not.

All calculations exclude households that were above the 95th percentile or below the 2nd percentile within each round in terms of calories per adult-equivalent.

"Incl. HH" excludes households in urban areas as well as in the small Likoma district, which was only visited in more recent survey rounds.

For the latter three rounds of data, average consumption levels appear to be reasonably in line with what we might expect for a relatively poor country like Malawi. Specifically, about half of the households surveyed are consuming less than 2,000 calories per adult on average. If anything, these metrics might still be somewhat overestimated, as there is a relatively large number of households reported consuming 3,000 or more calories per adult-equivalent per day.[22] The difference between the distributions among all households[23] versus those included in the analysis is also reasonably in line with expectations: the analysis includes only rural households, so it makes sense that when we add in urban households it generates a small increase in the level of calories consumed at all points along the distribution, and in all rounds of data.

Table 2.2 provides a sanity check as to the novel consumption metrics that I constructed, by comparing them to those used by other researchers. Across all 3 rounds for which other re-

[21] I remain perplexed as to why the Round 2 calorie measures are so much higher, and why this is also true of the replication data, which in my understanding comes from the *same constructed calorie counts* from which other researchers have noted an increase in calorie consumption from Round 2 to Round 3.

[22] On the other hand, caloric requirements are higher for people working in agriculture and with limited access to mechanization or labor-saving household devices (see Deaton and Drèze 2009), so these figures might be a bit low even for the very poor.

[23] Note that these figures do not actually represent *all* interviewed households, as they exclude "censored" calorie counts – those below the 2nd percentile or above the 95th percentile for each survey round.

searchers' metrics were available, my metrics are significantly lower on average than the replication data, particularly so for Rounds 3 and 4. However, I do not necessarily take this as evidence of the measures I used being less accurate. I used similar methodologies in constructing these metrics to other researchers, but also incorporated various consistency checks in the unit and calorie conversions that were not used in constructing alternative metrics. I also specifically investigated some of the large outliers in the replication data to identify improvements I might make to the calorie conversion procedures, and compared the details of some of the largest discrepancies to look for potential errors in my own conversions. This gives me a reasonable degree of confidence that for the Round 5 data, for which there were no replication data available, my constructed calorie metrics are sensible and in line with other rounds of data.[24]

Table 2.2: Mean calories compared to replication data

	Constructed calories	Replication calories	Difference
Round 2 (04-05)	2,603	2,659	−55.4***
	(9.36)	(9.83)	(4.78)
Round 3 (10-11)	2,120	2,569	−449.3***
	(8.48)	(9.52)	(5.81)
Round 4 (16-17)	2,086	2,355	−269.0***
	(7.90)	(8.17)	(4.86)
Round 5 (19-20)	2,270	–	–
	(9.65)		

$^{***}p < .01$. All means listed in this table are for censored measures of calories per adult-equivalent and calculations are unweighted. Standard errors of means (unweighted) are in parenthesis. Differences for all three rounds with replication data are signficant with a p-value < .0001.

Table 2.3 shows the results of an analysis testing the identifying assumption of the paper by looking at the difference in terms of various observable characteristics between Muslim and non-Muslim households observed during Ramadan versus at other times of year on various characteris-

[24] I also ran all of my main analyses for Rounds 2–4 using calorie measures from these replication data, and the findings were quite similar to the ones presented here. Results from these analyses are available from the author upon request.

tics. In general, on most characteristics, the identifying assumption seems to hold. While there are several significant differences between Muslims and non-Muslims, those differences are, for the most part, relatively similar for households observed during Ramadan compared to those observed at other times of year. Particularly encouraging is that in most cases where there does appear to be a large Muslim versus non-Muslim differential, that differential appears to be relatively stable.

Table 2.3: Balance in Observable Household Characteristics

	During Ramadan			Rest of Year			Diff-in-Diff					
	(1)	(2)	(3) Diff. (within	(4)	(5)	(6) Diff. (within	(7) **All**	(8) **p-value**	(9)	(10) Within Rounds:	(11)	(12)
	Muslim	Non-Mus	Distr×Round)	Muslim	Non-Mus	Distr×Round)	**Rounds**		2	3	4	5
No. HH members	4.83	4.53	0.13	4.50	4.44	0.17***	0.24**	.050	0.68**	0.01	0.16	0.22
	(0.11)	(0.05)	(0.16)	(0.04)	(0.02)	(0.05)	(0.12)		(0.33)	(0.32)	(0.30)	(0.17)
No. of children (< 16 yrs.)	2.56	2.19	0.08	2.35	2.15	0.19***	0.16	.152	0.41	−0.14	0.31	0.11
	(0.09)	(0.04)	(0.12)	(0.03)	(0.01)	(0.04)	(0.11)		(0.30)	(0.21)	(0.27)	(0.14)
No. working-age adults	2.10	2.18	−0.004	2.01	2.16	−0.023	0.08	.227	0.21	0.11	−0.02	0.07
	(0.06)	(0.04)	(0.06)	(0.02)	(0.01)	(0.025)	(0.06)		(0.15)	(0.18)	(0.11)	(0.10)
Female-headed (%)	39.8	27.4	7.65**	34.9	27.6	1.58	4.40	.126	3.1	3.5	4.0	5.5
	(3.5)	(1.2)	(3.56)	(1.1)	(0.36)	(1.07)	(2.87)		(4.6)	(4.6)	(4.0)	(6.0)
Age of HH head	43.1	44.1	1.71	43.4	43.4	0.27	−0.98	.496	1.5	1.9	−7.1***	0.68
	(1.5)	(0.48)	(1.17)	(0.32)	(0.11)	(0.38)	(1.44)		(3.1)	(1.6)	(1.7)	(2.3)
HH head married (%)	71.7	72.0	−2.52	70.4	71.4	2.92***	0.53	.903	1.2	−9.3**	14.1*	−3.7
	(3.7)	(1.0)	(3.62)	(0.80)	(0.35)	(1.12)	(4.36)		(5.7)	(4.5)	(8.3)	(5.2)
HH head always lived in village (%)	74.8	66.3	10.24**	69.8	66.7	4.79***	5.94*	.070	−5.4	−5.0	8.7**	12.5**
	(4.1)	(2.1)	(4.17)	(1.2)	(0.57)	(1.43)	(3.3)		(9.6)	(6.9)	(4.2)	(5.1)
Literate (% age ≥12)	52.2	68.1	−16.80***	51.6	66.8	−9.89***	−2.18	.413	1.1	2.1	−7.7	−2.5
	(3.1)	(1.2)	(3.40)	(1.0)	(0.41)	(1.13)	(2.67)		(4.4)	(4.6)	(5.1)	(3.4)
Ever attended school (% age ≥ 5)	72.7	82.2	−8.87***	70.2	80.9	−6.73***	0.40	.837	1.6	8.4**	−0.4	−4.3*
	(2.0)	(0.94)	(2.21)	(0.92)	(0.28)	(0.75)	(1.92)		(3.4)	(4.1)	(4.3)	(2.5)
In school (% 5≤age≤17)	88.7	91.2	−0.58	90.6	92.2	−0.79	−1.14	.612	−2.4	−0.7	−3.8	0.0
	(2.1)	(0.90)	(2.57)	(0.54)	(0.21)	(0.68)	(2.24)		(5.1)	(3.9)	(4.0)	(3.8)
Completed primary school (% age ≥14)	24.4	35.1	−14.26***	24.4	35.5	−8.23***	−1.44	.573	5.0	−4.5	−3.4	−0.6
	(2.7)	(1.5)	(3.66)	(0.76)	(0.44)	(1.11)	(2.55)		(3.8)	(4.0)	(5.3)	(4.1)
Completed secondary school (% age ≥18)	3.1	9.8	−10.41***	4.5	9.0	−5.00***	−2.89**	.035	−2.6	−0.5	−5.7*	−2.0
	(1.1)	(1.1)	(2.79)	(0.39)	(0.23)	(0.69)	(1.37)		(2.2)	(2.2)	(3.1)	(2.0)
N (households)†	355	2,141	2,496	4,465	32,242	36,707	39,203		9,840	10,038	10,015	9,310

Difference within each category (Columns 3 and 6) represents the coefficient on "Muslim" in a regression of the observable variable on "Muslim" after controlling Round × District Fixed Effects. For this reason, the Difference will not precisely equal the difference between the Muslim and non-Muslim means presented in the previous two columns. Diff-in-diff (Columns 7-12) represents the coefficient on Muslim *times* Ramadan in an analagous regression (that also controls for an "IsRamadan" dummy). Standard errors clustered by Enumeration Area within each round (the primary sampling unit) are in parentheses. Urban households excluded from analysis. All means and regressions are weighted based on sampling weights. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$.

† Respresents the number of households in the given category for each column. For some variables, means and regressions are based on a smaller number of observations as the variable was missing in some households.

We should, however, note two points of caution. First, some of the differences in characteristics are significant, such that it is possible that there are some underlying differences between households observed during Ramadan versus the rest of the year, even if those differences are coincidental. Second, we should be particularly cautious in reading into cross-round differences in estimated effects, as they might be purely idiosyncratic. For example, if we look at the "household head married" variable, the overall difference is very close to zero; however, in one round, the difference is negative and significant, while in another, the difference is positive and nearly significant. This could indicate some meaningful difference between the populations observed across these rounds, in violation of our identifying assumption, or it could be largely statistical noise. Thus, when it comes to our main outcome variables of interest (calorie consumption and labor supply), though we do observe some potentially interesting variation from round to round, we should be cautious in attributing differences that we observe between rounds to any particular explanation.

Table 2.4 shows results from the main analysis of the effects of Ramadan on consumption. Columns (1) and (2) show the results using all four rounds of data together. We can see that Muslim households overall have slightly lower levels of calorie intake on average, even after controlling for the district in which they live and seasonal consumption effects. This effect is small – about 40 calories per day, approximately a 2 percent decrease – and only marginally significant. As expected, the Ramadan effect for non-Muslims is quite close to zero on average.[25] Somewhat surprisingly, I detect no significant effect of Ramadan overall on calorie consumption among Muslims. The 95% confidence interval for the average Muslim×Ramadan effect across rounds places rough bounds on the effect of Ramadan on calorie consumption for Muslim households from a decrease of 120 calories per person per day to an increase of 267 calories (or between a 5 percent decrease and a 12 percent increase, based on the regression on log-calories).[26] This would seem to allow us to rule out a very large decrease in calories consumed on average across the four rounds of data. However, it should be emphasized that because we do not observe calories consumed at the individual level, we cannot rule out the presence of household members who would not be expected to fast (such as children and the elderly) potentially obscuring a reduction in calories among healthy, working-age adults.

It also seems to be the case that this observed effect varies substantially across the survey rounds; looking only at the average effect across rounds ignores these potentially interesting (or potentially spurious) differences. Looking at Columns (3) and (4), we see that in Round 2 of

[25] One benefit of studying this question in Malawi is that, because Muslims are a relatively small minority, changes in their consumption behavior should not have spillover effects into food prices or consumption levels for non-Muslims, which may be more of a concern in a country with a Muslim majority or a relatively larger Muslim minority population.

[26] Interestingly, these bounds look somewhat similar at the lower end to the estimated reduction of 150-200 calories per person that Sadeghirad et al. (2012) report as an average for people studied in the Middle East and East Asia, and also very similar at the higher end to the estimated average increase of about 260 calories per person per day in North Africa.

the survey, which took place in 2004–2005, there is a highly significant decrease in consumption among Muslim households during Ramadan. I find a decrease in calorie consumption for Muslim households of about 370 calories per adult-equivalent per day during Ramadan or about a 10% decrease when using log-calories as the outcome variable. This estimate is still somewhat less than, for example, the estimated reduction of around 600 calories per person per day among Muslim households reported by Schofield (2020). Again, this could partly be explained by the fact that some household members are not expected to fast. Additionally, Ramadan took place from mid-October to mid-November during this survey round, which in Malawi is generally just before the beginning of the hunger season. By this point in the season, most households would have run out of their own stockpile of grain, and often many will have already begun to cut back somewhat on consumption. In other words, the baseline household may already be consuming less at this time of year, such that fasting might not make as large of a difference compared to relatively more abundant times of year. Finally, it could simply be that the estimates in Schofield (2020) are context-specific and larger than the effects seen elsewhere; indeed, the meta-analysis by Sadeghirad et al. (2012) suggests an average daily individual-level calorie reduction of 150–200 calories per day in the Middle East and East Asia – by this standard, a reduction of 370 calories per day is quite large.

Table 2.4: Main Analysis: Calories

	All Rounds		Round 2		Round 3		Round 4		Round 5	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Cal.	ln(Cal)	Cal.	ln(Cal)	Cal.	ln(Cal)	Cal.	ln(Cal)	Cal.	ln(Cal)
Muslim	−39.85	−.021*	1.40	−.008	−38.69	−.018	−67.13	−.028	−42.38	−.024
	(24.60)	(.012)	(50.83)	(.020)	(47.34)	(.022)	(49.15)	(.025)	(47.40)	(.022)
Is Ramadan	−28.13	−.014	62.01	.008	−88.12	−.017	258.36***	.123***	−252.48***	−.118***
	(51.65)	(.022)	(80.09)	(.030)	(113.31)	(.057)	(75.22)	(.037)	(94.92)	(.039)
Muslim × Ramadan	74.01	.037	−369.75***	−.104**	−21.44	−.024	268.44	.105	180.61	.087
	(98.88)	(.043)	(109.29)	(.041)	(129.47)	(.066)	(182.46)	(.078)	(171.21)	(.077)
District × Round FEs	Y	Y								
District FEs			Y	Y	Y	Y	Y	Y	Y	Y
Month-Yr × Region FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	36,629	36,629	9,143	9,143	9,368	9,368	9,431	9,431	8,687	8,687

Standard errors clustered by Enumeration Area within each round (the primary sampling unit) are in parentheses. Urban households excluded from analysis. All regressions are weighted based on sampling weights, with weights adjusted within each district-rural area to account for observations excluded due to censoring the conusmption variable. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$.

Perhaps the most surprising result from this analysis is that in the remaining three rounds of survey data, despite the fact that Ramadan falls earlier in the harvest cycle, we see no decrease in caloric intake among Muslim households during Ramadan. In Round 3 (Columns 5 and 6), we see essentially no significant effect of Ramadan on calorie consumption. In Round 4 (Columns 7 and 8), we see that there is a highly significant and positive effect of Ramadan on calories consumed – but that this applies to both Muslim and non-Muslim households. If anything, Muslim households seem to experience an *even larger* increase in calorie consumption during Ramadan in this round compared to their non-Muslim counterparts. This increase is potentially quite large (the point estimate shows about 270 additional calories per person per day, on top of the 260 calorie increase in the general population), and while it is not statistically significantly different from zero, the lower bound of the 95% confidence interval is a decrease of approximately 90 calories, ruling out a decrease of similar magnitude to Round 2. In Round 5, we see a large and significant decrease in calorie consumption during Ramadan across all households, while Muslim households in this round seem to partially offset this decrease, consuming more than non-Muslims during Ramadan, though again this offset is not statistically significant. These results all remain qualitatively similar when using alternative fixed-effects measures to control for seasonality.[27]

How can we make sense of the above findings? While it is beyond the scope of this paper to provide a definitive answer to this question, I consider some possible explanations.

Understanding the timing of Ramadan in each survey round relative to the agricultural cycle provides helpful context. In Round 3, Ramadan falls between August and September, just as the harvest season is winding down in most parts of Malawi. In Round 4, Ramadan falls between June and July, which is generally the peak of the maize harvest, when prices are lowest and hunger is typically at its yearly minimum. In Round 5, Ramadan falls between May and June, which is on the early end of the main harvest season. Ideally, we would want to control for all seasonal effects so that our observed Ramadan effects would fully account for expected seasonal variations. It is clear that, at least in Round 4 and Round 5, we have not fully controlled for seasonality, as we see an increase in consumption during Ramadan in Round 4 among the general population above and beyond our seasonal controls, and a similar decrease in the general population in Round 5. That said, if the assumption of parallel trends holds, then we can still interpret the coefficient

[27]Specific results available from the author upon request.

on Muslim×Ramadan as the consumption effect of Ramadan, as whatever additional seasonal effect is detected among the general population is assumed to apply to Muslims and non-Muslims analogously.

