DOES WORKING FROM HOME WORK?
EVIDENCE FROM A CHINESE EXPERIMENT

Nicholas Bloom[a], James Liang[b], John Roberts[c] and Zhichun Jenny Ying[d]

July 2012

Abstract:
Over 10% of US employees now regularly work from home (WFH), but there is widespread skepticism over its impact highlighted by phrases like “shirking from home”. We report the results of a WFH experiment in a 13,000 employee NASDAQ listed Chinese multinational. Call center employees who volunteered to WFH were randomized into home or office working for 9-months. Home-working led to a 13% performance increase, of which about 9.5% is from working more minutes per shift (fewer breaks and sick-days) and 3.5% from more calls per minute (quieter working environment). Home workers also reported improved work satisfaction and their job attrition rate fell by 50%. After the experiment, the firm rolled the program out to all employees, letting them choose home or office working. Interestingly, only half of the volunteer group decided to work at home, with the other half changing their minds in favor of office working. After allowing employees to choose, the performance impact of WFH more than doubled, highlighting the benefits of choice alongside modern practices like home working.

Keywords: working from home, organization, productivity, field experiment, and China

Acknowledgements: We wish to thank Jennifer Cao, Mimi Qi, Maria Sun and Edison Yu for their help in this research project. We thank Chris Palauni for organizing our trip to Jet Blue, and David Butler, Jared Fletcher and Michelle Rowan for their time discussing the call-center and home-working industry. We thank in particular our discussants Mushfiq Mobarak, Rachael Heath and Sabrina Pabliona, and seminar audiences at the AEA, Brown, the London School of Economics, Stanford GSB and the World Bank for comments.

Conflict of interest statement: We wish to thank: our discussants Mushiq Mobarak and Rachel Heath; and Stanford Economics Department, Stanford GSB, and the Toulouse Network for Information Technology (which is supported by Microsoft) for funding for this project. No funding was received from CTrip. James Liang is the co-founder, ex-CEO and current Chairman of Ctrip. No other co-author has any financial relationship (or received any funding) from CTrip. The results or paper were not pre-screened by anyone.

[a] Stanford Economics, CEPR and NBER; [b] Ctrip, [c] Stanford GSB, [d] Stanford Economics

I. INTRODUCTION

Working from home (WFH) is becoming an increasingly common practice. In the United States, over 10% of the workforce reports working from home at least one day a week, while the proportion primarily WFH has almost doubled from 2.3% in 1980 to 4.3% in 2010 (Figure 1a). At the same time, the wage discount (after controlling for observables) from working exclusively at home has fallen, from 30% in 1980 to zero by 2000 as WFH moved from being prevalent in only low-skilled jobs to becoming a more widespread practice (Oettinger, 2010). Home-based workers now span a wide spectrum of occupations, ranging from sales assistants to managers and software engineers (Figure 1b).

The trade-off between home-life and work-life has also received increasing attention as the number of households in the US with all parents working has increased from 25% in 1968 to 48% by 2008 (Council of Economic Advisors, 2010). These rising work pressures are leading governments in the US and Europe to investigate ways to promote work-life balance. For example, the Council of Economic Advisers (CEA) published a report launched by Michelle and Barak Obama at the White House in the summer of 2010 on policies to improve work-life balance. One of the key conclusions in the executive summary concerned the need for research to identify the trade-offs in work-life balance policies, stating:

> "*A factor hindering a deeper understanding of the benefits and costs of flexibility is a lack of data on the prevalence of workplace flexibility and arrangements, and more research is needed on the mechanisms through which flexibility influences workers' job satisfaction and firms' profits to help policy makers and managers alike*" (CEA, 2010)

Not surprisingly, given this lack of research, many firms are also uncertain about what policies to adopt concerning working from home. As a result, firms in the same industry adopt different practices. For example, in the U.S. airline industry, Jet Blue allows all regular call-center employees to work from home while Delta and Southwest allow no home working and United has a mix of practices. More generally, Bloom, Kretschmer and Van Reenen (2010) report 39% of US and 37% of European manufacturing firms offer home working, with wide variation in adoption rates within every 3-digit SIC code surveyed. They find similar variation in the adoption of other modern work-life balance practices like job-sharing, part-time working, flexi-time and extended maternity leave within every industry, with no consensus around what defines a "best-practice".

Given the uncertainty over the impact of working from home, CTrip – China's largest travel agency with 13,000 employees and a $5bn valuation on NASDAQ – wanted to experiment before deciding whether to implement a policy across the firm. The motivation was both to reduce office costs, which were becoming increasingly onerous due to rising rental rates at the Shanghai headquarter, and also to reduce their 50% annual rate of attrition among call center workers. The executives' concern was that allowing employees to work at home, away from the supervision of their shift managers, would have an extremely negative impact on their performance.

This experiment is unique both as the first randomized experiment on working from home, and also because one of the co-authors of this paper (James Liang) was the co-founder of CTrip, first CEO and current Chairman. This has provided excellent access not just to the experimental data, but also to the Ctrip managements' thinking about the experiment and its results. This affords detailed insight into the adoption of a modern management practice by a large publicly listed multinational firm.

In summary, the firm decided to run a nine-month experiment on working from home. They asked the employees in the Airfare and Hotel divisions whether they would be interested in working from home four days a week. Approximately half of the employees (508) were interested. Of these, 255 were qualified to take part in the experiment by virtue of having at least six months tenure, broadband access and a private room at home (in which they could work). After a lottery draw, those with even birthdays were selected to work at home while those with odd birthdates stayed in the office to act as the control group. The home and office employees in each team had to work the same shift because they worked under a common team manager. The two groups also used the same IT equipment, faced the same work order flow from a common central server, and bonus pay system. Hence, the only difference between the two groups was the location of work. This allows us to isolate the impact of working-from-home (flexi-place) versus other practices that are commonly bundled alongside this practice, such as flexible work hours (flexi-time).

We found four main results. First, the performance of the home workers went up dramatically, increasing by 13% over the nine month experiment. This improvement came mainly from a 9.5% increase in the number of minutes they worked during their shifts (i.e., the time they were logged in taking calls). This was due to a reduction in breaks and sick-days taken by the home workers. The remaining 3.5% improvement came from home workers increasing calls per minute worked, due to the quieter working conditions at home. Second, there were no spillovers to the rest of the group – interestingly, those remaining in the office had no drop in performance despite losing the treatment lottery. Third, attrition fell sharply among the home workers, dropping by 50% versus the control group. Home workers also reported substantially higher work satisfaction and attitudinal survey outcomes. Finally, at the end of the experiment, the firm estimated they saved about $2,000 per employee from working at home, leading them to roll the option to work-from-home out to the entire firm. This allowed the treatment and control groups to re-select their working arrangements. Almost half of the treatment group changed their minds and returned to the office, while two thirds of the control group (who initially had all requested to work from home) decided to stay in the office. This selection led to much larger long-run impacts from working at home as workers with relatively better performance at home remained at home while those performing relatively poorly at home returned to the office.

More generally, this experiment also highlights the extensive learning by both the firm and employees around the adoption of a modern management practice like working from home. Ex-ante, both groups were unsure about its impact, and the 9-month experiment and subsequent roll-out process was essential for their ability to evaluate this. These gradual learning effects are one factor behind the slow adoption of modern management practices, and we see them as similar to the adoption process for other types of innovations, like hybrid seed-corn as emphasized in Griliches' (1957) classic article.

The paper connects to two literatures. First, there is a literature on the adoption of work-life balance practices, based primarily on case-studies and surveys across firms. These tend to show large positive associations with lower employee turnover and absenteeism, and higher productivity and profitability (for example, see the surveys in CEA 2010, Bloom, Kretschmer and Van Reenen 2010, Bloom and Van Reenen 2011, and Oyer and Lazear 2012). But these studies are hard to evaluate because of the non-randomized nature of these programs. This is both true in terms of the selection of firms into working-from home programs, and also the selection of employees to work at home. For example, as we show in Table 7 when CTrip allowed a general roll-out of home-working, we see high-performing employees choosing to move home and low-performing employees choosing to return to the office, so that the non-experimental impact of working from home looks two times larger than the experimental impact.

More generally, there is a long literature on the puzzling dispersion of productivity between firms (see the literature from Walker 1887 to Leibenstein 1966 to Syversson 2011). This paper provides one rationale for this dispersion, which is the spread of modern work-life balance practices like working-from home. Their adoption is highly variable across firms in the US and Europe due to the limited consensus on their impact, but they have potentially large impacts on measured productivity. For example, based on the type of Census data that is usually used to measure productivity, CTrip would have increased productivity by 13% after introducing working from home.[1]

Section II describes the experiment in more detail, while section III presents the firm results, and section IV the impact on employees, while section V discussed the roll-out and finally section VI provides a set of concluding comments.

II. THE EXPERIMENT

II.A. The Company

Our experiment takes place in Ctrip, a leading travel service provider for hotel accommodations, airline tickets and packaged tours in China. Ctrip aggregates information on hotels and flights, and generates revenue through commissions from travel suppliers. The services provided by Ctrip are comparable to Expedia, Orbitz or Travelocity. Ctrip was established in 1999 and was quoted on NASDQ in 2003, and is currently worth about $5bn. It is the largest travel agent in China for number of rooms in terms of hotel nights and airline tickets booked, with over 50% market share in 2010. The co-founder of Ctrip, first CEO and the current Chairman is James Liang, who was also a Stanford PhD student and co-author on this paper. This has provided us with unparalleled access to the company, both in terms of data and experimental design, but also is terms of understanding the management decision making behind the experiment and roll-out.

To provide some background on the company Exhibition A displays photos of the Ctrip headquarters and call center in Shanghai. This is a modern multi-story building that houses the

[1] Census data measures labor input in terms of usual shift hours (i.e. 9am to 5pm) rather than actual minutes worked. So the full 13% increase in output of home workers would be attributed to increased productivity.

call center which is running the experiment, as well as several other CTrip divisions and its top management team. The firm also operates a second larger call center in Nan Tong, outside Shanghai. Call center employees are organized into small teams of around 10 to 15 people (mean of 11.7 and median of 11), grouped by department and the type of work.

Teams sit together in one area of the floor, typically occupying an entire aisle. Each team member works in a cubical with equipment including a computer, a telephone and a headset. When team members are ready to start work, they log on and calls are automatically dropped into their headsets. When they want to take a break, they log out of the system. If demand is low, calls are routed to call center employees on a first-in first-out basis (so employees receive calls in the order they entered the queue). The team leaders patrol the aisles to monitor employees' performance as well as help resolve issues with reservations and provide ongoing training.

Employees are paid a flat wage of ¥1300 per month (about $200) and a performance-related bonus which is about another ¥1300 on average. The bonus is primarily a linear function of call and order volumes, but with small adjustments for call quality (penalties are applied for call quality scores below certain thresholds) and shift type (night shifts, for example, are paid a higher flat rate). Promotion to team-leader is also based on performance, so both salary and career concerns provide strong incentives for employees to perform well.

Ctrip was interested in running the experiment to investigate the impact of allowing employees to work from home. They believed allowing employees to work from home would allow them to save on office space, cut down turnover, and reduce labor costs by tapping into a wider pool of workers, such as people living too far outside Shanghai to commute in on a daily basis but close enough to commute in on a weekly basis. But they were uncertain of the impact of allowing employees to work from home on their performance. Their workforce is primarily younger employees, many of whom may struggle to remain focused working from home.

Since no other Chinese firm had moved to allowing home-working amongst its call center employees, there was no local precedent. In the US, the decision to allow employees in call centers to work from home varies across firms, even those within the same industry, suggesting a lack of any consensus on its impact. For example, in the airline industry while Jet Blue and American Airlines allow home-working, British Airways, Continental, Delta and Southwestern do not, and United is experimenting with a mixed model. The prior academic literature on call centers also offered limited guidance, being based on case-studies of individual firm-level interventions.

II.B. The Experimental Design

Ctrip employs about 13,000 employees, of which 7,500 work at two large call centers as customer service representatives in Shanghai and Nan Tong. Our experiment takes place in the airfare and hotel booking departments in the Shanghai call center. The representatives' main job is to answer phone calls, make reservations, and work to resolve issues on existing bookings. They typically work 5 shifts a week, scheduled by the firm ahead of time. Employees are organized by teams, and a team works on the same schedule so individuals do not choose their shifts. The firm adjusts the length of the shift depending on volume of the bookings.

