Diversity and Team Performance in a Kenyan Organization*

Benjamin Marx† Vincent Pons‡ Tavneet Suri §
Sciences Po and CEPR *Harvard University and NBER* *MIT and NBER*

October 2020

Abstract

We present the results from a field experiment on team diversity. Individuals working as door-to-door canvassers for a non-profit organization were randomly assigned a teammate, a supervisor, and a list of individuals to canvass. This created random variation within teams in the degree of horizontal diversity (between teammates), vertical diversity (between teammates and their supervisor) and external diversity (between teams and the individuals they canvassed). We observe team-level measures of performance and find that horizontal ethnic diversity decreases performance, while vertical diversity often improves performance, and external diversity has no effect. The data on time use suggests that horizontally homogeneous teams organized tasks in a more efficient way, while vertically homogeneous teams exerted lower effort.

Keywords: ethnic diversity, organizational behavior, labor management, performance.

JEL Classification: D22, J24, L22, M54, O12.

*We are extremely grateful to Suleiman Asman, Eleanor Wiseman, Geoffrey Onyambu and Bonnyface Mwangi for exceptional research management and assistance in the field. We benefited from helpful comments and suggestions from Anna Aevarsdottir, Evan Apfelbaum, Abhijit Banerjee, Nick Bloom, Ray Fisman, Christina Patterson, Jonathan Petkun, Otis Reid, Michael Stepner, as well as seminar audiences at Bocconi University, Ecole Polytechnique, the Econometric Society, Georgetown, IIES Stockholm, IPEG, MIT, MIT Sloan, the NBER Summer Institute, Novafrica, the Rotman School of Management, University of Capetown, and UVA Batten School. We gratefully acknowledge financial support from MIT Sloan School of Management and the Program on Innovation in Markets and Organizations at MIT Sloan. Suri is the corresponding author: E62-517, 100 Main Street, Cambridge MA 02142.

†Sciences Po Department of Economics. Email: benjamin.marx@sciencespo.fr.

‡Harvard Business School. Email: vpons@hbs.edu.

§MIT Sloan School of Management. Email: tavneet@mit.edu.

1 Introduction

A central question in organizations is how to strike an optimal balance between diversity and homogeneity within teams of workers. While diversity brings in a wider range of skills and ideas, it may also create communication costs and other frictions inside the organization. Many studies (Lazear, 1999; Prat, 2002; Lyons, 2017) have highlighted this trade-off. Which level of diversity maximizes team performance? Does team diversity make management more effective, or facilitate relationships with clients and partners?

In this paper, we study these questions in the context of a field experiment conducted inside a Kenyan organization. A key contribution of our design is that we are able to study, within a single organization, the effect of three dimensions of diversity. First, we analyze the impact of horizontal diversity (diversity among individuals who hold the same position and collaborate with each other) on various measures of team performance. Second, we look at the effect of hierarchical or "vertical" diversity (between supervisors and supervisees) on performance. Third, we test whether similarity between team members and the individuals they interact with outside the organization affects performance. Our primary focus is on ethnic diversity, which is known to be an important determinant of productivity in Sub-Saharan Africa (Alesina and Ferrara, 2005) and Kenya specifically (Hjort, 2014). We benchmark the effects of ethnic diversity against those of another randomly assigned characteristic of teams: gender diversity.

Our experiment was conducted inside a non-profit research organization which conducted a large voter canvassing and registration exercise in 2012. The experiment focuses on a standard set of tasks that employees of this organization conduct and that we use to measure effort and performance: in particular, canvassing success rates and canvassing times, time use, and registration rates among target households. We find that ethnically homogeneous teams on the horizontal dimension perform better across most of these measures, and that this effect intensifies over the duration of the exercise. In particular, horizontally homogeneous teams conduct more and longer visits, spend more time in the field, and are more likely to divide up tasks. However, vertical homogeneity has the opposite effect on team performance. Teams with a supervisor whose ethnicity matches that of one of the canvassers conduct shorter canvassing visits and spend less time in the field, suggesting that effort levels are lower in those teams. Canvassing outcomes are unaffected by ethnic homogeneity between canvassers and the residents which staff members were expected to canvass.[1]

While we are underpowered to detect effects on voter registration, an outcome which we could only measure in a subset of our experimental sample, the effects of horizontal, vertical and external homogeneity on registration are broadly consistent with the results we find in terms of

[1]This result, while perhaps surprising, is consistent with the findings in Berge et al. (2020) who show that ethnic bias among Kenyans may not be as extensive as previously thought. We show that canvassers did not strategically target households of the same ethnicity and that they did not spend more time canvassing these households.

visit completion and duration. Vertical homogeneity, in particular, has a negative effect on registration. Overall, the effects of horizontal and vertical ethnic homogeneity, which are opposite in sign, are large in magnitude and robust to a range of specifications. These effects are corroborated by self-reported estimates of the staff members' own performance, working hours, and socialization outcomes—horizontally homogeneous teams are more likely to socialize outside fieldwork, while the opposite holds for vertically homogeneous teams. By comparison, evidence that gender diversity affects team performance is much more mixed. Horizontal gender homogeneity positively impacts performance only in some specifications, this impact is smaller than that of horizontal ethnic homogeneity, and vertical gender homogeneity has no effect.

The context of this study was a door-to-door canvassing campaign conducted in collaboration with the Independent Electoral and Boundaries Commission of Kenya (IEBC) in 2012. The canvassing experiment aimed to increase voter registration in Nairobi's largest informal settlement, called Kibera. The intervention targeted approximately 15,000 households across 300 enumeration areas (census tracts)[2] and provided residents with information on the registration process via door-to-door canvassing conducted by teams of two.

Subsumed within this registration experiment, we designed a field experiment to study how diversity within teams and along the organizational hierarchy affects effort and performance. This involved three levels of randomization. First, we randomly allocated the canvassers to teams of two. Second, we randomly allocated each team to a supervisor responsible for monitoring them in the field. This created variation in vertical homogeneity, i.e., in the degree of heterogeneity between managers and the workers they supervised. Third, each team was randomly allocated a set of enumeration areas (EAs) to visit in a mandatory random order. This introduced variation in the degree of similarity between canvassers and the residents which staff members were expected to canvass. An important feature of our experimental design is that we observe the outcome of a large number of tasks (approximately 30,000 canvassing visits conducted over three weeks) across a relatively small number of teams (30).

The organization which conducted the canvassing visits on behalf of the IEBC resembles other survey companies across the world. A sizeable fraction of its workforce consists of short-term staff whose main task is to canvass or interview households, and substantial resources are invested in the monitoring of this short-term staff. In the context of this organization, we measure team performance in several ways. First, we observe whether a particular canvassing visit was completed, i.e. whether the correct household was found and whether the pair of canvassers was able to complete the canvassing interview. Second, we look at the duration of each visit. As we discuss in Section 3, conditional on the same amount of time spent in the field, longer visit durations are a positive outcome in the context of our study since they

[2]This is the size of our treatment sample. There was an equal-sized control group where households did not receive any canvassing visits.

indicate both effort on the part of canvassers and active involvement of the target households. This is consistent with insights from the get-out-the-vote literature, in which discussions with voters are found to be more effective when campaigners take the time to connect with them (Nickerson, 2007; Gerber and Green, 2015). The total duration of visits conducted by a team also correlates positively with various measures of self-reported voter registration in our data. Third, we look at (self-reported) registration outcomes measured in a phone survey conducted at the end of the registration period. Fourth, we analyze aggregate measures of time use and team organization, namely the total amount of time spent each day in the Kibera settlement, the total amount of time spent canvassing residents, and the fraction of visits conducted alone. Fifth, we look at individual assessments of the staff's own performance from a survey of the canvassers conducted after the experiment.

Our empirical findings shed light on a simple mechanism behind the diversity-homogeneity trade-off in organizations: diversity has different effects along different dimensions of organizational structure. On the one hand, diversity within teams of workers who collaborate with each other may reduce efficiency and performance by inducing communication costs, disutility from collaboration, or other frictions. On the other hand, homogeneity along the organization's hierarchy can also undermine effort and performance, due to preferential treatment and a loosening of discipline between managers and co-ethnic employees,[3] or other distortions of managerial effort, as in Bandiera et al. (2009). Given the growing amount of empirical evidence on the role of management in firms (Bloom and Van Reenen, 2007; Giuliano et al., 2009; Lazear et al., 2015), including firms in developing countries (Bloom et al., 2013, 2020), addressing the question of diversity in organizations requires evaluating the differential gains diversity brings within teams of co-workers and across firm hierarchies.

1.1 Contribution to Literature

There is a large literature on diversity in teams and organizations across economics, psychology, and organizational behavior. Reviewing the literature outside economics is beyond the scope of this paper, but excellent discussions can be found in Apfelbaum et al. (2014) and Phillips and Shim (2011). In economics, the question of whether and how diversity matters has been studied across various fields. In this section, we highlight some of the studies in these areas and discuss the contributions of our paper to the existing literature.[4]

[3] In a related paper, Barr and Oduro (2002) find large earnings differentials between ethnic groups in Ghanaian manufacturing, which they argue result from employers favoring their co-ethnics in terms of pay and job allocation.

[4] Aside from the evidence we discuss on firms and organizations, many studies explore the role of ethnic diversity in other economic spheres, including credit markets (Fafchamps, 2000; Fisman et al., 2017), financial markets (Hjort et al., 2019), education (Fairlie et al., 2014), justice (Anwar et al., 2012; Shayo and Zussman, 2011), professional sports (Price and Wolfers, 2010) and public goods provision (Alesina et al., 1999; Miguel and Gugerty, 2005; Algan et al., 2016). At the macro level, a large literature documents the strong negative correlations between diversity and

The early literature on diversity and performance was largely theoretical and descriptive. Becker (1957) argues that employees of a firm sometimes incur disutility from working with individuals from other groups. This taste-based discrimination can explain why wage differentials arise between ethnic groups: discriminators are compensated for working with the groups they do not like. Arrow (1973) proposes a model of statistical discrimination in which employers have beliefs about the ability of different groups, while Lazear (1999) and Prat (2002), among others, discuss the conditions under which homogeneity is optimal.

At the same time, there exists a "business case for diversity" positing that diverse teams can serve a broader spectrum of customers and solve a wider range of problems (Hamilton et al., 2012). This argument posits that ethnic diversity brings in diversity of skills—different groups have access to different information and skill sets, generating complementarities across groups. For such complementaries to arise, the skills possessed by different groups must be relevant to one another, and there must exist opportunities for learning (Lazear, 1999). Consequently, the relationship between diversity and productivity should become more positive as groups become more diverse in terms of skills, and more negative when groups are more diverse along other dimensions, such as language (Hamilton et al., 2012).[5] Furthermore, diversity can improve the functioning of markets by reducing herding behavior (Levine et al., 2014).

Hjort (2014) studies the effect of ethnic diversity on team productivity in the Kenyan flower industry. He finds that diversity adversely affects productivity (consistently with preferences à la Becker), and that ethnic conflict accentuates this negative effect. His results, however, are at odds with those of Hamilton et al. (2003) and those in Hoogendoorn et al. (2014) and Hoogendoorn and van Praag (2012), who study the impact of ability dispersion and diversity on team performance in a population of MBA students and young entrepreneurs in the Netherlands. In a related paper, Lyons (2017) studies diversity by nationality in teams within oDesk (an online labor market) in a field experiment. She finds that allowing teamwork in homogeneous teams improves performance, while teamwork in diverse teams has the opposite effect.

Furthermore, Lang (1986) argues that preference-based arguments are not borne out empirically. He frames the question of ethnic diversity in terms of the costs that communication between members of different groups imposes to the firm. Zenger and Lawrence (1989) similarly show that age differences increase communication costs between employees. In a related experiment, Reinhard and Warglien (2007) investigate the conditions under which teammates develop a common language. They find that a functional working language (code) is more likely to arise if teammates possess *ex ante* a larger repertoire of common linguistic symbols. These technological advantages are augmented with the higher ability of homogeneous teams

measures of economic performance across and within countries (Alesina and Ferrara, 2005).

[5]The findings of Hamilton et al. (2012) are broadly in line with these predictions: using personnel records from a Californian garment factory, they find that teams heterogeneous in ability are more productive, while more diverse teams in terms of age and ethnicity are less productive after controlling for ability.

to prevent free-riding through sanctions—Kandel and Lazear (1992) show this formally. Diversity, on the other hand, tends to weaken social ties and to reduce peer pressure (Hamilton et al., 2012), which in turn hampers the ability of teams to prevent free-riding (Mas and Moretti, 2009).

In this paper, we contribute to this literature along several dimensions. First, the literature has generally focused on the impact of horizontal homogeneity between teammates, rather than the impact of homogeneity across the firm hierarchy. This is at odds with the large literature highlighting the key role managers and the organization of management play in determining form performance (see Lazear et al. (2015) and Caliendo et al. (2015), among others). One important exception is Bandiera et al. (2009), who find that social connections between managers and workers increase individual productivity, but also that an increase in the average level of connections can decrease firm productivity by inducing distortions in the allocation of managerial effort. Our contribution is to study both horizontal as well as vertical diversity within the same organization and along two dimensions—ethnicity and gender.[6] In addition, we study how external homogeneity, i.e. similarities between workers and the individuals they interact with outside the organization, affects performance. In the context of information campaigns, there is some interest in understanding how team composition and incentives can improve performance. For example, Berg et al. (2019) show that monetary incentives can mitigate the effects of "social distance" as part of an information campaign on public health insurance in India. We present an experimental setting in which the pairing of target respondents with the agents hired to spread information was explicitly randomized.

Second, we contribute to the identification of the effects of team diversity on effort and performance by using a field experiment to generate random variation in team diversity outside the laboratory. Because of this randomized design, we are able to identify the causal impacts of diversity along the three dimensions discussed above and on a range of outcomes. Our detailed data on time use, team cohesiveness and team organization in the field, visit outcomes and voter registration outcomes, and our ability to track these outcomes over time enable us to clarify the mechanisms responsible for these effects.

The rest of the paper is organized as follows. Section 2 describes our experiment and section 3 describes our data. We present the econometric framework in section 4 and our main results in section 5. Section 6 concludes.

[6] Hjort (2014) uses quasi-experimental variation in both "upstream" and "downstream" ethnic interactions, but in his context the upstream/downstream dimension refers to the supply chain, rather than an actual reporting hierarchy within the firm. Karaca-Mandic et al. (2013) exploit the conditional random assignments of army recruits to study the impact of group composition on time to promotion. They find that an increase in the fraction of own-group (e.g., female) peers decreases the time to promotion for members of other groups (e.g., men), while the fraction of own-group members in leadership has the opposite effect. Our empirical framework is slightly different, since we are interested in the effect of homogeneity/diversity on joint (team) output.