Table 2.5 helps to further elucidate some of this variation between rounds. First, in columns (1) -- (4), we conduct analyses similar to those shown in Table 2.4, but only including Rounds 4 and 5, in which Ramadan overlapped with the harvest. Here we can see more clearly that there is some suggestive evidence of a marginally significant increase in Muslim households' calorie intake during Ramadan when we pool the Muslim×Ramadan effect over these two rounds. We also see wide variation in the Ramadan effect for the population as a whole (i.e. including non-Muslims), which serves as an important reminder that I may not have completely accounted for seasonal effects, and that we should be cautious about any interpretation of round-to-round heterogeneity, as there could be important factors driving these findings that I have not fully accounted for. Columns (5) – (8) help to further illustrate this point. Columns (5) and (6) use Round 2 as a base year, and compare the effects in other rounds to this base. We see a highly significant negative coefficient on the base Muslim × Ramadan interaction, and we see a significant effect in the opposite direction in the other 3 rounds, particularly in Rounds 4 and 5. This affirms that the differences in the observed Ramadan effect on Muslims between rounds are significant. However, Columns (7) and (8) caution that the Muslim × Ramadan coefficients for Rounds 4 and 5 are not significantly different from the corresponding coefficient for Round 3, which is itself quite close to 0. The evidence for Ramadan causing an *increase* in consumption in these rounds is thus not definitive.

Table 2.5: Differences between Rounds: Calories

	Rounds 4 and 5 only, Round 4 as base				All Rounds, Round 2 as base		All Rounds, Round 3 as base	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Cal.	ln(Cal)	Cal.	ln(Cal)	Cal.	ln(Cal)	Cal.	ln(Cal)
Muslim [base round]	−54.4	−0.026	−63.9	−0.028	1.4	−0.008	−38.7	−0.018
	(34.1)	(0.017)	(49.0)	(0.025)	(50.8)	(0.020)	(47.3)	(0.022)
Muslim ×Round effects			17.6	0.003	Y	Y	Y	Y
			(67.2)	(0.03)				
Is Ramadan [base round]	−42.9	−0.019	266.0***	0.125***	62.0	0.008	−88.1	−0.017
	(79.5)	(0.035)	(74.5)	(0.037)	(80.0)	(0.030)	(113.3)	(0.057)
Ramadan ×Round effects			−524.9***	−0.244***	Y	Y	Y	Y
			(120.9)	(0.053)				
Muslim × Ramadan [base]	211.3	0.091	216.1*	0.094*	−369.8***	−0.104**	−21.4	−0.024
	(130.0)	(0.057)	(127.4)	(0.056)	(109.2)	(0.041)	(129.4)	(0.064)
Muslim × Ramadan×R2							−348.3**	−0.080
							(169.3)	(0.078)
Muslim × Ramadan×R3					348.3**	0.080		
					(169.3)	(0.080)		
Muslim × Ramadan×R4					638.2***	0.208**	289.9	0.128
					(212.6)	(0.088)	(223.6)	(0.103)
Muslim × Ramadan×R5					550.4***	0.190**	202.0	0.110
					(203.0)	(0.087)	(214.5)	(0.102)
District × Round FEs	Y	Y	Y	Y	Y	Y	Y	Y
Month-Yr × Region FEs	Y	Y	Y	Y	Y	Y	Y	Y
Observations	18,118	18,118	18,118	18,118	36,629	36,629	36,629	36,629

Standard errors clustered by Enumeration Area within each round (the primary sampling unit) are in parentheses. Urban households excluded from analysis. All regressions are weighted based on sampling weights, with weights adjusted within each district-rural area to account for observations excluded due to censoring the conusmption variable. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$. Muslim (Ramadan) × Round effects refers to Muslim (Ramadan) *times* Round 5 (compared to the Round 4 base) in Columns (3) and (4); for Columns (5)–(8), it indicates that interactions with all non-base round dummies are included in the regression, but not shown in the table for the sake of brevity/legibility.

One possible explanation as to why calorie intake could increase (or at least not decrease) for Muslim households during Ramadan is the importance of the post-fast *Iftar* meal. In many Muslim communities, this meal is observed as a nightly gathering in which members of the community come together to share food and celebrate. It seems plausible that the size and scope of this meal could depend on what resources the community has available. Perhaps then, in Malawi, *Iftar* could be far more extravagant during the peak harvest season, when food is relatively abundant, and much more limited as the hunger season approaches.[28] It is also the case that household members who would not be expected to fast, such as children, the elderly, and the ill would still presumably be invited to the post-fast meal, despite having eaten normally during the day. Thus, if that meal is quite substantial, increases in calorie consumption among non-fasting individuals might outweigh

[28]Of course, I have no direct evidence for this, at this point it is purely speculative.

the decreases among fasting individuals. The idea that the size of the post-fast meal might depend on the overall availability of resources could account for the observation of a decrease in calorie intake when Ramadan falls in October and a potential increase when it falls in May or June.[29]

Another intriguing possibility is that consumption behavior during Ramadan can be explained, at least in part, by myopic decision making with regard to the hunger season, which in turn could help us understand seasonal consumption behavior more broadly. Specifically, if indeed households are preparing a large feast when Ramadan falls during the peak harvest season and decreasing their overall consumption when the hunger season is about to begin, this would be consistent with the idea that households are simply failing to take measures to smooth their consumption levels over the year. We might also see a similar effect if Muslim households simply have less access to smoothing mechanisms, such as informal insurance or bank accounts, and thus are more prone to seasonal consumption fluctuations. The results from Round 5 provide some suggestive evidence against this possibility, as the Ramadan effect for Muslims, though it is not significant, appears to partially counteract the effect observed in the overall population, but certainly does not seem to exaggerate the effect.

If it is true that Muslim households are using Ramadan as a way to save resources to mitigate the effects of the hunger season, but only when Ramadan falls shortly before the hunger season, this might suggest that a similar salience effect is precisely what's driving seasonal variation in consumption more broadly. That is, the same decision-making process leading Muslims households to consume more on Ramadan when it overlaps with the peak harvest season, while starting to prepare for the hunger season only when it is imminently approaching, may be leading other households to consume at high levels around the peak season, leaving all of them with too little savings by the end of the year. In contrast, the standard economic model of intertemporal consumption decision-making would suggest that if Ramadan is to serve as a vehicle for increased savings, it would be *most* likely to do so when it overlaps with the peak season, when baseline consumption levels are

[29]It is also plausible that charitable contributions, either from wealthier community members or from the international community, are helping the average Muslim household to consume more than they otherwise would during Ramadan.

relatively high and the marginal utility of consumption is lower, while households would be least likely to increase savings when consumption is already relatively low in the lead-up to the hunger season. The pattern we observe in our data is *precisely the opposite* of what we might expect if this standard economic model of savings were correct – we can essentially rule out the possibility that this model would produce a result consistent with our empirical observations. However, we do not have sufficient evidence to conclude that this myopia-based explanation in particular is the correct one. More research would be required to establish that these results do indeed represent a consistent pattern and are not simply one-off idiosyncrasies of these specific years, and also to establish whether this explanation can account for such a pattern.[30]

It is also worth keeping in mind that measurement error and/or omitted variable bias could be affecting the presented results. As previously discussed, if certain food types are more prevalent among Muslims during Ramadan and are measured incorrectly when converting to calories, this would skew our observed effect of Ramadan. In addition, I am relying on district and month fixed effects to control for inter-household variation in (counterfactual) consumption, and those predictions surely will not provide a perfect counterfactual for the levels of consumption that we would expect from Muslim households in the absence of Ramadan. Any noise or bias in these counterfactual predictions will make it more difficult to determine the true effect of Ramadan. I cannot rule out either of these potential sources of bias.

Table 2.6 shows the effects of Ramadan on hours worked per week for Muslims and non-Muslims. To construct this variable, I take the average hours worked across household members of working age (which I define as 16–69 years old) who reported working in the last week. I exclude lodgers, servants, and any of their family members who may be listed in the household roster but whose labor decisions are likely to be separate from those of other household members. Excluding

[30] Another potential explanation that would be difficult to distinguish from the myopic planning mechanism is that households do not have access to adequate saving technology to allow them to transfer funds and/or food stocks from high-consumption times of year to the hunger season, or that such a technology is so costly that using it would result in a net welfare loss. If, for example, spoilage rates of stored food start to increase exponentially after a few months, it could help explain why Ramadan might allow some increased savings into the hunger season when it falls shortly beforehand, but not when it falls during the peak season. Direct measures of access to and use of savings mechanisms could allow us to investigate this possibility further.

individuals who did not work from the household average might introduce some bias into this measure, particularly if people who otherwise might have worked some number of hours choose not to do so during Ramadan. While I am more interested in studying the effects of Ramadan on intensive labor supply, under the premise that fasting might make it more difficult for people to work as much as they otherwise would have, I cannot rule out the possibility that some people would stop working altogether for parts of the month of Ramadan.[31]

Table 2.6: Main Analysis: Labor Supply

	All Rounds		Round 2		Round 3		Round 4		Round 5	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Hrs	ln(Hrs)	Hrs	ln(Hrs)	Hrs	ln(Hrs)	Hrs	ln(Hrs)	Hrs	ln(Hrs)
Muslim	−0.18	−0.001	−0.93	−0.040	0.12	0.009	−0.61	−0.011	0.52	0.033
	(0.44)	(0.022)	(0.99)	(0.039)	(0.76)	(0.043)	(0.99)	(0.051)	(0.72)	(0.039)
IsRamadan	0.48	0.031	0.53	0.037	1.18	0.114	0.72	0.005	0.27	0.026
	(1.09)	(0.054)	(1.20)	(0.045)	(2.94)	(0.156)	(1.63)	(0.109)	(2.45)	(0.121)
Muslim × Ramadan	−2.91**	−0.177**	−3.54	−0.174*	−2.02	−0.265	−3.54	−0.355**	−3.00	−0.070
	(1.27)	(0.079)	(2.48)	(0.099)	(3.44)	(0.214)	(2.56)	(.165)	(1.96)	(0.113)
District × Round FEs	Y	Y								
District FEs			Y	Y	Y	Y	Y	Y	Y	Y
Month-Yr × Region FEs	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	33,956		9,057		8,618		7,996		8,285	
Mean hrs	22.71		28.21		21.91		19.51		21.88	

Standard errors clustered by Enumeration Area within each round (the primary sampling unit) are in parentheses. Urban households excluded from analysis. Households with no working-age members completing any work in the last week also excluded. All regressions are weighted based on sampling weights. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$.

As we would expect, Ramadan has no significant effect on the labor supply of non-Muslims. Consistent with other research, on the other hand, Ramadan does appear to have a significant effect on labor supply for Muslims. In particular, Table 2.6 shows that Muslims on average work approximately 3 fewer hours per week during Ramadan than we would otherwise expect, corresponding to approximately a 20 percent decrease in the regression on log-hours.[32] Of course, the number of hours worked may not fully reflect the productivity of those hours. If hourly output decreases as well, the overall productivity effects could be larger, in line with the 20–40 percent productivity reduction that Schofield (2020) estimates. Interestingly, Schofield (2020) does not find any

[31] Indeed, it could be a potentially useful extension of my analysis to study Ramadan effects on the extensive margin of labor supply by using a binary variable for "any work" in the last week as the outcome variable.

[32] Compared to the mean household across the whole sample, 3 hours corresponds to about a 13% reduction in the average worker's labor supply. This slight discrepancy in estimated percentage differences is likely attributable to a baseline counterfactual predicted non-Ramadan labor supply for Muslim households that is somewhat smaller than the full sample average after including district effects, month effects, etc.

significant reduction in labor supply among Muslim workers.

Compared to the effects on calories, there is little variation from round to round in the measured effect of Ramadan on labor supply. It is generally encouraging that we are able to detect an effect of Ramadan on labor supply for Muslims, as it suggests that our measures of whether interviews took place during Ramadan and whether households were Muslim are reasonably accurate. However, the fact that this effect is more difficult to detect within individual rounds gives reason to be cautious about the interpretation of null effects within each round of data: we may not have sufficient power to reliably uncover true variation in the size of the effect of Ramadan from one round to the next.

The fact that we do observe an overall negative effect of Ramadan on labor supply among Muslim households, while we did not see a consistent negative effect calories on calories, merits discussion. One possible explanation is that the decrease in labor supply is not in fact due to a reduction in calories, but perhaps due to abstaining from water, tending to religious or social obligations, or sleep loss from waking up before sunrise to eat. Another possibility could simply be that my measure of calories is noisy, and that that noise masks the true effects of Ramadan on calories. It is also possible, as previously discussed, that there is indeed a reduction in calories among working-aged people, which is causing them to reduce their labor supply, but that this effect is being masked by measuring calories at the household level. Finally, it is possible (consistent with the findings in Hu and Wang 2019) that hunger during daylight hours could make it more difficult to work, even if that hunger is temporary and fully counteracted via a large post-fast meal.

2.5 Conclusion

In this paper, I investigate the effects of Ramadan on calorie consumption and labor supply among Muslim households in rural Malawi. I find no evidence of a decrease in calorie consumption during Ramadan on average. I do, however, find evidence that working-age people reduce their weekly work by about three hours, or nearly 20 percent, on average. This could potentially indicate that a reduction in labor supply during Ramadan is not contingent upon an overall reduction in caloric intake. It could also be that, because I measure calories at the household level, which could include several household members who would not be expected to fast, I am underestimating the reduction in calories among working-age adults.

Another potentially interesting finding is that, in one round of survey data in which Ramadan fell just before the onset of the annual hunger season, I do see a significant decrease in caloric intake among Muslim households during Ramadan. I also find some evidence of an increase in caloric intake during Ramadan when it overlaps with the main harvest. While I cannot be sure that this finding is reflective of a larger pattern, and cannot offer evidence for any particular explanation, this evidence is, if nothing else, consistent with the possibility that the post-sunset meal that is traditionally consumed each night of Ramadan might have important implications for the overall effects of the fast on caloric intake. It is also consistent with the idea that this meal might be quite large during times of abundance and scaled back in times of scarcity.

This observation could also tentatively offer insight into the nature of seasonal food insecurity. A purely rational actor might be expected to use the opportunity of the Ramadan fast as a way to increase savings and mitigate the harm of the hunger season. We might expect this to be the case particularly when Ramadan overlaps with the harvest, and there is simply more food available to be saved. The fact that we instead see no decrease in consumption during Ramadan when it overlaps with the harvest, and some decrease in relatively lean times could indicate some potential barriers, either institutional or psychological, that make it difficult for households to smooth consumption throughout the year.

There are several reasons to be cautious about drawing any firm conclusions from these findings. For one, it is possible that there is indeed a reduction in calories, but that because calories are measured at the household level and not every household member necessarily would be expected to fast, I am underestimating the true reduction in calories among individuals who are, in fact, fasting. It is also possible that the identifying assumption underlying this analysis is violated, and that there are meaningful differences between Muslim households and/or non-Muslim households observed during Ramadan versus those observed at other times of year. While I provide some evidence for the validity of this identifying assumption, it is difficult to prove conclusively that it is upheld. Finally, while there is possible evidence of a seasonal pattern in the effects of Ramadan, we have only four rounds of survey data to draw from, limiting our ability to reliably establish any particular seasonal pattern.

One other key contribution of this paper is a novel set of calorie metrics that I constructed from the Malawi IHS survey data. In comparing my constructed metrics to replication data supplied by other researchers, I specifically looked into some of the more extreme and unlikely measures generated in the replication data in order to identify and correct for notable sources of likely inaccuracy in the existing conversions. While there are still certainly some inaccurate measures in the data I generated, I made an effort to correct any major inaccuracy that I was able to identify, which could be useful to other researchers who wish to work with these data.

The findings that I have presented indicate a number of potentially interesting avenues for future research. First, it could be useful to look not just at calories consumed, but also different types of food consumed, in order to develop a clearer picture of what is actually happening during Ramadan or at other times of year. For example, if particular foods are consumed as part of the post-Ramadan meal, we might be able to uncover that in the data, and then to test whether we do in fact observe variation from year to year in the amount of those particular foods consumed. This could potentially help us to understand if this post-fast meal is changing in size and scope over time.