The treatment in our experiment is to work 4 shifts at home and to work on the 5th shift in the office on a fixed day of the week determined by the team leader. Treatment employees still work on the same schedule as their teammates because they have to work under the supervision of the team leader (who is always office based), but operate from home for 4 of their five shifts. For example, in a team the treatment employees might work from home from 9am to 5pm on Monday, Tuesday, Wednesday and Friday and from the office from 9am to 5pm on Thursday. The control employees would work from the office from 9am to 5pm on all five days. Hence, the experiment only changes the location of work, not the type of work or the hours of work. Because all incoming phone-calls and work orders are distributed by central servers, the work flow is also identical between work and home locations.

Importantly, individual employees are not allowed to work overtime outside their team shift as they require their team leader to supervise their work. Hence, entire teams can have their hours changed – for example all teams had their shifts increased during the week before Chinese New Year – but no individual is able to work overtime on their own. So, the impact of eliminating commuting time, which is 80 minutes a day for the average employee, on home-workers ability to work overtime is not a factor directly driving the results.[2] Home workers also use the same CTrip provided computer terminal and phone equipment and software, face the same pay and promotions structure, and undertake the same training as office workers.

In early November 2010, employees in the airfare and hotel booking departments were informed of the working from home program. They all took an extensive survey on demographics, working conditions and their willingness to join the program. Employees who are both willing and qualified to join the program were recruited for the experiment. To qualify, an employee needed to have tenure of at least 6 months, have broadband Internet at home to connect to the network, and to have an independent workspace at home during their shift (such as their own bedroom). 51% of the 996 employees in the airfare and hotel booking departments qualified for the experiment. Of those, 49% were interested in joining the experiment, with those with a more expensive and longer commute, with less tenure in the firm, with less education and with their own bedroom significantly more likely to want to work from home (see Table 1). Importantly, prior-performance (measured by the gross-wage given that almost 50% of salary is performance related pay) was not predictive for the take-up of working from home. This helped to assuage one concern of the firm that lower performing employees would be more tempted to work from home to avoid the direct supervision of their managers.

Interestingly 51% of employees did not opt to work-from-home despite the considerable saving in commuting time and cost. Apparently, the major reason given for this was the loneliness of working from home. In the end, 255 eligible employees volunteered to join the experiment.

The treatment and control groups were then determined from this group of 255 employees through a public lottery. Employees with an even birthdate (a day ending 2, 4, 6, 8, etc.) were selected into the treatment and those with an odd birthdate were in the control group. This selection of even birthdates into the treatment group was randomly chosen by the Chairman, James Liang, by drawing a ping pong ball from an urn in a public ceremony one week prior to

---

[2] It could indirectly matter if, for example, employees at home can run household errands in the time saved by not commuting that employees working from the office have to take breaks to perform.

the experiment start date (see Exhibit B).[3] Even birthdate employees who had chosen to be in the experiment group are notified and equipment is installed at each treatment participant's home the following week. Odd birthdate employees who had chosen to be in the experiment acted as the control group. The experiment commenced on December 6, 2010.

The experiment lasted about 9 months, and all treatment employees had to remain at home for this period even if they changed their mind and wanted to return to the office. On August 15, 2011, employees were notified that the experiment had ended and Ctrip would roll out the experiment to those who were qualified and wanted to work at home in the Airfare and Hotels divisions on September 1st, 2011.

Throughout the experiment, employees were told the experiment would be evaluated to guide future company policies, but they did not learn the actual policy roll-out decision until August 15th. Because of the large scale of the experiment and the lack of dissemination of experimental results beyond the management team, employees were uncertain as to the long-run decision of the firm on roll-out prior to the decision. Employees in the treatment group who wished to come back to work in the office full-time were only allowed to do so after August 15th, while control workers had to stay in the office for the full duration of the experiment. Hence, the treatment and control assignments were fixed for the full 9 months.

Figure 2 shows compliance with the experiment throughout the experimental period, and after the general roll-out through May 2012. During the experiment, the percentage of treatment group working at home hovered between 80% and 90%. The compliance did not reach 100% because the broadband speed was sometimes not fast enough for home working, or because employees moved apartments and lost access to their own room.[4] Since compliance was not perfect, our estimators take even birthdate status as the treatment status, so we estimate an intention to treat result. Given we are interested in evaluating the impact of a policy of allowing home-working, this seems appropriate.

After the experiment, we see about 50% the treatment group immediately decided to return to the office. They do this despite having to incur the financial and time costs of commuting, with the main reason given for this being the loneliness of working from home. Strikingly, only about 35% of the control employees – who also all initially volunteered to work from home – actually move home when they were allowed to after the end of the experiment. Again, the main reason they gave for changing their mind is concerns over being lonely at home. Finally, we also see that about 10% of the workers that did not initially volunteer changed their minds after the experiment and decided to work from home.

Interestingly, the firm's management was surprised by about two things in these numbers. First, over how many employees changed their minds about working from home. About 50% of the

[3] It was important to have this draw in an open ceremony so that managers and employees could not complain of "favoritism" in the randomization process. The choice of odd/even birthdate was made to ensure the randomization was straightforward and transparent.

[4] In all estimations, we use the even birthdate as the indicator for working-at-home so these individuals are treated as home workers. In a probit for actually working from home during the experiment, none of the observables are significant, suggesting that returning to the office was effectively random. One reason is that the IT group policed this heavily to prevent employees fabricating stories to enable them to return to the office.

volunteer group and 10% of the non-volunteer group switched preferences after the 9-month experiment. This matches the anecdotal evidence from home-working companies in the US like Arise Virtual Solutions, which reports that monthly turn-over rates for new home-worker employees are about 80% in the first month, but rapidly drop to about 5% after six months, consistent with the sorting after initial experimentation by new employees.

Second, despite the time and financial savings from having no commute, more than half of the eligible workers decided to work from the office, suggesting they place a high value on social interactions at work (Hamermesh, 1990). This is particularly striking because as we note below we find no negative impact of home working on any other outcomes like call quality or promotions.

II.C. Data Collection

Ctrip has an extremely comprehensive central data collection system, as its founding team came from Oracle with extensive database software experience. The majority of data we use in our paper are directly extracted by from the firms' central database, providing extremely high data accuracy. The data we collected can be categorized in 6 fields: performance, labor supply, attrition, promotions, reported employee work satisfaction, detailed demographic information and surveys on attitudes towards the program.

Performance measures vary by the broad type of workers – 137 order takers and 118 order placers (details in Appendix 1). Order takers' main tasks are to answer phone calls and record travel and hotel orders in the Ctrip system. Their key performance measures are the number of phone calls answered and number of orders taken. Order placers process the orders by contacting the hotels and airlines, then notifying clients of confirmed reservations. Their key measures are the numbers of different types of confirmation phone calls made.

For order takers, we can also accurately measure time spent working in terms of minutes on the phone. We have logs of phone calls and call lengths from the central database of Ctrip. The firm also uses this measure to monitor the work of their employees. We also calculate phone calls answered by minutes on the phone as a measure of labor productivity for this type of workers.

We have daily key performance measures of all employees in the airfare and hotel booking departments from January 1st, 2010 onwards. We also have daily minutes on the phone for order takers during the same period. We also have daily records of hours of leave for the airfare department, and the date and reason of employees in the experiment quitting the firm. The firm also ran internal surveys of the employees during the experiment on work exhaustion, positive and negative attitudes (see details in Appendix A2). Finally, we conducted two rounds of surveys in November 2010 and August 2011, to collect detailed information on all the employees in the two departments including basic demographics, income, attitudes toward the Program.

III. IMPACT ON THE FIRM

We analyze the effect of home-working in terms of impact on the firm, which we cover in this section, and the impact on the employees, which we cover in the next section.

III.A. Individual Employee Performance

We start by estimating the intention to treat impact on employee performance via equation (1)

$$\text{EMPLOYEE\_PERFORMANCE}_{i,t} = a\text{TREAT}_i \times \text{EXPERIMENT}_t + b_t + c_i + e_{i,t} \tag{1}$$

where TREAT is a dummy variable that equals 1 if an individual belongs to the treatment group defined by having an even-numbered birthday; EXPERIMENT is a dummy variable that equals 1 for the experimental period December 6th to August 15th; and EMPLOYEE_PERFORMANCE is one of the key measures of work performance, including an overall performance z-score measure, log of weekly phone calls answered, log of phone calls answered per minute on the phone, and log of weekly sum of minutes on the phone. Finally, b_t incudes a series of week dummies to account for seasonal variation in traveling demand such as the World Expo in 2010 and the Chinese New Year, and c_i are a full set of individual fixed effect.

To make performance of different types of workers comparable, we use performance z-scores. First, we generate the weekly sum of key measures of performance for each type of workers. For example, order takers have two key measures of performance - phone calls answered and orders placed. To obtain z scores of each key measure, we subtract the weekly sum by the pre-experiment mean by department of the key measure, and divide it by the pre-experiment standard deviation. Then, we average the key measure z-scores within each type to generate an overall performance z-score measure. Finally, we normalize this measure again by dividing by the pre-experiment standard deviation to create the final double z-scored overall performance measure. This measure has mean 0 and standard deviation 1 over the pre-experiment period for each type of worker.

In column (1) of Table 2, overall performance of the treatment group is found to be 0.226 standard deviations higher than the control group after the experiment started, significant at the 1% level. The largest type of workers we have in our sample are the 137 order takers. If we limit the sample to the order takers, we can use phone calls answered as the key performance measure for all the order takers. The z-scores of phone calls account for different volume and average length of phone calls in two departments. In column (2), we look at just the phonecalls performance and find this is 0.263 standard-deviations better for the treatment group. In column (3), we look at the log of phonecalls and find these are 0.122 higher, so that treatment employees were making 13% (noting that 13%=exp(0.122)) more phone-calls.

We can also see these results in Figure 3a where we plot the raw number of phone calls per week for the treatment and control groups from Jan 1st 2010 until the end of the experiment in August 15th 2011. Before the experiment started, the treatment group trends closely together with the control group, both of which bounce around due to seasonal fluctuations in demand. But once the experiment begins, the treatment group starts to outperform the control group, answering about 40 more phone-calls per week. Figure 3b plots the cross-sectional distribution of performance for treatment and control groups at 3 months, displaying a broad distributional improvement from working-from-home (rather than the results being driven by a few outliers).

We further decompose the difference in performance observed in column (3) into phone calls answered per minute on the phone (a measure of productivity), and minutes on the phone (a measure of high-frequency labor supply). In column (4), we find treatment employees are making 3.4% (note that 3.4%=exp(0.033)) more phone calls per minute, which is primarily driven by the fact that working from home is quieter. The home workers told us this meant it was easier to hear the customers, so they did not have to ask them to repeat themselves as often and could process the information more quickly. This matches the psychology literature which has shown that background office noise can reduce cognitive performance (see, for example, Banbury and Berry, 1998).

But, the biggest factor increasing the home-workers performance is that, as shown in column (5), they work 9.4% (9.4%=exp(0.089)) more minutes per day. This is despite the fact that home and office workers both work the same nominal shift –for example 9am to 5pm on Monday to Friday – as they both work in the same team under the same team manager. The reason home-workers can increase minutes on the phone is within their shift they are logged on for more time, meaning they are taking less time-off during their shift.

III.B. Individual Employee Labor Supply

In Table 3, we investigate the factors driving this increase in minutes worked within each shift. Because we have accurate records of hours of leave from the airfare booking department only, we limit the sample further to 89 order takers in the airfare department. Column (1) repeats the results from the final column of Table 2, while Column (2) of Table 3 shows that these order takers show a very similar increase to of the full group.

Columns (3)-(5) break this difference in minutes on the phone down into three buckets based. In column (3), we look at whether treatment workers spend more minutes on the phone per hour logged in[5], column (4) looks at whether they are logged in for more hours per day worked, and column (5) looks at whether they work for more days.

What we find is that in column (3), there is no difference between the number of minutes on the phone while logged-in for the treatment and control employees. This is not surprising because both groups operate using the same call routing server and on the same queuing system.[6]

Column (4) shows that about two thirds of the difference in the time on the phone is accounted for by home-workers logging in for more hours per day worked. This is because: (a) they start work more punctually and leave early less often since they avoid commuting delays from events like the heavy snow in Shanghai in February 2011; and (b) they take shorter lunch breaks because they are usually eating on their own rather than with colleagues in the canteen. Finally, in column (5) we see that the other third of the difference in time worked between treatment and

[5] Note that sometimes employees will not be taking calls when they are logged in if demand is low, so that time on logged in and time on the phone are not necessarily the same (the former is higher when demand is low).