2 Experimental Design

2.1 Context and Experimental Site

A general election was held in Kenya in March 2013. Prior to this election, in the aftermath of a heavily contested election in 2007, a new electoral commission (called the IEBC) undertook to re-register the Kenyan electorate using biometric equipment. Registration took place across the entire country in November-December 2012. We collaborated with the IEBC during this period to implement a door-to-door canvassing experiment in Kibera, the largest of Nairobi's slums.[7]

Kibera was chosen as the site of the experiment for two reasons. First, it is host to a large migrant population, and there is evidence that migrant voters suffer the most from the various obstacles to registration (Braconnier et al., 2017). Kibera had also been a focal point in the 2007-2008 election violence. Second, building on earlier fieldwork conducted in the slum (Marx et al., 2019), we had data available on the households residing in Kibera, including a geo-localized census of residents and maps for each enumeration area (EA) in the slum. A more detailed description of this data is provided in Section 3.

2.2 Ethnic Diversity in Kenya

There are over 30 different ethnic groups in Kenya, as defined in the official census, including five main groups representing over sixty percent of the population: the Kikuyus (17%), Luhyas (14%), Kalenjins (13%), Luos (10%), and Kambas (10%) (figures from the 2009 Kenyan census). These groups differ in terms of their linguistic proximity with each other—the Kikuyus, Luhyas and Kambas are Bantu-speaking peoples, while the Luos and Kalenjins belong to the Nilotic linguistic family. Political coalitions between these groups have also varied over time. Three of the four Kenyan presidents hailed from the majority Kikuyu group, but the longest-serving Kenyan president, Daniel Arap Moi, was a Kalenjin. The ethnic composition of the site of the experiment, the Kibera slum, is not reflective of that of Kenyan society overall. Luos and Luhyas are the most prevalent groups in Kibera (together constituting 63% of the slum population, based on our 2012 census), while Kikuyus are a minority locally (6%).

2.3 Canvassing Experiment

2.3.1 EA Randomization

The voter registration experiment was designed to tease apart the effects of pure information on the registration process versus those of specific information about the new electoral commission, both of which were disseminated via door-to-door canvassing. The experimental design

[7] This experiment was designed to contribute to the literature on voter education campaigns in developing countries (related work includes Aker et al. (2017) and Banerjee et al. (2010), among others).

included two treatment groups. In the first group, households were encouraged to register and given information about the location of nearby registration centers, and what documents were required to register. In the second group (henceforth the IEBC treatment), households were, in addition, given detailed information about the IEBC, in particular about its record in organizing successful by-elections and its efforts to establish an accurate register based on biometrics.

The experimental treatment was clustered at the level of the EA (EAs in Kibera typically correspond to a compact block of dwellings). There are 643 EAs in the slum, 603 of which were part of our sampling frame. 303 EAs were randomly chosen to be part of the control group, after stratifying by village and by being above or below the median EA population (there are 16 villages in our 2012 census of Kibera). Half of the remaining EAs were allocated to each treatment group. Overall, the experiment covered 31,646 households across 603 EAs, with a total of 15,676 households living in the 300 treatment EAs.

2.3.2 Implementation

The fieldwork for this experiment was entrusted to a Kenyan organization affiliated with a U.S.-based non-profit research institution. The Kenyan organization has conducted field operations in Kenya and other African countries for over fifteen years. The organization typically hires long-term research coordinators and research assistants, as well as short-term survey staff. Some of these staff are hired again to work on other projects, based on performance. The effort and performance of the survey staff are a major determinant of the organization's overall productivity and operating costs. Monitoring the performance of this staff requires high-frequency supervision and quality controls, as is typically the case in survey companies across the world.

To implement the experiment, the organization recruited sixty canvassers for a period of three weeks. Sixteen of these canvassers had some prior fieldwork experience with the organization (and in Kibera) and the rest were new staff members hired specifically for the canvassing exercise. The canvassers were paid by the day, and wages were not tied to specific indicators of performance. Prior to the beginning of the door-to-door exercise, the canvassers were carefully trained on canvassing goals and methods. Teams were also given specific guidelines on the details of each treatment, along with a cheat sheet of the list of issues to be covered in each treatment group and a list of questions that were expected to frequently come up in the field.

Three layers of supervision were established over the sixty canvassers. The first layer was a team of seven "group leaders". Group leaders were assigned between four and five teams to accompany in the field. If teams fell behind schedule or if a team member was absent on a given day, the group leader was to step in and conduct canvassing visits instead. In practice, due to staff absences and replacements, group leaders ended up being mostly involved in implementation (canvassing) rather than supervision or management; hence we do not consider them in the analysis of vertical diversity. The second layer was a team of seven supervisors, all of whom

were previous employees of the organization with more advanced fieldwork experience. The responsibility of the supervisors was to spend each day in the field monitoring the teams under their responsibility. They actively watched canvassers in the field and independently checked with a subset of households whether the canvassing visit had taken place as planned. Finally, the entire field exercise was overseen by a director of operations, a research assistant, and the principal investigators (the authors of this paper). Team supervisors reported back to the director of operations and the research assistant on a daily basis.

To maximize contact with households, each EA was covered twice during implementation. Given our previous work in Kibera (in particular the slum census we conducted in 2012), we expected that any member of a given household would be found at home a little less than half the time. Many household members spend the day outside their dwelling, either working or looking for employment. Activities were therefore planned so that each household would be visited twice over the two-week period, once each week, and by a different team each week.

2.4 Diversity Experiment

Over the canvassing experiment, we collaborated with the organization to overlay an experiment on team diversity, with three layers of randomization. The canvassing exercise was conducted by teams of two. We randomly paired canvassers with each other, and in doing so, introduced random variation in the degree of horizontal diversity within each team. In this randomization, we ensured that junior canvassers (those with no previous experience working in Kibera) were generally paired with a more experienced canvasser, though this was not always feasible since only 16 senior canvassers were available to work on this campaign. Throughout our analysis, we show intent-to-treat estimates based on the initial random composition of teams, thus ignoring occasional staff replacements.

We then randomly allocated four to five teams to a supervisor, so as to induce random variation in vertical diversity. In the analysis below, "vertically diverse" teams are defined as teams where the ethnicity of the supervisor differs from that of both team members. The seven supervisors belonged to the following six ethnicities: Embu, Kamba, Kikuyu, Kisii, Luo, and Teso. Finally, we randomly allocated EAs (and hence residents) to each team to create variation in external diversity, i.e. in how different teams were from the households on their canvassing list.

Each team was allocated to one of the two treatment groups, and assigned a random set of twenty EAs to cover (ten per week of canvassing). On average, any given team had to complete 1,045 visits over the two week period. The set of EAs assigned in the first week was different from the set of EAs assigned in the second week, so that a household would not be visited by the same canvassing team twice. The order in which canvassing teams visited the EAs allocated to them was also randomly determined.

Teams used identifiable information (GPS coordinates and names of household heads) provided by the management to locate treatment households within the slum. Figure 1 shows a sample tracking sheet, with respondent identifiers left intentionally blank. Once they located a target household, canvassers were instructed to cover the relevant script for each treatment group and to collect basic information about visits (Section 3 provides additional details). During training, canvassers were also encouraged to disseminate information to as many household members and neighbors as possible, since treatment was defined at the EA level.

3 Data

3.1 Baseline and Randomization Data

The baseline data for this experiment came from two sources. First, we conducted a census of Kibera as part of prior work in 2012 (Marx et al., 2019). The census covered more than 30,000 households over two rounds of visits in the slum and was conducted using a short survey module. Second, also for the purpose of this earlier work, we were granted access to de-identified micro data for the entire slum from the 2009 national census, as well as EA maps of the area. This allowed us to aggregate the 2009 census data to the EA level and to locate households visited in our 2012 census in specific EAs. Figure 2 shows a map of the Kibera area with the EAs demarcated in yellow. Since the EAs were randomly allocated to the canvassing teams, we can test randomization balance by looking at correlations between staff characteristics and EA characteristics. We rely on our 2012 census, the 2009 national census, and related geospatial data for these balance checks (presented in Section 5).

3.2 Performance and Effort Data

Canvassing outcomes. Each canvassing team was provided with tracking sheets pre-filled with information on households from the 2012 census data. These tracking sheets included the names of target household heads, along with their GPS coordinates and village names. The canvassers were required to fill out information about each visit, in particular information on whether the household was found or not, the time when each visit began and ended, and other visit details.[8]

In what follows, we use data from the tracking sheets to create several measures of performance. First, we use a dummy variable for whether the canvassing visit was completed. This means the household was found and agreed to the canvassing visit (only 0.01% of the households successfully located by the staff refused to be canvassed). A second outcome is the dura-

[8]Canvassers were also required to fill out the name of the person in the household spoken to and their relationship to the household head; whether that person was already registered and whether they had a national ID card (needed to register); whether anyone else was present during the canvassing visit and if so, who they were.

tion of the visit. Based on the extensive training given to the canvassers on what was expected from them during a canvassing visit, we expected a typical successful visit to last between five and ten minutes. In the data collected from the tracking sheets, successful visits (i.e conditional on the household being found) lasted for 4.8 minutes on average. The median duration for a successful visit was 5 minutes and the 95th and 99th percentiles were 9 minutes and 15 minutes, respectively. Most visits therefore fell within the range we expected from the training.

In our empirical analysis, we use an unconditional measure of time spent with each household, ranging from zero minutes (for unsuccessful visits) to an hour.[9] We interpret visit duration as a positive measure of performance, as longer canvassing visits reflect greater engagement of both respondent households and team members themselves. This seems reasonable given the range of durations we see in the data. The main challenge canvassers face in an urban environment such as Kibera is retaining the attention of respondents (who can become impatient as the conversation unfolds) so that every aspect of the voter registration process is covered. In addition, the canvassers were trained to collect data on whether additional members of the household or other people people were at the canvassing visit. In our data, the duration of a visit correlates positively with whether any additional person attended the visit and with the number of additional people attending the visit. Visit duration also correlates positively with several self-reported measures of voter registration: namely whether the head of household reported to be registered, whether the entire family was registered, and whether registration occurred in Kibera (we provide more details on these variables below).

Organization of work. Based on the tracking sheets, we also observe whether teams conducted several visits simultaneously. This effectively indicates that team members split to canvass households individually. Two visits are coded as simultaneous if they were conducted by the same team, on the same day, and completed within one minute of each other. We aggregate this measure by team and by day to compute the fraction of visits conducted individually in each team. This gives us a proxy for how different canvassing pairs organized their work on a daily basis. This variable could also be capturing whether teams gathered households to conduct one visit with multiple households at the same time. This is consistent with our interpretation that the variable captures a more efficient organization of work within the team.

Time use. Finally, we use the data from the tracking sheets to create measures of effort. We compute daily measures of how much time the team spent in the field, and how much time was spent canvassing households. The time spent in the field was calculated as the difference between the start time of the first canvassing visit and the end time of the last canvassing visit

[9] We treat as enumeration errors visits recorded to have lasted more than 60 minutes. These visits represent approximately 0.05% of the raw tracking data.

for each team and each day,[10] thus variation in this measure is entirely driven by the times at which teams decided to start and to stop working. The time spent canvassing is the sum of the durations of all canvassing visits completed on a given day by each team. These measures of time use provide a reasonable proxy for effort and performance. Indeed, Table 1B shows that the time spent in the field and the time spent canvassing are strongly correlated with measures of voter registration (described below).

Although measures of performance are computed from the tracking sheets, we are not concerned about strategic misreporting by the canvassers driven by the demographic composition of teams. Indeed, canvassers had little incentive to engage in strategic misreporting in the first place. The organization's incentive structure was not tied to the number of successful canvassing visits or the duration of visits: staff members were paid fixed daily wages regardless of the number of households successfully canvassed, and the staff was not given specific targets in terms of visit completion rates. Furthermore, falsifying data was cause for immediate dismissal. We cannot entirely exclude that career concerns led some staff to alter their numbers, but these simple metrics did not carry any weight in subsequent promotion decisions (which were instead driven by staff behavior as observed by supervisors), and the staff had no reason to expect they would.

Summary statistics. Table 1A shows summary statistics from the performance data and the staff composition, and Table 1B reports correlations between team performance indicators and self-reported measures of voter registration. 23% of the canvassing pairs (7 teams out of 30) were ethnically homogeneous on the horizontal dimension. The ethnicity of the manager matches that of one of the team members in 23% of cases, and there is no instance where the ethnicity of the manager matches that of the two team members. Each team was allocated 1,045 visits on average, and managed to complete 426. On a daily basis, teams conducted an average of 33 visits and spent approximately two and a half hours canvassing households—the rest of the time was spent walking between dwellings and searching for target households. Overall, 41% of all scheduled visits were successfully completed, in line with comparable experiments in the literature (see, for example, Pons and Liegey (2018)). Table 1B shows that on average, one additional successful visit led to 0.33 more registrations, 0.28 more registrations of the entire family, and 0.26 more registrations in Kibera. An additional hour spent in the field led to 1.7 more registrations, and an additional hour spent canvassing led to 3.2 more registrations.

3.3 Registration Data

Once the canvassing experiment was completed, we conducted a short phone call experiment in the final two days before the registration deadline. 8,100 households were randomly selected

[10] We do not have a measure of the time spent searching for the first household every day.

from an eligible pool to receive a phone call reminder about registration. These households represent (a random) half of households with a valid phone number in our 2012 census of Kibera. In the census, about 51% of the household heads reported a phone number. Unlike the canvassing experiment, which was randomized at the EA level, this randomization was conducted at the household level. The phone calls were made by the same team of canvassers. Households reached as part of this exercise were asked whether they had already registered for the upcoming election. 85% of respondents with a valid ID reported they had already registered at the time of the call. 62% reported they registered within Kibera, while 73% said the entire household was registered. We use these self-reported variables to measure final registration outcomes.

For the original canvassing experiment, we were also able to access the official voter register for the two constituencies that cover the Kibera area. Unfortunately, since we were not granted access to individual phone numbers or ID numbers, and since many names do not uniquely identify individuals in our data, we are not able to confidently match participants in our experiment to the database of registered voters. We therefore rely on self-reported data collected from the phone calls to measure treatment effects on registration.

3.4 Staff Survey Data

We conducted a survey of the canvassing staff 18 months after the completion of the experiment in order to clarify the mechanisms behind our main experimental results. The survey questionnaire was conducted individually and collected data on social interactions between teammates (e.g., did teammates interact socially after work or after the experiment), working methods (e.g., what language was primarily used within the team), as well as self-reported assessments of performance and hours worked.

4 Empirical Framework

4.1 Effects of Ethnic Diversity on Performance

Horizontal diversity. We now describe the empirical specifications we use to study the impact of team composition on team performance. Our first specification is the following regression:

$$y_{ijt} = \alpha + \beta Horizontal_{jt} + \delta_{jt}^{jun} + \delta_{jt}^{sen} + \Omega X_{ijt} + \epsilon_{ijt} \tag{1}$$

where y_{ijt} is a measure of the outcome of a canvassing visit for household i visited by team j in week t; $Horizontal_{jt}$ is a measure of horizontal ethnic homogeneity, i.e. a dummy equal to one if the two canvassers in team j belong to the same ethnic group; δ_{jt}^{jun} and δ_{jt}^{sen} are ethnicity dummies for junior canvassers and senior canvassers, respectively (out of 30 teams, 15 teams

have at least one senior canvasser); and X_{ijt} are a set of controls and fixed effects. We report estimates from equation (1) for two measures of team performance: canvassing success rates (visit completion) and canvassing times (visit duration).

