An Experimental Test of Whether Policies Scale: Evidence from Delhi's Free Bus Scheme

Girija Borker Gabriel Kreindler Dev Patel

September 10, 2026*

Abstract

Testing whether a policy successfully performs at scale is often difficult because of the absence of a plausible counterfactual. We show that even when the policy itself cannot be randomized, its scale-up fidelity can still be evaluated by randomizing additional access to a pilot run of the same treatment among a sample of people already covered by the at-scale version. Under policy idempotence—receiving a policy twice is equivalent to receiving it once—this experiment tests whether the at-scale implementation delivers the effects of a pilot version. We apply this test to Delhi's landmark reform granting access to free bus travel for all women citywide, a treatment with clear idempotence. We randomized delivery of bus passes among 746 low-income women and collected panel mobility data spanning the pass distribution and the citywide rollout. Before the full-scale launch, passes increased bus trips by 64%. After the city made buses universally free, including for the control group of the pilot experiment, the treatment effect nevertheless persisted with no detectable attenuation. We show this persistence is not due to the citywide policy not being implemented, a lack of awareness, or crowding. Instead, the evidence suggests increased personalization in the pilot made free travel more salient than the government's version. Overall, we reject that the at-scale implementation of the free buses policy in Delhi delivered the same mobility effects as the pilot.

*Borker: Development Impact, World Bank (gborker@worldbank.org), Kreindler: Harvard University (gkreindler@fas.harvard.edu), Patel: Brown University (dev_patel1@brown.edu). We thank Emily Breza, Robert Garlick, Claudia Goldin, Seema Jayachandran, Florence Kondylis, Michael Kremer, Matthew Lowe, Gautam Rao, Manisha Shah, Alessandra Voena, and Kate Vyborny for helpful feedback, and Mayara Felix and Chen Lin for discussing the paper. We thank Sanya Sareen for extraordinary research assistance throughout the project, and Akshay Choudhary, Ida Grigoryeva, Daniel Lasso Jaramillo, Xinyu Ren, Daniela Shuman, Sanjeevan Singh, and Kartik Srivastava for research assistance help at various stages. We are grateful to J-PAL South Asia for implementation. We acknowledge generous and timely funding from the Weiss Family fund, the International Growth Centre, and the Warburg fund, and Patel acknowledges financial support from a National Science Foundation Graduate Research Fellowship under grant DGE1745303 and the Bill and Melinda Gates Foundation, awarded through the NBER. We are grateful to Patrick Baylis, Michael Greenstone, Kenneth Lee, and Harshil Sahai for collecting and sharing data used in this study. This project has human subjects approval from Harvard CUHS (protocol number IRB19-1313). The experiment was registered in the AEA RCT Registry (AEARCTR-0004660).

1 Introduction

Evaluating policies at scale often presents a significant challenge because of a lack of a valid counterfactual—a control group that did not experience the reform. Pilot versions of the treatment may not always translate to larger stages; in fact, in several notable cases with credible designs, promising results from small-scale trials often weaken or vanish altogether as the number of units treated grows. In the absence of the gold-standard method of randomization at scale (Muralidharan and Niehaus, 2017), how can we assess the efficacy of policy reforms?

In this paper, we propose a policy evaluation method when randomizing the full policy is infeasible yet researchers can nevertheless randomize a smaller pilot version of the same treatment. By overlapping the pilot experiment in a context where an entire population has already experienced a scaled-up policy, we can test whether the at-scale version subsumes the pilot implementation. Under an assumption about the degree of policy idempotence—the extent to which receiving a treatment twice affects the outcome beyond the impact of receiving it once—this design recovers how policy efficacy changes between the small- and large-scale versions of the treatment. For example, a policy providing access to a good or service at a given price is idempotent, provided access is not quantity-rationed. The same logic applies to a binary entitlement to a government benefit. Policy idempotence does not hold when dosage matters, as with cash transfers or training hours.