[6] Moreover, it shows that home-workers are not picking busier times to log-in to the system (i.e. they are not timing their breaks to coincide with quiet periods when demand is lower). I thank Wouter Dessein for pointing this out.

control is because treatment employees work more days because they take less sick-days. Employees explained this was because they worked at home when they felt mildly ill but would not have felt up to commuting into work.

Hence, in summary, avoiding the daily commute, taking shorter breaks, and avoiding days-off enables home-working employees to increase their hours by about 9.5% (while also being more productive per hour worked). Interestingly, while we define this as an increase in labor supply, in most census data this would typically end-up being allocated to an increase in productivity, as Census data on hours data typically only records mandated shift hours rather than actual working hours. Hence, if we were to examine CTrip using the type of US census data used to generate productivity numbers by, for example Foster, Haltiwanger and Syversson (2008), we would record the 13% increase in output from WFH as a 13% increase in total-factor productivity. This highlights how even management practices like working-from-home that may not primarily impact worker-level productivity can help to explain potentially large differences in measured firm-level productivity.

III.C. Quality and Spillovers

One question is that whether quality of the service has been compromised as a tradeoff for the increase in productivity in the treatment group. We construct two quality measures: conversion rate and weekly recording scores. Conversion rate is calculated as the percentage of phone calls answered resulting in orders, while the weekly recording scores comes from the 1% of phone-calls that are randomly evaluated by an external monitoring team. In summary (with the full details in table A3 in the appendix), we find no impact of working from home on call quality using either measure.

Another related question is whether the improvement in working from home comes from an improvement in the treatment group or a deterioration in the control group. Maybe the gap between treatment and control is caused not by the treatment group performing better but by the control group performing worse after they "lost" the randomization lottery? The group winning the treatment lottery saved themselves 9 months of commuting time and costs, a substantial gain worth about 17% of their salary evaluated at their CTrip wage rate.[7]

In Table 4, we collect data on two other "quasi" control groups to answer this question. The first group is the eligible employees in the Nan Tong call center. This is CTrip's other large call center, located in Nan Tong, a city about 1 hour drive outside of Shanghai. This call center also has airfare and hotel departments, and calls are allocated across the Shanghai and Nan-Tong call centers randomly from the same central server. The second group is the 253 eligible employees that did not volunteer to participate in the WFH experiment in the Shanghai call center. These are the individuals that were eligible for the experiment (own room, 6+ months of tenure and broadband), but did want to work from home. We think these two groups are comparable to the treatment and control groups for two reasons. First, all four groups face the same demand for their service. Second, they all meet the requirements for eligibility to participate in the experiment.

[7] The average employee makes about $100 per week for a 40 hour week. The commuting time is 40 minutes each way and cost $0.5 on average. Hence, the saving in time is about $13 a week in costs about $4 per week.

Figure 4 shows that, first, the performance of the eligible group in the Nan Tong call center tracks that of the treatment and control well before the experiment. After the experiment started, the performance of the Nan Tong group is similar to that of the control group. Results in the top panel of Table 4 confirm this finding in a regression setting. Differences in overall performance, efficiency and labor supply between the control group and the Nan Tong eligible group is statistically insignificant from zero. The bottom panel compares treatment and control group to the eligible non-experimental group in Shanghai. Again, we find no difference between the control group and the eligible non-experimental group. These results suggest that the gap between the treatment and control group reflects an improvement in the performance of the treatment group rather than any deterioration of the control group. That is, although the control group and the treatment group work in the same team, we find little evidence of the control group being discouraged by losing the working-from-home lottery.

We also look for spillovers by examining the variation in the number of individuals randomly assigned to treatment across the groups within the Shanghai office. Because groups are small, random variations in the number of employees with even and odd birthdays generates variations in the number of employees who get to work at home. We use this (the share of evens in the eligible volunteered group) to instrument for the share of all employees working from home, and investigate the impact of this on the team's performance. As we show in Table A4, we again find no evidence for spillovers across individuals from home-working.

III.D. Post-Experiment Selection

In August 2011, the management evaluated the experiment to have saved about $2,000 per home-working employee, and decided to immediately roll-out the experiment to the whole Airfare and Hotel division. So, employees in these divisions were notified that the experiment had ended and they were entitled to choose their location of work – control employees that still wished too could move home, and treatment employees that wanted to return to the office could do so.

As shown in Figure 5 – which plots the difference in normalized phone-calls between home and office workers - post-experiment selection substantially increased the performance increase from working from home. The differential increase in phonecalls (versus the pre-experiment baseline) from home working was about 0.2 standard-deviations during the experiment, rising to about 1 standard deviation within 6 months after the experiment.

This dramatic increase in the impact of working from home after the roll-out of the experiment was driven by treatment workers that had performed relatively badly at home returning to the office. This is shown in columns (1) to (4) of Table 5, which runs a probit on whether a treatment worker returns to the office. The results show that treatment workers that performed relatively worse at home versus the office returned to the office. This is despite the fact that all treatment workers had initially volunteered to work from home, suggesting that many of them subsequently discovered home working was not as beneficial for performance as they initially believed.

In columns (5) to (8), we examine the reverse flow of control workers moving home and discover no correlation with prior office performance. This is consistent with Table 1 that the decision to move home is not driven by *absolute* performance in the office (but is driven by *relative* office-vs-home performance as shown in columns (1) to (4)).

IV. IMPACT ON THE EMPLOYEES

III.A. Employee' self-reported outcomes

Ctrip management was also interested to find out how employee self-reported well-beings are impacted by the Program. They ran two sets of surveys: the satisfaction survey and the emotion survey. Details of survey questions and methodology are listed in Appendix A2, but in summary these are standard employee satisfaction tests developed by Christina Maslach and Susan Jackson in the 1970s (see for example Maslach and Jackson, 1981). The satisfaction survey was conducted five times throughout the experimental period. Once in early November before the randomization took place and four times after the experiment had started. Because the employees were unaware of the assignment at the initial survey, the first survey is a credible baseline. The first three columns of Table 6 show three different satisfaction measures. The treatment group reports no different satisfaction level from the control group at the first survey, but the treatment group reports statistically significantly higher satisfaction level throughout the experiment.

The emotion survey is conducted every week. The first week was conducted in late-November 2010, before the experiment began but after the randomization so that individuals had been informed of their status in the treatment or control groups. Interestingly, the treatment group already reports higher positive attitude (significant at the 10% level), less negative attitude and less exhaustion from work. This group has yet to move home, so this difference is entirely due to the control group learning they lost the randomization, and highlights the importance of comparing our treatment groups with other controls groups like Nan-tong and the non-volunteer group. After starting the experiment, the gap between the treatment and control group rose further, so that the treatment group reported statistically significantly higher positive attitude and less work exhaustion as their total work plus commute time was lower on average than the control group.

IV.B. Attrition

One of the key initial reasons Ctrip was interested in running the experiment was to retain workers. The turnover rate among Ctrip call center representatives has historically hovered around 50% per year, which is typical of the call center industry in China[8]. Management estimates that hiring and training a representative costs on average $2000, about 6 months' salary of an average employee. Figure 6 plots the cumulative attrition rate of treatment and control group separately over the experimental period. Shortly after the commencement of the experiment, cumulative attrition rates diverged between the two groups and the difference is statistically significant. By the end of the experiment, attrition rate in the treatment group (17%) is nearly half of that in the control group (35%).

---

[8] Note that CTrip can in principle fire employees, but this is rare and no employees in these two divisions were fired over this period as far as we are aware.

We further test whether selective attrition exists by running probit regressions in Table 7. The dependent variable is whether an employee quits the job during the experimental period between December 6^{th} 2010 and August 15th 2011. Column (1) in Table 5 confirms the finding in Figure 6, that treatment employees rate of attrition is about half that of the control group. In columns (2), we test whether employees with worse performance are more likely to attrite in the treatment group compared to the control group, but find no supporting evidence. Not surprisingly, we do find, however, that younger employees and those with higher commuting costs are more likely to quit.

In column (3), we use the same specifications as in column (2), but replace the pre-experiment performance with post-experiment performance. Post-experiment performance is the average of individual weekly performance z-scores during the post-experimental period from December 6^{th} 2010 to August 15th 2011. We find that low performers are significantly more likely to quit, particularly those in the control group. In columns (4) and (5), we estimate the impact of experimental period performance on quitting in the treatment and control groups separately and find only a significant impact for the control groups. From interviewing the employees, we heard that control employees who underperformed tended to quit for other call-center that they believe would pay better. Treatment employees, however, were much less likely to quit because no other home-working jobs existed, substantially reducing selection from the treatment group.

This differential attrition, of course, also raises the question of whether our estimated impact of WFH is biased. To address this issue, we use the Lee (2010) bounds estimator. This provides upper and lower bounds on the differential selection on performance across groups, assuming that selection into the control group monotonically increases attrition. This allows us to generate two bounds – the upper bound that assumes that the extra attrition in the control group is based on a negative correlation between performance (as we saw in Table 7) while the lower bound assumes a positive correlation (the reverse of what we see in Table 7, but included for completeness). We see that the upper bound lies above the actual treatment-control estimated impact, suggesting that the actual treatment impact is if anything larger than we expected, because the attrition of the worst performers from the control group biases our results down.

IV.C. Promotions

One possible negative impact from working at home is that long-run career performance is damaged by less “face-time” in the office, making it harder for home-based workers to achieve a promotion. To investigate this, we collected promotions data on the 255 employee experimental sample. In summary, during the period from the start of the experiment in December 2010 until May 2012, a total of 8 employees from the treatment group received promotions and 6 from the control group. Neither this raw difference nor the coefficient on treatment in promotion probits including or excluding demographic controls was significant. Thus, at least over the period of 18 months from the beginning of the experiment until May 2012, we find no negative impact of working from home 4 days a week on employees’ ability to get promoted.

V. FIRM LEARNING AND ROLL-OUT

One of the most interesting aspects of the experiment was the learning process for both the firm and the individual employees on the costs and benefits of working from home. Both groups were initially unsure about its impact as a practice given this had never previously been adopted by other Chinese travel-agents or call-centers. However, fortunately we were able to monitor their learning over the course of the experiment because of our extensive access to the CTrip's management team, and frequent employee surveys and interviews.

V.A. Firm learning

The firm saw working from home as a way to save on office costs and reduce employee turnover, but was initially worried about its impact on employee performance. They feared employees would shirk at home or that call quality would as employees multi-task on other activities which are prohibited in the office (like playing computer games, watching TV or using the internet). Running the experiment revealed, however, that working from home actually generated an improvement in employee performance, worth about $375 per employee. This was evaluated using the 13% performance improvement from the Table 3 intention to treat estimates, and would be several times larger if evaluated using the longer-run impact shown in Figure 5 selection effects. In addition, they estimated office cost savings of about $1250 per employee and reduced turnover savings of about $400 per employee. Hence, given the saving of at least $2000 per employee, the firm rolled the program out in August 2011, accompanied by an aggressive poster campaign to persuade employees to take-up the home working.

Interestingly, the firm learnt three important results from running the formal experiment versus the non-randomized pilot that they were initially considering. First, they learned that working-from-home improves performance. Without running a formal experiment, their view is they could have interpreted the drop in treatment performance shown in Figure 3 as a negative treatment effect. The period of the experiment (December 2010 to August 2011) coincided with a business slow-down for CTrip due to a combination of the (predicted) end of Shanghai Expo 2010 and an (unpredicted) increase in competition from other travel agency firms. As a result, the *difference* in performance for the treatment group was negative, and is only positive when evaluated as a *difference of differences* against the control group. This highlights the importance of having a well matched (ideally randomized) control group to strip out these kinds of seasonal effects.

Second, ex ante there was very little discussion of selection effects on employee performance, but by running the experiment and then rolling this out it is clear that allowing employee choice generates a far higher impact than requiring work from home. The impact of working from home is positive, on average, but appears to have a large variance, so that employee choice leads to a much higher impact as shown in Figure 5.