Another potentially interesting question would be to look at the distributional effects of Ramadan. The analysis conducted in this paper relied entirely on average consumption levels. However, there could potentially be interesting differences in the effects of Ramadan on people at the lower or higher ends of the income distribution. Because Ramadan is supposed to be a time of charitable giving, we might see that the poorest households are increasing consumption at this time, while in richer households consumption is decreasing.

It could also be potentially interesting to look not just at consumption but also at income and savings to see how these are affected by Ramadan, and to see if these follow similar seasonal patterns throughout each round of survey data. Finally, it could be interesting to test whether any of the patterns and mechanisms that I have discussed in this paper, particularly surrounding the relationship between Ramadan and the agricultural cycle, would hold up in other sources of data, whether it be a future round of the Malawi IHS survey, or data coming from other countries with similar seasonal consumption patterns. Specifically, incorporating additional data of this sort could give us a way to test whether Ramadan causes an increase or decrease in calorie intake in other contexts and whether this lines up with seasonal patterns of consumption in a similar way to the data analyzed in this paper.

Chapter 3: Difficulties with DOTS: Can Direct Observation by Lay Providers Improve Tuberculosis Treatment Adherence? (authored jointly with Kaivan Munshi and Nancy K. Luke)

3.1 Introduction

The ongoing COVID-19 pandemic has highlighted the potentially devastating economic impact of contagious respiratory disease. In developing countries especially, this pandemic has been described as "wiping out" a decade or more of economic progress The Economist 2020. But the scourge of tuberculosis (TB) has taken a substantial toll on both human life and economic development: along with the approximately 10 million people estimated to become sick with TB and almost 1.5 million estimated to die each year (World Health Organization 2019), researchers have estimated that TB causes a 4–7 percent reduction in GDP in the highest-burden countries, can cause household earnings – especially for the poorest households – to fall by 30 percent or more, can impose catastrophic costs of seeking treatment for the poorest households even when medications are provided free of charge, and can trap families in a generational cycle of poverty by forcing children out of school and into the workforce to cover household and treatment expenses and make up for lost earnings (Laxminarayan et al. 2007; Rajeswari et al. 1999; Shete et al. 2018). India, where our study is based, has the most TB cases per year of any country, accounting for over 25 percent of cases worldwide. It also accounts for over 25 percent of the world's reported cases of drug-resistant TB (World Health Organization 2019), which initially emerges when patients fail to properly adhere to the recommended treatment regimen, but can spread to others who were not previously infected or treated for TB (Keshavjee and Farmer 2012).

There is substantial evidence that Directly Observed Treatment (DOT), in which TB patients are assigned a DOT Provider who is expected to meet with them each time they are scheduled to take their medications and ensure that they do so, is more effective than relying on patients to administer their own treatment (see for example: Chaulk and Kazandjian 1998; Frieden and Sbarbaro 2007; Sivaraj et al. 2014). Despite efforts in India to implement a public health system based on a DOTS (Directly Observed Treatment, Short-Course) regimen,[1] the country continues to have high rates of TB infections and mortality, and a growing prevalence of Multi-Drug-Resistant TB (MDR-TB). One possible explanation for this failure is that the government-run system generally designates a low-level public health official, the Village Health Nurse (VHN), to serve as a DOT provider for all TB patients in his/her local area. Though VHNs are knowledgeable and well-trained, they have a host of other public health responsibilities, and little to no enforcement of or incentive to fulfill the responsibility of meeting each TB patient three times per week to oversee them taking their medications. It is thus unrealistic to assume that VHNs would be able to closely monitor TB patients' treatment adherence.

Our study investigates an intervention with great potential for scalability throughout the country. Specifically, we look at the effect of asking patients to choose a trusted member of their community to serve as a DOT provider, meeting the patient each time they are scheduled to take their medication treatment and directly observing as they take each prescribed dose. While it may seem reasonable to assume that a layperson with a strong prior relationship to the patient would outperform the overburdened VHN, it is not obvious that community members without formal medical training would have a full appreciation of the importance of ensuring strict treatment adherence. Further, it is possible that cultural attitudes, in particular stigma associated with TB, could introduce complications into the selection and performance of community DOT providers. Such concerns indicate that the effectiveness of community DOT providers is not a forgone conclusion, and merits empirical inquiry.

To this end, over the course of two and a half years, we approached every adult patient diagnosed with TB who entered the public health system in our study area in rural South India to enroll them in our study. Ultimately, we enrolled just over 1,000 patients. Once a patient agreed to

[1] "Short-Course" refers to the relatively short duration of treatment, around 6 months for most patients, as opposed to longer treatment regimens that may last a year or more.

participate in our study, we randomly assigned him or her to one of three study arms. In the control arm, patients would receive standard treatment administered through the public health system, with the VHN usually serving as the DOT provider. In the two treatment arms, patients received the same course of treatment, but were asked to select a member of their community to serve as a DOT provider. In one of the two treatment arms, we required that the chosen DOT provider be a member of the patient's caste, while in the other arm, we allowed them to choose anyone from their community. The purpose of this was to investigate whether requiring patients to choose someone from their own caste would overcome concerns about stigma and encourage them to choose a DOT provider who would be more intrinsically motivated to support them. Ultimately, around 85% of patients assigned to the "open" arm, in which they could choose any community DOT provider, selected a member of their own caste.

On various metrics of DOT provider performance, we find that community DOT providers outperform the standard government system: patients report seeing their selected DOT provider significantly more often and report higher rates of satisfaction with their DOT providers' performance. However, we did not find a significant improvement in rates of successful completion of treatment. It is unclear why more frequent DOT provider visits failed to translate into improved treatment outcomes, and while we can suggest possible explanations, these are largely speculative, and this remains a crucial question for future research.

The primary contribution of this paper is our finding about the effectiveness of community-based DOT provision. One legitimate concern about assigning lay persons to serve as DOT providers is that they lack professional training and might therefore not fully appreciate the importance of ensuring compliance with the treatment regimen. For example, if a patient were to complain about side effects from medications and say that they are no longer experiencing TB symptoms, an untrained DOT provider may not insist on the importance of continuing with treatment. Furthermore, in a context such as India in which there is significant stigma associated with TB, it might be difficult to enforce the expectation that DOT providers meet with patients regularly to ensure treatment compliance. Our findings suggest that community DOT provider adherence is high, but that concerns about the effectiveness of a lay provider may still be valid.

Caste plays an important role in almost all aspects of life in rural India, [2] and there is good reason to think that this might be true for treatment of TB. Taking the role of caste into account could be crucial for successful design and implementation of a TB control system in India. Various papers have examined the role of caste in finding jobs (de Haan 1997; Luke and Munshi 2011), providing informal insurance (Mobarak and Rosenzweig 2012; Mazzocco and Saini 2012; Munshi and Rosenzweig 2016), and establishing trust when entering an industry or for other interactions with asymmetric information (Banerjee and Munshi 2004; Anderson 2011; Munshi 2011). This paper considers the potential role of caste in supporting, but also potentially hindering, successful completion of treatment. In particular, members of the same caste might have a greater intrinsic motivation to provide support to a fellow caste member. On the other hand, patients facing concerns about stigma surrounding TB might not want other caste members to find out about their diagnosis, and this in turn may lead to suboptimal DOT-provider selection, with patients potentially prioritizing discreteness or lack of connection to members of one's own network over expected performance in the role. We provide some suggestive evidence about the association between perceived stigma and TB treatment outcomes, and also look at the potential interactions between stigma and the efficacy of community DOT Providers. Interestingly, we do not find a significant association between initial perceptions of stigma and treatment outcomes, DOT provider performance, or the assigned treatment arm, though in the present analysis we are only able to look at such associations descriptively and cannot make causal claims.

3.2 Background

3.2.1 Epidemiology of Tuberculosis

Tuberculosis, despite the broad availability of highly effective and relatively inexpensive treatments, and despite being fully curable in most cases, kills nearly 1.5 million people every year, making it the 10th leading cause of death worldwide (World Health Organization 2019). While

[2] See Munshi (2019) for an in-depth look at the many ways in which caste influences life in India.

TB infections are reported in almost every country worldwide, the vast majority of cases and almost all of TB deaths occur in developing or middle income countries; approximately 87 percent of cases worldwide come from the 30 highest-burden countries (Lopez et al. 2006; World Health Organization 2019). Though TB is often considered a disease of poverty, the increasing spread of drug resistant strains presents a potential risk to rich and poor countries alike. In our highly interconnected world, the threat of a global pandemic of untreatable TB must be taken seriously (Kim et al. 2005).

TB is caused by the *Mycobacterium tuberculosis* bacterium. These bacteria spread through the air when a person with TB disease of the lungs or throat coughs, speaks, or sings (CDC 2018). People nearby may breathe in these bacteria and become infected. Though this infection persists for life unless treated, many infected people will never experience TB symptoms, and indeed many may never know they were infected. TB infection without symptoms is referred to as latent TB: the bacteria remain in the lungs in a dormant state, but the immune system prevents them from reproducing and multiplying. Fortunately, TB cannot be spread to others in this dormant state. If, however, the immune system fails to prevent TB bacteria from multiplying, the infection will progress to TB disease, which has a very high mortality rate when left untreated. Often this occurs when the immune system is weakened due to factors such as poor nutrition, diabetes, HIV, or cigarette smoking (Lawn and Zumla 2011).[3] At this point, the infected person will start to experience symptoms of TB disease, and will be capable of spreading the infection to others. Researchers have estimated that around 25%, or by some estimates even up to one third, of the world population is infected with latent TB, and perhaps up to 80% in high burden TB countries (Kim et al. 2005; Kumar and Robbins 2007), though it is also estimated that only about 5–10% of those infected will ever develop TB disease in their lifetimes (World Health Organization 2019).

[3]It is also possible for a person who is otherwise healthy to become sick with TB disease.

DOTS: The Standard Treatment for TB

Although both latent and active TB are fully treatable and curable, effective treatment requires adherence to a long antibiotic chemotherapy regimen that often generates unpleasant side effects while causing symptoms to remit long before treatment is complete. Patients thus often (incorrectly) assume that they have been cured and decide to stop taking their medications before completing the full course of treatment. Failure to fully adhere to the prescribed treatment increases the likelihood of relapse and development of resistance to the first-line antibiotic medications.

The WHO has approved several treatment regimens for new cases of standard drug-susceptible TB, all under the umbrella of DOTS,[4] which vary in their duration (generally at least 6 months) and the frequency (either daily or thrice-weekly, every other day). To improve adherence to the full course of treatment, the WHO recommends the use of Directly Observed Treatment (DOT) Providers – individuals who are assigned to monitor the patient and ensure that each dose of the medication is taken as prescribed (World Health Organization 2017). They further recommend that the individual be chosen from outside the household so as to avoid tensions of intra-household dynamics, such as a patient's spouse or child having to force them to take their medications against their will (Frieden and Sbarbaro 2007).

DOT providers have been shown in numerous studies to be an essential component of effective treatment due to the difficulty of compliance for many patients under self-administered treatment (Chaulk and Kazandjian 1998; Frieden and Sbarbaro 2007; Sivaraj et al. 2014). In visiting the patient to ensure they take each dose as recommended, well-trained DOT providers can potentially substantially improve adherence to treatment.

Though the evidence is not entirely clear-cut, more recently the WHO has begun to recommend a daily regimen over the thrice-weekly option. Of course, from the perspective of the DOT provider, it is likely to be easier to visit the patient three times per week rather than daily, so DOT providers may be somewhat less effective under a daily regimen. On the other hand, there is some

[4] Directly Observed Treatment, Short-Course, where the "Short-Course" refers to an approximately 6-month treatment regimen, as opposed to the year-long regimens that were more common historically.

evidence that sporadically missing a few doses is less risky, i.e. less likely to result in relapse or development of drug resistance, under a daily treatment regimen compared to a thrice-weekly schedule. That is to say, a daily regimen may be more robust to human error, though it might also be more difficult to fully comply with. A similar argument might be made for a shorter treatment regimen: there have been some suggestions that alternate regimens lasting, for example, 4 months rather than 6 months could be successful, but it is possible that such a regimen, even if generally effective, might also carry a greater risk of relapse or complications if not adhered to strictly (van den Boogaard et al. 2011).

Drug Resistance

One of the biggest concerns surrounding TB is the increasing prevalence of treatment-resistant strains of the bacteria, leading to multi-drug-resistant (MDR-) TB. The medicines used for first-line treatment are highly effective (when used correctly), inexpensive, and widely available. While using a cocktail of different antibiotics makes it substantially less likely that the bacteria to develop resistance, when resistance does develop, it can make the disease substantially more difficult to treat. Though advances have been made in recent years in the so-called "second-line" medications used to treat MDR-TB, and even for Extensively Drug Resistant (XDR-) TB that fails to respond to several of these second-line medications (see Dheda et al. 2014; Seung et al. 2015; Park et al. 2019), these medications tend to require much longer courses of treatment, are often much more expensive, less widely available, less effective (cure rates are lower), and often induce very severe side effects (Seung et al. 2015). Drug resistant strains are also capable of spreading and infecting people who had not been previously infected with or treated for TB, so there are potentially large negative externalities, and substantial risks to broader public health, in non-adherence to treatment (Keshavjee and Farmer 2012). As drug-resistant strains become increasingly common, the likelihood increases of a global pandemic of TB that is very difficult if not impossible to treat, which could affect both rich and poor countries (Kim et al. 2005). An outbreak of a strain of TB that is unresponsive to existing treatments could quickly become a global pandemic, and as the

COVID-19 pandemic has demonstrated, even strong public health systems may struggle to control the spread.[5]

While it is known that failure to strictly adhere to the full course treatment increases the risk of relapse and of developing drug resistance, the exact relationship between missed doses of medication and the probability of adverse outcomes is not well understood (van den Boogaard et al. 2011; Stagg et al. 2020). It is broadly agreed that a single missed dose is unlikely to substantially increase the likelihood of relapse, while a treatment default, defined as failing to take medications for two or more months, does increase the risk. In between these extremes, it is unclear how the risk of relapse or drug resistance might depend on the timing of missed doses (whether earlier or later in the course of treatment) and whether doses are missed consecutively or sporadically (e.g. is it riskier to miss 3 doses consecutively or 6 sporadically throughout treatment?), and the like.[6] This is important to note because we would ideally want to measure patients' treatment adherence or non-adherence in a way that identifies those at greater risk of relapse or drug-resistance, while allowing for minor non-adherence that does not significantly increase the risk of an adverse outcome.

TB in Developing Countries

While TB infections occur worldwide, they are particularly concentrated in developing countries. Along with the human toll, TB has a massive economic cost. Between the cost of treatment, time spent seeking treatment, lost wages of patients and their caretakers, and children leaving school to enter the labor force when a parent becomes sick, estimates have shown that TB could account for losses of up to 7% of GDP per year in the most severely affected countries (Laxminarayan et al. 2007).

Furthermore, those infected with TB who are unable to work are at increased risk of falling

[5]One important difference is that there is no evidence that TB, even a highly drug-resistant strain, would be spread by people with asymptomatic or latent infection.

[6]It is perhaps not surprising that these details are not well understood, as it would be impractical and quite possibly unethical to randomly assign patients various patterns of treatment non-adherence in order to accurately identify the precise causal relationships to the likelihood of relapse or other treatment failure.

into a poverty trap. The inability to work can lead to malnutrition, which increases the severity of disease and can allow a latent TB infection to turn into active TB disease. Thus, the compound of TB infection and poor nutrition/health can make it substantially more difficult to reenter the workforce, and thus to recover from both the physical and economic burden of disease.

On an individual level, though anyone can become sick with TB, the disease disproportionately affects adults rather than children, and men rather than women (World Health Organization 2019). In many developing countries, in which men are the primary earners for a household, TB often prevents people from working, which can result in a substantial loss of household income, and can mean that children are taken out of school in order to support household earnings, leaving them much more likely to grow up in poverty and to remain poor as adults (Rajeswari et al. 1999; Laxminarayan et al. 2007).

On a societal level, countries with weaker public health systems are less able to effectively treat and contain the spread of TB, thereby reinforcing the disparate impact of TB on developing countries.