The key idea of our method is to compare participants with access to the scaled up policy (the control group) to those with access to both the scaled up policy and the additional pilot version (the treatment group). If the policy is idempotent and the at-scale version delivers the same direct effects as the pilot, the two groups have the same outcomes. Any effect of the additional pilot access therefore rejects the joint hypothesis of idempotence and successful scaling.

What does the magnitude of the additional pilot impact measure? Suppose the pilot and the at-scale policy differ only through their effect on latent *activation*—whether the policy actually enters a person's decision. Then the additional pilot effect is the average treatment effect among those the pilot activates and the at-scale policy does not, multiplied by the share of such people. In other words, we recover the effect of the policy for people whom the policy reaches but the scale-up does not.

We apply this method to evaluate Delhi's landmark reform making bus travel free for all women. The government implemented this policy across the entire city at once, making at-scale randomization impossible. Instead, we enacted the same core policy lever at a much smaller scale by randomly providing half of 746 women a free bus pass enabling them to ride

city buses for free for one month. We focus our analysis on women with low socioeconomic status—the primary intended beneficiaries of the government policy (Jamba et al., 2025)—restricting our sampling frame to women living in low-income, high-density neighborhoods, who reported some mobility at baseline. We collect rich panel data on this sample, tracking their travel behavior over six months.

We launched this experiment just before the city-wide roll-out, allowing us to both test the impact of the pilot version itself prior to the full-scale launch and evaluate whether the at-scale implementation ultimately subsumes our experimental treatment.[1] Prior to the implementation of the government's reform, the pilot bus pass intervention caused a large and immediate increase in low-income women's mobility. Using detailed trip diaries, we show that women took 0.16 more trips per weekday using the bus (a 64% increase), 0.13 more trips out of their neighborhood per weekday, and made 35% more trips on their own after having received a pass.[2]

We next turn to the evaluation of the at-scale policy: did Delhi's city-wide reform subsume the pilot experiment we ran? We find that women who received a monthly bus pass as part of the experiment continue to travel relatively more than those in the control group even after the policy made bus travel free for everyone. In fact, even up to three months after the passes we distributed had expired, women in the treatment group exhibit significantly higher rates of mobility. We do not detect any change in the magnitude of the treatment effect on the main mobility outcomes, nor any trend over this period. We can therefore reject that the at-scale policy subsumed the pilot version, indicating a failure of the citywide policy in the sense that it did not render the similar pilot policy obsolete.

We find other supporting evidence that the citywide policy did not affect mobility in this sample. We do not detect any effect on bus usage or overall mobility among our sample of low-income women in Delhi, neither in a simple pre-post analysis, nor when comparing to women from other cities or men in Delhi. Although these designs offer imperfect comparisons and yield estimates with substantial uncertainty, point estimates are predominantly negative, allowing us to reject moderate positive effects in the initial three month period.

One advantage of our approach is that it allows us to make a comparison between the pilot and the at-scale policy in the same population, at the same time. Fixing the sample matters in our case. Using two other data sets collected in bus stops in the Delhi area, we find

[1] This "early access" version of our method requires an assumption on dynamic treatment effects, which we discuss further in Section 2.

[2] Our mobility outcomes are self-reported. While we cannot fully rule out experimenter demand effects, they are unlikely: participants were explicitly told that we neither encouraged nor discouraged bus travel, demand would need to generate the precise patterns and heterogeneity we document, and results do not vary by measured social desirability (Dhar et al., 2022) or by participants' beliefs about the study's purpose.

that the policy led to increases in *overall* female ridership, and that this was concentrated among highly educated women.