Finally, having the large sample of treatment and control employees allowed the firm to evaluate the impact on different types of employees. Somewhat surprisingly, they found a very homogeneous impact across all types of employees. For example, in Figure 7, we plot the impact on the top half of the treatment versus control distribution and the bottom half of the treatment vs control distribution. To calculate this, both groups were split in half by the pre-experiment median performance and then compared. What we see is a similar improvement in performance for home working for both groups. CTrip's ex-ante expectation was that the bottom-half of employees were the less motivated employees, and would perform far worse at home. Table A5 shows a similar result that the impact of working-from-home was homogeneous across a wide range of other characteristics like gender, commute time, age, prior experience and living arrangements. These results have led the firm to offer working-from-home to all employee groups going forwards rather than any selected sub-samples (such as high-performers), which they were initially intending to target.

V.B. Employees' learning

One direct measure of the extent of employee learning is the number of employees that changed their minds over working from home. Figure 2 shows that after the experiment about 50% of the initial treatment and control volunteers changed their minds and decided to work in the office after the end of the experiment, while 10% of the initial non-volunteer group opted to work from home.

We also designed a survey to inquire into employees' evolving views toward the Program from across all 996 Airfare and Hotel division employees. We administered the same survey with the help of the Ctrip management in November 2010 and August 2011. Employees are asked specifically whether they are interested in participating in the Work-at-Home Program if they were eligible. They can choose from three answers: yes, no or undecided.

In Table 8, we tabulate employees' answers in November 2010 against August 2011. The sample includes 568 employees who answered both surveys. In November 2010, 51% of the employees are willing to work at home, compared to 40% in August 2011. We find 47% of the employees changed their positions across the two surveys, evidenced by the weight on the off-diagonals. About 20% of those who answered yes in the first survey decided they were not interested in the second survey while 12% of those who initially were not interested showed interest in the second survey.

VI. CONCLUSIONS

The frequency of working from home has been rising rapidly in the US, with over 10% of the work-force now reporting regular home working. But there is uncertainty and skepticism over the effectiveness of this, highlighted by phrases like "shirking from home". We report the results of the first randomized experiment on home-working, run in a 13,000 employee NASDAQ listed Chinese firm. Employees that volunteered to work from home were randomized into 9-months of home-working by even/odd birth-date. We find a highly significant 13% increase in performance from home-working, of which 9% is from working more minutes of their shift period (fewer breaks and sick-days) and 3.5% from higher performance per minute. We find no negative spillovers onto workers left in the office. Home workers also reported substantially higher work satisfaction and psychological attitude scores, and their job attrition rates fell by over 50%.

This experiment highlights how complex the process of learning about new management practices is. For the CTrip, having no precedent in terms of similar Chinese firms that had adopted working from home for their employees led them to run this extensive field experiment. Given their success, other firms are now likely to copy this, generating the type of gradual adoption of a new management practice that Griliches (1957) highlighted. More generally, given the large impact of this practice on employee performance – a 13% direct increase in output, with potentially longer run increases from selection and reduction in capital from reductions in office-space – this also provides a management-practice based explanation for heterogeneous firm performance.

V. BIBLIOGRAPY (to be completed)

Banbury, Simon and Berry, Dianne. (1998). “Disruption of office related tasks by speech and office noise”. British Journal of Psychology, 89, 494-517.

Bloom, Nick and Van Reenen, John, (2011), “Human resources and management practices”, Handbook of Labor Economics, Volume 4, edited by Orley Ashenfelter and David Card.

Bloom, Nick, Tobias Kretschmer and John Van Reenen, (2009), “Work-life Balance, Management Practices and Productivity’, in International Differences in the Business Practice and Productivity of Firms, Richard Freeman and Kathy Shaw (eds). Chicago: University of Chicago Press.

Foster, Lucia, John Haltiwanger and Chad Syverson (2008), “Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?” American Economic Review, 98(1), 394-425

Council of Economic Advisors (2010), “Work-life balance and the economics of workplace flexibility”, http://www.whitehouse.gov/files/documents/100331-cea-economics-workplace-flexibility.pdf

Griliches, Zvi (1957), “Hybrid Corn: An Exploration in the Economics of Technological Change”, Econometrica, volume 25 (4), pp. 501-522.

Hamermesh, Daniel (1990), "Shirking or Productive Schmoozing: Wages and the Allocation of Time at Work," Industrial and Labor Relations Review, January

Lazear, Edward, and Paul Oyer (2012), “Personnel Economics,” in Robert Gibbons and John Roberts, eds. *Handbook of Organizational Economics*, Princeton University Press, forthcoming.

Lee, David (2008), “Training, wages and sample selection: estimating sharp bounds on treatment effects”. *Review of Economic* Studies, 76(3), pp 1071-1102.

Leibenstein, H. (1966) “Allocative Efficiency vs." X-efficiency," *American Economic Review, 56(3): 392-415*

Maslach, C., & Jackson, S.E. (1981). Maslach Burnout Inventory. Research edition. Palo Alto, CA: Consulting Psychologist Press.

Oettinger, Gerald (2012), “The Incidence and Wage Consequences of Home-Based Work in the United States, 1980-2000”, *Journal of Human Resources* forthcoming

Syverson, Chad (2011), ‘’What Determines Productivity at the Micro Level?”, forthcoming *Journal of Economic Literature*.

DATA APPENDIX

Appendix A1: Table for different types of workers and their key performance measures

Types of Workers	Department	Key Performance Measures	Number of Workers
Order Takers	Airfare	Phone Calls Answered Orders Taken	89
	Hotel		48
Order Placers	Airfare	Notifications Sent	46
	Hotel	Reservation Phone Calls Made	25
Order Correctors	Hotel	Orders Corrected	36
Night Shift Workers	Hotel	Reservation Phone Calls Made Orders Corrected	11

Appendix A2: Explanations on the Work Satisfaction Survey

Work Exhaustion: CTrip's in-house psychology counselors use an adapted excerpt from the Maslach Burnout Inventory Survey to measure the emotional exhaustion of the employees from work. The MBI survey was developed by Berkeley psychologist Christina Maslach and Susan Jackson in the 1970s (see Maslach and Jackson, 1981).

Each employee is asked to evaluate his or her "emotional exhaustion" at the end of the work week. The survey contains 6 questions. Each employee is asked to report how often he has felt the way described at work during the week: feel this way every day, almost all the time, most of the time, half of the time, a few times, rarely, never. The survey questions are listed below:

1. I feel emotionally drained from my work.
2. I feel used up at the end of the work day.
3. I dread getting up in the morning and having to face another day on the job.
4. I feel burned out from my work.
5. I feel frustrated by my job.
6. I feel I am working too hard on my job.

Positive and Negative Attitudes: CTrip's in-house psychology counselors use an adapted 16-item Positive and Negative Affect Schedule (PANAS) developed by Clark and Tellegen in 1988 to measure the positive and negative attitudes of the employees.

The survey comprises two mood scales, one measuring positive affect and the other measuring negative affect. Each item is rated on a 5-point scale ranging from 1 = *very slightly or not at all* to 5 = *extremely* to indicate the extent to which the employee feels this way the day he takes the survey. To evaluate the positive affect, psychologists sum the odd items. In cases with internally missing data (items not answered), the sums were computed after imputation of the missing values: # items on scale / # actually answered, multiplied by the sum obtained from the answered items. A higher score indicates more positive affect, or the extent to which the individual feels enthusiastic, active, and alert. The negative affect is evaluated similarly by summing up the even items.

The 16 items are listed below. Cheerful, Jittery, Happy, Ashamed, Excited, Nervous, Enthusiastic, Hostile, Content, Guilty, Relaxed, Angry, Proud, Dejected, Active, Sad

Appendix A3: Quality did not change in the experiment

	(1)	(2)	(3)	(4)
Dependent Variable	recording grade	recording grade	conversion (z score)	conversion (z score)
Individual FE	No	Yes	No	Yes
Week fixed-effects	Yes	Yes	Yes	Yes
Experiment*Treatment	-0.007	-0.006	-0.026	-0.026
	(0.008)	(0.008)	(0.071)	(0.065)
Treatment	0.000		-0.011	
	(0.005)		(0.091)	
Number of Employees	89	89	135	135
Number of Weeks	87	87	87	87
Observations	5689	5689	9815	9815

Notes: Sample in the first two columns includes 89 order takes in the airfare department (for which we can obtain recording grade information). The sample in the last two columns includes 135 order takers in airfare and hotels (the group for which conversion rate data exists). Clustered standard errors. *** denotes 1% significance, ** 5% significance and * 10% significance.

Appendix A4. Lack of any obvious cross-sectional Spillover effects

	(1)	(2)	(3)	(4)
Dependent variable	Overall Performance	Overall Performance	Overall Performance	Overall Performance
Sample	Non-experiment	Control	Treatment	Non-experiment + Control
Specification	IV	IV	IV	IV
Treat/Total	-0.221	-0.574	-0.523	-0.263
	(0.398)	(0.392)	(1.039)	(0.357)
Week FE	Yes	Yes	Yes	Yes
No. of Teams	79	59	56	79
Observations	36660	8218	9587	44846
R-squared	0.410	0.359	0.467	0.398

	IV first stage	IV first stage	IV first stage	IV first stage
Sample	Non-experiment	Control	Treatment	Non-experiment + Control
Dependent variable	Treat/Total	Treat/Total	Treat/Total	Treat/Total
Treat/(Treat+Control)	0.253***	0.390***	0.219***	0.264***
	(0.0226)	(0.0295)	(0.0484)	(0.0236)
Week FE	Yes	Yes	Yes	Yes
No. of Teams	79	59	56	79
Observations	36660	8218	9587	44846
R-squared	0.881	0.903	0.891	0.874

Notes: "Treat/total" is the number of employees in treatment divided by the number of employees in each team. A team is composed of 10 to 20 employees who specialize in the same type of tasks and work the same schedule of shifts. Each team is monitored by the same team leader. "Treat/(Treat+Control)" is the number of employees in treatment divided by the number of employees in treatment and control group within each team. Both "Treat/total" and "Treat/(Treat+Control)" are set to zero before the experiment started on December 6, 2010. "Treat/(Treat+Control)" is fixed at the beginning of the experiment. "Non-experiment", "Control" and "Treatment" refer to employees from each group. The sample includes data from January 1, 2010 to August 15, 2011. Clustered standard errors. *** denotes 1% significance, ** 5% significance and * 10% significance.

Appendix A5. Panel A: Treatment Effects Seem Homogeneous across Characteristics

Performance	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(12)
	Child	Female	Commute >120min	renter	young	short prior experience	short tenure	live w/ parents	live w/ spouse	live w/ friends	pre-exper performance
experiment x treat x "characteristic"	0.0788	-0.106	0.155	-0.111	-0.0430	0.0559	-0.0544	0.0127	-0.0132	-0.126	0.0963
	(0.184)	(0.130)	(0.146)	(0.148)	(0.132)	(0.134)	(0.135)	(0.141)	(0.178)	(0.253)	(0.104)
experiment x	-0.0395	0.105	-0.0612	0.0764	0.00864	0.0493	0.0730	0.0171	-0.0244	0.213	-0.312***
"characteristic"	(0.133)	(0.0919)	(0.0955)	(0.109)	(0.0946)	(0.0973)	(0.0971)	(0.105)	(0.120)	(0.210)	(0.0812)
experiment x	0.216***	0.278***	0.171**	0.243***	0.249**	0.208**	0.257**	0.208*	0.217***	0.223***	0.221***
Treatment	(0.0711)	(0.101)	(0.0831)	(0.0781)	(0.106)	(0.0974)	(0.112)	(0.120)	(0.0688)	(0.0692)	(0.0616)
Observations	17611	17611	17603	17526	17611	17611	17611	17526	17526	17526	17611
R-squared	0.417	0.417	0.416	0.416	0.417	0.417	0.417	0.416	0.416	0.417	0.423

Notes: The performance z-scores are constructed by taking the average of normalized performance measures (normalizing each individual measure to a mean of zero and standard deviation of 1 across the sample). The sample includes data from January 1, 2010 to August 15, 2011. "young" equal 1 if an employee is under 24. "Short prior experience" equals 1 if an employee with less than 6 months of experience before joining Ctrip. "Short tenure" equals 1 if an employee has worked in Ctrip for less than 24 month by December 2010. "Pre-exper performance" is the average z-score of performance between Jan 1, 2010 and Oct 1, 2010 for each employee. Clustered standard errors. *** denotes 1% significance, ** 5% significance and * 10% significance.