In this specification, visit outcomes are indexed by time since each household was to be visited twice, so the data is at the household-week level in our preferred specifications. Throughout, we present intent-to-treat estimates, i.e. we use the value of the match variable $Horizontal_{jt}$ from the initial random allocation of teams. Controls include a dummy variable for being allocated to the IEBC treatment, the random order in which household i was visited by team j in week t, as well as strata fixed effects used in the randomization for the canvassing experiment (described in section 2.3).[11] We show robustness to removing controls and strata fixed effects in the Appendix. Standard errors are clustered by team, which is the level at which diversity is observed. Since there are only 30 teams, we also report the p-values from (i) a wild bootstrap clustering with 1,000 replications at the team level and (ii) a randomization inference procedure with 1,000 replications implemented following Young (2018).[12]

Vertical diversity. We then augment equation (1) by including additional dimensions of team diversity. First, we estimate the effect of vertical diversity in the following equation:

$$y_{ijt} = \alpha + \beta_1 Horizontal_{jt} + \beta_2 Vertical_{jt} + \delta_{jt}^{jun} + \delta_{jt}^{sen} + \delta_{jt}^{sup} + \Omega X_{ijt} + \epsilon_{ijt} \tag{2}$$

where $Vertical_{jt}$ is a dummy variable equal to one if the ethnicity of the manager matches the ethnicity of either team member, and δ_{jt}^{sup} are ethnicity dummies for the team supervisor. Since there is no instance where the ethnicity of the manager matches that of the two team members, we do not interact the effects of horizontal and vertical homogeneity. We discuss how this affects our interpretation of mechanisms in Section 5. Note that the correlation between $Horizontal_{jt}$ and $Vertical_{jt}$ is -0.32 in our data. We also report the wild-bootstrap p-values and randomization inference p-values for β_1 and β_2 in equation (2).

External diversity. Finally, we estimate the effect of external homogeneity on performance, i.e. we ask whether the ethnicity of households randomly assigned to each team affects the performance of canvassers. For this, we use the following specification:

$$y_{ijt} = \alpha + \beta_1 Horizontal_{jt} + \beta_2 Vertical_{jt} + \beta_3 External_{ijt} + \delta_{jt}^{jun} + \delta_{jt}^{sen} + \delta_{jt}^{sup} + \delta_{ijt}^{h} + \Omega X_{ijt} + \epsilon_{ijt} \tag{3}$$

where $External_{ijt}$ is a dummy equal to 1 if the household has the same ethnicity as either can-

[11] Since canvassing was only conducted in treatment EAs (and not in control EAs), outcome data is only available for households belonging to treatment EAs. Hence we only use households allocated to either one of the two canvassing treatment groups to analyze the performance impacts of ethnic homogeneity.

[12] We report the p-value from the "randomization-t" procedure in Young (2018).

vasser, and δ_i^h are ethnicity dummies for household i.[13]

4.2 Effects of Gender Diversity on Performance

In order to compare the effects of ethnic diversity with those of gender diversity we estimate the following specification:

$$y_{ijt} = \alpha + \beta_1 Horizontal_{jt}^G + \beta_2 Vertical_{jt}^G + \lambda_{jt}^{jun} + \lambda_{jt}^{sen} + \lambda_{jt}^{sup} + \Omega X_{ijt} + \epsilon_{ijt} \quad (4)$$

where $Horizontal_{jt}^G$ and $Vertical_{jt}^G$ are measures of horizontal and vertical gender homogeneity, respectively (defined as before); and λ_{jt}^{jun}, λ_{jt}^{sen}, and λ_{jt}^{sup} are gender dummies for the junior canvasser, senior canvasser, and manager. We also estimate a version of equation (4) including the effects of horizontal and vertical ethnic homogeneity—defined as in equation (2)—as well as the relevant main effects of ethnicity.

Beyond gender, we note that equations (1), (2), and (3) cannot distinguish the effect of ethnic homogeneity itself from that of other characteristics correlated with ethnic matches. This is analogous to other papers in the literature on dyadic matches, with the partial exception of Goldberg-paradigm experiments and correspondence studies such as Bertrand and Mullainathan (2004)—Bertrand and Duflo (2016) provide a relevant discussion. To address this issue, in Appendix Table 3 we further show that our estimates of the effects of ethnic homogeneity and gender homogeneity are not driven by homogeneity along two other sociodemographic characteristics, namely age and education.

4.3 Learning Over Time

The canvassing experiment took place for approximately three weeks between December 1 and December 24, 2012, with 99% of visits taking place between December 3 and December 15. In this period of time, ethnic diversity may have affected the teams' ability to learn about their assigned tasks and improve their performance. To test for this, we interact the main effects of ethnic diversity included in equations (1), (2) and (3) with the day (entered linearly) in which a particular canvassing visit was conducted, in the following specification:

$$\begin{aligned} y_{ijt} = \alpha &+ \beta_1 Horizontal_{jt} * Day_{ijt} + \beta_2 Vertical_{jt} * Day_{ijt} + \beta_3 External_{jt} * Day_{ijt} + \gamma_1 Day_{ijt} \\ &+ \gamma_2 Horizontal_{jt} + \gamma_3 Vertical_{jt} + \gamma_4 External_{ijt} + \delta_{jt}^{jun} + \delta_{jt}^{sen} + \delta_{jt}^{sup} + \Omega X_{ijt} + \epsilon_{ijt} \quad (5) \end{aligned}$$

[13] We only consider external matches for households belonging to one of the seven ethnic groups for which we collected ethnicity data in the 2012 census (Kalenjins, Kambas, Kikuyus, Kisiis, Luhyas, Luos, and Nubis). The remaining ethnicities were lumped together (only 2% of households in the 2012 census data do not belong to any of these ethnic groups). For these households, the external match dummy is unobserved (coded as missing).

where β_1, β_2, and β_3 are the main coefficients of interest. From a theoretical standpoint, the sign of these coefficients is ex ante ambiguous. On the one hand, the effects of horizontal ethnic homogeneity on performance may dissipate over time if diversity creates frictions (for example in terms of communication) that can be overcome with time. Learning could also be facilitated by diversity if different groups bring diverse sets of skills that directly contribute to the completion of tasks assigned to the team. On the other hand, if group-specific skills are irrelevant or if socialization between similar teammates decisively affects learning, then ethnic homogeneity would instead improve the teams' ability to learn how to complete their tasks more efficiently. Equation (5) allows us to adjudicate between these channels empirically.

4.4 Sample Size

Before presenting our experimental results, here we discuss sample size and inference issues. While we have variation in outcomes for the 15,676 thouseholds treated in the canvassing experiment, the random variation we use to estimate the effects of horizontal and vertical ethnic homogeneity is at the team level. Since this randomization is clustered at the team level with 30 teams, following standard practice for clustered randomized designs we report standard errors clustered at the team level. In addition, we worry about a small number of clusters biasing standard errors down: to quote from Cameron et al. (2008), "Standard asymptotic tests can over-reject, however, with few (five to thirty) clusters." We therefore use the tools proposed by Cameron et al. (2008) to deal with a small number of clusters, the wild bootstrap, which we also report in all tables. In addition, we systematically report the results from a randomized inference procedure following Young (2018).

A small number of clusters in an RCT is not uncommon in the development and labor economics literatures. For example, Bloom et al. (2013) report on an experiment involving 17 firms; Cohen and Dupas (2010) report on an experiment with 20 clinics; Fryer (2014) reports on an experiment in 16 schools; Wantchekon et al. (2015) analyze a policy across 8 regions; and Fujiwara (2015) analyzes a policy across Brazil's 27 states. As we do in this paper, these studies also use the wild bootstrap to account for the small number of clusters in their design.

Finally, note that this discussion is only relevant for testing the effects of horizontal and vertical homogeneity. The effective sample size we have to estimate the effects of external homogeneity is very large since the variation we use comes from the random matching of teams with households spread across 300 enumeration areas. Nevertheless, the standard errors we report for this test are also conservatively clustered at the team level.

5 Results

5.1 Randomization Balance Checks

We first present a set of checks to verify that the randomization produced a balanced sample. In particular, we check whether the measures of horizontal, vertical, and external homogeneity correlate with any observable characteristics of staff members, target areas, and target households. Tables 2A and 2B show these results. Column (1) in this table reports the sample mean of the relevant dependent variable. In column (2), we show the coefficient obtained from regressing each relevant staff, household, and EA level characteristic on the horizontal homogeneity team dummy (equal to one for teams with co-ethnic canvassers) and ethnicity dummies for each junior and senior team member. In column (3), we report the coefficient from a similar specification using the measure of vertical homogeneity (coded as one if the manager belongs to the same ethnicity as any one of the two canvassers) as the main regressor. In column (4) of Table 2B, we show results for the external match with a team member.[14] All regressions include the relevant main effects, and standard errors are clustered by team.

The outcomes of these regressions are pre-determined staff characteristics, household characteristics from the 2012 census, and EA level aggregates of the 2009 census micro data. Staff characteristics include the gender, age, years of college education, and seniority (previous experience working in Kibera) of each staff member. For household characteristics, the corresponding regressions are run at the team-day level. Household characteristics include household size, the number of children in the household, whether the household pays rent for their housing, the (log) amount paid in rent, whether the household was ever evicted from their dwelling in the past, the number of years spent in the same dwelling, the number of years spent in Kibera, and a dummy for households with a phone number in our data. The corresponding regressions are at the household level. Finally, EA-level variables include log consumption per capita, the EA-level poverty rate,[15] the number of households in the EA, average age of the household head, years of education of the household head, whether the household head owns a business, a radio, and TV, whether the household head is employed, the fraction of youth unemployed in the EA, the average fraction of household dwellings with a cement floor, a mixed mud-cement wall, piped water, an uncovered pit latrine for sanitation, and electricity, and whether the household cooks with paraffin. These regressions are run at the EA level.

Tables 2A and 2B illustrate that the randomization of the composition of canvassing teams produced a balanced experimental sample. Only two coefficients (out of 27) are statistically significant at conventional levels.

[14] Here note that since the external match is defined at the household level, the only checks we can conduct are those with household characteristics from the 2012 census.

[15] The poverty indicator was computed as part of a poverty mapping exercise. See Marx et al. (2019) for details.

5.2 Ethnic Composition and Team Performance

5.2.1 Horizontal Homogeneity

Table 3 reports estimates from equations (1), (2), and (3) for two outcomes: visit completion and unconditional visit duration. We first estimate the effect of horizontal homogeneity, and then introduce dummies for vertical homogeneity and external homogeneity. As a consequence of the experimental design, these coefficients estimate the causal impact of a team's horizontal, vertical, and external ethnic homogeneity on canvassing outcomes, respectively. In all specifications, we control for the IEBC Treatment, the order in which the team was instructed to visit that household,[16] and strata fixed effects. Standard errors are clustered by team, and we also report the p-values from a wild bootstrap and a randomization inference procedure, both using 1,000 replications.

Columns (1) through (3) of Table 3 show that ethnically homogeneous teams on the horizontal dimension are approximately 8 percentage points more likely to complete a canvassing visit—about a 20% effect size. Such teams also conduct visits that are 0.8 to 1.3 minutes longer on average (columns (4) through (6)). These results are similar across all columns, large in magnitude, and robust to using the wild bootstrap and randomization inference in 4 out of 6 specifications. Overall, a team conducts more and longer visits if it is composed of two co-ethnic canvassers. In Figure 3, we provide further evidence that the fraction of co-ethnic pairs is a monotically increasing function of visit duration across all visits lasting less than 10 minutes.

5.2.2 Vertical Homogeneity

Looking at both canvassing outcomes, we find some evidence that the effects of vertical homogeneity are opposite in sign to those of horizontal homogeneity. If a manager and any one team member belong to the same ethnicity, the probability that the canvassing visit is completed decreases by about 3 percentage points, but this estimate is not statistically significant (columns (1)-(3)). The duration of visits decreases by 1.4 minutes (columns (4)-(6)). The effects on duration are sizeable and statistically significant, even when we rely on wild bootstrap or randomization inference p-values for inference. Recall that since there are no instances of a "triple match" between both canvassers and the manager, these effects are identified from canvassing pairs that are not ethnically homogeneous themselves.[17] In Figure 4, we show that vertical homogeneity

[16] Across specifications, a position at the bottom of the canvassing order reduces the chance of a successful visit, and reduces visit duration.

[17] Consequently, an equivalent interpretation is that teams where the supervisor and both canvassers belong to 3 different ethnic groups are on average more productive than teams where the supervisor and one canvasser belong to the same ethnic group. Since vertically homogeneous teams are never horizontally homogeneous while vertically diverse teams are sometimes horizontally homogeneous, the comparison between both types of teams captures variation in vertical homogeneity as well as some variation in horizontal homogeneity. However, the regressions that control for both vertical and horizontal homogeneity allow us to correctly estimate the effect of vertical homogeneity.

is a monotically decreasing function of visit duration for all visits under 10 minutes.

Mechanisms. Since no team is ethnically homogeneous along both the horizontal and the vertical dimension, our estimate of β_2 in equations (2) and (3) combines two distinct effects. First, this estimate captures differences in the *degree* of vertical diversity, i.e. whether teams where the supervisor is non-co-ethnic with both employees outperform teams with one vertical connection. Second, this estimate captures differences in the *nature* of the vertical relationship between supervisors and a team with at least one co-ethnic employee compared to the vertical relationship between supervisors and a team with no co-ethnic employee. These two effects can equivalently be understood as the intensive and the extensive margin effects of vertical diversity—equations (2) and (3) estimate the total effect of vertical diversity across these two margins.

Furthermore, two potential mechanisms could be driving the performance differential between teams with and without a vertical ethnic match. First, vertical homogeneity could drive a further wedge between already struggling horizontally diverse pairs. Alternatively, managers could manage pairs with at least one co-ethnic team member less intensively because they are, in essence, cutting their co-ethnic employee more slack. While our empirical setup does not allow us to disentangle between these two mechanisms, the results from our survey of the field staff (discussed in Table 8 and section 5.7) suggest that the latter mechanism is most likely driving the vertical diversity result. Columns (2) and (4) of Table 8 show that vertical homogeneity undermines the (self-assessed) performance of the co-ethnic employee, without affecting the performance of the non-co-ethnic employee. If the negative impact of vertical diversity came from an increase in the wedge between already struggling horizontally diverse pairs, then the performance of both team members should have been equally affected. Furthermore, columns (5)-(6) and (9)-(10) of Table 8 show that vertical diversity does not affect language use or the likelihood that team members stay in contact after the experiment. Again, one would have expected an effect on these variables if the former mechanism were at play. Overall, these estimates provide suggestive evidence that vertical diversity undermines team performance through a loosening of discipline and effort in those teams—a channel supported by the evidence in column 1 of Table 7, where vertical homogeneity reduces the time spent by teams in the field. We elaborate on this explanation in sections 5.6 and 5.7.

