Appendix: Optimal random sampling experiments vs. constrained random sampling experiments

Our study includes 352 participants, for which we conducted 1000 experiments following the optimal random sampling (RS) approach as presented in the paper. Using one-day sampling as an example, for each of the 1000 RS experiments, we have calculated the fraction of the sample that is Monday, Tuesday, Wednesday,…, Sunday. Table A-1 shows the means and standard deviations of the fractions among the 1000 experiments. The average fractions of Monday, Tuesday, Wednesday,…, Sunday are 14.24% - 14.37%. Overall, the distributions of the fractions are smooth with small spreads.

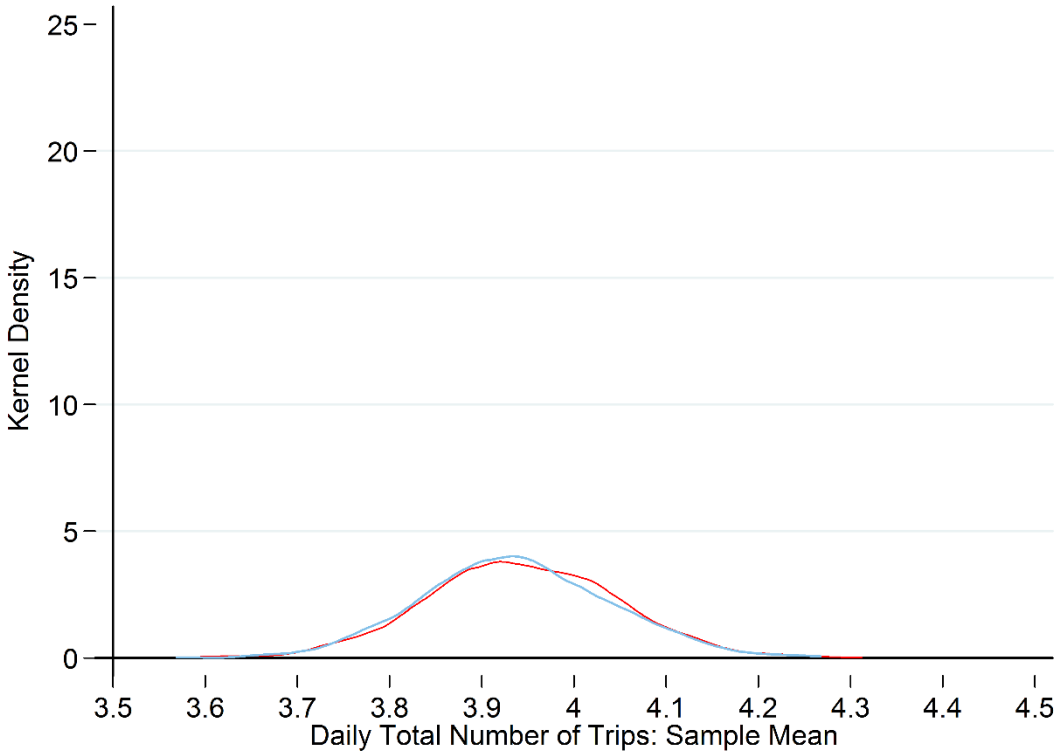
Table A-1: Each day's fraction in a random sampling experiment