Table 1. Characteristics of employees who volunteer to join WFH

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	Sample mean
Children	0.123**		0.075	0.065	0.084	0.090		0.092	0.09
	(0.055)		(0.081)	(0.080)	(0.080)	(0.079)		(0.080)	
Married		0.095**	0.054	-0.002	0.039	0.037		0.040	0.15
		(0.044)	(0.063)	(0.064)	(0.064)	(0.064)		(0.065)	
Cost of commute (Yuan)				0.005**	0.005**	0.005**		0.005**	5.54
				(0.002)	(0.002)	(0.002)		(0.002)	
Bedroom				0.097***	0.089***	0.090***		0.092***	0.60
				(0.032)	(0.032)	(0.033)		(0.034)	
Tertiary education and above					-0.087***	-0.090***		-0.089***	0.41
					(0.032)	(0.032)		(0.032)	
Tenure (months)					-0.002***	-0.003***		-0.003***	24.9
					(0.001)	(0.001)		(0.001)	
Gross wage (Yuan)						-0.003	-0.019	0.032	2872
						(0.001)	(0.017)	(0.023)	
Age								-0.001	23.2
								(0.007)	
Male								0.000	0.32
								(0.035)	
Number of Employees	996	996	996	996	996	996	996	996	996

Notes: The total sample covers all CTrip employees in their Shanghai Airfare and Hotel group. Willingness to participate is based on the initial survey in Nov 2010. Employees were not told the eligibility rules in advance of the survey (i.e.: own room, 6+ months tenure, internet connect etc). Gross wage is calculated as a monthly average of salary from Jan 2010 to Sep 2010 (note that 1 Yuan is about 0.15 Dollars). The t-stat in the second column tests whether differences between volunteered employees and all employees are significant, while those in the last column tests whether differences within the volunteered group between eligible and all employees are significant.

Table 2: The performance impact of working from home

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	Overall Performance	Phonecalls	Phonecalls	Phonecalls Per Minute	Minutes on the Phone
Dependent Normalization	z-score	z-score	log	log	log
Period: 11 months pre-experiment and 9 months of experiment					
Experiment*Treatment	0.226***	0.263***	0.122***	0.033**	0.089***
	(0.064)	(0.064)	(0.026)	(0.013)	(0.028)
Number of Employees	255	137	137	137	137
Number of Weeks	85	85	85	85	85
Observations	17778	9503	9503	9503	9503

Notes: The regressions are run at the individual by week level, with a full set of individual and week fixed effects. Experiment*treatment is the interaction of the period of the experimentation (December 6th 2010 until August 20th 2011) by an individual having an even birthdate (2nd, 4th, 6th, 8th etc day of the month). The pre period refers to January 1st 2010 until December 5th 2010. The z-scores are constructed by taking the average of normalized performance measures (normalizing each individual measure to a mean of zero and standard deviation of 1 across the sample). Since all employees have z-scores but not all employees have phonecall counts (because for example they do order booking) the z-scores covers a wider group of employees. Minutes on the phone is recorded from the call logs. Standard errors are clustered at the individual level. *** denotes 1% significance, ** 5% significance and * 10% significance.

Table 3: Decomposition of the change in labor supply

VARIABLES	(1) Minutes on the Phone	(2) Minutes on the Phone	(3) Minutes on the Phone/ Hours Worked	(4) Hours Worked/ Days Worked	(5) Days Worked
Sample	All	Airfare	Airfare	Airfare	Airfare
Period: 11 months pre-experiment and 9 months of experiment					
Experiment*Treatment	0.089***	0.090**	-0.017	0.068**	0.039**
	(0.028)	(0.044)	(0.033)	(0.028)	(0.015)
Number of Employees	137	89	89	89	89
Number of Weeks	85	85	85	85	85
Observations	9,503	3531	3531	3531	3531

period of the experimentation (December 6th 2010 until August 20th 2011) by an individual having an even birthdate (2nd, 4th, 6th, 8th etc day of the month). The pre period refers to January 1st 2010 until December 5th 2010. Only employees in the Airfare group provides full holiday and leave data so the breakdown by hours and days in the office is only undertaken for this group. Standard errors are clustered at the individual level. *** denotes 1% significance, ** 5% significance and * 10% significance. Minutes on the phone is recorded from the call logs. Hours worked is measured by the phone system log-in and log-out data.

Table 4: The treatment performance also looked good benchmarked against non-experimental and Nantong employees

VARIABLES	(1) Overall Performance	(2) Overall Performance	(3) Phone calls	(4) Phone calls
Comparison to Nan Tong				
	Treatment Vs. Nan Tong	Control Vs. Nan Tong	Treatment Vs. Nan Tong	Control Vs. Nan Tong
Experiment*treatment	0.191***		0.241***	
	(0.047)		(0.049)	
Experiment*control		-0.032		-0.032
		(0.048)		(0.044)
Observations	92181	90825	83242	81770
Comparison to Eligible Non-experiment group				
	Treatment Vs. Non-experiment	Control Vs. Non-experiment	Treatment Vs. Non-experiment	Control Vs. Non-experiment
Experiment*treatment	0.209***		0.198***	
	(0.049)		(0.052)	
Experiment*control		-0.021		-0.06
		(0.056)		(0.047)
Observations	48542	47186	31032	30278

Notes: Nan-Tong is CTrip's other large call center, located in Nan-Tong, a city about 1 hour drive outside of Shanghai. This call center also has airfare and hotel departments, and calls are allocated across the Shanghai and Nan-Tong call centers randomly. The "Eligible non-experimental group" are the individuals that were eligible for the experiment (own room, 6+ months of tenure and broadband) but did not participate in the two departments in Shanghai. The regressions are run at the individual by week level, with a full set of individual and week fixed effects. Experiment*treatment is the interaction of the period of the experimentation (December 6th 2010 until August 20th 2011) by an individual having an even birthdate (2nd, 4th, 6th, 8th etc day of the month), while Experiment*control is the interaction of the period of the experimentation by an individual having an odd birthdate. All performance measures are z-scores (constructed by taking the average of normalized performance measures, where these are normalizing each individual measure to a mean of zero and standard deviation of 1 across the sample). Standard errors are clustered at the individual level. *** denotes 1% significance, ** 5% significance and * 10% significance.

Table 5: Employee switches after the end of the experiment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Switch	Home to Office	Home to Office	Home to Office	Home to Office	Office to Home	Office to Home	Office to Home	Office to Home
Performance during the experiment	-0.221		-0.532**	-0.776***	0.00273		-0.118	-0.123
	(0.182)		(0.264)	(0.298)	(0.243)		(0.345)	(0.361)
Performance before the experiment		0.0126	0.442	0.696**		0.0909	0.185	0.0938
		(0.202)	(0.305)	(0.333)		(0.275)	(0.383)	(0.411)
Age				0.00169				0.0983*
				(0.0432)				(0.0552)
Married				-0.955*				-0.0961
				(0.499)				(0.397)
Live with parents				-0.629*				0.132
				(0.324)				(0.405)
Cost of commute				-0.0340				0.0166
				(0.0273)				(0.0317)
Constant	-0.660***	-0.644***	-0.666***	0.0723	-0.332**	-0.338**	-0.352**	-2.977**
	(0.135)	(0.133)	(0.135)	(1.039)	(0.152)	(0.152)	(0.158)	(1.320)
Observations	104	104	104	104	73	73	73	73

Notes: Sample for returning to the office includes the 104 treatment works still at CTrip at the end of the experiment in September 2011. Out of the 104 treatment workers, 27 opt to come back to work in the office full-time. Pre-experiment performance is the average of individual weekly performance z-score during the pre-experimental period from January 1st 2010 to December 5th 2010. Post-experiment performance is the average of individual weekly performance z-score during the post-experimental period from December 6th 2010 to August 20th 2011. The sample for moving home includes the 75 control group employees still in the experiment by September 1st, 2011. Out of 73 control workers, 27 employees petitioned to work at home, and 25 successfully installed the equipment. Robust standard errors. *** denotes 1% significance, ** 5% significance and * 10% significance.

Table 6: Employee self-reported work outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
Variables:	Satisfaction	General Satisfaction	Life Satisfaction	Exhaustion	Positive Attitude	Negative Attitude
Data source:		Satisfaction survey			Emotion Survey	
Experiment *treatment	0.155***	0.072***	0.168***	-0.564***	0.160***	-0.183***
	(0.052)	(0.021)	(0.047)	(0.168)	(0.040)	(0.058)
Announcement*treatment				-0.102	0.080*	-0.095
				(0.167)	(0.042)	(0.058)
Treatment	-0.015	-0.012	-0.043			
	(0.048)	(0.020)	(0.066)			
Observations	855	855	855	5109	5109	5109

Notes: The satisfaction survey was conducted five times throughout the experimental period. See details of survey questions and methodology in Appendix A2. Once in early November before the randomization took place and four times after the experiment had started. The emotion survey is conducted every week. The first week was conducted in late-November 2010, before the experiment begun but after the randomization so that individuals had been informed of their status in the treatment or control groups. All the dependent variables are logged values. The regressions are run at the individual level with a full set of time-dummies. Experiment*treatment is the interaction of the treatment group with the period of the experimentation. Announcement*treatment is the interaction with the treatment group with the period of post-announcement but pre-experiment. Standard errors are clustered at the individual level. *** denotes 1% significance, ** 5% significance and * 10% significance.

Table 7. Attrition

	(1)	(2)	(3)	(4)	(5)
Dependent variable	Quit	Quit	Quit	quit	quit
Performance Measure Period	Baseline	Pre-experiment	Post-experiment	Post-experiment	Post-experiment
Sample	Total	Total	Total	Control	Treatment
Performance		-0.315	-1.044***	-1.093***	-0.374
		(0.225)	(0.217)	(0.223)	(0.242)
Performance*Treatment		0.214	0.635*		
		(0.300)	(0.328)		
Treatment	-0.565***	-0.550***	-0.142		
	(0.184)	(0.186)	(0.241)		
Age	-0.114***	-0.107***	-0.0940***	-0.0574	-0.142***
	(0.0332)	(0.0330)	(0.0348)	(0.0469)	(0.0538)
Men	0.190	0.0959	-0.0540	-0.249	0.205
	(0.182)	(0.198)	(0.203)	(0.278)	(0.297)
Married	-0.167	-0.140	-0.290	-0.169	-0.332
	(0.333)	(0.335)	(0.381)	(0.565)	(0.578)
Cost of Commute	0.0288***	0.0291***	0.0296***	0.0305	0.0289**
	(0.0110)	(0.0111)	(0.0111)	(0.0249)	(0.0120)
Children	0.558	0.595	0.930**	0.622	1.259*
	(0.369)	(0.374)	(0.423)	(0.549)	(0.688)
Constant	1.949**	1.795**	1.070	0.298	1.908
	(0.761)	(0.756)	(0.799)	(1.073)	(1.196)
Observations	255	254	254	122	132

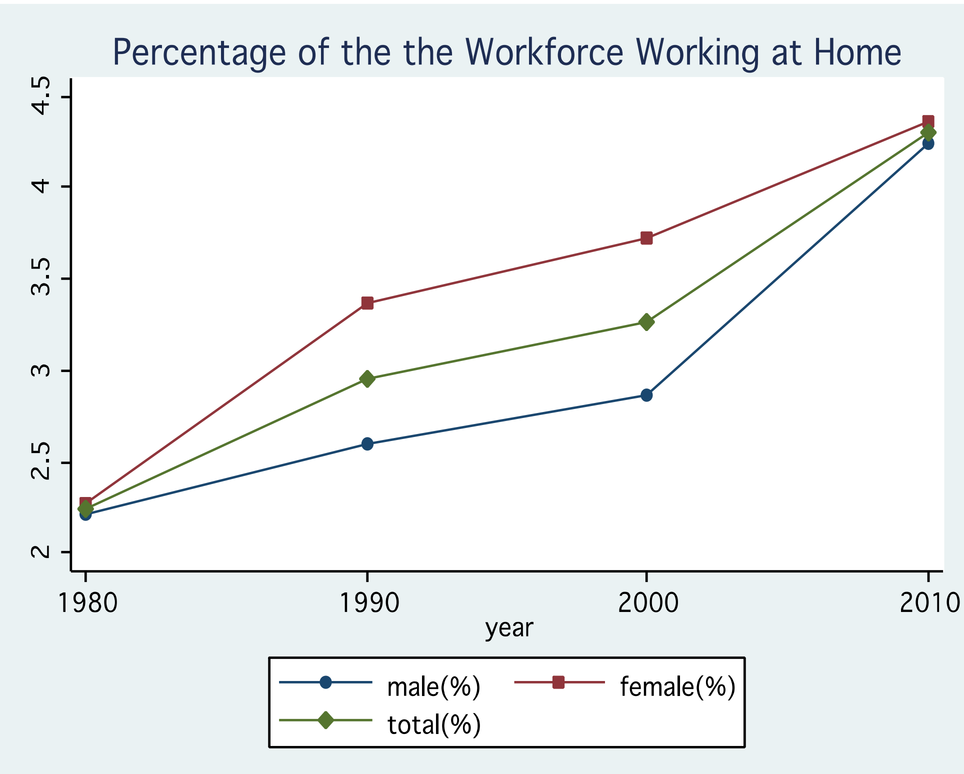
Notes: The regressions are all probits at the individual level. The dependent variable is whether the employee quit over the experimental period between December 6th 2010 and August 20th 2011. Pre-experiment performance is the average of individual weekly performance z-score during the pre-experimental period from January 1st 2010 to December 5th 2010. Post-experiment performance is the average of individual weekly performance z-score during the post-experimental period from December 6th 2010 to August 20th, 2011. Performance*treatment is the interaction of the performance measure by an individual having an even birthdate (2nd, 4th, 6th, 8th etc day of the month). Cost of commute is measured at daily level in Chinese Yuan (note that 1 Yuan is about 0.15 Dollars). Standard errors are clustered at the individual level. *** denotes 1% significance, ** 5% significance and * 10% significance.