3.2.2 TB in India

The Revised National TB Control Program (RNTCP)

India has the largest number of TB cases per year of any country (World Health Organization 2019). Between the 1960s and the 1990s, the program for TB control in India relied on self-administered treatments, in which medication was provided to TB patients free of cost and patients were entrusted to complete the treatment regimen as prescribed. Recognizing this as an abject failure, in the late 1990s, public health officials devised the Revised National Tuberculosis Control Programme (RNTCP) with help from the WHO. The primary innovation of this program was the implementation of a DOT provider-based system of patient observation (RNTCP 2005).

In practice, implementation of this system is flawed. While guidelines allow for either the selection of a Community DOT provider or assignment to a government worker, in practice, nearly all patients are assigned to a government worker, typically a Village Health Nurse (VHN) (Nirupa et al. 2005). VHNs, while highly trained and well-qualified, are assigned many other responsibilities for public health in their communities and were given no additional incentive to perform Directly Observed Treatment. Further, because VHNs would not necessarily have any previous social connection to their patients, they may be less likely to make extra effort to ensure treatment compliance, compared to a patient-selected community member. Though official statistics showed the RNTCP system reaching the WHO targets of reaching 85% cure rates, independent audits have found that, in reality, actual cure rates were much lower than reported, and relapse rates substantially higher, suggesting that the system was not being implemented as designed (Santha et al. 2002; Velayutham et al. 2018).

The standard treatment regimen under the RNTCP classifies a patient who has not previously been diagnosed with TB as Category I. These patients are prescribed a 6-month chemotherapy antibiotic regimen every other day (thrice weekly). Treatment is broken down into a 2-month Intensive Phase (IP), followed by a 4-month Continuation Phase (CP) with a slightly less extensive medication regimen. Symptoms usually remit by the end of the IP, but in order for treatment to be successful, medications need to be continued for the entirety of the CP. If the patient is known to have already defaulted or relapsed after previous TB treatment, they are considered Category II, have a 3-month IP that includes supplemental antibiotic injections, and a 5-month CP with the standard thrice-weekly drug cocktail (RNTCP 2005).

Caste and Stigma

In theory, assigning Directly Observed Treatment to community DOT providers could be more effective than the VHN-based system, because the patient would be able to choose somebody they trust and who may feel some personal responsibility for their care and successful recovery. However, there are a few reasons to be cautious about the idea of replacing the VHN with community DOT providers. One major concern is that VHNs are well-trained professionals, whereas a community DOT provider would require separate training, and even with training may not fully appreciate the importance of ensuring strict adherence to treatment.

A second major concern, particularly in the context of rural India, is the role of stigma. As a contagious disease that has been endemic for millennia, cultural practices naturally developed to contain the spread. Observing that families were often affected together, TB was traditionally viewed as a potentially inheritable disease. As a result, people who fell ill with TB were sometimes stigmatized and socially isolated, and it could make it more difficult to find a marital partner for them or even for their children.

As the body of knowledge regarding transmission of TB has grown, we have a better understanding of how TB spreads and how the spread can be prevented. Stigma has moved from being a somewhat effective disease control mechanism to now being a hindrance to effective treatment: making people avoid getting tested and reluctant to let others know about their diagnosis. This also makes it more difficult to enforce treatment compliance and potentially even to reduce spread if people try to disguise their illness. Infected TB patients may even be reluctant to continue treatment once symptoms abate so as to reduce suspicions about their illness.

It is particularly important to consider the role of caste in this dynamic. The caste system plays a role in nearly every aspect of life in India, particularly rural India. People's social connections and closest friends will often be members of the same caste, and almost all marriages, especially in rural areas, are arranged between two members of the same caste. The caste as a community also provides social and economic support, for example, helping with finding jobs and lending money for large expenses (Luke and Munshi 2011; Mobarak and Rosenzweig 2012; Munshi and Rosenzweig 2016; Munshi 2019). Being cut off from this network of support due to a TB diagnosis could be devastating. At the same time, provision of Directly Observed Treatment would generally seem to fit squarely within the realm of social support typically provided by one's caste members, and a fellow caste member may be more likely to be willing to perform DOT provider duties, despite the personal inconvenience, than a villager from a different caste. However, choosing a caste member to serve as a DOT provider may also risk an individual exposing his/her TB condition to other fellow caste members. These countervailing forces might cause people to make suboptimal choices about who they would be willing to have serve as their DOT provider, and this is an important consideration for effective implementation of DOTS in India.

3.3 Experimental Design

3.3.1 Recruitment

For our study, we selected an area with a population of about 1.2 million people in Vellore District in rural South India. Specifically, we chose three "Tuberculosis Units" (TUs – geographical units designated through the RNTCP) to work with that were far enough from our collaborating institution to not be directly affected by its various community health outreach efforts, but close enough to be relatively easily reached by members of our team at the Christian Medical College (CMC). Anyone from these areas who sought treatment in the public health system would be approached to be enrolled in our study. [7] Everybody over the age of 18 was eligible as long as they weighed at least 25 kilograms, did not have significant comorbidities, were not initially diagnosed with MDR-TB or extrapulmonary TB, and had not previously failed a Category II treatment regimen.

3.3.2 Study Procedures

All individuals who consented to be a part of our study were randomly assigned to one of three arms. In the first arm, patients would be asked to list potential DOT providers all from their own caste. In the second arm, patients were asked to list potential DOT providers from any caste. In the third arm, patients would be enrolled in the government system in which in most scenarios, a VHN would be assigned as their DOT provider. Randomization was implemented by researchers at our collaborating institution in the CMC. Patients assigned to one of the first two treatment arms (i.e., assigned to a community DOT provider) were then asked to list three potential community members whom they would trust to serve as their DOT provider. The order of this list would then be randomly reshuffled, and a member of our study team would approach the first selected DOT

[7] This excludes people who were treated exclusively through private providers, which we believe to be a small percentage of the population.

provider for recruitment. In the rare event that the person could not be found or refused to take part in the study, they would then move to the next person on the list. Community DOT providers would participate in a basic training session according to RNTCP guidelines which would emphasize the importance of making sure that providers observed patients taking their complete cocktail of medications three times per week. Outside of the DOT provider assignment, all aspects of treatment were the same across the three study arms and based in RNTCP protocols. In all treatment arms, the VHNs continue to be responsible for monitoring all TB patients, conducting a pill count every two weeks, and maintaining a treatment card for each patient. The study team was trained to ensure that the VHN's authority would not be undermined. Patients were interviewed on several occasions throughout the course of their treatment: the beginning, at the end of the Intensive Phase, at the end of treatment, and finally 6 months after the end of treatment. At these interviews, our field investigators would ask various questions about patients' subjective wellbeing, their experience with the treatment, their social interactions, their employment, and their experience with their DOT provider.

3.3.3 Measuring Outcomes[8]

Our study team took various measures to assess treatment progress. Patients were asked to show their pill cases to the interviewer so that the interviewer could conduct a pill count. Patients were also asked to self-report if they missed doses of their medication and how often this happened. In addition, at each interview, a sputum smear test was taken. Of note, many patients can be sick with TB and nonetheless return a negative test result. However, a positive test result is a strong indicator of an active case of TB. If patients were found to have defaulted, which is defined as not taking medication for two months, they would be returned to the public health system, restart treatment, and this would be considered a failure. Patients who died, defaulted, relapsed, or were withdrawn due to complications during the course of the treatment would not receive a follow-up interview. Importantly, all patients whom we were able to reach for the follow-up

[8] See Farmer (1999) for a general discussion of measuring treatment adherence, and Valencia et al. (2016) for a review of methodologies for studying adherence to TB treatment specifically.

interview after the close of treatment would receive an additional sputum test that was not part of the RNTCP protocol. Thomas et al. 2005 found that among TB patients who (presumably) did not fully comply with treatment and ultimately relapsed within an 18-month post-treatment window, a large percentage (over 75%) of these people relapsed and tested positive for TB within the first 6 months after their treatment was deemed complete, even if they had previously tested negative. Thus, we would expect with our 6-month-post-treatment sputum test to catch a large percentage of relapsed patients who might otherwise have been deemed cured.

3.4 Results

3.4.1 Randomization

To begin, we analyze pre-treatment characteristics that should not be affected by the treatment, to test that patients appear to be similar across the three study arms on observable characteristics, suggesting that the randomization was implemented appropriately. Table 3.1 shows the results of this analysis.

Table 3.1: Balance in initial patient characteristics

	Randomized Arm						
	Govt. DOT (Control)	Community DOT (Open)	Community DOT (Within Caste)	p-value† (all arms)	Community DOT (Combined)	Difference: (Treatment – Control)	p-value‡ (Difference)
Female	.217	.204	.215	.91	.210	−.008	.78
	(.022)	(.022)	(.022)		(.016)	(.027)	
Age	45.6	47.7	46.8	.17	47.2	1.61	.08*
	(.760)	(.775)	(.740)		(.535)	(.929)	
Weight (kg)	44.7	45.4	44.2	.23	44.8	.097	.86
	(.443)	(.532)	(.398)		(.331)	(.553)	
Height (cm)	160.7	160.4	160.4	.87	160.4	−.30	.60
	(.47)	(.47)	(.46)		(.33)	(.57)	
Category II	.143	.113	.127	.50	.120	−.023	.31
	(.019)	(.012)	(.018)		(.013)	(.023)	
Hindu	.918	.945	.943	.33	.944	.026	.14
	(.015)	(.013)	(.013)		(.009)	(.017)	
Has Ration Card	.950	.945	.964	.47	.955	.004	.77
	(.012)	(.013)	(.010)		(.008)	(.014)	
Married	.838	.893	.881	.09*	.897	.049	.04**
	(.020)	(.017)	(.012)		(.012)	(.838)	
Reports Stigma	.615	.598	.601	.89	.599	−.016	.63
	(.026)	(.027)	(.027)		(.019)	(.033)	
Education (yrs.)	5.25	4.94	4.58	.12	4.76	−.487	.08*
	(.230)	(.230)	(.226)		(.161)	(.281)	
Sputum pos. at baseline	.801	.798	.814	.86	.806	.006	.83
	(.022)	(.023)	(.022)		(.016)	(.027)	
N	345	328	335		663		

Standard errors in parentheses. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$.
† Calculated from F-statistic resulting from regression of outcome variable on binary variables for each of the two treatment arms, using heteroskedasticity-robust standard errors.
‡ Calculated from regression of outcome variable on a consolidated binary treatment variable, using heteroskedasticity-robust standadard errors.
Excluded from this table are anyone who attrited from treatment or were never fully enrolled in the study.
Category II refers to patients who are being treated after a previous default or relapse.
Stigma refers a composite measure of patients' desire to keep their diagnosis secret from the initial interview.

Table 3.1 suggests that randomization was implemented reasonably effectively. There are very few metrics along which there are any significant differences across the different arms. Patients in the treatment arms do appear to be slightly older and perhaps slightly more likely to be married, each of which might have an influence on treatment prognosis. However, these differences are small, and we do not think that there is any reason to suspect a failure to properly implement randomization across treatment arms.

3.4.2 Treatment outcomes

Our primary outcome of interest is whether patients successfully adhered to and completed their prescribed treatment regimen. Because symptoms will often abate long before treatment is complete, and failure to continue taking medications through the end of the prescribed regimen puts patients at risk of relapse and development of drug resistant disease, it is important to consider not only whether or not patients *appear* to be cured, but also whether or not they actually adhered closely to the prescribed regimen. Of course, treatment adherence is difficult to strictly monitor, so our ability to measure adherence is limited to the information we could gather from interviews, medical tests, and patient records.

Patients are considered to have failed to fully adhere to treatment if any of the following criteria are met:

1. A reported adverse outcome in which the patient withdraws from the study due to medical complications, default (failure to take medication for two months), relapse, or death.

2. Sputum microscopy test results. While a negative test result is not a strong indicator of treatment success, a positive result is a strong indicator of failure. If a patient tests positive at the end of the Intensive Phase, at the end of treatment, or after our follow-up interview six months beyond the end of treatment, we treat this as a failure on this metric.

3. Self-reported missed doses. If a patient endorses missing doses for at least two weeks during either the Intensive Phase or Continuation Phase, or misses a combined three weeks or more across the two phases, we treat this as an adherence failure.

4. Pill counts. Field investigators are asked to independently count any pills remaining at the end of the Intensive Phase and at the end of treatment. Similar to self-reported missed doses, if an equivalent of two weeks' worth of medications remain at the end of either the Intensive Phase or the Continuation Phase, or a combined number of doses exceeds three weeks' worth of medications across the two phases, we categorize this as an adherence failure.

5. If the DOT provider migrates during treatment or all eligible DOT providers listed by the patient refuse to enroll, this is also treated as an adverse outcome.

Table 3.2 shows the main results of our intervention: the effects of assigning a Community DOT Provider on treatment success. When we look at our preferred measure of treatment success, it shows no statistically significant difference between the control group and either of the treatment groups (i.e., community DOT providers). Perhaps interestingly, there is some suggestion of a difference when patients who failed extremely early on in their treatment are included. These patients would have failed within approximately the first two weeks of treatment, before many patients would even have a DOT provider assigned to them, so this would appear to be more of a statistical fluke than indicative of any greater issue with the treatment. Indeed, when we look just at early failure as an outcome, there is a significant difference in the treatment arm in the likelihood of failing early on, which would most likely indicate a statistical oddity or a failure of randomization. Another finding of potential interest is that patients assigned to a community DOT provider were more likely to voluntarily withdraw from the study or to be lost to follow-up (i.e., our study team was unable to locate the patient) *before the initial interview*, which was supposed to take place approximately two weeks after initial contact with the patient. This may indicate that once patients realized that they would have to disclose their diagnosis to a member of their community, they became reluctant to continue with the study, suggesting a possible role of stigma in treatment compliance and success of a community DOT provider assignment. Importantly, we do not find such an effect for later voluntary withdrawals (See Table 3.6, Columns (1) and (4)).

Table 3.2: Treatment Outcome

	Treatment failure (any)		Early failure		Early migration or voluntary withdrawal		**Treatment failure (excl. early)**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant (control mean)	$.405^{***}$	$.405^{***}$	$.008^{*}$	$.008^{*}$	.006	.006	$.400^{***}$	$.400^{***}$
	(.026)	(.026)	(.005)	(.005)	(.004)	(.004)	(.026)	(.026)
Community DOT (either)	.032		$.018^{**}$		$.017^{**}$		.021	
	(.032)		(.008)		(.007)		(.033)	
Community DOT (within caste)		.002		$.030^{***}$		.011		−.018
		(.037)		(.011)		(.008)		(.037)
Community DOT (open)		$.063^{*}$		.006		$.023^{**}$		.060
		(.038)		(.008)		(.010)		(.038)
Observations	1,030	1,030	1,078	1,078	1,078	1,078	1,008	1,008

Robust standard errors in parentheses. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$.
Early failure (Column 2) refers to withdrawal from the study due to severe treatment complications or death prior to the initial interview.
Early migration or voluntary withdrawal (Column 3) refers to patients who chose to withdraw from the study or were lost to follow up before the initial interview.
The primary treatment failure variable (Column 4) excludes cases of early failure, as well as voluntary withdrawal or loss to follow up at any point.

3.4.3 Potential explanations

Next, we look at potential explanations for why community DOT providers were not more effective than the government system. One possible explanation is that government DOT providers relieved of some portion of their TB assignments and understanding that they were under observation may have been more likely to exert effort in their role as DOT providers. A close examination of our results make this type of Hawthorne Effect seem unlikely. We find that failure rate in the control group is 40%, which is substantially higher than official statistics would suggest. Though the official statistics may not be reliable, a 40% failure rate would not seem to indicate that government workers were putting in extra effort at ensuring successful treatment.

Next, we consider whether community DOT providers were poorly trained or otherwise not fulfilling their assigned responsibilities. We look at patient reports of DOT provider satisfaction as well as how often they met with their DOT providers during treatment. Our measure of satisfaction with the DOT provider is based on a series of questions at the end of treatment about patients' DOT provider's performance, including their overall satisfaction and whether their DOT provider provided adequate support/information. DOT provider frequency is measured after both the Intensive Phase and the Continuation Phase. In the Intensive Phase, we ask if the DOT provider met the patient three times per week, if they met at least once per week, or less frequently than that. In the Continuation Phase, we ask if patients saw their DOT provider at least once per week. We note that because these are self-reported measures, patients may be reluctant to disclose if their DOT provider failed to fulfill their responsibilities, especially in the treatment arms in which the DOT provider was a person of the patient's own choosing.