5.2.3 External Homogeneity

We also estimate whether households' ethnicities affect the success and duration of canvassing visits. There are at least two distinct reasons why such "external" homogeneity could affect our measures of performance. Canvassers could target their efforts towards finding households of the same ethnicity if they feel responsible for mobilizing co-ethnic voters in the upcoming election. Second, conditional on finding households, external homogeneity may also affect the

time canvassers spend with respondents.

However, the effect of an external ethnic match with any one canvasser is a precisely estimated zero—we are able to rule out very small effects. We do not report effects of a "triple match" between the household's ethnicity and both canvassers on the team, since this configuration occurs in only 3% of all cases. This effect is negative, small in magnitude, and not significantly different from zero (results available upon request).

5.2.4 Robustness

In Appendix Tables 1 and 2 we check the robustness of these results to alternative specifications. For comparison we report the main specification (equation 3) in column (1) of these tables, using visit completion and visit duration as the dependent variable in Appendix Tables 1 and 2, respectively. In column (2) we remove controls and strata fixed effects, while in column (3) we report estimates obtained with controls, without strata fixed effects. In column (4) we include EA fixed effects instead of strata fixed effects, and in column (5) we include supervisor fixed effects. Finally, in column (6) we use a different parametrization of the ethnicity main effects where instead of dummy variables for the ethnicity of the junior and the senior canvasser we use dummy variables equal to 1 if the relevant ethnicity is represented on the team (for instance, the *Kikuyu* dummy is equal to 1 if either of the two canvassers is a Kikuyu).

Across all these specifications, horizontal homogeneity has a positive, significant effect on both visit completion (Appendix Table 1) and visit duration (Appendix Table 2). Vertical homogeneity has a negative, significant effect on visit duration. Results for the effect of vertical homogeneity on visit completion are more mixed: the coefficient on this variable is positive, non-significant in the specification without controls (column (2)) and negative, generally non-significant in other specifications.

In Appendix Table 6 we also ensure that our results on visit duration are not driven by outliers in this variable. We report the main effects of horizontal and vertical homogeneity in columns (1) and (3) in this table, where duration trimmed at the 99th percentile (10 minutes) is the dependent variable in column (1) and log of (1+duration) is the dependent variable in column (3). Our estimates are robust throughout.

Finally, in Appendix Table 7 we compare our baseline estimates of the effects of horizontal and vertical diversity on visit duration with estimates obtained after making various assumptions on the nature of selection into visit completion. Column (1) of this table reproduces our baseline estimates, where the dependent variable is unconditional visit duration. In columns (2) through (4) of Appendix Table 7, we then show that the positive (negative) effect of horizontal (vertical) homogeneity is robust to alternative assumptions about the potential duration of incomplete visits. In column (2), we impute visit duration at the mean of conditional visit duration for all incomplete visits. In column (3), we impute visit duration at the 10th percentile

of conditional visit duration (2 minutes), and in column (4) we impute visit duration at the 90th percentile of conditional of visit duration (8 minutes). These estimates show that the effects of horizontal and vertical homogeneity hold independently of the effect on visit completion.

5.3 Ethnic Composition and Learning

In Table 4, to test for the presence of learning effects, we also look at whether the effects of horizontal and vertical homogeneity differ across the duration of the canvassing exercice. We interact horizontal, vertical, and external homogeneity with the day in which a particular canvassing visit was completed, as described in equation (5).[18] We report additional robustness checks on these estimates in Appendix Tables 4 and 5, using the same alternative specifications as in Appendix Tables 1 and 2. In Appendix Table 6, we show estimates obtained using trimmed duration and log of (1+duration) as the dependent variables in equation (5).

Across all specifications, we find robust evidence that the effect of horizontal ethnic homogeneity on performance increases over the duration of the exercise. Horizontal homogeneity increases completion rates by approximately 1 percentage point for every additional day of canvassing, suggesting that co-ethnic pairs become more efficient at finding target households as they spend more time together in the field. Visit duration also increases by 0.08 minutes for every additional day of canvassing for these teams.[19] This suggests that learning occurs primarily within homogeneous pairs: learning is not improved by the blending of group-specific skills, but it is facilitated by communication or increased socialization between co-ethnic teammates.

In the second row of Table 4, we find no similar evidence that the effects of vertical ethnic homogeneity accentuate over time. If the negative effect of vertical homogeneity on performance is caused by preferential treatment or distortions of managerial effort, then these distortions are not a function of time spent together as a team.

5.4 Gender Composition and Team Performance

We estimate the effects of gender composition and compare them with the effects of ethnicity in Table 5. As in Table 3, we first look at the effects of (gender) horizontal homogeneity separately, and then introduce a dummy for vertical homogeneity. Columns (2) and (6) report the coefficients of interest from equation (4). The specifications in columns (3)-(4) and (7)-(8) include measures of both gender and ethnic homogeneity on the right-hand side, in order to compare the effects of both dimensions.

The results we find are much more nuanced than those we find for horizontal and vertical ethnic homogeneity. The effect of gender horizontal homogeneity is positive in all specifica-

[18] The sample size is smaller in Table 4 because the visit date was not recorded in the canvassing data for 21% of attempted visits.

[19] This implies that homogeneous teams don't simply rush to complete their tasks as the exercise evolves.

tions, but is relatively small in magnitude and significant in only 3 out of 8 specifications. We also show that gender vertical homogeneity does not significantly affect either measure of performance. On the other hand, the effects we estimate for horizontal and vertical ethnic homogeneity are almost identical to those reported in Table 3.

In Appendix Table 3, we further show that our results on horizontal and vertical ethnic and gender homogeneity are not driven by homogeneity along two other dimensions, namely age and education. We separately control for (i) a dummy for both canvassers being within 2 years in age of each other, and (ii) a dummy for both canvassers having the same education level (undergraduate or graduate; as all staff members have completed some university education). To preserve our statistical power to measure these effects, and because we do not observe the age and education levels of supervisors, here we do not control for the main effects of age and education and we do not study vertical homogeneity along these dimensions. Overall, our main results are robust to including these controls.

5.5 Registration Outcomes

In Table 6, we look at the impacts of the three dimensions of ethnic homogeneity (horizontal, vertical, external) on measures of voter registration collected with a restricted subset of voters in our sample. As described in Section 3, this measure is only available for households we could reach by phone in the last days before the registration deadline. The sample used to compute the estimates presented in Table 6 is restricted to 3,151 households out of 4,019 who were randomly selected for both the canvassing intervention and the phone survey.[20] Here we regress household-level registration outcomes on a specification analogous to equation (3) which includes 3 dummy variables indicating whether at least one of the two teams instructed to canvass the household was homogeneous on the horizontal, the vertical, or the external dimension, as well as controls and ethnicity main effects.

For comparison with our earlier results, in columns (1) and (2) we reproduce our estimates for visit completion and duration in this limited sample. As in Tables 3 through 5, these regressions are conducted on a dataset at the household-week level; hence there are approximately twice more observations in columns (1)-(2) than in the other columns. The effects of horizontal and vertical homogeneity in this sample remain positive and negative, respectively, but the effects on visit completion are no longer significant.

The effects we find on registration outcomes are broadly consistent with these patterns, with the caveat that we are underpowered to detect such effects. In columns (3) through (5), we look at three alternative measures of registration: whether the respondent to the phone call reported being registered (column (3)), whether all household members were reported to be registered

[20]The response rate in our phone survey was 84%. Attrition was uncorrelated with the canvassing treatments.

(column (4)), and whether the respondent registered in a polling station in Kibera (column (5)). Registration within Kibera is a relevant outcome of our canvassing intervention since registration outside of Kibera required much greater effort and preparation and was therefore likely to have occurred without any canvassing (voters registered outside Kibera were likely inframarginal in terms of the canvassing campaign).[21]

External ethnic homogeneity has no significant impact on these outcomes, which is perhaps the most surprising result. The horizontal ethnic match dummy is positive and significant in column (4) (the entire family registered), but not in columns (3) and (5). Vertical ethnic homogeneity has a negative effect across the board, significant in columns (4) and (5). The signs of these coefficients are broadly consistent with our earlier results: in particular, vertical homogeneity negatively affected the performance of teams.

5.6 Time Use and Team Organization

Next, we show evidence on the organization of work within teams in Table 7. We look at several outcomes in this table. First, we look at the total daily amount of time spent in the field. We interpret this variable as a measure of effort, since it is entirely driven by the times at which teams started and stopped working each day in Kibera. We then look at the daily amount of time spent canvassing and the total number of visits conducted each day. We interpret these variables as measures of productivity, since the time spent canvassing and the number of successful visits are determined by the speed at which teams could successfully locate target households. This required quickly acquiring knowledge on how to move within the Kibera slum and how to gather the relevant information from local elders and neighbors. Furthermore, Table 1B shows that these measures of time use are all strongly correlated with registration outcomes. Finally, we look at a measure of team organization: whether teams split to conduct canvassing visits individually. Visits conducted individually provide a measure of team organization and decision-making, since no clear instructions were given as to whether visits should always be conducted by the pair, or whether team members were advised to split.[22]

In columns (1) and (2), we show that ethnic homogeneity affects both the amount of time spent in the field and the time spent canvassing. Horizontal homogeneity increases team effort and productivity, while vertical homogeneity has the opposite effect. In column (3), we look at the effects of horizontal and vertical homogeneity on the total number of visits completed each day. Horizontal homogeneity has a positive, significant impact on this variable, while vertical homogeneity has a positive, non-significant effect. Finally, in column (4), we find that

[21] In columns (3) to (5), standard errors are two-way clustered by team 1 and team 2 (the two teams by which they were visited).

[22] The effect of splitting the team on performance is ex ante ambiguous. On the one hand, keeping the team together at all times might have promoted synergies and learning. On the other hand, splitting visits obviously allowed the team to conduct a larger number of visits.

horizontally homogeneous team complete more visits alone (we do not report results using the fraction of visits conducted alone, which are similar).

In addition to these regression estimates, we also report in Figure 5 the distributions of these measures of time use, separately for teams that are ethnically homogeneous (along the horizontal dimension) and those that are not. This figure provides additional support to the results we report in Table 7.

5.7 Staff Survey Results

In Table 8, we present results from the survey conducted with the canvassing staff. This survey covered a number of aspects of the team exercise, including self-reported measures of effort and performance, team communication, and whether canvassers socialized with their teammates outside or after the study. In odd-numbered columns of Table 8, we regress these outcomes on the horizontal and the vertical ethnic match dummies, as well as ethnicity main effects and basic demographic controls (age, gender, and years of education). In even-numbered columns, we look separately at whether the canvasser was herself a co-ethnic of the manager, as opposed to her coworker being a co-ethnic of the manager.

In columns (1) and (2), we look at the canvasser's evaluation of their own performance, measured on a 1-5 scale (from very bad to very good). In column (1), homogeneous teams on both the horizontal and the vertical dimension report lower own performance, but these effects are not significant. In column (2), canvassers who were themselves co-ethnics with the supervisor report significantly lower performance. In columns (3) and (4) we look at a self-reported measure of effort – daily hours worked. Hours worked are significantly lower for vertically homogeneous teams, and this effect is driven by canvassers with a co-ethnic manager.

In columns (5) and (6), we look at whether homogeneity affected the choice of language spoken within the team. Only one team out of thirty spoke their native tongue, while all others communicated in either English or Swahili. We therefore test whether homogeneity influenced whether a canvasser spoke her most preferred language (the language she reports speaking most often). We do not find significant results on this variable, suggesting that communication (language) frictions were not a critical factor for our results.

Finally, in columns (7) through (10) we look at two measures of socialization between teammates: whether the canvassers socialized outside fieldwork (columns (7)-(8)) and whether they stayed in contact after the experiment (columns (9)-(10)). Teammates in horizontally homogeneous teams are 23-25 percentage points more likely to socialize outside of work (significant in column (8)) and 32-34 percentage points more likely to be in contact 18 months after the experiment, significant at 1%. The evidence on vertical homogeneity is more mixed. In column (8), canvassers with a co-ethnic manager are significantly less likely to socialize with their teammate outside of work.

Overall, we interpret these results as evidence that the horizontal homogeneity acted as a monitoring and disciplining device and improved work organization within the team, which facilitated learning and possibly led to increased socialization after the experiment. Meanwhile, vertical homogeneity led to less stringent norms and discipline and hence lower effort and performance. This finding is consistent with Bandiera et al. (2009), who show that an increase in the average level of social connections between managers and workers can reduce average firm productivity by inducing distortions in managerial effort (under fixed wages).[23] These results highlight the importance of looking at different dimensions of diversity when analyzing the effect of diversity on performance. While there may be costs to diversity in terms of the horizontal structure of the firm, these costs must be balanced against the gains from diversity we find in the vertical dimension.

6 Conclusion

In this paper, we study the effects of team diversity on performance. We use a field experiment implemented within a non-profit research organization based in Kenya as part of a door-to-door canvassing exercise where pairs of canvassers were assigned a list of households to visit. We use the data from this canvassing exercise, including data on time use in the field and data from a survey of the staff, to compute measures of effort and performance. Our most innovative contribution is that we are able to study the causal effects of diversity along various dimensions—within teams, along the hierarchy of the organization, and between workers and clients—within a single organization.

We find that ethnic homogeneity between teammates (the horizontal dimension) improves team performance, measured in terms of the completion and duration of each canvassing visit assigned to the team, as well as some measures of time use and household-level voter registration. This effect intensifies over the duration of the canvassing exercise, suggesting that horizontal homogeneity (instead of diversity) promotes learning in the context of our experiment. The magnitude of the effect of ethnic homogeneity on performance is economically sizeable. For example, ethnically homogeneous teams on the horizontal dimension complete about a third more visits every day. However, vertical ethnic homogeneity between team members and the manager often has the opposite effect: teams with this configuration report poorer performance on average. Finally, we find no effect of an ethnic match between households and canvassing teams or managers. This null result suggests that teams did not strategically direct their efforts towards co-ethnic households.

Our analysis suggests that while horizontally homogeneous teams do not spend more time in the field, they organize their time more efficiently and spend more time actually canvassing

[23] The introduction of managerial incentives can undo this effect, but such incentives were absent in our context.

households. Conversely, vertical homogeneity reduces both the time spent in the field and the time spent canvassing, suggesting a lower level of effort in these teams. We interpret these results in light of evidence from a survey of the canvassing staff. First, the survey helps us rule out simple language explanations for the different performances of teams—homogeneous teams are no more likely to use their tribal language, or their preferred language as their primary working language. However, there is evidence that horizontally homogeneous teams are more cohesive, interact more outside of fieldwork, and are more likely to divide up tasks.

Our findings suggest much of the trade-off between diversity and homogeneity in organizations may come from the different effects diversity has along different dimensions of organizational structure. On the one hand, diversity may reduce efficiency within teams of workers, by creating communication costs or other frictions, leading to a worse division of tasks and lower performance. On the other hand, diversity along the organization's hierarchy has the opposite effect in our context, since it improves both effort and performance.

References

Aker, Jenny C., Paul Collier, and Pedro C. Vicente, "Is Information Power? Using Mobile Phones and Free Newspapers During an Election in Mozambique," *Review of Economics and Statistics*, 2017, *99* (2), 185–200.