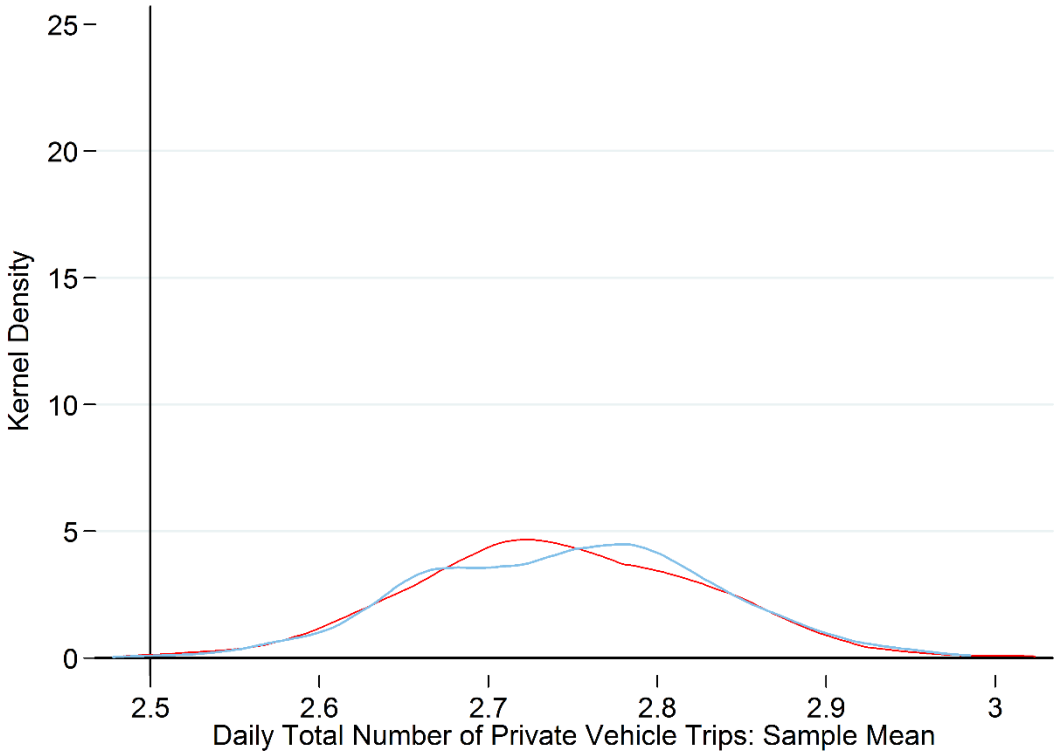
Variable	Obs.	Mean	Std. Dev.
% of Mondays	1000	0.1430	0.0184
% of Tuesdays	1000	0.1425	0.0196
% of Wednesdays	1000	0.1433	0.0185
% of Thursdays	1000	0.1424	0.0180
% of Fridays	1000	0.1425	0.0184
% of Saturdays	1000	0.1437	0.0188
% of Sundays	1000	0.1426	0.0187

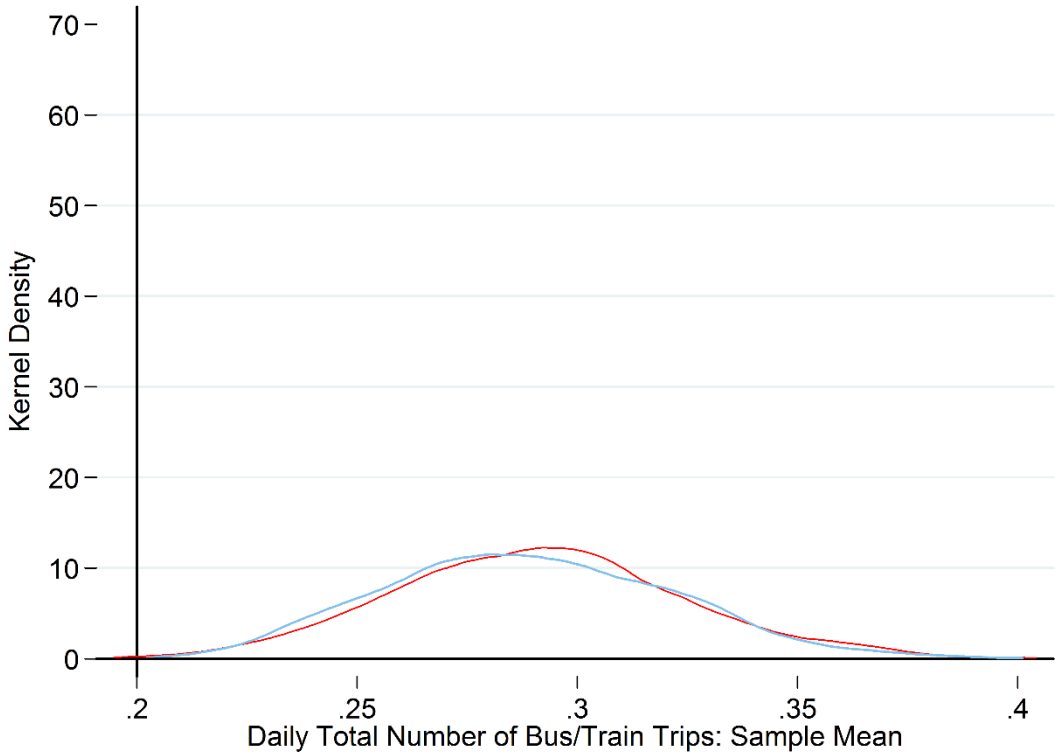
We have also designed a constrained random sampling procedure, using one-day sampling as an example. As there are 352 participants in our study, we randomly delete two participants to have 350 participants. The constrained random sampling procedure ensures that, in each experiment, the fraction of participants drawn for each day of a week is the same. That is to say, in each of the 1000 constrained random sampling experiments, the fraction of survey days for Monday, Tuesday, Wednesday,…, Sunday is