Table 3.3: DOT Provider Performance

	Meet at least once a week - whole Tx		Meet 3x/week during IP		Meet at least once a week during CP		Patient satisfied with DOT provider	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant (control mean)	.525***	.525***	.199***	.199***	.532***	.532***	.851***	.851***
	(.027)	(.027)	(.022)	(.022)	(.029)	(.029)	(.021)	(.021)
Community DOT (either)	.342***		.478***		.324***		.108***	
	(.030)		(.029)		(.033)		(.023)	
Community DOT (within caste)		.337***		.498***		.321***		.106***
		(.033)		(.034)		(.036)		(.024)
Community DOT (open)		.346***		.458***		.326***		.110***
		(.033)		(.035)		(.036)		(.024)
Observations	977	977	937	937	849	849	854	854

Robust standard errors in parentheses. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$.

From Table 3.3, we see that patients in the two Community DOT provider arms did report significantly higher satisfaction with DOT provider performance (an 11 percentage point difference), though the vast majority of patients in all arms, including the control, responded affirmatively to questions about this. On measures of DOT provider frequency,Community DOT providers also significantly outperform the Government DOT providers, though both groups received relatively high marks. Only about 20% of Government DOT providers met with patients three times per week, compared to nearly 70% of Community DOT providers in both treatment arms. Across the whole treatment, at least 50% of Government DOT providers met with patients at least once per week, while the corresponding figure was around 85% in the treatment arms. This represents a substantial difference. Of note, the relatively high frequency with which Government DOT providers met with patients might be suggestive of a Hawthorne Effect, though nonetheless, patients in this arm failed to adhere to treatment at relatively high rates. Additionally, as noted, we should be cautious in interpreting patients' self-report of frequency meeting their Community DOT providers.

Despite this relatively promising result for DOT provider performance, higher frequency of meeting with Community versus Government DOT providers failed to translate to higher rates of measured treatment adherence. Under the assumption that our treatment randomization could only affect treatment outcomes via an improvement in DOT provider performance, this lends itself to an Instrumental Variables analysis. Table 3.4 shows the results of this analysis using the randomly assigned treatment arm as an instrument for the various measures of DOT provider performance, and looking at the effects on treatment success. We find that despite a highly significant first-stage relationship between arm assignment and DOT provider performance, there is no evidence that this translates to an improvement in treatment outcomes.

Table 3.4: Instrumental Variable Analysis: DOT Provider Performance → Treatment Success

	Treatment success (main)	Treatment success (main)	Treatment success (IP)	Treatment success (CP)
	(1)	(2)	(3)	(4)
First stage:	Instrument used:			
	DOT meet weekly	Pt. satis. w/ DOT	DOT meet 3x/week during IP	DOT meet weekly during CP
Community DOT	.352***	.108***	.478***	.324***
	(.031)	(.023)	(.029)	(.033)
F-stat	131.21	22.97	267.03	97.58
Second stage:				
$\widehat{Success}$	−.038	.137	−.046	.035
	(.094)	(.318)	(.045)	(.104)
Observations	957	854	937	849

Robust standard errors in parentheses. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$.
Treatment success IP and CP refer to measures of treatment adherence in the Intensive Phase and Continuation Phase, respectively.
Treatment success is equal to 1 − the treatment failure variable used in 3.2

There are a few potential explanations for this: (1) community DOT providers, despite their best efforts, were not actually ensuring that patients took their treatment as prescribed; (2) patients in the Community DOT provider arms, having a closer relationship to their DOT providers, were less likely to report a poor DOT provider performance, perhaps fearing the potential consequences for their friend; (3) we lack sufficient power to detect a modest improvement in treatment outcomes.[9]

The last explanation is testable, in that given our sample size, we can ascertain a minimum detectable effect size under a standard set of statistical parameters. Using a standard power level

[9]Another possibility is that treatments, even when taken as prescribed, are not as effective as we assumed, though given the ample body of evidence on the effectiveness of these medications, this explanation seems unlikely.

of 80% and a .95 significance level, and just considering the difference between the control arm and the two treatment arms pooled together, we find a minimum detectable effect for treatment success of .091, meaning that there would need to be an almost 10 percentage point improvement in success in order for us to be reasonably likely to detect it statistically. The confidence interval for the pooled Community DOT provider effect ranges from $-.04$ to .08, indicating that it is indeed unlikely that the true effect, if any, is large enough to be able to detect statistically from our data.

Finally, we consider the potential role of stigma. If stigma interfered with selection of effective DOT providers, or made patients less cooperative or willing to meet with DOT providers regularly, then we would expect a negative interaction between assignment to one of the two Community DOT provider arms and measures of stigma. That is to say, we would expect patients with greater concerns about stigma to benefit less from the assignment of a community DOT provider than those with fewer concerns about stigma. We measure stigma by constructing a composite variable based on whether (1) patients report going to a specific treatment facility specifically for the purpose of keeping their diagnosis hidden from their community, and (2) whether they would prefer to keep their diagnosis hidden from friends and relatives living in or outside of their village. If a patient endorsed either of these criteria, we categorize them as "stigma-sensitive." In case the DOT provider assignment itself affected patients' perceptions of stigma surrounding TB, we only consider for the purposes of this measure patients' *initial* responses to these questions, indicating their perceptions at the beginning of their treatment regarding the stigma they might face due to their TB diagnosis. We do not consider the corresponding *realized* experiences reported at the end of treatment for this purpose.

Table 3.5 incorporates this measure of stigma into the analysis of treatment outcomes. In this table, we use "Treatment Success" as the primary outcome of interest, which is simply the complement to the "Treatment Failure" variable used above (i.e. Treatment Success = 1 $-$ Treatment Failure). From Column (1), we can see, perhaps surprisingly, that treatment success or failure is not significantly correlated with our stigma measure. Column (2) shows that this finding holds even after controlling for the treatment arm assignment. Column (3) shows that the interaction effect is also not significantly different from zero: we observe no significant correlation between expectations of stigma and the effectiveness of the Community DOT Provider treatment. Column (4) shows that this finding holds when considering the two Community DOT Provider arms and their interaction effects with stigma perceptions separately. Interestingly, we do see that our stigma measure is slightly negatively correlated with the reported frequency of meeting of DOT providers, but this effect is not statistically significant.

Table 3.5: Interaction: Community DOT and Stigma

	Treatment Success				Meet DOT Provider Weekly				Patient satisfied with DOT provider			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Stigma	.028	.028	.016	.016	−.013	−.008	.039	.039	.020	.023	.039	.039
	(.032)	(.032)	(.055)	(.055)	(.028)	(.026)	(.056)	(.056)	(.019)	(.019)	(.045)	(.045)
Community DOT (either)		−.019	−.030	−.068		.342***	.386***	.398***		.111***	.126***	.130***
		(.033)	(.053)	(.061)		(.030)	(.048)	(.052)		(.023)	(.039)	(.041)
Community DOT (within caste)				.076				−.024				−.007
				(.061)				(.040)				(.028)
Community DOT × Stigma			.019	.013			−.072	−.082			−.024	−.031
			(.067)	(.078)			(.062)	(.067)			(.048)	(.051)
Community DOT (within caste) × Stigma				.011				.021				.013
				(.079)				(.054)				(.034)
Observations	1,002				970				850			

Robust standard errors in parentheses. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$.

One consideration to keep in mind for all of the above results is the possibility that null results could be attributable, at least in part, to attenuation bias due to measurement error. In particular, we construct our treatment outcome measures based in part on patient reports, which might not be fully reliable. Even pill-counts might not be a completely reliable measure if patients were reluctant to show leftover medications to our interviewers, or had hidden or disposed of unfinished medications. As previously noted, a negative sputum test result is not a reliable indicator of completed TB treatment, and many patients had a negative sputum test result even at the beginning of their treatment. Some patients are treated as failures due to an MDR-TB diagnosis, but it's possible that these patients could have contracted a drug-resistant strain of the bacterium rather than developing drug resistance due to treatment non-adherence.[10] Finally, even if pill counts and self-reported missed doses were recorded fully accurately, we would not be able to observe these measures for patients who chose to withdraw from the study before the end of treatment, migrated

[10]Even though this should not disproportionately affect any of the treatment arms, it could add noise to the measured outcome.

away from their homes and were lost to follow up (which is effectively a form of voluntary withdrawal), or were removed from the study due to relapse, death, or serious illness, all potentially biasing the outcomes that we were able to observe. Further, even if no doses were missed or there were no long stretches of missed doses, this does not necessarily mean that treatment compliance was perfect: we do not observe if patients strictly followed protocols of taking medications exactly 3 days per week every other day, or if they may have missed a few doses or taken them irregularly. It's possible that incorporating additional measures of treatment success, such as symptom improvement or weight gain, could ameliorate this concern to some extent.

Table 3.6 considers factors contributing to treatment attrition and whether those factors could have biased the results discussed above.

Table 3.6: Loss to Follow-Up: Community DOT and Stigma

	Loss to Follow-Up, incl. Voluntary Withdrawal			Loss to Follow-Up or Treatment Failure		
	(1)	(2)	(3)	(4)	(5)	(6)
Community DOT (either)	−.007		−.021	−.002		.005
	(.013)		(.023)	(.032)		(.052)
Stigma		−.002	−.021		−.029	−.035
		(.013)	(.024)		(.032)	(.054)
Community DOT × Stigma			.028			.009
			(.028)			(.067)
Observations	1,060	1,031	1,031	1,060	1,031	1,031

Robust standard errors in parentheses. $^{*}p < .1$, $^{**}p < .05$, $^{***}p < .01$.

From this table, we see that initial reports of stigma are not associated with patient attrition, nor is there any differential attrition across treatment arms. When we consider treatment non-adherence and attrition jointly, we similarly find no effect of the treatment assignment, nor any association with reported stigma. Of course, it should be noted that our measure of stigma is indirect and, particularly because these metrics were based on initial expectations, may not accurately capture patients' experiences.

Given the above, we are unable to show compelling evidence for *any* particular explanation as to why our Community DOT Provider intervention was not more successful than the existing system. It is important to note that though there is not evidence that Community DOT providers were more effective than the standard government system, there is also not sufficient evidence to suggest that the intervention resulted in worse treatment outcomes. It is also important to keep in mind that treatment outcomes in the control group were themselves quite poor. In fact, around 40 percent of all patients failed on at least one metric of treatment completion. This reaffirms the hypothesis that the status quo of the TB control program is not functioning particularly well; there is ample room for improvement. Our study suggests that if Community DOT Providers are to be part of that improvement, it might take a significant investment in DOT Provider performance, perhaps including more extensive training on the importance of strict treatment adherence, on building patient trust, and actually making sure to observe patients for each dose of their medication. Increasing incentives for successful completion and reducing costs for DOT providers to monitor patients may also help to improve outcomes, though further research would be required before we could reach any conclusions on any of these possibilities.

3.5 Conclusion

A crucial finding in this paper is that failure rates of TB treatment, broadly measured, were quite high across the board. While we had fairly stringent criteria for success, if it's true that even minor noncompliance with the prescribed treatment regimen for TB can greatly increase the risk of relapse and development of drug resistance, then it is clear that there is vast room for improvement in treatment compliance, which could be crucial for better controlling the spread of TB.

In our simple intervention, we allowed people to choose a member of their community to visit with them and ensure that they were taking their medications on the prescribed schedule. We find that this was effective at having patients meet more often with their DOT provider, at least based on self-reported frequency. However, this did not translate to a significant improvement in treatment compliance. We did not see substantial evidence that failure to adhere to the prescribed treatment regimen was associated with perceptions of stigma around TB.

We might conclude that either our measures of success were flawed or that the rate of improvement or worsening was too small to be detected statistically. Additionally, while we do not know for sure whether or not DOT providers were adherent to their responsibilities, a plausible explanation for our results is that more training was required to ensure that they actually observed patients taking their medications during their visits, although we do not have sufficient evidence to substantiate this claim.

Further study is required to elucidate what additional intervention could improve the efficacy of the DOT provider program (e.g., style or intensity of training, incentivization). Patient education could be another intervention worth exploring, for if patients understood the importance of adherence to treatment and what to expect of their DOT providers, it may generate greater accountability and better outcomes. This could be a simpler or less expensive intervention than rigorous DOT provider training. It might also well be the case that centralized DOT provision by trained professionals is the best system of ensuring patient compliance, and lay provision is not an effective intervention.

Another avenue that might merit exploration is to what extent there is heterogeneity across different subgroups in their performance as DOT providers. For example, one could investigate whether closer friends or family members perform their duties more effectively. It may also be worth exploring whether people of certain levels of education or from certain castes achieve higher levels of treatment success.

Finally, although we did not find evidence of a substantial role for stigma in the performance of DOT providers, we also cannot definitively rule out the presence of such an effect. It might be worth considering whether, in cases where stigma is particularly high, a DOT provider could be assigned from within the household in order to allow patients to keep their diagnosis hidden from their wider community.

References

ABLARD, J. D. (2017): "'The barracks receives spoiled children and returns men': Debating Military Service, Masculinity and Nation-Building in Argentina, 1901–1930," *The Americas*, 74, 299–329.

ACEMOGLU, D., S. NAIDU, P. RESTREPO, AND J. A. ROBINSON (2019): "Democracy Does Cause Growth," *Journal of Political Economy*, 127, 47–100.

ALBRECHT, J. W., P.-A. EDIN, M. SUNDSTRÖM, AND S. B. VROMAN (1999): "Career Interruptions and Subsequent Earnings: A Reexamination Using Swedish Data," *Journal of Human Resources*, 34, 294–311.

ALMOND, D. AND B. MAZUMDER (2011): "Health Capital and the Prenatal Environment: The Effect of Ramadan Observance During Pregnancy," *American Economic Journal: Applied Economics*, 3, 56–85.

ALMOND, D., B. MAZUMDER, AND R. VAN EWIJK (2014): "In Utero Ramadan Exposure and Children's Academic Performance," *The Economic Journal*, 125, 1501–1533.

ANDERSON, S. (2011): "Caste as an Impediment to Trade," *American Economic Journal: Applied Economics*, 3, 239–263.

ANGRIST, J. AND A. KRUEGER (1992): "Estimating the Payoff to Schooling Using the Vietnam-Era Draft Lottery," NBER Working Paper 4067.

ANGRIST, J. AND A. B. KRUEGER (1994): "Why Do World War II Veterans Earn More than Nonveterans?" *Journal of Labor Economics*, 12, 74–97.

ANGRIST, J. D. (1990): "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *American Economic Review*, 80, 313–336.

ANGRIST, J. D. AND S. H. CHEN (2011): "Schooling and the Vietnam-Era GI Bill: Evidence from the Draft Lottery," *American Economic Journal: Applied Economics*, 3, 96–118.

ANGRIST, J. D., S. H. CHEN, AND B. R. FRANDSEN (2010): "Did Vietnam Veterans Get Sicker in the 1990s? The Complicated Effects of Military Service on Self-Reported Health," *Journal of Public Economics*, 94, 824–837.

ANGRIST, J. D., S. H. CHEN, AND J. SONG (2011): "Long-term Consequences of Vietnam-Era Conscription: New Estimates Using Social Security Data," *American Economic Review*, 101, 334–338.

AUTOR, D. H., M. G. DUGGAN, AND D. S. LYLE (2011): "Battle Scars? The Puzzling Decline in Employment and Rise in Disability Receipt among Vietnam Era Veterans," *American Economic Review*, 101, 339–344.

BANERJEE, A. AND K. MUNSHI (2004): "How Efficiently is Capital Allocated? Evidence from the Knitted Garment Industry in Tirupur," *Review of Economic Studies*, 71, 19–42.

BARRO, R. J. (1999): "Ramsey Meets Laibson in the Neoclassical Growth Model," *The Quarterly Journal of Economics*, 114, 1125–1152.

BARTIK, A. W. (2018): "Moving Costs and Worker Adjustment to Changes in Labor Demand: Evidence from Longitudinal Census Data," Unpublished Manuscript, University of Illinois at Urbana-Champaign.