What explains the differences in efficacy between the small- and large-scale versions of making bus travel free for women in Delhi? Our framework points to three candidate explanations. First, this context seems well-suited for policy idempotence: making bus travel free a second time should not have an additional effect from making it free initially. Second, the government coupled the price reform with a massive expansion of guards on buses called "bus marshals," unarmed, uniformed officials with limited formal training intended to promote safety. This component of the policy was completely absent from our small-scale version. Third, like many pilots, our experiment mimicked the main treatment while differing along key dimensions of implementation. The free bus pass we distributed made bus travel free just like the policy but was hand-delivered to each participating woman at her doorstep in the form of a personalized card with their name, age and photograph. By contrast, when implemented at scale, women simply show up on the bus and receive a pink ticket allowing them a zero-fare ride.

Our results indicate neither general equilibrium effects nor the bus marshal expansion explain the low-efficacy of the at-scale policy. First, we find no change in bus crowding based on data we collected from bus stops. Second, we confirm in our sample that low-income women who used the bus *noticed* a sharp increase in female bus riders during their trips. This marks a key pre-requisite for the hypothesized "safety-in-numbers" virtuous cycle whereby more female riders would make riding buses feel safer for more women. Yet we find no evidence this increase (driven by higher-educated women outside of our main sampling frame) affected ridership in our sample. Combined with no detectable effect on low-income women's bus usage from the policy, this means that in this context we find no direct support for the "safety-in-numbers" argument that partly motivated the government reform.

Third, we examine how the sharp and sudden increase in bus marshals—uniformed guards with minimal training who were explicitly dedicated to monitoring safety onboard buses—impacted women's travel behavior. We start by confirming through data collected in 50 randomly drawn bus stops across Delhi that bus marshals were nearly always visible in buses after the policy launch, increasing in presence from fewer than 20 percent of the buses prior to the policy to close to 100 percent after the policy. We then use two difference-in-differences designs based on types of buses and times of the day that already had bus marshals before the policy. We find no effect of adding marshals on the number of women traveling along that route.[3] The point estimates are negative and moderately imprecise: we

[3]In this analysis, based on the number of women observed in each passing bus by the surveyor stationed in the bus stop, we cannot distinguish between women based on income and education.

can reject at the 95% level an increase in female ridership due to bus marshals equal around half of the overall policy impact on female ridership. These results are consistent with bus marshals not meaningfully affecting safety, or safety not being a decisive factor for ridership at the margin.

Instead, our preferred explanation is that implementation details played a key role in the success of our pilot over the government's roll-out. The bus passes were tangible, personalized with the beneficiary's photo and other details, and the research team delivered them to the study participants. These features likely made free bus travel more salient than the citywide policy.[4]

The persistence results are also consistent with the bus pass intervention successfully "activating" the policy for a subgroup of women. In this scenario, when these women ride the bus, they notice and benefit from the city implementation of the policy, which leads to further persistence in their behavior. Through the lens of our model and additional assumptions, *full* persistence means that none of the women induced to travel more by the bus pass are also induced to travel more by the city policy.

This paper contributes to the literature on the "science of scaling," which documents that the effects of interventions often shrink when programs move from pilot to scale and asks why (DellaVigna and Linos, 2022; Bold et al., 2018; Muralidharan and Singh, 2020; Vivalt, 2020; Allcott, 2015). Proposed explanations include non-representative samples and situations, implementation fidelity, and equilibrium effects that emerge only at scale (Banerjee et al., 2017; Al-Ubaydli et al., 2017; Davis et al., 2017; List, 2022, 2024). The main methodological response has been to experiment at scale—randomizing across large units or saturation levels so that equilibrium effects are internal to the experiment (Crépon et al., 2013; Muralidharan et al., 2016; Muralidharan and Niehaus, 2017; Baird et al., 2018; Egger et al., 2022). We address the complementary case in which at-scale randomization is unavailable. Our design has the advantage of evaluating the pilot and the at-scale policy in the same population at the same time.