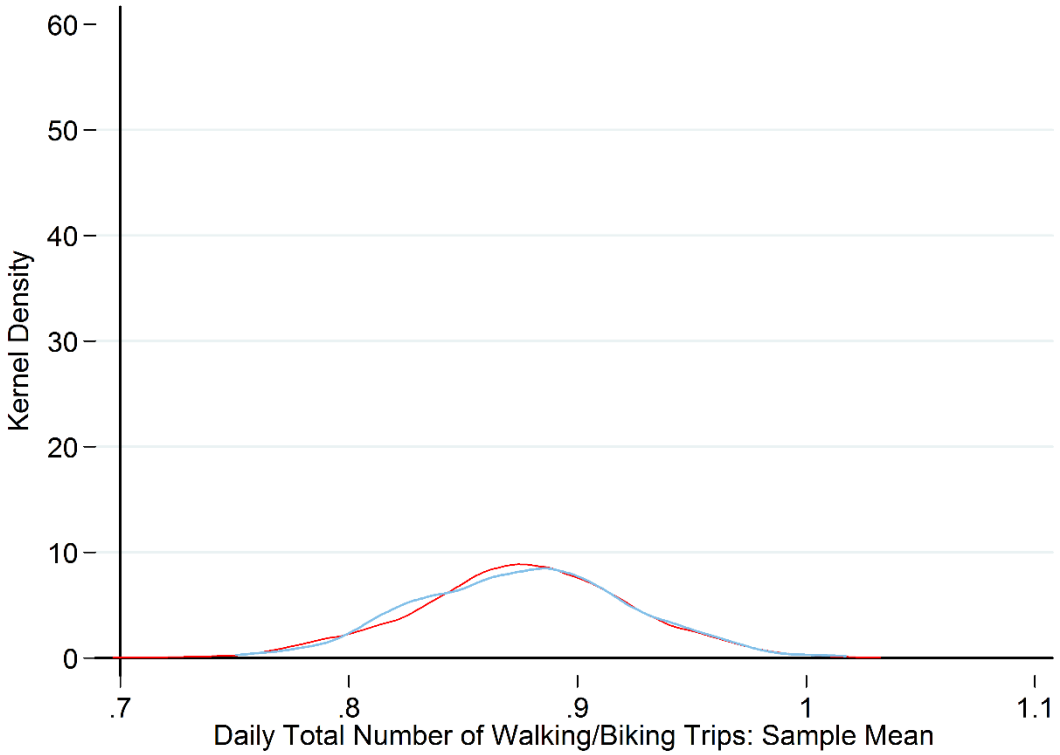
the same: 1/7; in other words, there are 50 survey days for each day of a week in each constrained random sampling experiment.

As illustrated in Figure A-1, the distributions of sample means from the constrained random sampling procedure (shown in red lines) align well with their counterparts from the optimal RS procedure (shown in blue lines) selected by researchers.




25
20
15
10
5
0
Kernel Density
2.5
2.6
2.7
2.8
2.9
3
Daily Total Number of Private Vehicle Trips: Sample Mean


70
60
50
40
30
20
10
0
Kernel Density
.2
.25
.3
.35
.4
Daily Total Number of Bus/Train Trips: Sample Mean


60
50
40
30
20
10
0
Kernel Density
.7
.8
.9
1
1.1
Daily Total Number of Walking/Biking Trips: Sample Mean