Table 8: Employee survey views before and after the experiment

Interested in in working from home: August 2011	Interested in working from home: November 2010: No	Yes	Undecided	Total
No	71	59	79	209
	12.5	10.39	13.91	36.8
Yes	12	181	55	236
	2.11	31.87	9.68	41.55
Undecided	17	43	51	123
	2.99	7.57	8.98	21.65
Total	100	295	173	568
	17.61	51.94	30.46	100

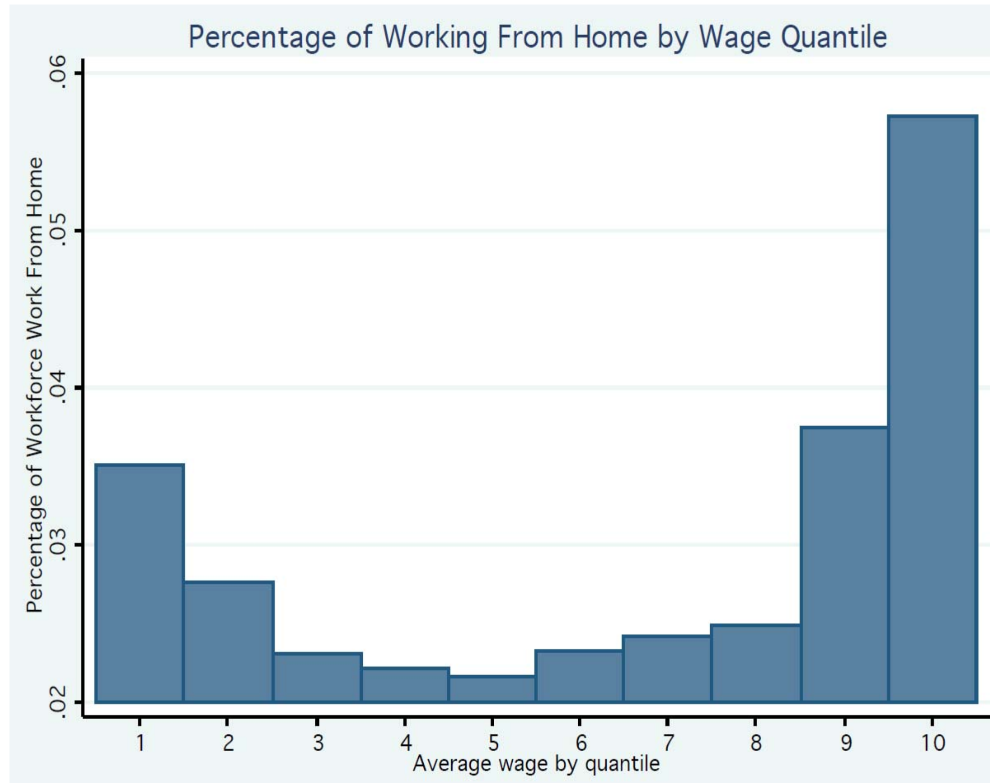
Notes: The total sample covers all CTrip employees in their Shanghai Airfare and Hotel group in November 2010 and August 2011. For the November 2010 survey employees were not told the eligibility rules in advance of the survey (i.e.: own room, 6+ months tenure, internet connect etc). For the November 2011 survey they were told the experiment was being rolled out to the company, but again not what the criteria for this would be.

Figure 1a: Proportion of the workforce working at home has doubled for both men and women since 1980



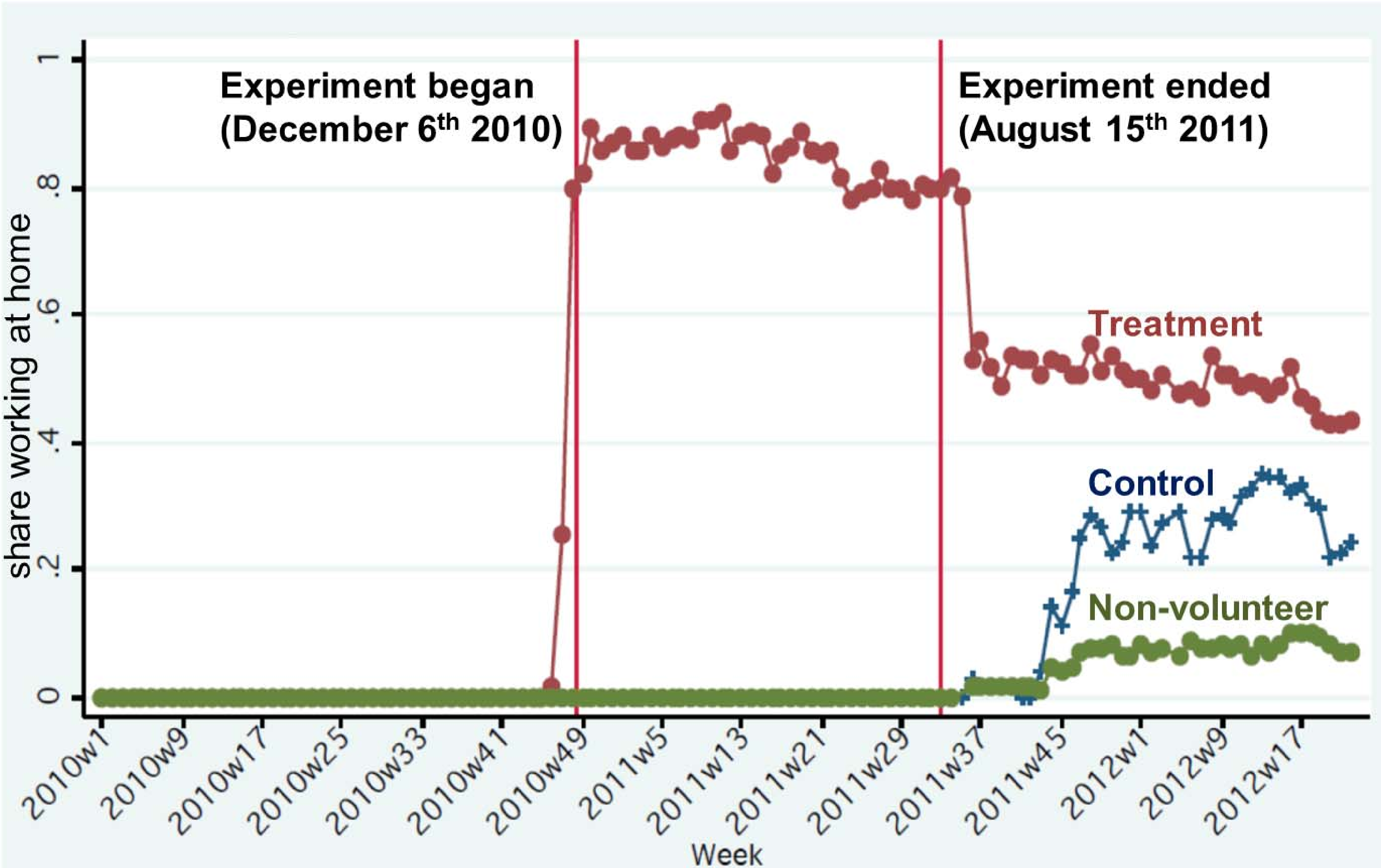
Notes: Source: PUMS census data for 1980, 1990, 2000 and 2010. We classify workers as mainly working from home if they answer "work from home" to the census question "How did you get to work last week?" Percentage of the workforce working at home equals number of workers reporting working from home divided by number of employed workforce.

Figure 1b: Home workers are bi-modally distributed across the earnings spread



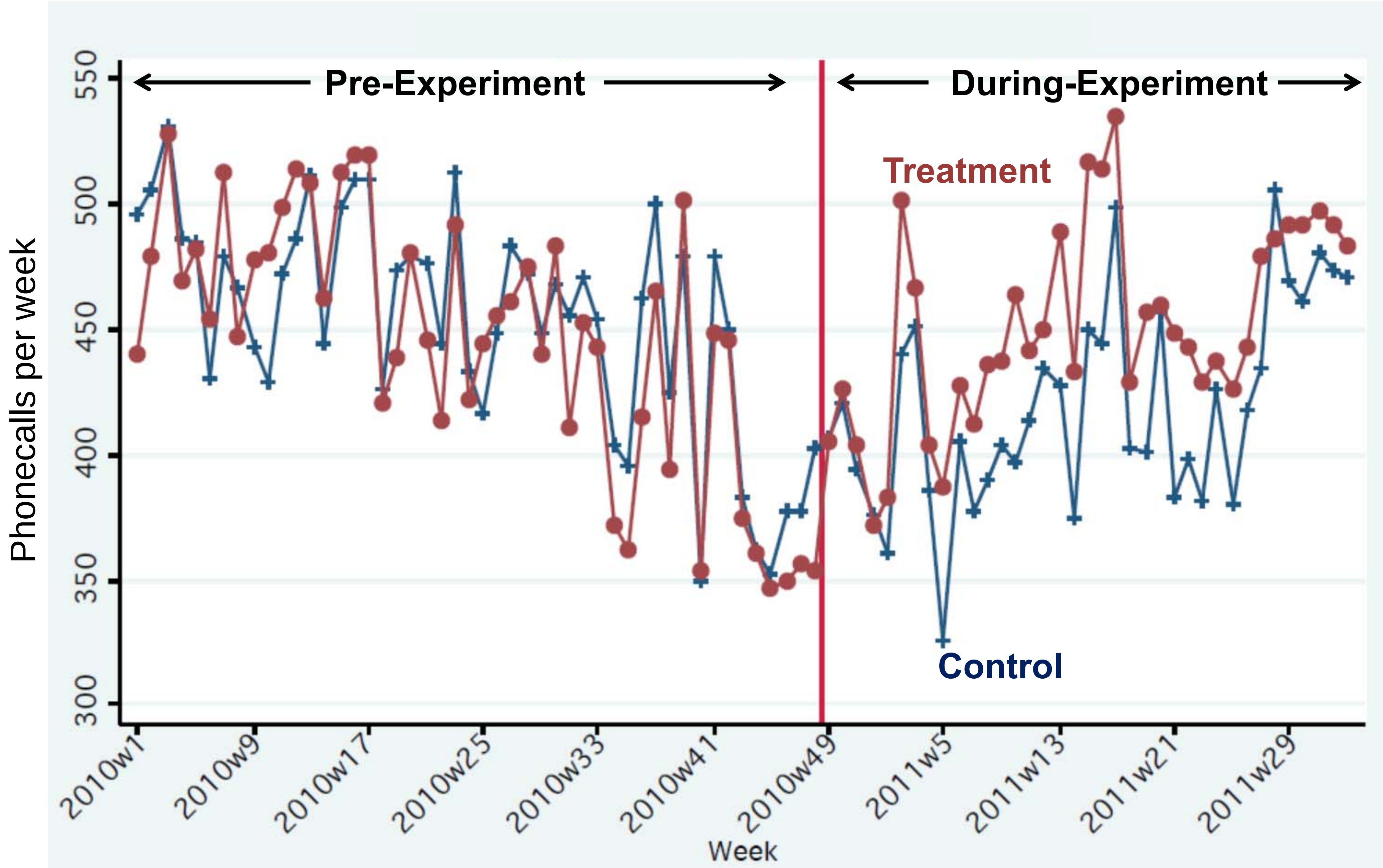
Notes: Source: PUMS census data for 2010. We classify workers as mainly working from home if they answer "work from home" to the census question "How did you get to work last week?" All employees are divided into 10 bins by wage. Percentage of workforce working from home is then calculated within each bin.