BASSO, G. AND G. PERI (2020): "Internal Mobility: The Greater Responsiveness of Foreign-Born to Economic Conditions," *Journal of Economic Perspectives*, 34, 77–98.

BASU, K. AND M. WONG (2015): "Evaluating Seasonal Food Storage and Credit Programs in East Indonesia," *Journal of Development Economics*, 115, 200–216.

BAUER, T. K., S. BENDER, A. R. PALOYO, AND C. M. SCHMIDT (2012): "Evaluating the Labor-Market Effects of Compulsory Military Service," *European Economic Review*, 56, 814–829.

BAUER, T. K., S. BRAUN, AND M. KVASNICKA (2013): "The Economic Integration of Forced Migrants: Evidence for Post-War Germany," *The Economic Journal*, 123, 998–1024.

BAZZI, S., A. GADUH, A. D. ROTHENBERG, AND M. WONG (2016): "Skill Transferability, Migration, and Development: Evidence from Population Resettlement in Indonesia," *American Economic Review*, 106, 2658–2698.

BECK, U., K. PAUW, AND R. MUSSA (2015): "Methods Matter: The Sensitivity of Malawian Poverty Estimates to Definitions, Data, and Assumptions," WIDER Working Paper 2015/126.

BECKER, S. O., I. GROSFELD, P. GROSJEAN, N. VOIGTLÄNDER, AND E. ZHURAVSKAYA (2020): "Forced Migration and Human Capital: Evidence from Post-WWII Population Transfers," *American Economic Review*, 110, 1430–1463.

BECKERMAN, A. AND L. FONTANA (1989): "Vietnam Veterans and the Criminal Justice System," *Criminal Justice and Behavior*, 16, 412–428.

BEDARD, K. AND O. DESCHÊNES (2006): "The Long-Term Impact of Military Service on Health: Evidence from World War II and Korean War Veterans," *American Economic Review*, 96, 176–194.

BENJAMIN, D. J., J. J. CHOI, AND G. FISHER (2016): "Religious Identity and Economic Behavior," *Review of Economics and Statistics*, 98, 617–637.

BERGER, M. C. AND B. T. HIRSCH (1983): "The Civilian Earnings Experience of Vietnam-Era Veterans," *The Journal of Human Resources*, 18, 455.

BOUFFARD, L. A. (2003): "Examining the Relationship between Military Service and Criminal Behavior during the Vietnam Era: A Research Note," *Criminology*, 41, 491–510.

BRUNE, L., X. GINÉ, J. GOLDBERG, AND D. YANG (2011): "Commitments to Save: A Field Experiement in Rural Malawi," World Bank Policy Research Working Paper No. 5748.

——— (2016): "Facilitating Savings for Agriculture: Field Experimental Evidence from Malawi," *Economic Development and Cultural Change*, 64, 187–220.

BRYAN, G., S. CHOWDHURY, AND A. M. MOBARAK (2014): "Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh," *Econometrica*, 82, 1671–1748.

BRYAN, G. AND M. MORTEN (2019): "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia," *Journal of Political Economy*, 127.

BUONANNO, P. (2006): "Long-term Effects of Conscription: Lessons from the UK," Working Paper 0604.

BURKE, M., L. F. BERGQUIST, AND E. MIGUEL (2018): "Sell Low and Buy High: Arbitrage and Local Price Effects in Kenyan Markets," *The Quarterly Journal of Economics*, 134, 785–842.

CADENA, B. C. AND B. K. KOVAK (2016): "Immigrants Equilibrate Local Labor Markets: Evidence from the Great Recession," *American Economic Journal: Applied Economics*, 8, 257–290.

CAMPANTE, F. AND D. YANAGIZAWA-DROTT (2015): "Does Religion Affect Economic Growth and Happiness? Evidence from Ramadan," *The Quarterly Journal of Economics*, 130, 615–658.

CARD, D. AND A. R. CARDOSO (2012): "Can Compulsory Military Service Raise Civilian Wages? Evidence from the Peacetime Draft in Portugal," *American Economic Journal: Applied Economics*, 4, 57–93.

CARD, D. AND T. LEMIEUX (2001): "Going to College to Avoid the Draft: The Unintended Legacy of the Vietnam War," *American Economic Review*, 91, 97–102.

CAVALLO, A. (2013): "Online and Official Price Indexes: Measuring Argentina's Inflation," *Journal of Monetary Economics*, 60, 152–165.

CAVALLO, A., G. CRUCES, AND R. PEREZ-TRUGLIA (2016): "Learning from Potentially Biased

Statistics," *Brookings Papers on Economic Activity*, 2016, 59–108.

CAVALLO, A. AND R. RIGOBON (2016): "The Billion Prices Project: Using Online Prices for Measurement and Research," *Journal of Economic Perspectives*, 30, 151–178.

CDC (2018): "Tuberculosis (TB)," Website, accessed at https://www.cdc.gov/tb/default.htm.

CHAULK, C. P. AND V. A. KAZANDJIAN (1998): "Directly Observed Therapy for Treatment Completion of Pulmonary Tuberculosis: Consensus Statement of the Public Health Tuberculosis Guidelines Panel," *JAMA*, 279, 943.

CHETTY, R., N. HENDREN, AND L. F. KATZ (2016): "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment," *American Economic Review*, 106, 855–902.

CHIRWA, E. W., A. DORWARD, AND M. VIGNERI (2012): *Seasonality, Rural Livelihoods, and Development*, Routledge, chap. Seasonality and Poverty: Evidence from Malawi, 97–116, 1st ed.

CHYN, E. (2018): "Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children," *American Economic Review*, 108, 3028–3056.

CIA (2019): "The World Factbook," Online, Field Listing: Military Service Age and Obligation, accessed at https://www.cia.gov/library/publications/the-world-factbook/fields/333.html.

CIPOLLONE, P. AND A. ROSOLIA (2007): "Social Interactions in High School: Lessons from an Earthquake," *American Economic Review*, 97, 948–965.

COLLINS, D., J. MORDUCH, S. RUTHERFORD, AND O. RUTHVEN (2010): *Portfolios of the Poor*, Princeton University Press.

CONLEY, D. AND J. HEERWIG (2012): "The Long-Term Effects of Military Conscription on Mortality: Estimates From the Vietnam-Era Draft Lottery," *Demography*, 49, 841–855.

DE HAAN, A. (1997): "Unsettled Settlers: Migrant Workers and Industrial Capitalism in Calcutta," *Modern Asian Studies*, 31, 919–949.

DEATON, A. (1997): *The Analysis of Household Surveys : A Microeconometric Approach to Development Policy*, Baltimore, MD: Johns Hopkins University Press.

DEATON, A. AND J. DRÈZE (2009): "Food and Nutrition in India: Facts and Interpretations," *Economic and Political Weekly*, 44, 42–65.

DEATON, A. AND S. ZAIDI (2002): *Guidelines for Constructing Consumption Aggregates for Welfare Analysis*, Washington, DC: World Bank, lSMS Working Paper No. 135.

Dercon, S. and P. Krishnan (2000): "Vulnerability, Seasonality and Poverty in Ethiopia," *Journal of Development Studies*, 36, 25–53.

Deryugina, T., L. Kawano, and S. Levitt (2018): "The Economic Impact of Hurricane Katrina on Its Victims: Evidence from Individual Tax Returns," *American Economic Journal: Applied Economics*, 10, 202–233.

Devereux, S., R. Sabates-Wheeler, and R. Longhurst, eds. (2012): *Seasonality, Rural Livelihoods, and Development*, Routledge, 1st ed.

Devereux, S., B. Vaitla, and S. Hauenstein Swan (2008): *Seasons Of Hunger: Fighting Cycles of Starvation Among the World's Rural Poor*, Pluto Press.

D'Haene, E., S. Desiere, M. D'Haese, W. Verbeke, and K. Schoors (2019): "Religion, Food Choices, and Demand Seasonality: Evidence from the Ethiopian Milk Market," *Foods*, 8, 167.

Dheda, K., T. Gumbo, N. R. Gandhi, M. Murray, G. Theron, Z. Udwadia, G. B. Migliori, and R. Warren (2014): "Global control of tuberculosis: from extensively drug-resistant to untreatable tuberculosis," *The Lancet Respiratory Medicine*, 2, 321–338.

Dobkin, C. and R. Shabani (2009): "The Health Effects of Military Service: Evidence from the Vietnam Draft," *Economic Inquiry*, 47, 69–80.

Duh, J. and D. Spears (2017): "Health and Hunger: Disease, Energy Needs, and the Indian Calorie Consumption Puzzle," *The Economic Journal*, 127, 2378–2409.

Ecker, O. and M. Qaim (2011): "Analyzing Nutritional Impacts of Policies: An Empirical Study for Malawi," *World Development*, 39, 412–428.

Eisenberg, D. and B. Rowe (2009): "The Effect of Smoking in Young Adulthood on Smoking Later in Life: Evidence based on the Vietnam Era Draft Lottery," *Forum for Health Economics & Policy*, 12.

Eli, S. and N. Li (2015): "Caloric Requirements and Food Consumption Patterns of the Poor," Working Paper No. 21697.

——— (2020): "Caloric Intake and Energy Expenditures in India," *The World Bank Economic Review*.

Ertola Navajas, G., P. A. López Villalba, M. A. Rossi, and A. Vazquez (2020): "The Long-Term Effect of Conscription on Personality and Beliefs," Working Paper No. 132, Universidad de San Andrés, Departamento de Economía.

Farmer, K. C. (1999): "Methods for Measuring and Monitoring Medication Regimen Adherence in Clinical Trials and Clinical Practice," *Clinical Therapeutics*, 21, 1074–1090.

FARMER, P. AND J. Y. KIM (1998): "Community Based Approaches to the Control of Multidrug Resistant Tuberculosis: Introducing 'DOTS-Plus'," *British Medical Journal*, 317, 671–674.

FRIEDEN, T. R. AND J. A. SBARBARO (2007): "Promoting Adherence to Treatment for Tuberculosis: The Importance of Direct Observation," *Bulletin of the World Health Organization*, 85, 407–409.

GALIANI, S., M. A. ROSSI, AND E. SCHARGRODSKY (2011): "Conscription and Crime: Evidence from the Argentine Draft Lottery," *American Economic Journal: Applied Economics*, 3, 119–136.

GIBBONS, M. A. AND M. A. ROSSI (2020): "Military Conscription, Sexist Attitudes, and Intimate Partner Violence," Working Paper No. 140, Universidad de San Andrés, Departamento de Economía.

GIBSON, J., K. BEEGLE, J. D. WEERDT, AND J. FRIEDMAN (2014): "What does Variation in Survey Design Reveal about the Nature of Measurement Errors in Household Consumption?" *Oxford Bulletin of Economics and Statistics*, 77, 466–474.

GILBERT, R., T. BENSON, AND O. ECKER (2019): "Are Malawian Diets Changing? An Assessment of Nutrient Consumption and Dietary Patterns Using Household-level Evidence from 2010/11 and 2016/17," IFPRI Malawi Strategy Support Program Working Paper 30, Washington, DC.

GINÉ, X. AND D. YANG (2009): "Insurance, Credit, and Technology Adoption: Field Experimental Evidence from Malawi," *Journal of Development Economics*, 89, 1–11.

GOLDBERG, J., M. S. RICHARDS, R. J. ANDERSON, AND M. B. RODIN (1991): "Alcohol Consumption in Men Exposed to the Military Draft Lottery: A Natural Experiment," *Journal of Substance Abuse*, 3, 307–313.

GRENET, J., R. A. HART, AND J. E. ROBERTS (2011): "Above and Beyond the Call: Long-Term Real Earnings Effects of British Male Military Conscription in the Post-War Years," *Labour Economics*, 18, 194–204.

GULEK, A. (2021): "Driving While Hungry: The Effect of Fasting on Traffic Accidents," *SSRN Electronic Journal*.

HALLAK, M. H. AND M. Z. A. NOMANI (1988): "Body Weight Loss and Changes in Blood Lipid Levels in Normal Men on Hypocaloric Diets during Ramadan Fasting," *The American Journal of Clinical Nutrition*, 48, 1197–1210.

HARRIS, J. R. AND M. P. TODARO (1970): "Migration, Unemployment and Development: A

Two-Sector Analysis," *American Economic Review*, 60, 126–142.

HARUVY, E. E., C. A. IOANNOU, AND F. GOLSHIRAZI (2018): "The Religious Observance of Ramadan and Prosocial Behavior," *Economic Inquiry*, 56, 226–237.

HASTINGS, J. AND E. WASHINGTON (2010): "The First of the Month Effect: Consumer Behavior and Store Responses," *American Economic Journal: Economic Policy*, 2, 142–162.

HAYTOWITZ, D. B., J. K. AHUJA, X. WU, M. SOMANCHI, M. NICKLE, Q. A. NGUYEN, J. M. ROSELAND, J. R. WILLIAMS, K. Y. PATTERSON, Y. LI, AND P. R. PEHRSSON (2019): "USDA National Nutrient Database for Standard Reference, Legacy Release," Available at data.nal.usda.gov/dataset/usda-national-nutrient-database-standard-reference-legacy-release.

HEADEY, D. AND O. ECKER (2013): "Rethinking the Measurement of Food Security: From First Principles to Best Practice," *Food Security*, 5, 327–343.

HEARST, N., T. B. NEWMAN, AND S. B. HULLEY (1986): "Delayed Effects of the Military Draft on Mortality," *New England Journal of Medicine*, 314, 620–624.

HODLER, R., P. RASCHKY, AND A. STRITTMATTER (2020): "Religion and Terrorism: Evidence from Ramadan Fasting," Unpublished manuscript.

HOVERD, W. J. AND C. G. SIBLEY (2013): "Religion, Deprivation and Subjective Wellbeing: Testing a Religious Buffering Hypothesis," *International Journal of Wellbeing*, 3, 182–196.

HSIEH, C.-T. AND E. MORETTI (2019): "Housing Constraints and Spatial Misallocation," *American Economic Journal: Macroeconomics*, 11, 1–39.

HU, Z. AND Z. WANG (2019): "Nutrition, Labor Supply, and Productivity: Evidence from Ramadan in Indonesia," *SSRN Electronic Journal*.

IMBENS, G. AND W. VAN DER KLAAUW (1995): "Evaluating the Cost of Conscription in the Netherlands," *Journal of Business and Economic Statistics*, 13, 207–215.

INTERNATIONAL ASTRONOMICAL CENTER (2021): "Islamic Crescents Observation Project (ICOP): Crescent Observation Results," Accessed at www.astronomycenter.net/res.html.

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE (IFPRI) (2020): "Replication Data for: Estimates of Nutrient Availability From the Integrated Household Surveys in Malawi," Harvard Dataverse.

IYER, S. (2016): "The New Economics of Religion," *Journal of Economic Literature*, 54, 395–441.

JAYACHANDRAN, S. (2006): "Selling Labor Low: Wage Responses to Productivity Shocks in

Developing Countries," *Journal of Political Economy*, 114, 538–575.

KARAAĞAOĞLU, N. AND S. YÜCECAN (2000): "Some Behavioural Changes Observed among Fasting Subjects, Their Nutritional Habits and Energy Expenditure in Ramadan," *International Journal of Food Sciences and Nutrition*, 51, 125–134.

KAUR, S., M. KREMER, AND S. MULLAINATHAN (2015): "Self-Control at Work," *Journal of Political Economy*, 123, 1227–1277.

KELLER, K., A. WAGENER, AND P. POUTVAARA (2010): "Does a Military Draft Discourage Enrollment in Higher Education?" *FinanzArchiv*, 66, 97.

KESHAVJEE, S. AND P. E. FARMER (2012): "Tuberculosis, Drug Resistance, and the History of Modern Medicine," *New England Journal of Medicine*, 367, 931–936.

KHANDKER, S. R. AND W. MAHMUD (2012): *Seasonal Hunger and Public Policies: Evidence from Northwest Bangladesh*, The World Bank.

KIM, J. Y., A. SHAKOW, K. MATE, C. VANDERWARKER, R. GUPTA, AND P. FARMER (2005): "Limited Good and Limited Vision: Multidrug-Resistant Tuberculosis and Global Health Policy," *Social Science & Medicine*, 61, 847–859.

KUMAR, V. AND S. L. ROBBINS (2007): *Robbins Basic Pathology*, Philadelphia, PA: Saunders/Elsevier, 8th ed.