Alesina, Alberto, Reza Baqir, and William Easterly, "Public Goods and Ethnic Divisions," *Quarterly Journal of Economics*, 1999, *114*, 1243–1284.

Alesina, Roberto and Eliana La Ferrara, "Ethnic Diversity and Economic Performance," *Journal of Economic Literature*, 2005, *43* (3), 762–800.

Algan, Yann, Camille Hémet, and David Laitin, "The Social Effects of Ethnic Diversity: a Natural Experiment with Exogenous Residential Allocation," *Journal of Political Economy*, 2016, *124* (3), 696–733.

Anwar, Shamena, Patrick Bayer, and Randi Hjalmarsson, "The Impact of Jury Race in Criminal Trials," *Quarterly Journal of Economics*, 2012, *127* (2), 1017–1055.

Apfelbaum, Evan P., Katherine W. Phillips, and Jennifer A. Richeson, "Rethinking the Baseline in Diversity Research: Should We Be Explaining the Effects of Homogeneity," *Perspectives on Psychological Science*, 2014, *9* (3), 235–244.

Arrow, Kenneth, "The Theory of Discrimination," in O. Ashenfelter and A. Rees, eds., *Discrimination in Labor Markets*, Princeton, NJ: Princeton University Press, 1973.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul, "Social Connections and Incentives in the Workplace: Evidence From Personnel Data," *Econometrica*, 2009, *77* (4), 1047–1094.

Banerjee, Abhjit V., Selvan Kumar, Rohini Pande, and Felix Su, "Do Informed Voters Make Better Choices? Experimental Evidence from Urban India," *Working Paper*, 2010.

Barr, Abigail and Abena Oduro, "Ethnic Fractionalization in an African Labour Market," *Journal of Development Economics*, 2002, *68*, 355–379.

Becker, Gary, *The Economics of Discrimination*, University of Chicago Press, 1957.

Berg, Erlend, Maitreesh Ghatak, R Manjula, D Rajasekhar, and Sanchari Roy, "Motivating Knowledge Agents: Can Incentive Pay Overcome Social Distance," *Economic Journal*, 2019, *129* (617), 110–142.

Berge, Lars Ivar Oppedal, Kjetil Bjorvatn, Simon Galle, Edward Miguel, Daniel Posner, Bertil Tungodden, and Kelly Zhang, "Ethnically Biased? Experimental Evidence from Kenya," *Journal of the European Economic Association*, 2020, *18* (1).

Bertrand, Marianne and Esther Duflo, "Field Experiments on Discrimination," NBER Working Papers 22014, National Bureau of Economic Research February 2016.

_ **and Sendhil Mullainathan**, "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *American Economic Review*, September 2004, *94* (4), 991–1013.

Bloom, Nicholas and John Van Reenen, "Measuring and Explaining Management Practices Across Firms and Countries," *Quarterly Journal of Economics*, 2007, *122* (4), 1351–1408.

_ , Aprajit Mahajan, David McKenzie, and John Roberts, "Do Management Interventions Last? Evidence from India," *American Economic Journal: Applied Economics*, April 2020, *12* (2), 198–219.

_ , Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts, "Does Management Matter? Evidence from India," *Quarterly Journal of Economics*, 2013, *128* (1), 1–51.

Braconnier, Céline, Jean-Yves Dormagen, and Vincent Pons, "Voter Registration Costs and Disenfranchisement: Experimental Evidence from France," *American Political Science Review*, 2017, *111* (3), 584–604.

Caliendo, Lorenzo, Giordano Mion, Luca David Opromolla, and Esteban-Rossi-Hansberg, "Productivity and Organization in Portuguese Firms," *NBER Working Paper No. 21811*, 2015.

Cameron, Colin, Jonah Gelbach, and Douglas Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 2008, *90*, 414–427.

Cohen, Jessica and Pascaline Dupas, "Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment," *Quarterly Journal of Economics*, 2010, *125*, 1–45.

Fafchamps, Marcel, "Ethnicity and Credit in African Manufacturing," *Journal of Development Economics*, 2000, *61*, 205–235.

Fairlie, Robert W., Florian Hoffmann, and Philip Oreopoulos, "A Community College Instructor Like Me: Race and Ethnicity Interactions in the Classroom," *American Economic Review*, 2014, *104* (8), 2567–2591.

Fisman, Raymond, Daniel Paravisini, and Vikrant Vig, "Cultural Proximity and Loan Outcomes," *American Economic Review*, 2017, *107* (2), 457–492.

Fryer, Roland Jr, "Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments," *Quarterly Journal of Economics*, 2014, *129*, 1355–1407.

Fujiwara, Thomas, "Voting Technology, Political Responsiveness, and Infant Health: Evidence From Brazil," *Econometrica*, 2015, *83*, 423–464.

Gerber, Alan S. and Donald P. Green, *Get Out the Vote: How to Increase Voter Turnout*, Washington, D.C.: Brookings Institution Press, 2015.

Giuliano, Laura, David I. Levine, and Jonathan Leonard, "Manager Race and the Race of New Hires," *Journal of Labor Economics*, October 2009, *27* (4), 589–631.

Hamilton, Barton H., Jack A. Nickerson, and Hideo Owan, "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation," *Journal of Political Economy*, 2003, *111* (3), 465–497.

Hamilton, Barton H, Jack A Nickerson, and Hideo Owan, "Diversity and Productivity in Production Teams," *Advances in the Economic Analysis of Participatory and Labor-Managed Firms*, 2012, *13*, 99–138.

Hjort, Jonas, "Ethnic Divisions and Production in Firms," *Quarterly Journal of Economics*, 2014, *129* (4).

_ , Changcheng Song, and Christopher Yenkey, "Ethnic Investing and the Value of Firms," *Working Paper*, 2019.

Hoogendoorn, Sander and Mirjam van Praag, "Ethnic Diversity and Team Performance: A Field Experiment," *Tinbergen Institute Discussion Paper 2012-068/3*, 2012.

_ , Simon C. Parker, and Mirjam van Praag, "Ability Dispersion and Team Performance: A Field Experiment," *Tinbergen Institute Discussion Paper 2014-053/VII*, 2014.

Kandel, Eugene and Edward P. Lazear, "Peer Pressure and Partnerships," *Journal of Political Economy*, 1992, *100* (4), 801–817.

Karaca-Mandic, Pinar, Nicole Maestas, and David Powell, "Peer Groups and Employment Outcomes: Evidence Based on Conditional Random Assignment in the U.S. Army," 2013.

Katherine, S.Y. Kim-Jun Phillips and S. Shim, "The Value of Diversity in Prganizations: A Social Psychological Perspective," in R. van Dick David De Cremer and K. Murnighan, eds., *Social Psychology and Organizations*, New York, NY: Routledge, 2011.

Lang, Kevin, "A Language Theory of Discrimination," *The Quarterly Journal of Economics*, 1986, *101* (2), 363–382.

Lazear, Edward P., "Globalisation and the Market for Team-Mates," *The Economic Journal*, 1999, *109*, C15–C40.

_ , Kathryn L. Shaw, and Christopher T. Stanton, "The Value of Bosses," *Journal of Labor Economics*, 2015, *33* (4), 823–861.

Levine, Sheen S., Evan P. Apfelbaum, Mark Bernard, Valerie L. Bartelt, Edward J. Zapac, and David Stark, "Ethnic diversity deflates price bubbles," *PNAS*, 2014, *111* (52), 18524–18529.

Lyons, Elizabeth, "Team Production in International Labor Markets: Experimental Evidence from the Field," *American Economic Journal: Applied Economics*, 2017, *9* (3), 70–104.

Marx, Benjamin, Thomas M. Stoker, and Tavneet Suri, "There Is No Free House: Ethnic Patronage in a Kenyan Slum," *American Economic Journal: Applied Economics*, October 2019, *11* (4), 36–70.

Mas, Alexandre and Enrico Moretti, "Peers at Work," *American Economic Review*, 2009, *99* (1), 112–145.

Miguel, Edward and Mary Kay Gugerty, "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya," *Journal of Public Economics*, 2005, *89*, 2325–2368.

Nickerson, David W., "Quality Is Job One: Professional and Volunteer Voter Mobilization Calls," *American Journal of Political Science*, 2007, *51* (2), 269–282.

Pons, Vincent and Guillaume Liegey, "Increasing the Electoral Participation of Immigrants: Experimental Evidence from France," *forthcoming Economic Journal*, 2018.

Prat, Andrea, "Should a Team be Homogeneous?," *European Economic Review*, 2002, *46*, 1187–1207.

Price, Joseph and Justin Wolfers, "Racial Discrimination among NBA Referees," *The Quarterly Journal of Economics*, 2010, *125* (4), 1859–1887.

Reinhard, Selten and Massimon Warglien, "The Emergence of Simple Languages in an Experimental Coordinate Game," *PNAS*, 2007, *104* (18), 7361–7366.

Shayo, Moses and Asaf Zussman, "Judicial Ingroup Bias in the Shadow of Terrorism," *Quarterly Journal of Economics*, 2011, *126*, 1447–1484.

Wantchekon, Leonard, Marko Klasnja, and Natalija Novta, "Education and Human Capital Externalities: Evidence from Colonial Benin," *Quarterly Journal of Economics*, 2015, *129*, 703–757.

Young, Alwyn, "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results*," *The Quarterly Journal of Economics*, 11 2018, *134* (2), 557–598.

Zenger, Todd R. and Barbara S. Lawrence, "Organizational Demography: The Differential Effects of Age and Tenure Distributions on Technical Communication," *Academy of Management Journal*, 1989, *32* (2), 353–376.

Table 1A: Summary Statistics

	Mean	SD	N
Household-Week Level Data			
Visit Completed	0.41	0.49	30947
Duration of Visit	1.82	3.02	29302
Duration of Visit, Trimmed at 99th pctile	1.67	2.48	29017
Duration, Conditional on Finding Household	4.79	3.13	11135
External Ethnic Match	0.30	0.46	27543
Team-Level Data			
Total Visits Allocated to Team	1045.07	117.72	30
Total Visits Completed by Team	426.00	73.27	30
Horizontal Ethnic Match	0.23	0.43	30
Vertical Ethnic Match	0.23	0.43	30
Total EAs Allocated to Team	20.00	0.37	30
Team-Day Level Data			
# of Visits Conducted by Team-Day	32.85	18.32	279
% of Alone Visits Conducted by Team-Day	0.14	0.09	279
Time Spent Canvassing	155.97	97.88	279
Time Spent in Field	319.41	132.78	274
Staff Level Data			
Gender (Dummy for Male)	0.43	0.50	60
Age	26.73	3.08	59
Years of College Education	3.04	1.24	58
Staff of Kikuyu Tribe	0.37	0.49	60
Staff of Kisii Tribe	0.10	0.30	60
Staff of Luhya Tribe	0.28	0.45	60
Staff of Luo Tribe	0.12	0.32	60
Staff of Other Tribe	0.13	0.34	60
Registration Data (Household Level)			
Respondent Registered (self-report)	0.80	0.40	3391
Entire Family Registered	0.69	0.46	3339
Registered in Kibera	0.58	0.49	3391

Note: Each of the 30 teams was composed of two staff members.
Each team covered 10 EAs each week for a total of two weeks.
Each household was visited twice, each time by a different team.
Duration data is missing for 1,645 completed visits.
The external match is missing when household ethnicity is unobserved.

Table 1B: Correlations between Team Performance and Voter Registration

	Total Registered			Family Registered			Registered in Kibera		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Number of visits	0.334***			0.281***			0.255***		
	(0.022)			(0.018)			(0.019)		
Time spent in field		1.668***			1.371***			1.194***	
		(0.254)			(0.209)			(0.186)	
Time spent canvassing			3.247***			2.713***			2.476***
			(0.250)			(0.200)			(0.207)
R^2	0.497	0.182	0.374	0.488	0.169	0.361	0.486	0.155	0.364
Dep Var Mean	14.81	14.77	14.81	12.52	12.48	12.52	10.89	10.86	10.89
Observations	279	274	279	279	274	279	279	274	279

Note: Correlations are computed from the team-day level data. Robust standard errors in parentheses.

Registration was self-reported and collected via phone with a subset of our sample (see text for details).

'Number of visits' is the number of successful (completed) canvassing visits completed by a team in a given week.

'Time spent in field' and 'time spent in canvassing' are measured in hours and aggregated by team-week.

'Total registered' indicates the total number of registrations among the target households allocated to a team in a given week.

'Family Registered' aggregates all instances where the entire family registered (as reported by households).

'Registered in Kibera' aggregates all instances where the household registered in a Kibera polling station .

Table 2A: Randomization Checks

	Mean	Horizontal Match	Vertical Match
Staff Level Variables			
Gender (Dummy for Male)	0.4333	-0.1925	0.087
		(0.1251)	(0.1646)
Same Gender as Teammate	0.5333	-0.1366	0.0497
		(0.2192)	(0.2196)
Age	26.7288	-0.0468	1.2921
		(0.876)	(1.2313)
Years of College Education	3.0431	0.0373	0.1786
		(0.3379)	(0.3409)
Senior Canvasser in Team	0.5	0.0932	-0.0932
		(0.2196)	(0.2196)
EA Level Variables			
Log Consumption per Capita	10.4168	0.0052	0.0155
		(0.0224)	(0.0534)
EA Poverty Rate	0.1467	-0.0047	-0.019
		(0.0084)	(0.0232)
Number of Households in EA	52.2533	2.0954	7.2829
		(2.6427)	(6.0092)
Average Age of Household Head	35.4021	-0.1605	0.2986
		(0.2733)	(0.682)
Head Years of Education	9.2715	-0.0151	0.2899
		(0.1992)	(0.3367)
Household Head Owns a Business	0.1496	-0.0307**	-0.0281
		(0.0144)	(0.0336)
Household Head Unemployed	0.3157	-0.0228	-0.0454
		(0.0152)	(0.0364)
Fraction of Youth Unemployed	0.1682	0.0252*	-0.0168
		(0.0151)	(0.0324)
Households with a Radio	0.7542	0.0055	0.0176
		(0.0157)	(0.0354)
Households with a TV	0.4143	0.029	0.0167
		(0.023)	(0.0682)
Households with Mud Cement Walls	0.511	0.0326	0.0034
		(0.0474)	(0.0744)
Households with Cement Floor	0.6331	0.0343	0.0766
		(0.0312)	(0.0778)
Households with Piped Water	0.6813	0.0528	0.0784
		(0.0588)	(0.095)
Households with Covered Pit	0.6582	0.0577	-0.0847
		(0.0485)	(0.0838)
Households Cooking with Paraffin	0.5325	0.028	-0.0296
		(0.0244)	(0.0462)
Households with Electricity	0.5135	0.037	-0.0168
		(0.0305)	(0.087)

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team in parentheses. Column (1) shows the mean of each row variable. In columns (2) and (3), each cell shows estimates from a regression of the row variable on the column variable.