Figure 2. Share of treatment, control and non-volunteer working at home



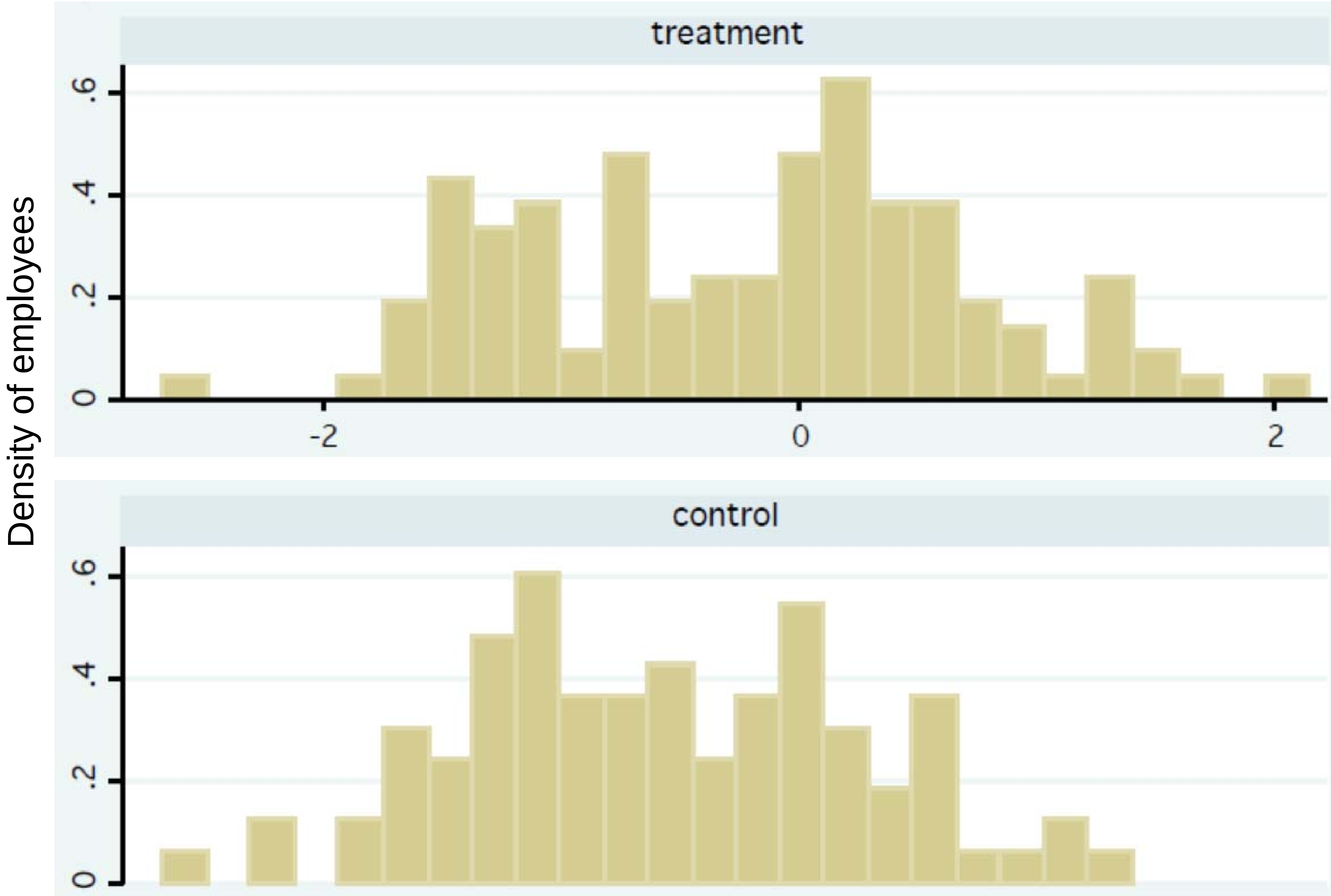
Note: Data from January 4th 2010 until June 3 2012. Percentage of workers working at home = (number of workers working at home / number of workers still employed) calculated separately for treatment (even birthdays), control (odd birthdays) and non-volunteer workers (those either that did not volunteer to work from home). First red line indicates the beginning of the experiment on December 6, 2010 and second red line indicates the end of the experiment on August 15, 2011, after which the working-from-home was available to all employees. Sample is all employees in airfare and hotel departments in Ctrip's Shanghai headquarter.

Figure 3a. Performance of treatment and control employees: phone calls



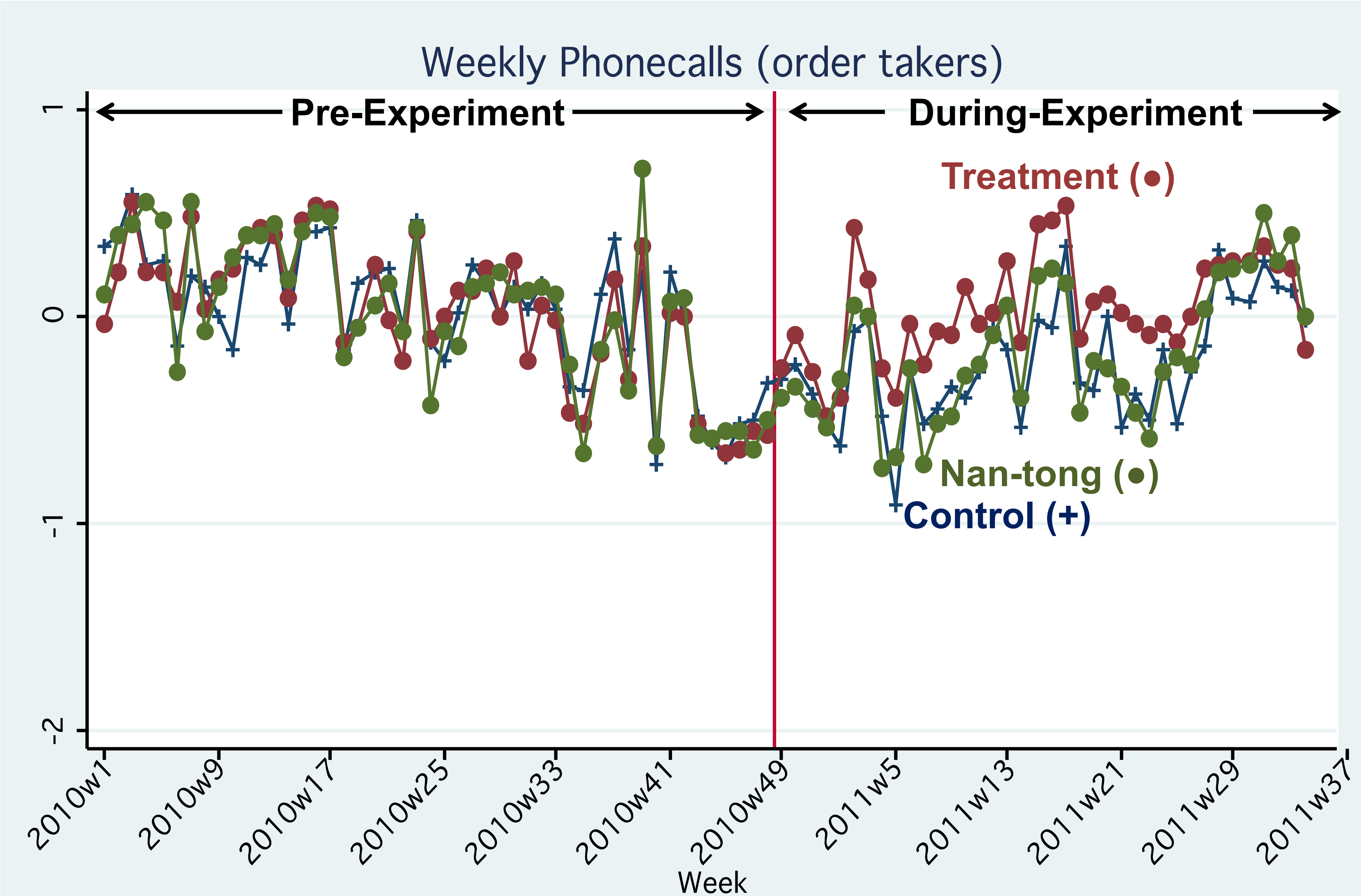
Note: Data from January 4th 2010 until August 15 2011. Number of phone calls made for order takers (the group for whom taking phone-calls is a performance metric) calculated separately for treatment (even birthdays) and control (odd birthdays).

Figure 3b. The cross-sectional improvement in working from home performance



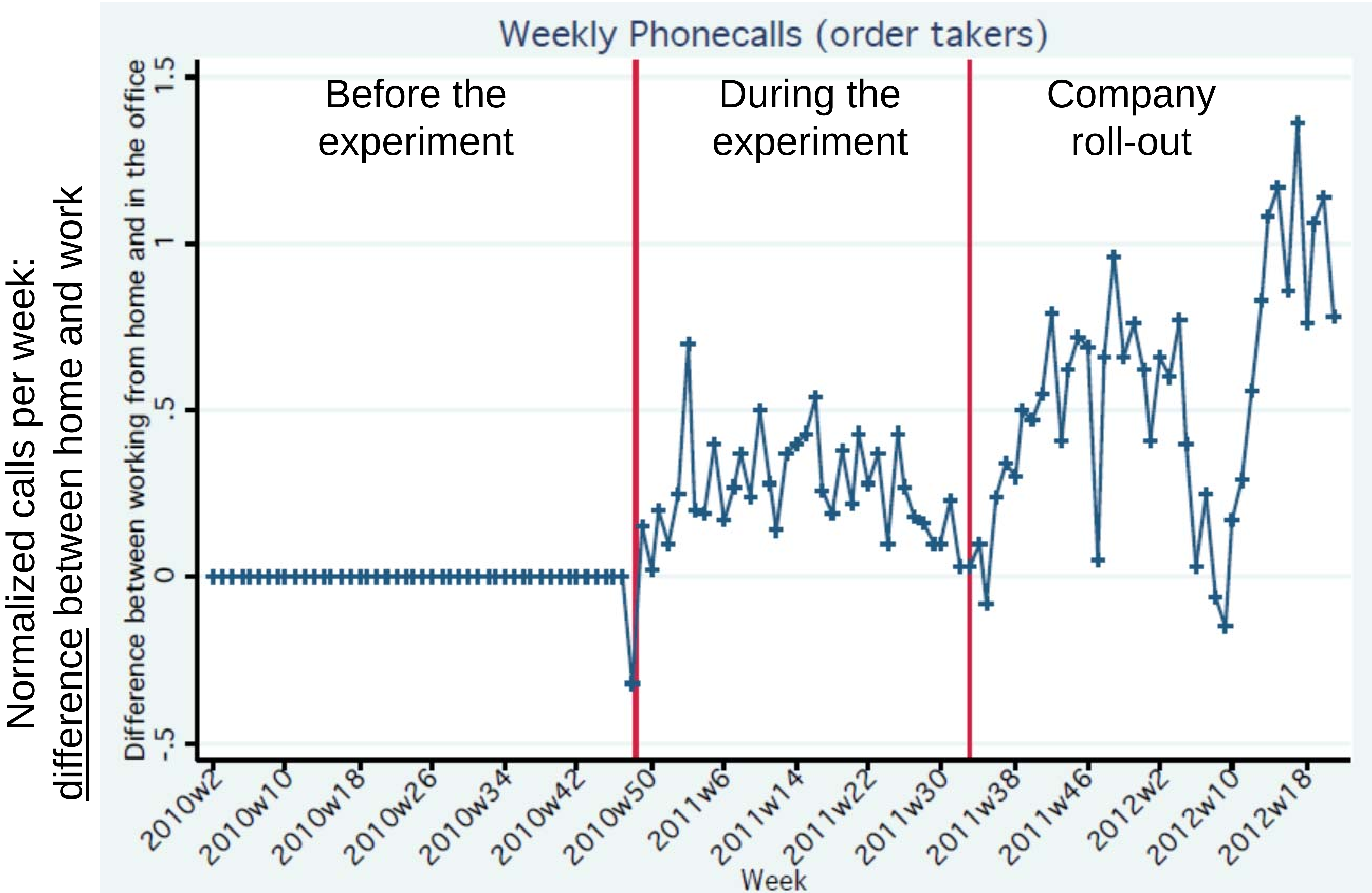
Notes: The histogram of the performance z-score for the treatment and control groups at 3 months into experiment (SD=1 across individuals in pre-experimental data)

