ONLINE DATA APPENDIX

Appendix Table O1: Comparison between Volunteers and Non-volunteers

	Volunteer	Non-volunteer	t-stat
Number	503	491	
Children	0.10	0.07	2.17
	(0.30)	(0.25)	
Married	0.17	0.13	2.15
	(0.38)	(0.33)	
Daily commute	86.50	74.47	3.38
	(60.47)	(51.32)	
Own bedroom	0.66	0.54	3.90
	(0.47)	(0.50)	
Tertiary education and above	0.37	0.46	-2.84
	(0.48)	(0.50)	
Tenure (months)	23.22	26.84	-2.73
	(20.28)	(21.53)	
Gross Wage (1000 yuan)	2.83	2.90	-1.13
	(0.92)	(0.94)	
Age	23.22	23.23	-0.07
	(3.28)	(2.89)	
Male	0.34	0.30	1.30
	(0.47)	(0.46)	

Note: The total sample covers all CTrip employees in their Shanghai airfare and hotel departments. Willingness to participate was based on the initial survey in November 2010. 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).

Appendix Table O2: Explanations of the Work Satisfaction Survey

Work Exhaustion: CTrip's in-house psychology counselors used an adapted excerpt from the Maslach Burnout Inventory (MBI) 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 was asked to evaluate his or her "emotional exhaustion" at the end of the work week. The survey contained 6 questions. Each employee was asked to report how often he or she 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 used an adapted 16-item Positive and Negative Affect Schedule (PANAS) developed by Clark and Tellegen (1988) to measure the positive and negative attitudes of the employees.

The survey comprised two mood scales, one measuring positive affect and the other measuring negative affect. Each item was 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 felt this way the day he took the survey. To evaluate the positive affect, psychologists summed 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 were (1) Cheerful, (2) Jittery, (3) Happy, (4) Ashamed, (5) Excited, (6) Nervous, (7) Enthusiastic, (8) Hostile, (9) Content, (10) Guilty, (11) Relaxed, (12) Angry, (13) Proud, (14) Dejected, (15) Active and (16) Sad.

Appendix Table O3: Quality did not change in the experiment

Dependent Variable	(1) recording grade	(2) recording grade	(3) Conversion (z score)	(4) Conversion (z score)
Mean	3.647	3.647	0.102	0.102
SD	(0.139)	(0.139)	(1.048)	(1.048)
Individual FE	No	Yes	No	Yes
Week fixed-effects	Yes	Yes	Yes	Yes
Experiment*Treatment	0.008	-0.005	-0.012	-0.012
	(0.023)	(0.008)	(0.074)	(0.069)
Treatment	0.000		-0.003	
	(0.005)		(0.094)	
Number of Employees	125	125	134	134
Number of Weeks	85	85	85	85
Observations	6264	6264	9429	9429

Notes: Sample in the first two columns includes 125 order takers for whom we can obtain recording grade information. The sample in the last two columns includes 134 order takers in airfare and hotels (the group for which conversion rate data exists). Individual clustered standard errors. *** denotes 1% significance, ** 5% significance and * 10% significance.

Appendix Table O4. Treatment Effects Seem Homogeneous across Characteristics

Dep. Variable Performance z-score	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Child	Female	Female w/ Child	Commute >120min	Renter	Young	Short prior experience	Short tenure	Live w/ parents	Live w/ spouse	Live w/ friends	Pre-exper. performance
Experiment_t x treat_i x "characteristic"$_i$	0.055	0.068	0.057	0.124	-0.186	-0.128	0.046	-0.060	0.072	-0.056	-0.170	0.079
	(0.170)	(0.123)	(0.203)	(0.143)	(0.138)	(0.126)	(0.127)	(0.125)	(0.134)	(0.166)	(0.220)	(0.111)
Experiment_t x "characteristic"$_i$	0.019	-0.061	-0.024	-0.061	0.130	0.007	0.029	0.084	-0.040	0.005	0.274	-0.257***
	(0.128)	(0.088)	(0.174)	(0.090)	(0.103)	(0.092)	(0.092)	(0.090)	(0.101)	(0.113)	(0.181)	(0.083)
Experiment_t x Treatment_i	0.231***	0.198**	0.230***	0.192**	0.277***	0.308***	0.215**	0.264***	0.191*	0.247***	0.242***	0.230***
	(0.067)	(0.080)	(0.064)	(0.078)	(0.074)	(0.099)	(0.092)	(0.101)	(0.113)	(0.064)	(0.066)	(0.060)
Observations	17814	17814	17814	17814	17814	17814	17814	17814	17814	17814	17814	17814
R-squared	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.400	0.402

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. Individual clustered standard errors. *** denotes 1% significance, ** 5% significance and * 10% significance.