Table 2B: Randomization Checks

	Mean	Horizontal Match	Vertical Match	External Match
Household Level Variables				
Household Size	3.6692	-0.0073	0.0731	-0.0546
		(0.0583)	(0.13)	(0.0332)
Nb of children in HH	1.5524	-0.0032	0.0811	-0.0229
		(0.0412)	(0.102)	(0.0217)
Household pays rent	0.9231	-0.0007	-0.0009	0.0003
		(0.0036)	(0.0094)	(0.0045)
Housing Rent (log)	7.2695	-0.0892	0.0318	-0.0198
		(0.0531)	(0.0735)	(0.0175)
Household ever evicted in Kibera	0.1222	0.0036	-0.0162	0.0069
		(0.005)	(0.0145)	(0.0072)
Years in Same Structure	8.2269	0.002	0.0249	0.0588
		(0.2278)	(0.504)	(0.1739)
Years in Kibera	15.5485	-0.2448	-0.4171	-0.0185
		(0.286)	(0.6273)	(0.1968)
Household Has Phone	0.5125	0.0118	0.045	-0.0111
		(0.025)	(0.0328)	(0.0074)

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team in parentheses.

Column (1) shows the mean of each row variable.

In columns (2), (3) and (4), each cell shows estimates from a regression of the row variable on the column variable.

Table 3: Effect of Ethnic Composition on Team Performance

	Visit Completion			Visit Duration		
	(1)	(2)	(3)	(4)	(5)	(6)
Horizontal Ethnic Match	0.085***	0.075**	0.083**	0.703**	1.243***	1.280***
	(0.022)	(0.032)	(0.032)	(0.268)	(0.324)	(0.325)
Wild Bootstrap	[0.002]	[0.193]	[0.137]	[0.085]	[0.081]	[0.070]
Randomization Inference	[0.022]	[0.170]	[0.145]	[0.081]	[0.058]	[0.056]
Vertical Ethnic Match		-0.033	-0.035		-1.435***	-1.407***
		(0.042)	(0.045)		(0.374)	(0.394)
Wild Bootstrap		[0.564]	[0.504]		[0.048]	[0.050]
Randomization Inference		[0.643]	[0.653]		[0.053]	[0.070]
External Ethnic Match			-0.002			-0.035
			(0.007)			(0.051)
Random Visit Order	-0.004***	-0.004***	-0.004***	-0.034***	-0.035***	-0.032***
	(0.001)	(0.001)	(0.001)	(0.008)	(0.008)	(0.008)
IEBC Treatment	-0.055***	-0.110***	-0.124***	0.134	-0.470*	-0.554*
	(0.015)	(0.027)	(0.032)	(0.189)	(0.275)	(0.295)
R^2	0.022	0.023	0.031	0.031	0.036	0.040
Dep Var Mean	0.41	0.41	0.42	1.82	1.82	1.87
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	30	30	30	30	30	30
Observations	30947	30947	27543	29302	29302	26067

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team.

Wild bootstrap and randomization inference p-values reported in brackets for horizontal & vertical matches.

All specifications include ethnicity dummies for each staff member.

Columns (3) and (6) also include household ethnicity dummies.

Table 4: Ethnic Composition and Learning

	Visit Completion			Visit Duration		
	(1)	(2)	(3)	(4)	(5)	(6)
Horizontal Ethnic Match × Day	0.010***	0.008**	0.009**	0.083**	0.083**	0.090**
	(0.003)	(0.004)	(0.004)	(0.031)	(0.032)	(0.033)
Wild Bootstrap	[0.011]	[0.061]	[0.136]	[0.014]	[0.041]	[0.071]
Randomization Inference	[0.018]	[0.078]	[0.080]	[0.019]	[0.034]	[0.031]
Vertical Ethnic Match × Day		-0.005	-0.005		0.009	0.009
		(0.004)	(0.004)		(0.021)	(0.020)
Wild Bootstrap		[0.053]	[0.015]		[0.691]	[0.558]
Randomization Inference		[0.289]	[0.221]		[0.689]	[0.692]
External Ethnic Match × Day			0.001			0.001
			(0.002)			(0.013)
Horizontal Ethnic Match	0.007	0.019	0.016	0.014	0.636*	0.599*
	(0.033)	(0.036)	(0.038)	(0.290)	(0.343)	(0.348)
Vertical Ethnic Match		-0.024	-0.011		-1.949***	-1.798***
		(0.064)	(0.066)		(0.409)	(0.409)
External Ethnic Match			-0.016			-0.047
			(0.021)			(0.118)
Day	0.002	0.003	0.001	0.028	0.026	0.022
	(0.004)	(0.004)	(0.004)	(0.024)	(0.026)	(0.029)
R^2	0.028	0.030	0.037	0.038	0.045	0.049
Dep Var Mean	0.435	0.435	0.443	1.939	1.939	1.977
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	30	30	30	30	30	30
Observations	24360	24360	21842	22900	22900	20518

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team.

Wild bootstrap and randomization inference p-values reported in brackets.

All specifications include ethnicity dummies for each staff member.

All specifications also include controls for the daily visit order and IEBC treatment.

Columns (3) and (6) also include household ethnicity effects.

Day refers to the day in the two week period of canvassing (see text for details).

Table 5: Gender Composition and Team Performance

	Visit Completion				Visit Duration			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Horizontal Gender Match	0.036	0.060***	0.013	0.007	0.098	0.134	0.530**	0.691***
	(0.022)	(0.018)	(0.024)	(0.021)	(0.182)	(0.172)	(0.200)	(0.191)
Wild Bootstrap	[0.222]	[0.051]	[0.787]	[0.983]	[0.697]	[0.642]	[0.214]	[0.166]
Randomization Inference	[0.182]	[0.022]	[0.716]	[0.855]	[0.654]	[0.533]	[0.120]	[0.104]
Vertical Gender Match		0.050		-0.015		0.031		0.338
		(0.030)		(0.018)		(0.236)		(0.251)
Wild Bootstrap		[0.261]		[0.620]		[0.963]		[0.545]
Randomization Inference		[0.202]		[0.717]		[0.917]		[0.549]
Horizontal Ethnic Match			0.074***	0.064*			0.806***	1.136***
			(0.018)	(0.032)			(0.264)	(0.368)
Wild Bootstrap			[0.006]	[0.326]			[0.096]	[0.268]
Randomization Inference			[0.032]	[0.353]			[0.083]	[0.132]
Vertical Ethnic Match				-0.012				-2.031***
				(0.044)				(0.438)
Wild Bootstrap				[0.814]				[0.105]
Randomization Inference				[0.907]				[0.066]
Random Visit Order	-0.004***	-0.004***	-0.004***	-0.004***	-0.035***	-0.035***	-0.034***	-0.035***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.007)	(0.007)	(0.008)	(0.008)
IEBC Treatment	-0.046**	-0.037*	-0.065***	-0.106***	0.274	0.409**	-0.060	-0.833***
	(0.019)	(0.020)	(0.017)	(0.033)	(0.164)	(0.177)	(0.192)	(0.204)
R^2	0.015	0.017	0.023	0.024	0.017	0.019	0.036	0.042
Dep Var Mean	0.41	0.41	0.41	0.41	1.82	1.82	1.82	1.82
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Clusters	30	30	30	30	30	30	30	30
Observations	30947	30947	30947	30947	29302	29302	29302	29302

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team in parentheses.

Wild bootstrap and randomization p-values reported in brackets.

All specifications include gender dummies for each staff member.

Columns (3)-(4) and (7)-(8) also include staff ethnicity dummies.

Table 6: Self-Reported Registration Outcomes

	Visit Completion	Visit Duration	Registered	Family Registered	Kibera Registered
	(1)	(2)	(3)	(4)	(5)
Horizontal Ethnic Match	0.035	1.188***	-0.023	0.045*	0.002
	(0.033)	(0.396)	(0.034)	(0.025)	(0.021)
Wild Bootstrap	[0.620]	[0.125]	[0.670]	[0.301]	[0.946]
Randomization Inference	[0.520]	[0.097]	[0.697]	[0.285]	[0.957]
Vertical Ethnic Match	-0.030	-1.568***	-0.060	-0.068***	-0.130***
	(0.052)	(0.523)	(0.045)	(0.025)	(0.034)
Wild Bootstrap	[0.727]	[0.080]	[0.539]	[0.116]	[0.117]
Randomization Inference	[0.719]	[0.118]	[0.446]	[0.129]	[0.046]
External Ethnic Match	-0.024	-0.180**	0.011	0.011	0.014
	(0.016)	(0.082)	(0.015)	(0.017)	(0.027)
Random Visit Order	-0.002*	-0.027**	-0.001	0.001	0.006**
	(0.001)	(0.010)	(0.002)	(0.002)	(0.003)
IEBC Treatment	-0.089**	-0.425	0.031	0.013	0.038
	(0.036)	(0.371)	(0.080)	(0.057)	(0.056)
R^2	0.022	0.041	0.004	0.007	0.011
Clusters	30	30	30	30	30
Dep var mean	0.483	2.123	0.801	0.687	0.587
Observations	7387	6932	3151	3100	3151

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team in parentheses.

Wild bootstrap and randomization p-values reported in brackets for the horizontal and vertical ethnic matches.

The sample is restricted to households randomly selected to be canvassed and contacted by phone.

All specifications include ethnicity dummies for households and staff members.

Registered (column 3) indicates whether the head of household registered.

Family (column 4) indicates whether the entire household registered.

Kibera (column 5) indicates whether the head of household registered to vote in Kibera.

Table 7: Time Use

	Field Time	Canvassing Time	# Visits	# Visits Alone
	(1)	(2)	(3)	(4)
Horizontal Ethnic Match	88.202***	125.340***	14.444***	3.013*
	(31.899)	(33.815)	(4.483)	(1.574)
Wild Bootstrap	[0.140]	[0.060]	[0.054]	[0.283]
Randomization Inference	[0.115]	[0.030]	[0.051]	[0.225]
Vertical Ethnic Match	-140.513***	-52.613	5.465	1.772
	(48.144)	(39.910)	(4.948)	(1.450)
Wild Bootstrap	[0.092]	[0.399]	[0.405]	[0.381]
Randomization Inference	[0.095]	[0.419]	[0.423]	[0.445]
R^2	0.117	0.201	0.131	0.124
Clusters	30	30	30	30
Dep Var Mean	321.77	152.15	32.39	5.13
Observations	266	266	266	266

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team in parentheses.

Wild bootstrap and randomization p-values reported in brackets.

All regressions are at the team-day level and include ethnicity dummies for each staff member.

Table 8: Staff Survey Results

	Performance		Work Hours		Language		Socialize		In Contact	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Horizontal Ethnic Match	-0.138	-0.124	-0.199	-0.167	0.037	0.039	0.231	0.249*	0.323***	0.338***
	(0.194)	(0.195)	(0.286)	(0.287)	(0.211)	(0.213)	(0.149)	(0.144)	(0.103)	(0.108)
Wild Bootstrap	[0.875]	[0.847]	[0.587]	[0.612]	[0.989]	[0.936]	[0.051]	[0.040]	[0.016]	[0.014]
Randomization Inference	[0.508]	[0.556]	[0.477]	[0.556]	[0.841]	[0.839]	[0.132]	[0.095]	[0.005]	[0.005]
Vertical Ethnic Match	-0.156		-0.881*		-0.113		-0.149		0.157	
	(0.364)		(0.495)		(0.307)		(0.248)		(0.206)	
Wild Bootstrap	[0.954]		[0.325]		[0.874]		[0.263]		[0.699]	
Randomization Inference	[0.670]		[0.117]		[0.707]		[0.540]		[0.471]	
Vertical Ethnic Match, Self		-0.802**		-1.042**		-0.156		-0.483*		0.244
		(0.294)		(0.455)		(0.298)		(0.237)		(0.176)
Wild Bootstrap		[0.031]		[0.071]		[0.712]		[0.110]		[0.463]
Randomization Inference		[0.020]		[0.052]		[0.591]		[0.096]		[0.210]
Vertical Ethnic Match, Not Self		0.215		-0.472		-0.072		0.138		0.225
		(0.243)		(0.438)		(0.317)		(0.230)		(0.225)
Wild Bootstrap		[0.055]		[0.848]		[0.959]		[0.389]		[0.490]
Randomization Inference		[0.412]		[0.324]		[0.829]		[0.566]		[0.339]
R^2	0.351	0.456	0.174	0.186	0.217	0.219	0.244	0.336	0.441	0.453
Clusters	30	30	30	30	30	30	30	30	30	30
Dep Var Mean	4.351	4.351	9.105	9.105	0.509	0.509	0.544	0.544	0.351	0.351
Observations	57	57	57	57	57	57	57	57	57	57

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team in parentheses.

Wild bootstrap and randomization inference p-values reported in brackets.

All regressions include Staff ethnicity dummies and demographic controls (gender, age, and education).

Performance is a personal evaluation of the canvasser's own performance on a 1-5 scale.

Work hours are self-reported daily work hours.

Language is a dummy indicating the canvasser spoke his most preferred language with his coworker.

Socialize is a dummy indicating the canvasser socialized with his coworker outside the experiment.

In Contact is a dummy indicating the canvasser stayed in touch with his coworker after the experiment.