KURAN, T. (2018): "Islam and Economic Performance: Historical and Contemporary Links," *Journal of Economic Literature*, 56, 1292–1359.

LAGAKOS, D. (2020): "Urban-Rural Gaps in the Developing World: Does Internal Migration Offer Opportunities?" *Journal of Economic Perspectives*, 34, 174–192.

LAGAKOS, D., S. MARSHALL, A. M. MOBARAK, C. VERNOT, AND M. E. WAUGH (2020): "Migration Costs and Observational Returns to Migration in the Developing World," *Journal of Monetary Economics*, Forthcoming.

LAGAKOS, D., A. M. MOBARAK, AND M. WAUGH (2018): "The Welfare Effects of Encouraging Rural-Urban Migration," NBER Working Paper 24193.

LAIBSON, D. (1997): "Golden Eggs and Hyperbolic Discounting," *The Quarterly Journal of Economics*, 112, 443–478.

LAWN, S. D. AND A. I. ZUMLA (2011): "Tuberculosis," *The Lancet*, 378, 57–72.

LAXMINARAYAN, R., E. KLEIN, C. DYE, K. FLOYD, S. DARLEY, AND O. ADEYI (2007): "Economic Benefit of Tuberculosis Control," Policy Research Working Paper 4295.

LAXMINARAYAN, R., E. Y. KLEIN, S. DARLEY, AND O. ADEYI (2009): “Global Investments in TB Control: Economic Benefits,” *Health Affairs*, 28, w730–w742.

LEE, S., M. NAM, D. JEONG, AND W. LEE (2020): “Does Ramadan Harm Infant Health? Evidence from Ethiopia,” *International Economic Journal*, 34, 613–633.

LEWIS, W. A. (1954): “Economic Development with Unlimited Supplies of Labour,” *The Manchester School*, 22, 139–191.

LINDO, J. M. AND C. STOECKER (2014): “Drawn into Violence: Evidence on ’What Makes a Criminal’ from the Vietnam Draft Lotteries,” *Economic Inquiry*, 52, 239–258.

LOPEZ, A. D., C. D. MATHERS, M. EZZATI, D. T. JAMISON, AND C. J. L. MURRAY, eds. (2006): *Global Burden of Disease and Risk Factors*, The World Bank.

LUKE, N. AND K. MUNSHI (2011): “Women as Agents of Change: Female Income and Mobility in India,” *Journal of Development Economics*, 94, 1–17.

MAFOODS (2019): *Malawian Food Composition Table*, Lilongwe, Malawi: South African Medical Research Council, Biostatistics Unit, 1st ed.

MAJID, M. F. (2015): “The Persistent Effects of In Utero Nutrition Shocks Over the Life Cycle: Evidence from Ramadan Fasting,” *Journal of Development Economics*, 117, 48–57.

MALAMUD, O. AND A. WOZNIAK (2012): “The Impact of College on Migration,” *Journal of Human Resources*, 47, 913–950.

MAZZOCCO, M. AND S. SAINI (2012): “Testing Efficient Risk Sharing with Heterogeneous Risk Preferences,” *American Economic Review*, 102, 428–468.

MCCLEARY, R. M. AND R. J. BARRO (2006): “Religion and Economy,” *Journal of Economic Perspectives*, 20, 49–72.

MINISTERIO DE DEFENSA, ARGENTINA (2015): *Libro Blanco de la Defensa 2015*, Ministerio de Defensa.

MINNESOTA POPULATION CENTER (2019): “Integrated Public Use Microdata Series, International: Version 7.2, Census of Argentina: 1970–2010,” https://doi.org/10.18128/D020.V7.2.

MOBARAK, A. M. AND M. R. ROSENZWEIG (2012): “Selling Formal Insurance to the Informally Insured,” Economic Growth Center Discussion Paper No. 1007.

MOLLOY, R., C. L. SMITH, AND A. WOZNIAK (2011): “Internal Migration in the United States,” *Journal of Economic Perspectives*, 25, 173–196.

MORTEN, M. AND J. OLIVEIRA (2018): "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City," Working Paper.

MUNSHI, K. (2011): "Strength in Numbers: Networks as a Solution to Occupational Traps," *The Review of Economic Studies*, 78, 1069–1101.

——— (2019): "Caste and the Indian Economy," *Journal of Economic Literature*, 57, 781–834.

MUNSHI, K. AND M. ROSENZWEIG (2016): "Networks and Misallocation: Insurance, Migration, and the Rural-Urban Wage Gap," *American Economic Review*, 106, 46–98.

NAKAMURA, E., J. SIGURDSSON, AND J. STEINSSON (2019): "The Gift of Moving: Intergenerational Consequences of a Mobility Shock," Working Paper, University of California, Berkeley.

NIRUPA, C., G. SUDHA, T. SANTHA, R. PONNURAJA, R. FATHIMA, V. CHANDRASEKHARAN, K. JAGGARAJAMMA, A. THOMAS, P. G. GOPI, AND P. R. NARAYANAN (2005): "Evaluation of Directly Observed Treatment Providers in the Revised National Tuberculosis Control Programme," *The International Journal Of Tuberculosis And Lung Disease*, 52, 73–77.

NSO (NATIONAL STATISTICS OFFICE) (2005a): "Note on Construction of Expenditure Aggregate and Poverty Lines for IHS2," Unpublished documentation.

——— (2005b): "Second Integrated Household Survey 2004–05 (IHS2)." Zomba, Malawi: National Statistics Office.

——— (2012): "Third Integrated Household Survey 2010–11 (IHS3)." Zomba, Malawi: National Statistics Office.

——— (2018): "Fourth Integrated Household Survey 2016–17 (IHS4)." Zomba, Malawi: National Statistics Office.

——— (2021): "Fifth Integrated Household Survey 2019–20 (IHS5)." Zomba, Malawi: National Statistics Office.

OI, W. Y. (1967): "The Economic Cost of the Draft," *American Economic Review*, 57, 39–62.

PALOYO, A. R. (2010): "Compulsory Military Service in Germany Revisited," *SSRN Electronic Journal*, Ruhr Economic Paper No. 206.

PARK, M., G. SATTA, AND O. M. KON (2019): "An Update on Multidrug-Resistant Tuberculosis," *Clinical Medicine*, 19, 135–139.

PAUW, K., U. BECK, AND R. MUSSA (2014): "Did Rapid Smallholder-Led Agricultural Growth Fail to Reduce Rural Poverty? Making Sense of Malawi's Poverty Puzzle," WIDER Working Paper No. 123.

PAXSON, C. H. (1993): "Consumption and Income Seasonality in Thailand," *Journal of Political Economy*, 101, 39–72.

PINGLE, J. F. (2007): "A Note on Measuring Internal Migration in the United States," *Economics Letters*, 94, 38–42.

POTASH, R. A. (1969): *The Army and Politics in Argentina, 1928–1945*, Stanford University Press.

——— (1980): *The Army and Politics in Argentina, 1945–1962*, Stanford University Press.

——— (1996): *The Army and Politics in Argentina, 1962–1973*, Stanford University Press.

RAJESWARI, R., R. BALASUBRAMANIAN, M. MUNIYANDI, S. GEETHARAMANI, X. THRESA, AND P. VENKATESAN (1999): "Socio-economic Impact of Tuberculosis on Patients and Family in India," *The International Journal of Tuberculosis and Lung Disease*, 3, 869–877.

RNTCP (2005): *Technical and Operational Guide for Tuberculosis Control*, Indian Ministry of Health and Family Welfare.

ROBACK, J. (1982): "Wages, Rents, and the Quality of Life," *Journal of Political Economy*, 90, 1257–1278.

ROHLFS, C. (2010): "Does Combat Exposure Make You a More Violent or Criminal Person?" *Journal of Human Resources*, 45, 271–300.

ROSEN, S. (1979): *Current Issues in Urban Economics*, Baltimore, MD: Johns Hopkins University Press, chap. Wage-based Indexes of Urban Quality of Life, 74–104.

ROSENZWEIG, M. R. AND K. I. WOLPIN (1993): "Credit Market Constraints, Consumption Smoothing, and the Accumulation of Durable Production Assets in Low-Income Countries: Investments in Bullocks in India," *Journal of Political Economy*, 101, 223–244.

SADEGHIRAD, B., S. MOTAGHIPISHEH, F. KOLAHDOOZ, M. J. ZAHEDI, AND A. A. HAGHDOOST (2012): "Islamic Fasting and Weight Loss: A Systematic Review and Meta-analysis," *Public Health Nutrition*, 17, 396–406.

SANTHA, T., R. GARG, T. R. FRIEDEN, V. CHANDRASEKARAN, R. SUBRAMANI, P. G. GOPI, N. SELVAKUMAR, S. GANAPATHY, N. CHARLES, J. RAJAMMA, AND P. R. NARAYANAN (2002): "Risk Factors Associated with Default, Failure and Death among Tuberculosis Patients Treated in a DOTS Programme in Tiruvallur District, South India, 2000." *The International Journal Of Tuberculosis And Lung Disease*, 6, 780–788.

SARVIMÄKI, M., R. UUSITALO, AND M. JÄNTTI (2019): "Habit Formation and the Misallocation of Labor: Evidence from Forced Migrations," *SSRN Electronic Journal*.

SCHNEIDER, K. (2021): "Nationally Representative Estimates of the Cost of Adequate Diets, Nutrient Level Drivers, and Policy Options for Households in Rural Malawi," *SSRN Electronic Journal*.

SCHOFIELD, H. (2014): "The Economic Costs of Low Caloric Intake: Evidence from India," Job Market Paper, Harvard University.

——— (2020): "Ramadan Fasting and Agricultural Output," Unpublished manuscript.

SCHÜNDELN, M. (2013): "Are Immigrants More Mobile than Natives? Evidence from Germany," *Journal of Regional Science*, 54, 70–95.

SECRETARÍA DE MODERNIZACIÓN (2018): "Series de Tiempo (API)," Accessed at https://datos.gob.ar/dataset/jgm_3/archivo/jgm_3.13.

SEUNG, K. J., S. KESHAVJEE, AND M. L. RICH (2015): "Multidrug-Resistant Tuberculosis and Extensively Drug-Resistant Tuberculosis," *Cold Spring Harbor Perspectives in Medicine*, 5, a017863.

SHALIHIN, N., F. FIRDAUS, Y. YULIA, AND U. WARDI (2020): "Ramadan and Strengthening of the Social Capital of Indonesian Muslim Communities," *HTS Theological Studies*, 76.

SHAPIRO, J. M. (2005): "Is There a Daily Discount Rate? Evidence from the Food Stamp Nutrition Cycle," *Journal of Public Economics*, 89, 303–325.

SHETE, P. B., M. REID, AND E. GOOSBY (2018): "Message to World Leaders: We Cannot End Tuberculosis without Addressing the Social and Economic Burden of the Disease," *The Lancet Global Health*, 6, e1272–e1273.

SIVARAJ, R., S. UMARANI, P. MURALIDHAR, AND S. PARASURAMAN (2014): "Revised National Tuberculosis Control Program Regimens with and without Directly Observed Treatment, Short-course: A Comparative Study of Therapeutic Cure Rate and Adverse Reactions," *Perspectives in Clinical Research*, 5, 16.

SMITH, L. C., O. DUPRIEZ, AND N. TROUBAT (2014): "Assessment of the Reliability and Relevance of the Food Data Collected in National Household Consumption and Expenditure Surveys," IHSN Working Paper No. 008.

STAGG, H. R., J. J. LEWIS, X. LIU, S. HUAN, S. JIANG, D. P. CHIN, AND K. L. FIELDING (2020): "Temporal Factors and Missed Doses of Tuberculosis Treatment. A Causal Associations Approach to Analyses of Digital Adherence Data," *Annals of the American Thoracic Society*, 17, 438–449.

SUBRAMANIAN, S. AND A. DEATON (1996): "The Demand for Food and Calories," *Journal of Political Economy*, 104, 133–162.

THE ECONOMIST (2020): "The Great Reversal – Covid-19 Is Undoing Years of Progress in Curbing Global Poverty," *The Economist*, May 23, 2020 edition, accessed at https://www.economist.com/international/2020/05/23/covid-19-is-undoing-years-of-progress-in-curbing-global-poverty.

THOMAS, A., P. G. GOPI, T. SANTHA, V. CHANDRASEKARAN, R. SUBRAMANI, N. SELVAKUMAR, S. I. EUSUFF, K. SADACHARAM, AND P. R. NARAYANAN (2005): "Predictors of Relapse among Pulmonary Tuberculosis Patients Treated in a DOTS Programme in South India." *The International Journal of Tuberculosis and Lung Disease*, 9, 556–561.

TODA, M. AND K. MORIMOTO (2004): "Ramadan Fasting – Effect on Healthy Muslims," *Social Behavior and Personality*, 32, 13–18.

TOPEL, R. H. (1986): "Local Labor Markets," *Journal of Political Economy*, 94, S111–S143.

TOWNSEND, R. M. (1994): "Risk and Insurance in Village India," *Econometrica*, 62, 539.

VALENCIA, S., M. LEÓN, I. LOSADA, V. G. SEQUERA, M. FERNÁNDEZ QUEVEDO, AND A. L. GARCÍA-BASTEIRO (2016): "How Do We Measure Adherence to Anti-Tuberculosis Treatment?" *Expert Review of Anti-Infective Therapy*, 15, 157–165.

VAN DEN BOOGAARD, J., M. J. BOEREE, G. S. KIBIKI, AND R. E. AARNOUTSE (2011): "The Complexity of the Adherence-Response Relationship in Tuberculosis Treatment: Why Are We Still in the Dark and How Can We Get Out?" *Tropical Medicine & International Health*, 16, 693–698.

VAN EWIJK, R. (2011): "Long-term Health Effects on the Next Generation of Ramadan Fasting during Pregnancy," *Journal of Health Economics*, 30, 1246–1260.

VELAYUTHAM, B., V. K. CHADHA, N. SINGLA, P. NARANG, V. G. RAO, S. NAIR, S. RAMALINGAM, G. N. SIVARAMAKRISHNAN, B. JOSEPH, S. SELVARAJU, S. SHANMUGAM, R. NARANG, P. PACHIKKARAN, J. BHAT, C. PONNURAJA, B. B. BHALLA, B. A. SHIVASHANKARA, G. SEBASTIAN, R. YADAV, R. K. SHARMA, R. SARIN, V. P. MYNEEDU, R. SINGLA, K. KHAYYAM, S. K. MRITHUNJAYAN, S. P. JAYASANKAR, P. SANKER, K. VISWANATHAN, R. VISWAMBHARAN, K. MATHURIA, M. BHALLA, N. SINGH, K. B. TUMANE, A. DAWALE, C. P. TIWARI, R. BANSOD, L. JAYABAL, L. MURALI, S. D. KHAPARDE, R. RAO, M. S. JAWAHAR, AND M. NATRAJAN (2018): "Recurrence of Tuberculosis among Newly Diagnosed Sputum Positive Pulmonary Tuberculosis Patients Treated under the Revised National Tuberculosis Control Programme, India: A Multi-centric Prospective Study," *PLOS ONE*, 13, e0200150.

VERDUZCO-GALLO, I., O. ECKER, AND K. PAUW (2014): "Changes in Food and Nutrition Security in Malawi: Analysis of Recent Survey Evidence," IFPRI Malawi Strategic Support Program Working Paper 06.

WANG, A., S. WANG, AND X. YE (2020): "Religion and Motivated Cognition: When Ramadan Meets the College and Entrance Exam," Unpublished manuscript.

WORLD HEALTH ORGANIZATION (2017): *Guidelines for Treatment of Drug-Susceptible Tuberculosis and Patient Care: 2017 Update*, World Health Organization.

——— (2019): *Global Tuberculosis Report 2019*, World Health Organization.

YAGAN, D. (2014): "Moving to Opportunity? Migratory Insurance over the Great Reession," Job Market Paper.

YAGER, T., R. LAUFER, AND M. GALLOPS (1984): "Some Problems Associated With War Experience in Men of the Vietnam Generation," *Archives of General Psychiatry*, 41, 327.

YOUNG, A. (2013): "Inequality, the Urban-Rural Gap, and Migration," *The Quarterly Journal of Economics*, 128, 1727–1785.