Appendix Table O5: Robustness Check of Table 2 using IV regressions

| Dependent Variable	(1) Overall Performance	(2) Gross Wage	(3) Overall Performance	(4) Phonecalls	(5) Phonecalls	(6) Phonecalls Per Minute	(7) Minutes on the Phone
Period	Pre and during experiment	Pre and during experiment	During experiment	Pre and during experiment	Pre and during experiment	Pre and during experiment	Pre and during experiment
Dependent Normalization	z-score	Log	z-score	z-score	log	log	log
Second Stage							
Experiment$_t$*WFH$_i$	0.271***	0.077**		0.285***	0.138***	0.036**	0.101***
	(0.073)	(0.032)		(0.067)	(0.029)	(0.015)	(0.031)
WFH$_i$			0.214**				
			(0.100)				
First Stage							
Experiment$_t$*Treatment$_i$	0.856***	0.881***	0.858***	0.871***	0.871***	0.871***	0.871***
	(0.005)	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Number of Employees	249	249	249	134	134	134	134
Number of Weeks	85	20 (months)	37	85	85	85	85
Individual Fixed Effects	Yes	Yes	No	Yes	Yes	Yes	Yes
Observations	17806	4530	7464	9426	9426	9426	9426

Notes: The regressions are run at the individual by week level, with a full set of individual and week fixed effects except for Column (2) where the regression is run at the individual by month level with individual and month fixed effects. Experiment*treatment is the interaction of the period of the experimentation (December 6th 2010 until August 15th 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. Experiment*WFH is the interaction of the period of the experiment by a treatment employee working at home for at least one shift during that week (or month in Column(2)). Experiment*treatment is used as the instrument for experiment*WFH in all regressions. Pre-experiment and during the experiment period started on Jan 1st, 2010 and ended on August 15th, 2011. During the experiment period started on December 6th, 2010 and ended on August 15th, 2011. Overall performance is the z-score for each employee on their main task. 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 for overall performance covers a wider group of employees than for phonecall. Minutes on the phone is recorded from the call logs. Three employees have been excluded because they lack pre-experimental data. Standard errors are clustered at the individual level. *** denotes 1% significance, ** 5% significance and * 10% significance.

Appendix Table O6: Experimental Impact on Wages

VARIABLES	(1) Gross wage	(2) Base wage	(3) Bonus	(4) log(Gross wage)	(5) log(Base wage)	(6) log(Bonus)
Experiment*treatment	248.771***	-4.352	217.407**	0.094***	-0.000	0.200***
	(88.396)	(15.842)	(86.057)	(0.032)	(0.008)	(0.074)
Number of individuals	249	249	249	249	249	249
Number of months	20	20	20	20	20	20
Observations	4648	4648	4648	4648	4648	4551
R-squared	0.549	0.811	0.501	0.489	0.778	0.394

Notes: The regressions are run at the individual by month level, with a full set of individual and month fixed effects. Experiment*treatment is the interaction of the period of the experimentation (December 6th 2010 until August 15th 2011) by an individual having an even birthdate (2nd, 4th, 6th, 8th etc day of the month). The sample period is from January 2010 until August 2011. Bonus includes performance-based incentive pay, overtime pay and transportation subsidy. 97 observations where bonus are zero are dropped in column (6).