Figure 1: Sample Tracking Sheet

NYUMBA HADI NYUMBA: SECOND VISIT

WARD: **VILLAGE: Kichinjio** **EA NUMBER: 201008**

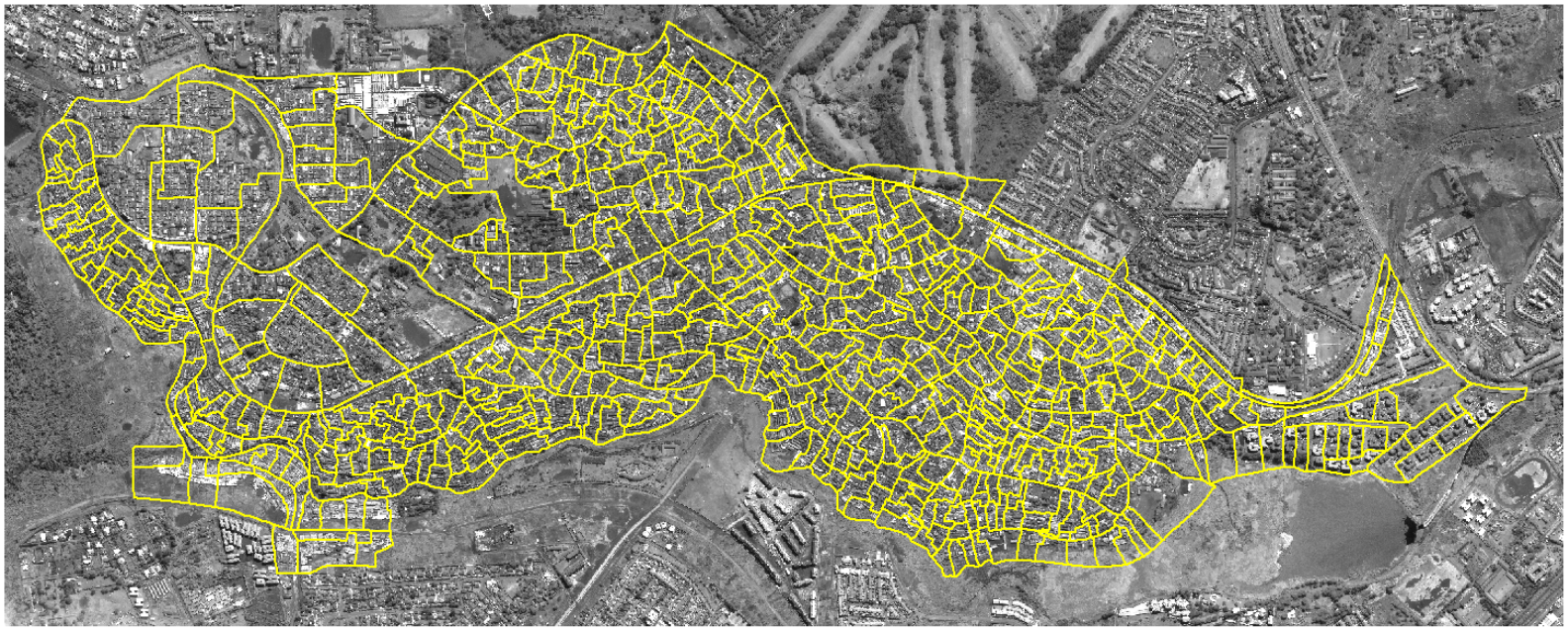
TREATMENT STATUS: Treatment 2
TEAM: 21
ORDER: 4

Enumerator Code: First and Last Name: _____ / _____

Date: |__|__| / |__|__|/ 2012 (DD/MM/YYYY)

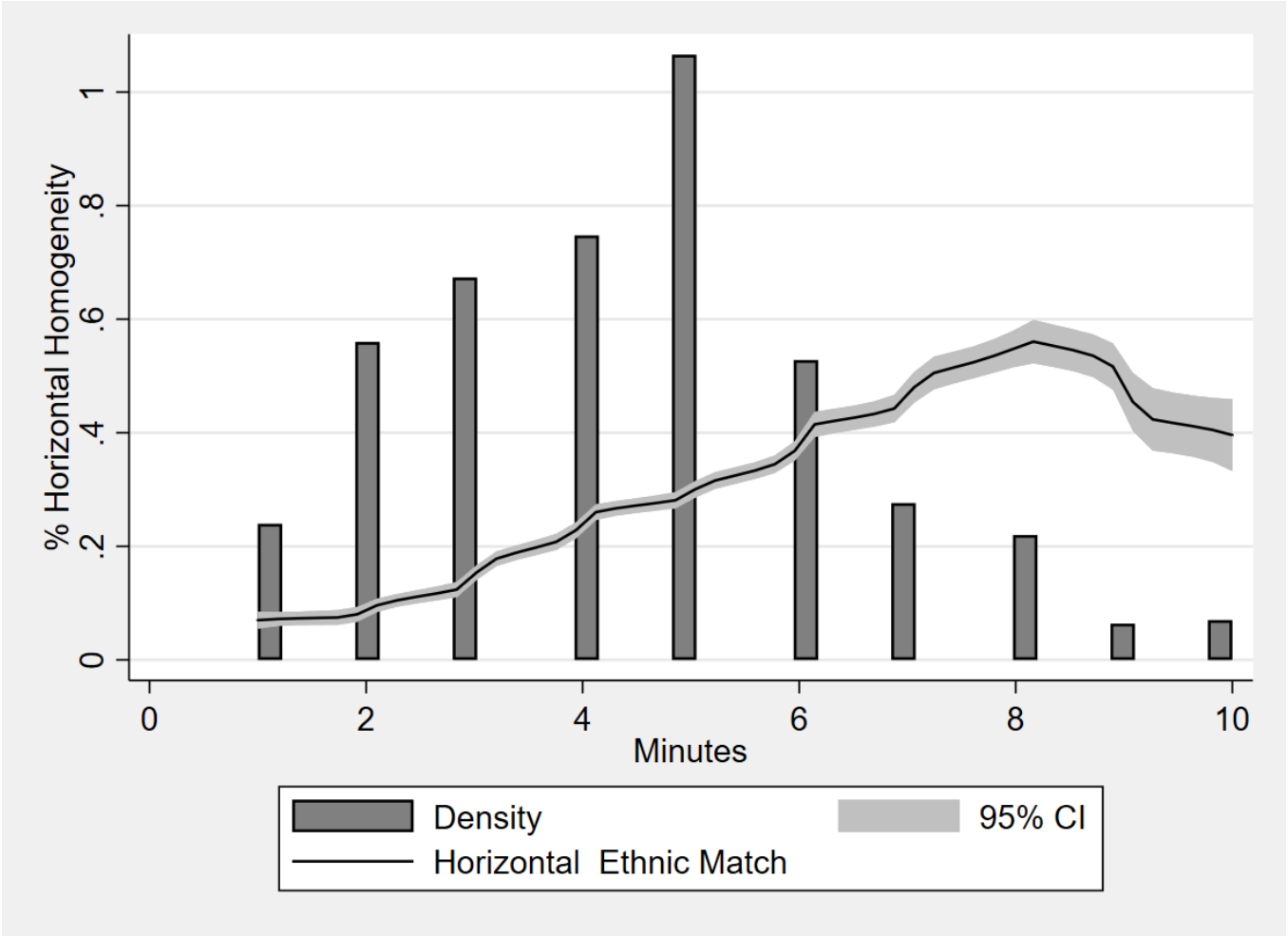
Household Head Name	HHID	GPS Lat.	GPS Long.	Result	First and Last name of individual spoken with	Relation to Head	Person is Registered 1=Yes 2=No	Person has ID 1=Yes 2=No	Who else was present? First and Last names	Relation to head	Start time	End time	Enumerator Likelihood that respondent registers w/in a week
				|__|	_____ / _____	|__|	|__|	|__|	|__|	|__
__
__			__	__	: 	__	__			__	__	 : 	__	__			__	
				|__|	_____ / _____	|__|	|__|	|__|	|__|	|__
__
__			__	__	: 	__	__			__	__	 : 	__	__			__	
				|__|	_____ / _____	|__|	|__|	|__|	|__|	|__
__
__			__	__	: 	__	__			__	__	 : 	__	__			__	
				|__|	_____ / _____	|__|	|__|	|__|	|__|	|__
__
__			__	__	: 	__	__			__	__	 : 	__	__			__	
				|__|	_____ / _____	|__|	|__|	|__|	|__|	|__
__
__			__	__	: 	__	__			__	__	 : 	__	__			__	
				|__|	_____ / _____	|__|	|__|	|__|	|__|	|__
__
__			__	__	: 	__	__			__	__	 : 	__	__			__	

Figure 2: Map of Kibera



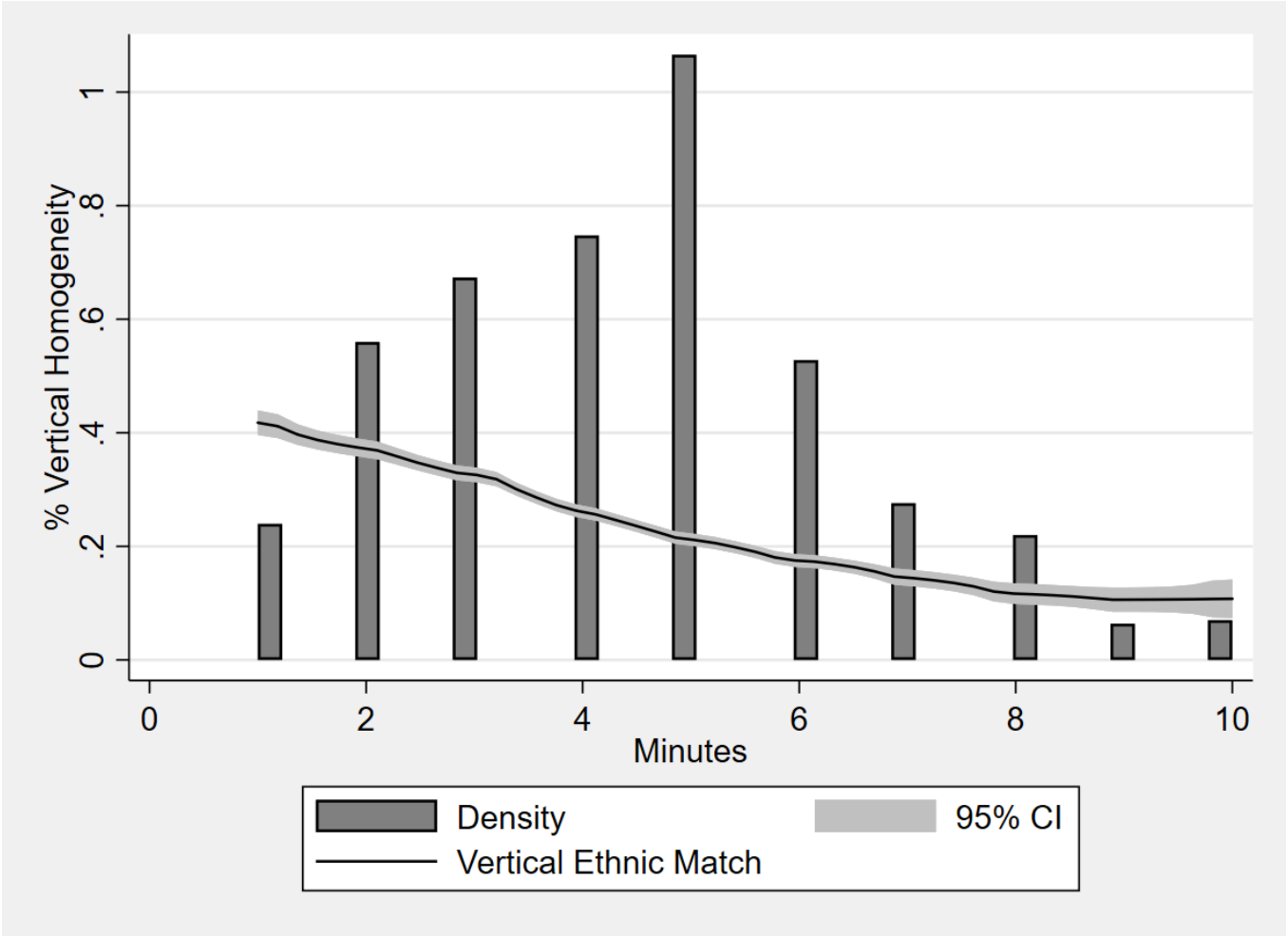
Note: Satellite image of the Kibera slum, with Enumeration Areas (EAs) outlined in yellow.
The Kibera slum in Nairobi covers about five square km of area and over 30,000 households across more than 600 Enumeration Areas.

Figure 3: Horizontal Homogeneity Over Distribution of Duration



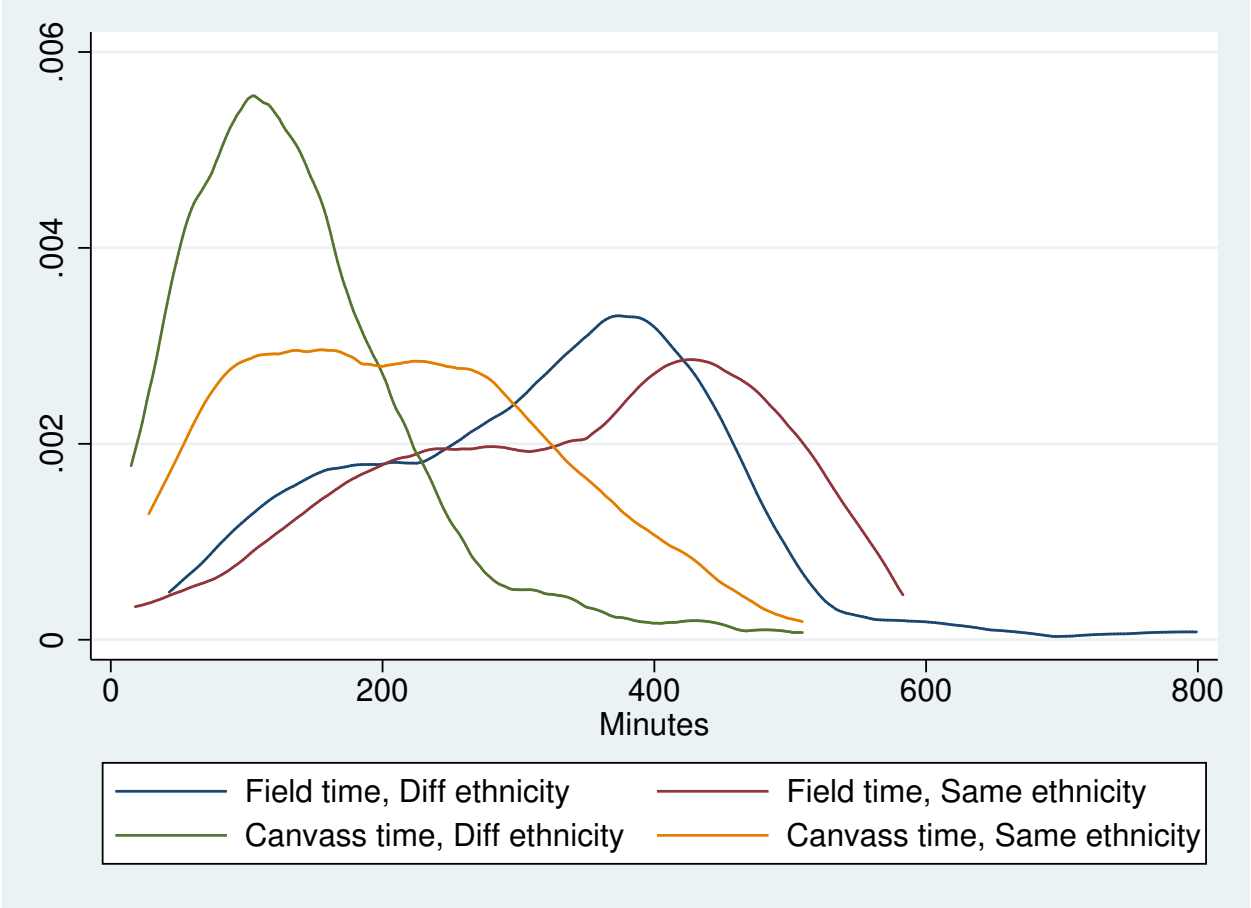
Note: The bar chart is a histogram of visit duration (for completed visits only) for durations 10 minutes or less. The solid black line is a kernel density of the horizontal ethnic match over values of visit duration.

Figure 4: Vertical Homogeneity Over Distribution of Duration



Note: The bar chart is a histogram of visit duration (for completed visits only) for durations 10 minutes or less. The solid black line is a kernel density of the vertical ethnic match over values of visit duration.

Figure 5: Time Use in the Field, by Horizontal Ethnic Diversity



Note: All plots are kernel densities. Time use is measured in minutes per day.
Canvassing time is measured as the sum of the durations of all visits conducted each day.
Field time is calculated as the difference between the end of the last visit and the beginning of the first visit each day.