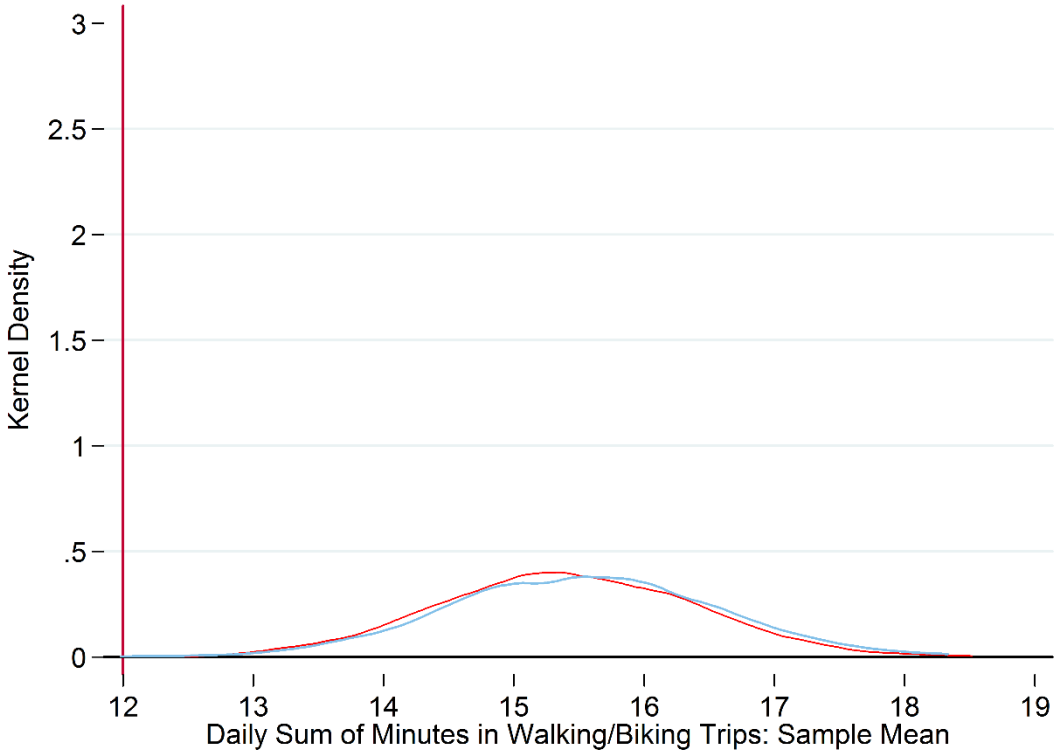

3
2.5
2
1.5
1
.5
0
Kernel Density
12
13
14
15
16
17
18
19
Daily Sum of Minutes in Walking/Biking Trips: Sample Mean

Figure A-1: Distribution of 1-day sample means:
Optimal random sampling (blue) vs. constrained random sampling (red)

Table A-2: Distribution of 1-day sample means:
Optimal random sampling vs. constrained random sampling

	Within $\pm 0.025*\mu_{total}$ (352 persons) Optimal Random Sampling	Within $\pm 0.05*\mu_{total}$ (352 persons) Optimal Random Sampling	Within $\pm 0.025*\mu_{total}$ (352 persons) Constrained Random Sampling	Within $\pm 0.05*\mu_{total}$ (352 persons) Constrained Random Sampling	Within $\pm 0.025*\mu_{total}$ (350 persons) Constrained Random Sampling	Within $\pm 0.05*\mu_{total}$ (350 persons) Constrained Random Sampling
Total Trip Count	66.9%	95%	67.6%	94.2%	67.4%	94.2%
Private Vehicle Trip Count	55.9%	90.1%	56.5%	90.3%	56.5%	90.1%
Bus/Train Trip Count	16.6%	32.6%	16.4%	34.9%	17.5%	35.1%
Walking/Biking Trip Count	36%	63.7%	37.0%	66.8%	36.9%	66.8%
Walking/Biking Duration	30.2%	55.4%	30.1%	55.9%	29.9%	56.0%

In Table A-2, we compare results for the five travel metrics from the optimal RS procedure with their counterparts from the constrained random sampling procedure. Columns 4-5 are based on 7-day means from the original 352 participants. Columns 6-7 are based on 7-day means from the 350 participants: 3.9395 (total trip count), 2.7422 (private vehicle rip count), 0.2909 (bus/train trip count), 0.8766 (walking/biking trip count), and 15.4748 (walking/biking duration in minutes). These 7-day means are extremely similar to their counterparts for the original 352 participants. In either case (7-day means based on 350 vs. 352 participants), results are very similar to their counterparts generated from the optimal random sampling procedure.