Online Appendix Figure O1. The Lee (2008) bounds for estimating the impact of non-random attrition
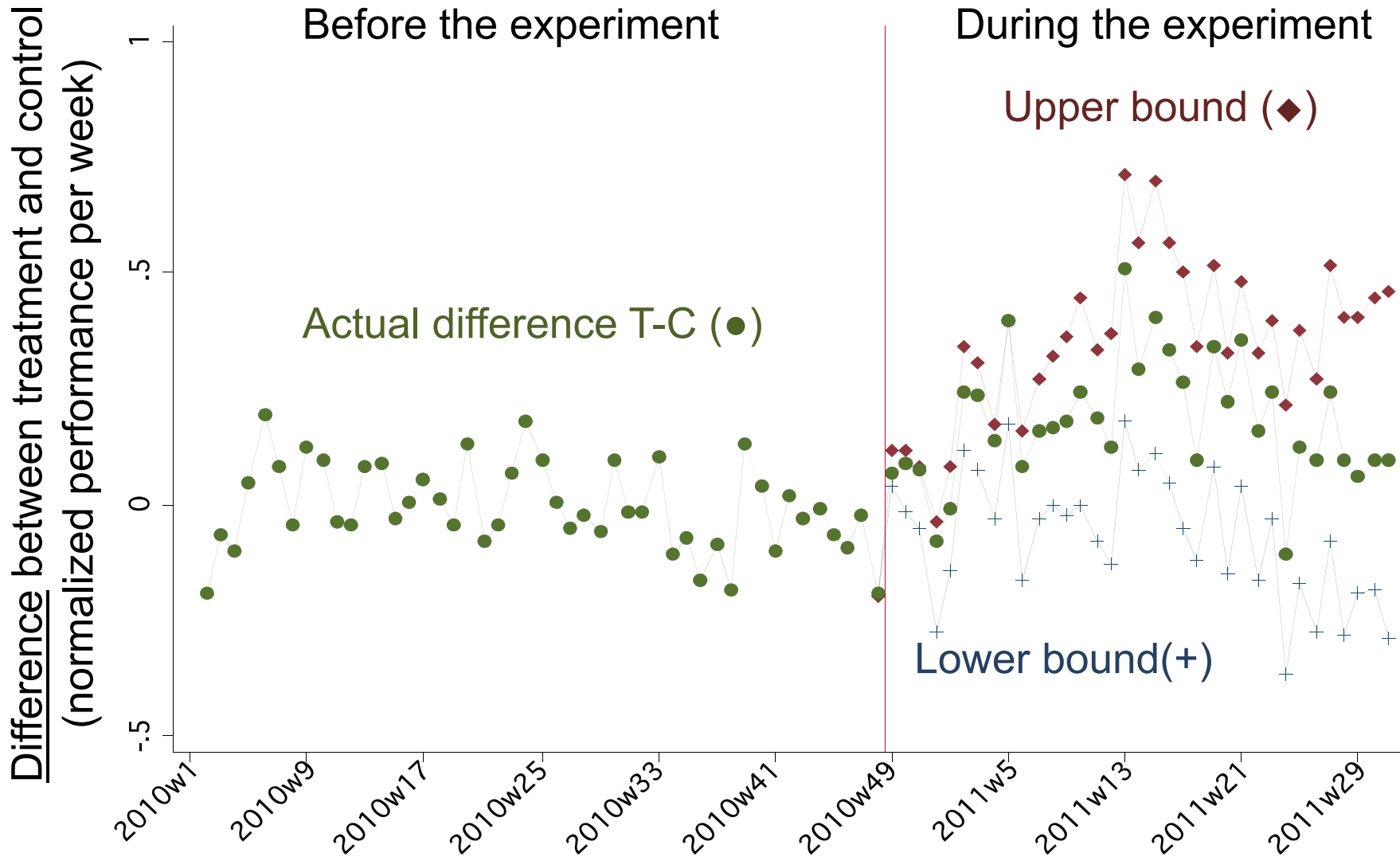
Note: Data from January 4th 2010 until August 14th 2011. Performance in z-scores (normalized so the pre-experiment values are mean zero and standard deviation 1; performance measures used depend on the type of employees). Upper bound is calculated assuming the worst performers of the treatment group would quit to create a equal attrition rate between treatment and control each week. Lower bound is calculated assuming the best performers for the treatment group would quit to create an equal attrition rate between treatment and control. Approach follows Lee (2008).