Appendix Table 1: Ethnic Composition and Team Performance, Alternative Specifications

	Visit completed					
	(1)	(2)	(3)	(4)	(5)	(6)
Horizontal Ethnic Match	0.083**	0.063*	0.088***	0.166***	0.090***	0.054**
	(0.032)	(0.032)	(0.029)	(0.027)	(0.031)	(0.023)
Wild Bootstrap	[0.137]	[0.325]	[0.142]	[0.204]	[0.067]	[0.168]
Randomization Inference	[0.145]	[0.242]	[0.087]	[0.078]	[0.121]	[0.146]
Vertical Ethnic Match	-0.035	0.047	-0.044	-0.072**	-0.052	-0.047
	(0.045)	(0.046)	(0.039)	(0.031)	(0.041)	(0.037)
Wild Bootstrap	[0.504]	[0.485]	[0.497]	[0.343]	[0.925]	[0.496]
Randomization Inference	[0.653]	[0.513]	[0.494]	[0.371]	[0.456]	[0.407]
External Ethnic Match	-0.002	-0.000	0.002	0.006	0.002	-0.004
	(0.007])	(0.008)	(0.008)	(0.008)	(0.008)	(0.007)
Random Visit Order	-0.004***		-0.004***	-0.003***	-0.004***	-0.003**
	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)
IEBC Treatment	-0.124***		-0.130***			-0.099***
	(0.032)		(0.028)			(0.032)
R^2	0.031	0.018	0.022	0.059	0.023	0.012
Dep Var Mean	0.42	0.42	0.42	0.42	0.42	0.42
Ethnicity effects	Yes	Yes	Yes	Yes	Yes	Alt.
Controls	Yes	No	Yes	Yes	Yes	Yes
Strata FE	Yes	No	No	No	No	No
EA FE	No	No	No	Yes	No	No
Supervisor FE	No	No	No	No	Yes	No
Clusters	30	30	30	30	30	30
Observations	27543	27543	27543	27543	27543	27543

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team in parentheses.
Wild bootstrap and randomization inference p-values reported in brackets.
In col. (6), staff ethnicity effects are parametrized as dummies for any ethnicity being represented on the team (see text for details).

Appendix Table 2: Ethnic Composition and Team Performance, Alternative Specifications

	Visit duration					
	(1)	(2)	(3)	(4)	(5)	(6)
Horizontal Ethnic Match	1.280***	1.146***	1.248***	1.902***	1.280***	0.480*
	(0.325)	(0.333)	(0.327)	(0.327)	(0.341)	(0.240)
Wild Bootstrap	[0.070]	[0.104]	[0.074]	[0.066]	[0.070]	[0.229]
Randomization Inference	[0.056]	[0.063]	[0.050]	[0.035]	[0.062]	[0.170]
Vertical Ethnic Match	-1.407***	-0.985**	-1.394***	-2.331***	-1.561***	-0.458
	(0.394)	(0.446)	(0.428)	(0.440)	(0.418)	(0.279)
Wild Bootstrap	[0.050]	[0.189]	[0.053]	[0.096]	[0.578]	[0.194]
Randomization Inference	[0.070]	[0.205]	[0.082]	[0.066]	[0.060]	[0.240]
External Ethnic Match	-0.035	-0.029	-0.023	0.044	-0.026	-0.039
	(0.051)	(0.052)	(0.050)	(0.051)	(0.051)	(0.043)
Random Visit Order	-0.032***		-0.033***	-0.029***	-0.033***	-0.032***
	(0.008)		(0.009)	(0.008)	(0.009)	(0.009)
IEBC Treatment	-0.554*		-0.551*			-0.015
	(0.295)		(0.303)			(0.247)
R^2	0.040	0.030	0.035	0.073	0.035	0.026
Dep Var Mean	1.87	1.87	1.87	1.87	1.87	1.87
Ethnicity effects	Yes	Yes	Yes	Yes	Yes	Alt.
Controls	Yes	No	Yes	Yes	Yes	Yes
Strata FE	Yes	No	No	No	No	No
EA FE	No	No	No	Yes	No	No
Supervisor FE	No	No	No	No	Yes	No
Clusters	30	30	30	30	30	30
Observations	26067	26067	26067	26067	26067	26067

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team in parentheses.
Wild bootstrap and randomization inference p-values reported in brackets.
In col. (6), staff ethnicity effects are parametrized as dummies for any ethnicity being represented on the team (see text for details).

Appendix Table 3: Ethnic Composition and Team Performance, Controlling for Other Demographic Homogeneity

	Visit Completion				Visit Duration			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Horizontal Ethnic Match	0.056*	0.073**	0.073*	0.094***	1.316***	1.559***	1.560***	1.382***
	(0.027)	(0.027)	(0.037)	(0.031)	(0.278)	(0.294)	(0.381)	(0.305)
Wild Bootstrap	[0.249]	[0.209]	[0.295]	[0.262]	[0.106]	[0.088]	[0.109]	[0.159]
Randomization Inference	[0.314]	[0.215]	[0.334]	[0.166]	[0.044]	[0.026]	[0.066]	[0.068]
Vertical Ethnic Match	-0.015	-0.020	-0.073	-0.085	-1.970***	-2.054***	-2.769***	-1.843***
	(0.044)	(0.045)	(0.070)	(0.055)	(0.455)	(0.428)	(0.647)	(0.616)
Wild Bootstrap	[0.781]	[0.744]	[0.515]	[0.363]	[0.112]	[0.101]	[0.037]	[0.145]
Randomization Inference	[0.871]	[0.839]	[0.612]	[0.443]	[0.067]	[0.027]	[0.058]	[0.176]
Horizontal Gender Match	0.018			0.020	0.459***			0.204
	(0.019)			(0.018)	(0.128)			(0.166)
Horizontal Age Match		0.035**		0.036*		0.455**		0.280
		(0.016)		(0.019)		(0.194)		(0.249)
Horizontal Education Match			-0.030	-0.039**			-0.432	-0.326
			(0.023)	(0.016)			(0.263)	(0.285)
IEBC Treatment	-0.105***	-0.104***	-0.128***	-0.139***	-0.850***	-0.786***	-1.118***	-0.710*
	(0.032)	(0.034)	(0.038)	(0.033)	(0.212)	(0.222)	(0.311)	(0.360)
Random Visit Order	-0.004***	-0.004***	-0.004***	-0.004***	-0.034***	-0.034***	-0.034***	-0.034***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.008)	(0.008)	(0.008)	(0.008)
R^2	0.024	0.025	0.025	0.025	0.041	0.041	0.040	0.039
Dep Var Mean	0.41	0.41	0.41	0.41	1.82	1.82	1.82	1.82
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	30947	30947	30947	30947	29302	29302	29302	29302

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team in parentheses.

Wild bootstrap and randomization inference p-values in brackets for the horizontal and vertical ethnic matches.

Teams have an age match if the age of both canvassers is within 2 years of each other.

Teams have an education match if both canvassers have completed the same education level (undergraduate or graduate).

Appendix Table 4:
Ethnic Composition and Learning, Alternative Specifications

	Visit completed					
	(1)	(2)	(3)	(4)	(5)	(6)
Horizontal Ethnic Match*Day	0.009**	0.008	0.008	0.005	0.008	0.009*
	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)
Wild Bootstrap	[0.136]	[0.062]	[0.074]	[0.137]	[0.149]	[0.044]
Randomization Inference	[0.080]	[0.136]	[0.152]	[0.351]	[0.159]	[0.098]
Vertical Ethnic Match*Day	-0.005	-0.011***	-0.010***	-0.001	-0.010***	-0.009**
	(0.004)	(0.003)	(0.004)	(0.005)	(0.004)	(0.003)
Wild Bootstrap	[0.015]	[0.018]	[0.018]	[0.011]	[0.011]	[0.030]
Randomization Inference	[0.221]	[0.011]	[0.022]	[0.880]	[0.022]	[0.027]
External Ethnic Match*Day	0.001	0.002	0.002	0.001	0.002	0.001
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
Horizontal Ethnic Match	0.016	0.015	0.032	0.079**	0.033	-0.028
	(0.038)	(0.040)	(0.042)	(0.031)	(0.045)	(0.035)
Vertical Ethnic Match	-0.011	0.122**	0.030	-0.010	0.027	0.049
	(0.066)	(0.057)	(0.060)	(0.064)	(0.063)	(0.044)
External Ethnic Match	-0.016	-0.019	-0.015	-0.001	-0.015	-0.011
	(0.021)	(0.024)	(0.024)	(0.019)	(0.024)	(0.024)
Day	0.001	-0.008**	0.005	0.009	0.004	0.004
	(0.004)	(0.003)	(0.005)	(0.006)	(0.005)	(0.004)
R^2	0.037	0.023	0.026	0.069	0.026	0.025
Dep Var Mean	0.443	0.443	0.443	0.443	0.443	0.443
Ethnicity effects	Yes	Yes	Yes	Yes	Yes	Alt.
Controls	Yes	No	Yes	Yes	Yes	Yes
Strata FE	Yes	No	No	No	No	No
EA FE	No	No	No	Yes	No	No
Supervisor FE	No	No	No	No	Yes	No
Clusters	30	30	30	30	30	30
Observations	21842	21842	21842	21842	21842	21842

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team in parentheses.
Wild bootstrap and randomization inference p-values reported in brackets.
Day refers to the day in the two week period of canvassing (see text for details).

Appendix Table 5:
Ethnic Composition and Learning, Alternative Specifications

	Visit duration					
	(1)	(2)	(3)	(4)	(5)	(6)
Horizontal Ethnic Match*Day	0.090**	0.079**	0.082**	0.159***	0.081**	0.078**
	(0.033)	(0.035)	(0.037)	(0.038)	(0.038)	(0.038)
Wild Bootstrap	[0.071]	[0.049]	[0.048]	[0.074]	[0.054]	[0.070]
Randomization Inference	[0.031]	[0.060]	[0.061]	[0.033]	[0.073]	[0.081]
Vertical Ethnic Match*Day	0.009	-0.017	-0.014	0.070**	-0.013	-0.021
	(0.020)	(0.021)	(0.022)	(0.034)	(0.022)	(0.019)
Wild Bootstrap	[0.558]	[0.538]	[0.663]	[0.491]	[0.505]	[0.399]
Randomization Inference	[0.692]	[0.453]	[0.572]	[0.303]	[0.594]	[0.322]
External Ethnic Match*Day	0.001	0.003	0.004	0.009	0.004	0.004
	(0.013)	(0.015)	(0.014)	(0.011)	(0.014)	(0.014)
Horizontal Ethnic Match	0.599*	0.626*	0.678*	0.845**	0.707*	-0.220
	(0.348)	(0.360)	(0.352)	(0.412)	(0.374)	(0.320)
Vertical Ethnic Match	-1.798***	-1.136**	-1.575***	-3.184***	-1.704***	-0.322
	(0.409)	(0.416)	(0.419)	(0.449)	(0.420)	(0.331)
External Ethnic Match	-0.047	-0.063	-0.055	-0.017	-0.058	-0.063
	(0.118)	(0.136)	(0.131)	(0.099)	(0.132)	(0.135)
Day	0.022	-0.072***	0.040	-0.006	0.038	0.032
	(0.029)	(0.018)	(0.027)	(0.052)	(0.027)	(0.028)
R^2	0.049	0.039	0.042	0.084	0.042	0.036
Dep Var Mean	1.977	1.977	1.977	1.977	1.977	1.977
Ethnicity effects	Yes	Yes	Yes	Yes	Yes	Alt.
Controls	Yes	No	Yes	Yes	Yes	Yes
Strata FE	Yes	No	No	No	No	No
EA FE	No	No	No	Yes	No	No
Supervisor FE	No	No	No	No	Yes	No
Clusters	30	30	30	30	30	30
Observations	20518	20518	20518	20518	20518	20518

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team in parentheses.
Wild bootstrap and randomization inference p-values reported in brackets.
Day refers to the day in the two week period of canvassing (see text for details).

Appendix Table 6: Robustness to Outliers in Duration

	Trimmed Duration		Log (1+Duration)	
	(1)	(2)	(3)	(4)
Horizontal Ethnic Match	1.160***	0.600*	0.324***	0.143
	(0.295)	(0.294)	(0.089)	(0.090)
Wild Bootstrap Horizontal	[0.052]		[0.057]	
Randomization Inference Horizontal	[0.051]		[0.054]	
Vertical Ethnic Match	-1.394***	-1.856***	-0.358***	-0.483***
	(0.337)	(0.385)	(0.101)	(0.118)
Wild Bootstrap Vertical	[0.023]		[0.036]	
Randomization Inference Vertical	[0.046]		[0.057]	
Horizontal Ethnic Match*Day		0.077***		0.025***
		(0.025)		(0.008)
Wild Bootstrap Horizontal		[0.035]		[0.045]
Randomization Inference Horizontal		[0.018]		[0.016]
Vertical Ethnic Match*Day		0.012		0.000
		(0.019)		(0.005)
Wild Bootstrap Vertical		[0.615]		[0.228]
Randomization Inference Vertical		[0.552]		[0.991]
Day		0.013		0.004
		(0.022)		(0.007)
R^2	0.050	0.062	0.040	0.053
Dep Var Mean	1.71	1.805	0.65	0.678
Strata FE	Yes	Yes	Yes	Yes
Clusters	30	30	30	30
Observations	25812	20291	26067	20518

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team in parentheses.

The dep var in cols. 1-2 is duration trimmed at the 99th percentile (10 minutes).

The dep var in cols. 3-4 is log of (1+duration).

Wild bootstrap and randomization inference p-values in brackets.

Day refers to the day in the two week period of canvassing (see text for details).

Appendix Table 7: Unpacking Visit Duration

	Visit Duration			
	(1) Baseline	(2) Mean	(3) p10	(4) p90
Horizontal Ethnic Match	1.243***	0.716***	1.023***	0.363*
	(0.324)	(0.187)	(0.255)	(0.190)
Wild Bootstrap	[0.081]	[0.110]	[0.084]	[0.355]
Randomization Inference	[0.058]	[0.060]	[0.059]	[0.273]
Vertical Ethnic Match	-1.435***	-0.947***	-1.231***	-0.620**
	(0.374)	(0.257)	(0.311)	(0.271)
Wild Bootstrap	[0.048]	[0.080]	[0.045]	[0.194]
Randomization Inference	[0.053]	[0.077]	[0.061]	[0.181]
Random Visit Order	-0.035***	-0.013**	-0.026***	0.001
	(0.008)	(0.005)	(0.006)	(0.006)
IEBC Treatment	-0.470*	0.055	-0.251	0.407***
	(0.275)	(0.159)	(0.223)	(0.118)
R^2	0.036	0.059	0.043	0.055
Dep Var Mean	1.82	4.79	3.06	6.78
Strata FE	Yes	Yes	Yes	Yes
Clusters	30	30	30	30
Observations	29302	29302	29302	29302

Note: * p<0.1, ** p<0.05, *** p<0.01. Standard errors clustered by team.

Column 1 reproduces the baseline estimates from Table 3, column 5.

In col.2, we impute duration at the mean of conditional duration for incomplete visits.

In col.3, we impute duration at the 10th percentile of conditional duration (2 minutes).

In col.4, we impute duration at the 90th percentile of conditional duration (8 minutes).