ZABEK, M. (2019): "Local Ties in Spatial Equilibrium," *Finance and Economics Discussion Series*, 2019.

ZIAEE, V., M. RAZAEI, Z. AHMADINEJAD, H. SHAIKH, R. YOUSEFI, L. YARMOHAMMADI, F. BOZORGI, AND M. J. BEHJATI (2006): "The Changes of Metabolic Profile and Weight during Ramadan Fasting," *Singapore Medical Journal*, 47, 409–414.

Appendix A: Appendix to Chapter 1

A.1 Data Appendix

A.1.1 Merging records across data sources

While in principle the DNI is meant to represent a unique individual identifier, which would in turn allow a perfect correspondence between the SIPA data and the voter rolls, there are several complicating factors. I briefly discuss these complications, and how they affect my analyses, here.

In the SIPA data, workers are identified by a number known as a "CUIL" (*Código Único de Identificación Laboral,* or Unique Labor Identification Code) rather than the DNI. CUILs are constructed by adding two digits before the DNI (typically 20 for men and 27 for women, 23 and 24 are also used in some cases), along with one check digit at the end. A small number of the CUILs included in the SIPA data did not follow this pattern, and so could not be converted to a DNI. Of the 21,222,006 unique CUILs included in the SIPA records from 1995 to 2016, 3,347 (0.02%) were in an invalid format and thus had to be dropped.

In addition, I found that in the Voter Roll data, there were relatively few DNIs listed with values greater than 50,000,000 (to be precise, 9,797 records out of 33,094,829 (0.03%) fall in this range. Because DNIs were quite sparsely assigned in this range, mostly assigned to people born in the 1990s or early 2000s, and because non-citizens are generally assigned DNIs somewhere in this range (though I was not exactly sure where that cutoff is generally supposed to fall), I eliminated all CUILs from the SIPA data corresponding to DNIs greater than 50,000,000[1]. This brought the total number of individuals from the SIPA data to try to match to the Voter Rolls down to 19,794,288, or around 93.3% of the original set of potentially valid DNIs.

If DNIs were truly unique, there would presumably be a one-to-one DNI-to-CUIL correspon-

[1] I also eliminated DNIs of 0

dence. However, DNIs assigned early on were intended to be unique *within sex* only. All DNIs that are 7 digits or fewer (below 10,000,000) are potentially duplicated across sex. In general, this corresponds to people born in 1951 or earlier – people born in 1952 or later were generally assigned an 8-digit DNI, which was intended to be unique regardless of sex. This means that in order to properly match such records, both the DNI and sex have to be recorded accurately. In addition, there are a small number of cases in which multiple people of the same sex were assigned the same 7-digit-or-lower DNI, and a small number of cases in which two people were assigned the same 8-digit DNI.

To match the remaining records from the SIPA data to the Voter Rolls, I divided both into two segments, one with 8-digit DNIs, and one with 7-or-fewer-digit DNIs. Starting with the 8-digit DNIs: we begin with 17,402,257 unique CUILs in this range. From this, I drop cases where two different CUILs correspond to the same DNI, and also cases with duplicated DNIs in this range in the Voter Roll data. This leaves us with 16,802,039 unique 8-digit DNIs in the SIPA data, about 96.6% of the valid DNIs in this range. Of these, 616,658 (3.67%) do not match up to any record in the Voter Rolls. This failure to match could be attributed to several potential explanations: (1) people may have died and been removed from the Voter Rolls, (2) the DNI might be recorded incorrectly in either data source, or (3) the person may be a non-citizen or otherwise ineligible to vote. For the sake of consistency across the analysis, I excluded all cases that I could not match to a Voter record.

Next, I consider DNIs below 10,000,000. These need to matched on both sex and DNI in order to be valid[2]. We begin with 2,392,031 observations in this range in the SIPA data. Of these, 52,501 (about 2.2%) are removed either because of the sex of the worker is not recorded in the SIPA data, or because of inconsistencies in the sex as recorded across different variables, workplaces, and years. Of these, 652,347 (27.9%) fail to match to a record of the corresponding sex in the voter rolls, whereas 1,687,183 records do match. This relatively high rate of attrition is likely due to

[2]It might not be necessary in all cases to match on sex, as some ranges of DNIs were only assigned to men or women, not both. However, identifying such ranges would have been quite costly and unlikely to generate a large quantity of additional matches.

many of these people having died and been removed from the Voter Rolls.

All told, this leaves us with 17,872,564[3] records (or about 90.3% of the 19,794,288 valid DNIs between 0 and 50,000,000) from the SIPA data matched with a reasonable degree of confidence to the Voter Rolls. In order to avoid misclassifying records that appear in the Voter Rolls but are dropped from the SIPA data due to duplication or other data inconsistencies as never appearing in the formal sector (which is one of the main outcome variables I look at), I drop 439,986 individuals (around 1.3% of the 33,094,829 records in the valid DNI range) from the Voter Roll data.

A.1.2 Imputation of province of origin

The imputation procedure I develop for province of origin relies on two observations: (1) that many people do not move from their province of origin, and (2) that DNIs were issued through centralized distribution of pre-filled forms of consecutive numbers. Thus, if we sort all individuals from the Voter Rolls by their DNI, we would expect to see high levels of correlation between adjacent observations in the *current* province of residence. We would also expect to see reasonably well-defined breaks, in which we go from a large cluster of individuals mostly residing in one province, to another cluster of individuals mostly residing in another province. Further, if we see, for example, three people in a row living in Province A, followed by one person in Province B, then another cluster of people in Province A, we can reasonably infer that the one person we identified in Province B is a migrant originally hailing from Province A. Based on this intuition, I implement the following procedure[4] (I also include an illustrative example from the actual Voter Roll data that is reasonably representative of the full dataset):

Step 1: Starting with the data sorted by DNI (see Figure A.1), define "groups" of consecutive observations *currently* living in same place, born around the same time. Specifically, I consider an individual to be the start of a new "group" if he lives in a different province from the previous

[3]A much smaller number are included in the final analysis, as for this paper I study only men, only those in cohorts in which some individuals were conscripted (and some not), generally only those who were of prime working age at some point between 1995 and 2016, and generally excluding those for whom I could not impute a province of origin.

[4]This description is slightly simplified for the sake of clarity. The full STATA do-file that implements the imputation procedure is available from the author upon request.

Figure A.1: Starting Data (Example)

	province	DNI	birthyr
249	Santa Fe	XXXX2951	1956
250	Buenos Aires	XXXX2952	1956
251	Santa Cruz	XXXX2953	1958
252	Santa Fe	XXXX2954	1958
253	Santa Fe	XXXX2957	1956
254	Santa Fe	XXXX2962	1956
255	Córdoba	XXXX2963	1958
256	Santa Fe	XXXX2966	1952
257	Buenos Aires	XXXX2969	1956
258	Santa Fe	XXXX2972	1958
259	Santa Fe	XXXX2973	1958
260	Santa Fe	XXXX2976	1958
261	Tierra del Fuego	XXXX2980	1958
262	Santa Fe	XXXX2982	1956
263	Santa Fe	XXXX2984	1958
264	Santa Fe	XXXX2985	1959
265	Buenos Aires	XXXX2986	1958
266	Santa Fe	XXXX2988	1958
267	Buenos Aires	XXXX2991	1956
268	Santa Fe	XXXX2993	1956
269	Santa Fe	XXXX2995	1957
270	Córdoba	XXXX2997	1956
271	Santa Fe	XXXX2998	1958
272	Tucumán	XXXX3001	1956
273	Tucumán	XXXX3002	1958
274	Tucumán	XXXX3005	1958
275	Santiago del Estero	XXXX3006	1956
276	Tucumán	XXXX3010	1956

Figure A.2: Initial Grouping of Example Data

	province	DNI	birthyr	grp_tag	num_grp
249	Santa Fe	XXXX2951	1956	0	2
250	Buenos Aires	XXXX2952	1956	1	1
251	Santa Cruz	XXXX2953	1958	1	1
252	Santa Fe	XXXX2954	1958	1	3
253	Santa Fe	XXXX2957	1956	0	3
254	Santa Fe	XXXX2962	1956	0	3
255	Córdoba	XXXX2963	1958	1	1
256	Santa Fe	XXXX2966	1952	1	1
257	Buenos Aires	XXXX2969	1956	1	1
258	Santa Fe	XXXX2972	1958	1	3
259	Santa Fe	XXXX2973	1958	0	3
260	Santa Fe	XXXX2976	1958	0	3
261	Tierra del Fuego	XXXX2980	1958	1	1
262	Santa Fe	XXXX2982	1956	1	3
263	Santa Fe	XXXX2984	1958	0	3
264	Santa Fe	XXXX2985	1959	0	3
265	Buenos Aires	XXXX2986	1958	1	1
266	Santa Fe	XXXX2988	1958	1	1
267	Buenos Aires	XXXX2991	1956	1	1
268	Santa Fe	XXXX2993	1956	1	2
269	Santa Fe	XXXX2995	1957	0	2
270	Córdoba	XXXX2997	1956	1	1
271	Santa Fe	XXXX2998	1958	1	1
272	Tucumán	XXXX3001	1956	1	3
273	Tucumán	XXXX3002	1958	0	3
274	Tucumán	XXXX3005	1958	0	3
275	Santiago del Estero	XXXX3006	1956	1	1
276	Tucumán	XXXX3010	1956	1	1

individual, or if his birth year is more than 2 years before or after the previous person's. The resulting groupings for the example data are shown in Figure A.2.

Step 2: Exclude singleton groups.[5] Re-sort included observations by DNI; all observations will be adjacent to at least one other from the same province, many will now have several more adjacent observations from the same province.

Step 3: Re-group observations by the same criteria as Step 1. Resulting groups for the example data after this step are shown in Figure A.3.

[5]Here, I introduce a slight modification for the Province of Buenos Aires, which is the largest province in the country by a wide margin; includes the urban and suburban areas outside of Buenos Aires city proper, but also extends far beyond this; receives the largest number of internal migrants (though it does *not* have the highest fraction born outside the province); and has relatively low out-migration rates for people born there. Because it is not uncommon that two adjacent observations living in this very large province are both migrants, I exclude groups smaller than 3 current Buenos Aires residents as part of this same step.

Figure A.3: Example Data: Re-grouped

	province	DNI	birthyr	grp_tag	num_grp	grp_tag_new	num_grp_new
180	Santa Fe	XXXX2951	1956	0	2	0	68
181	Santa Fe	XXXX2954	1958	1	3	0	68
182	Santa Fe	XXXX2957	1956	0	3	0	68
183	Santa Fe	XXXX2962	1956	0	3	0	68
184	Santa Fe	XXXX2972	1958	1	3	0	68
185	Santa Fe	XXXX2973	1958	0	3	0	68
186	Santa Fe	XXXX2976	1958	0	3	0	68
187	Santa Fe	XXXX2982	1956	1	3	0	68
188	Santa Fe	XXXX2984	1958	0	3	0	68
189	Santa Fe	XXXX2985	1959	0	3	0	68
190	Santa Fe	XXXX2993	1956	1	2	0	68
191	Santa Fe	XXXX2995	1957	0	2	0	68
192	Tucumán	XXXX3001	1956	1	3	1	3
193	Tucumán	XXXX3002	1958	0	3	0	3
194	Tucumán	XXXX3005	1958	0	3	0	3

Figure A.4: Example Data: Re-grouped (again)

	province	DNI	birthyr	grp_tag	num_grp	grp_tag_new	num_grp_new
180	Santa Fe	XXXX2951	1956	0	2	0	68
181	Santa Fe	XXXX2954	1958	1	3	0	68
182	Santa Fe	XXXX2957	1956	0	3	0	68
183	Santa Fe	XXXX2962	1956	0	3	0	68
184	Santa Fe	XXXX2972	1958	1	3	0	68
185	Santa Fe	XXXX2973	1958	0	3	0	68
186	Santa Fe	XXXX2976	1958	0	3	0	68
187	Santa Fe	XXXX2982	1956	1	3	0	68
188	Santa Fe	XXXX2984	1958	0	3	0	68
189	Santa Fe	XXXX2985	1959	0	3	0	68
190	Santa Fe	XXXX2993	1956	1	2	0	68
191	Santa Fe	XXXX2995	1957	0	2	0	68
192	Tucumán	XXXX3001	1956	1	3	1	6
193	Tucumán	XXXX3002	1958	0	3	0	6
194	Tucumán	XXXX3005	1958	0	3	0	6

Step 4: Analogous to excluding singleton groups in Step 2, now exclude groups of two. Then redo grouping as in Step 3. Results for the example data after this step are shown in Figure A.4.

Step 5: Keep only groups of 5 or more. Define the province, the min and max DNI, and the min and max birth year for all remaining groups.

Step 6: Reintroduce excluded observations and sort by DNI. Those falling within year and DNI range of a defined group are "absorbed" into the surrounding group. Absorbed observations from a different province are labeled migrants. Non-absorbed observations are not assigned to any

Figure A.5: Final Example Data with Province of Origin Imputations

	province	DNI	birthyr	grp_tag_new	num_grp_new	grp_prov	excl_final	absorbed
249	Santa Fe	XXXX2951	1956	0	68	Santa Fe	0	0
250	Buenos Aires	XXXX2952	1956	.	.	Santa Fe	0	1
251	Santa Cruz	XXXX2953	1958	.	.	Santa Fe	0	1
252	Santa Fe	XXXX2954	1958	0	68	Santa Fe	0	0
253	Santa Fe	XXXX2957	1956	0	68	Santa Fe	0	0
254	Santa Fe	XXXX2962	1956	0	68	Santa Fe	0	0
255	Córdoba	XXXX2963	1958	.	.	Santa Fe	0	1
256	Santa Fe	XXXX2966	1952	.	.	.	1	.
257	Buenos Aires	XXXX2969	1956	.	.	Santa Fe	0	1
258	Santa Fe	XXXX2972	1958	0	68	Santa Fe	0	0
259	Santa Fe	XXXX2973	1958	0	68	Santa Fe	0	0
260	Santa Fe	XXXX2976	1958	0	68	Santa Fe	0	0
261	Tierra del Fuego	XXXX2980	1958	.	.	Santa Fe	0	1
262	Santa Fe	XXXX2982	1956	0	68	Santa Fe	0	0
263	Santa Fe	XXXX2984	1958	0	68	Santa Fe	0	0
264	Santa Fe	XXXX2985	1959	0	68	Santa Fe	0	0
265	Buenos Aires	XXXX2986	1958	.	.	Santa Fe	0	1
266	Santa Fe	XXXX2988	1958	.	.	Santa Fe	0	1
267	Buenos Aires	XXXX2991	1956	.	.	Santa Fe	0	1
268	Santa Fe	XXXX2993	1956	0	68	Santa Fe	0	0
269	Santa Fe	XXXX2995	1957	0	68	Santa Fe	0	0
270	Córdoba	XXXX2997	1956	.	.	.	1	.
271	Santa Fe	XXXX2998	1958	.	.	.	1	.
272	Tucumán	XXXX3001	1956	1	6	Tucumán	0	0
273	Tucumán	XXXX3002	1958	0	6	Tucumán	0	0
274	Tucumán	XXXX3005	1958	0	6	Tucumán	0	0
275	Santiago del Estero	XXXX3006	1956	.	.	Tucumán	0	1
276	Tucumán	XXXX3010	1956	.	.	Tucumán	0	1

province of origin. The final groupings for the example data are shown in Figure A.5.[6]

[6]The keen observer might note that there appear to be two excluded observations right after the defined end of the "Santa Fe" group that are likely candidates for inclusion in that group, and that the switch-over from the "Santa Fe" to "Tucumán" groups happens around a X999 – X001 boundary, which in the context of the larger dataset appears to be fairly common (specifically, I see many groups ending in 50 or 00 (and starting with 01 or 51), and to a lesser extent ending in 25 or 75 (and starting with 26 or 76). Clearly, it could be useful to take such observations into account in future research using these data or other similar data. It is also worth noting that a potential shortcoming with the procedure described here is that it might do a better job identifying province of origin for places from which few people leave, and might generate misleading imputations for bordering provinces with high levels of cross-migration, or for provinces that in general receive large numbers of migrants.