Figure 4. Performance of treatment, control and Nantong employees



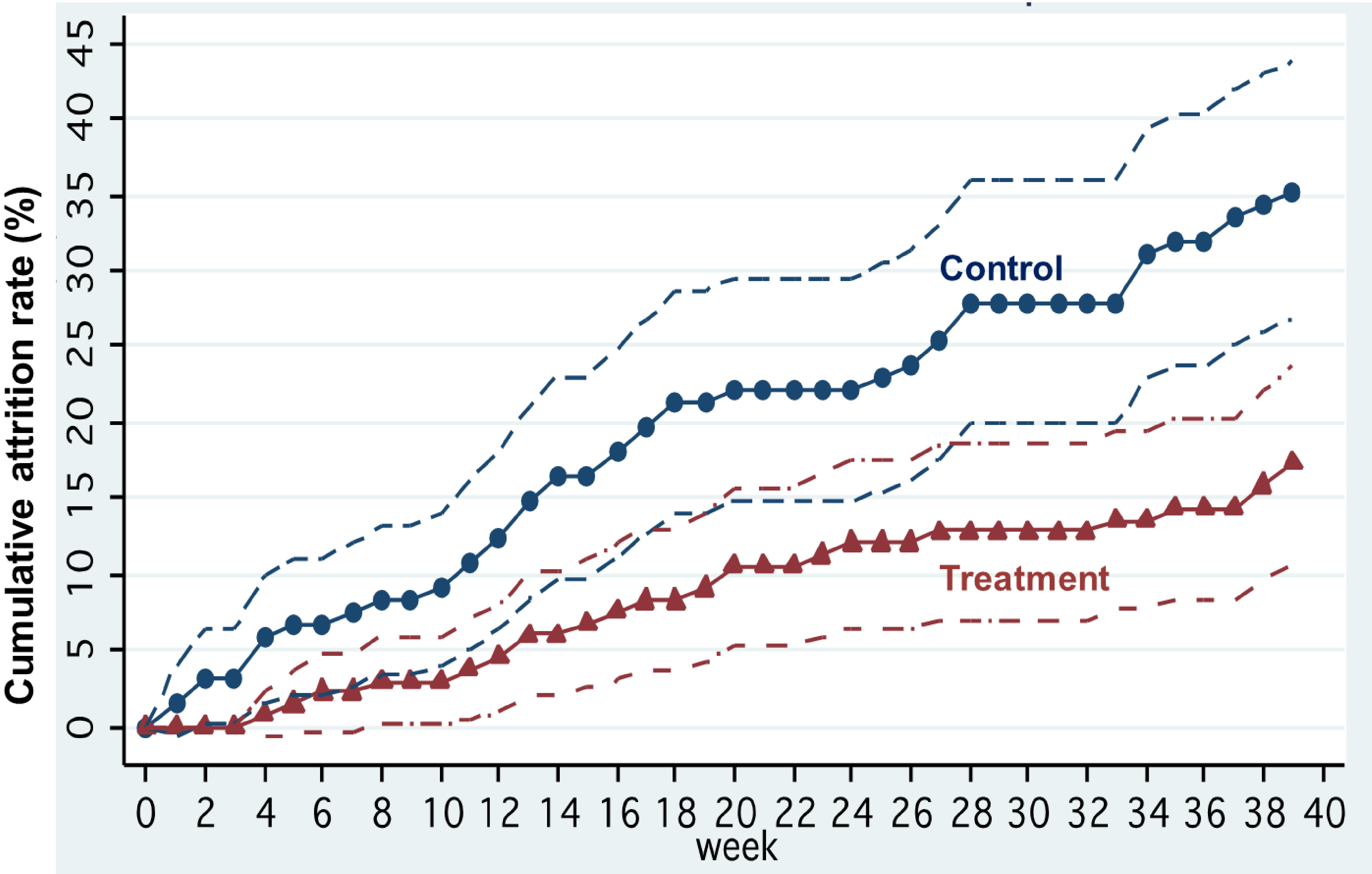
Note: Data from January 4th 2010 until August 15 2011. Phone calls in z-scores (normalized so the pre-experiment values are mean zero and standard deviation 1). Calculated separately for treatment (even birthdays), control (odd birthdays) and Nantong employees with the same eligibility requirements (6+ months tenure)

Figure 5. Selection further increased the performance impact of home working during the company roll-out



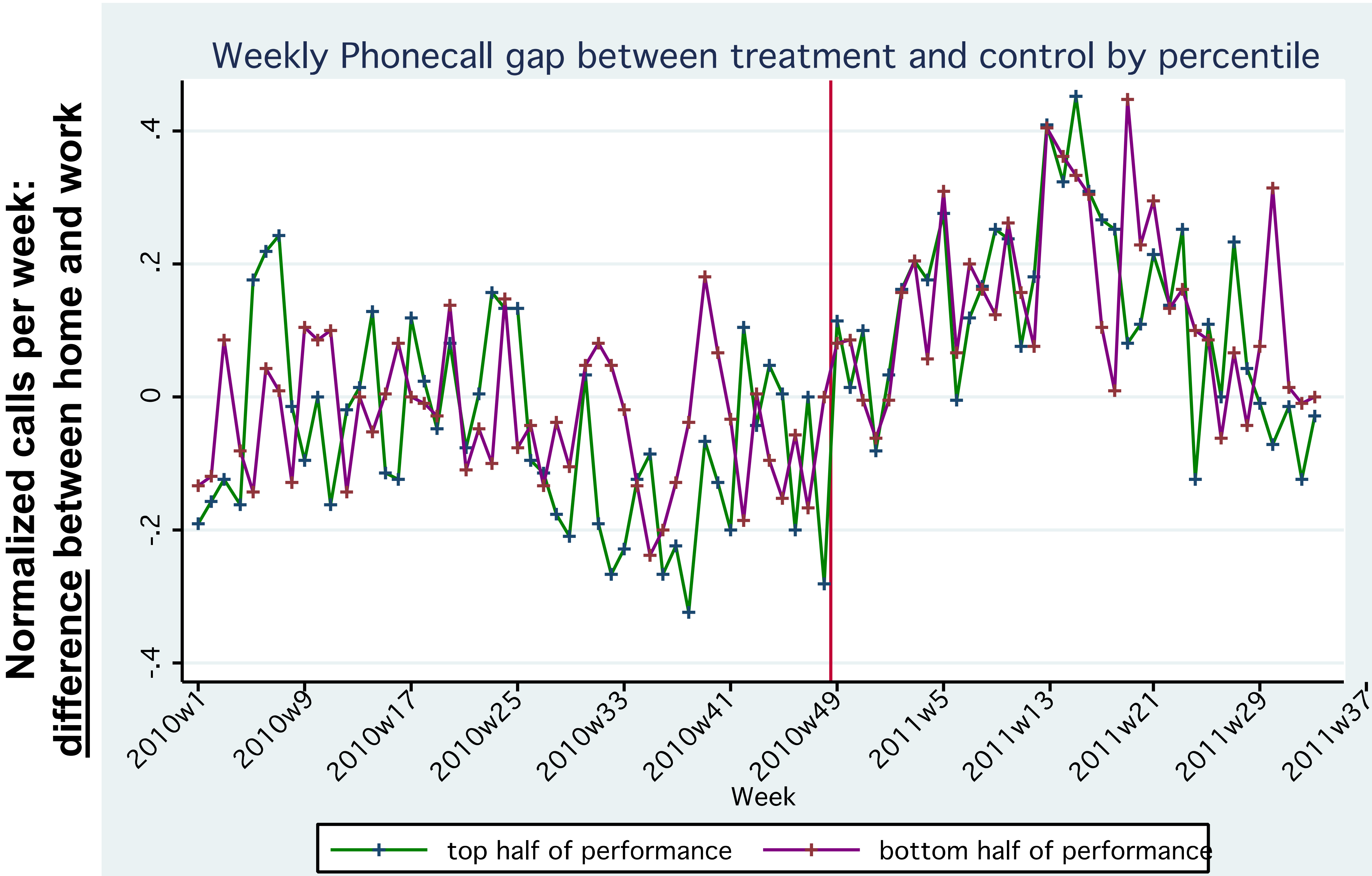
Note: Data from January 4th 2010 until June 1st 2012. Phone calls in z-scores (normalized so the pre-experiment values are mean zero and standard deviation 1) shown as the difference between home and office workers.

Figure 6. Attrition is halved by working from home



Notes: Cumulative attrition rate equals number of employees attrited by week x of the experiment divided by total number of employees at the beginning of the experiment, calculated separately by treatment and control group. Dashed lines represent 95% confidence interval. Experiment started at week 0 and ended at week 37.

Figure 7: The top and bottom half of employees by pre-experiment performance both improved from working at home



Note: Data from January 4th 2010 until August 15 2011. Phone calls in z-scores (normalized so the pre-experiment values are mean zero and standard deviation 1). Calculated separately for the difference between the top half of the treatment and control groups and the bottom-half of the treatment and control groups, where performance halves are based on pre-experiment performance.

Exhibit A: Ctrip is a large and modern firm in China

Headquarter in Shanghai

Main Lobby

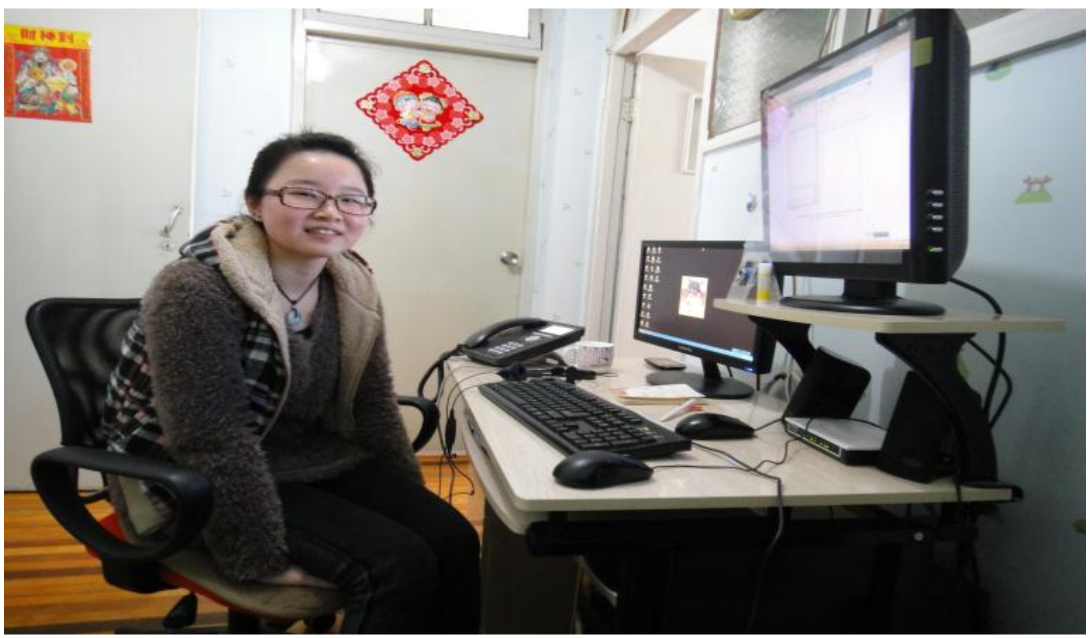
Call Center Floor

Team Leader Monitoring Performance

Exhibit B: The experimental randomization, and examples of home-workers

Treatment groups were determined by a lottery

Working at home



Working at home

Working at home