Our paper also relates to the literature on the role that program delivery plays in driving program impact, and how implementation affects key indicators of take-up and success (Duflo, 2017). For example, mailing cards with program information to beneficiaries increases the subsidy they receive (Banerjee et al., 2018) and door-to-door canvassing is significantly more effective at increasing voter turnout than direct mails or telephone calls (Gerber and

[4]As discussed above, we believe that misreporting due to experimenter demand effects is unlikely to explain our results. However, the results are potentially consistent with (real behavior) Hawthorne effect, whereby being part of the bus pass intervention made women travel more. We cannot tell whether such an effect is innate to the small scale of the policy, or whether other implementations of the citywide policy (e.g. where all women would receive a card in the mail) would also have similar impacts.

Green, 2000). Temporary incentives and nudges can change long-term behaviors (Thaler, 2018) in several settings (Hummel and Maedche, 2019; Hussam et al., 2022). This paper contributes to this literature by measuring how the personalized delivery of public transit subsidies can change longer-term travel behavior for women.

2 Econometric Framework

This section formalizes the method in a potential outcomes framework. We begin with a stylized description of the method in a setting where neither, one, or both of a small-scale (pilot) version and an at-scale (citywide) version of a policy are offered. We then outline a more general framework, define policy idempotency and scaling, and highlight the key assumptions and results. Finally, we introduce an additional assumption that the pilot and scaled up policy differ only in how they activate the same policy for individuals, and characterize further the impact of the pilot, with or without the scaled up policy.

2.1 Main Idea

Consider a population of agents i observed at times t. There are two versions of the same underlying policy: a *pilot* version (P) and a *scale-up* version (S). These versions may differ in terms of policy features and implementation.

Treatment status takes values in $\mathcal{D} = \{0, P, S, S\&P\}$, corresponding to no policy, the pilot only, the at-scale policy only, and both. Let $Y_i(d)$ denote potential outcomes under $d \in \mathcal{D}$, and $\mu_d \equiv E[Y_i(d)]$.

The substantive question is whether the at-scale policy makes the pilot redundant, that is, whether the at-scale policy subsumes the pilot. We formalize this as $Y_i(S\&P) = Y_i(S)$, i.e., adding the pilot on top of the at-scale policy has no effect on the outcome. A fortiori, this implies that the average treatment effect of the pilot on top of the scale-up is $\tau^{\text{post}} \equiv \mu_{S\&P} - \mu_S = 0$.

For identification, suppose S is implemented for the full target population and the researcher randomizes P on top of S within a random sample according to $T_i \in \{0, 1\}$. Under random assignment, and assuming no spillovers in the pilot sample, the difference in means $\hat{\tau}^{\text{post}} = \overline{Y}_{T=1} - \overline{Y}_{T=0}$ identifies τ^{post}. Rejecting $H_0 : \tau^{\text{post}} = 0$ implies that, in this population, the scaled-up policy S does not subsume the pilot version of the policy P.

Why might we reject $\tau^{\text{post}} = 0$? This can happen for two main reasons. First, it can happen when receiving two different instances of the policy is not the same as receiving a single one (policy idempotency fails). Second, it can happen when the at-scale version affects

outcomes differently (or not at all) for some individuals, relative to the pilot version (policy scaling fails). For example, there may exist a group of people who respond to the pilot but not the at-scale policy. Hence, $\tau^{\text{post}} = 0$ is a joint test of idempotency and scaling. If we assume on ex-ante grounds that the policy is idempotent, it becomes a test of scaling. The converse is not true, failed scaling does not always imply $\tau^{\text{post}} \neq 0$. Moreover, failed scaling does not pin down the average treatment effect of the at-scale policy.

We now set up a more detailed framework to explain the key assumptions and results.

2.2 General Framework

Treatment is a tuple of policy exposures, possibly repeated. Let $d_{it} = \big(N_{it}^P,\, N_{it}^S\big)$ denote the number of times individual i has been treated with P and with S, respectively, in period t. Throughout we focus on the case $N_{it}^S \leq 1$.

Let $\mathbf{d}_t = \{d_{jt}\}_j$ denote the vector of everyone's period-t treatments, and let $H_t = \{\mathbf{d}_s\}_{s<t}$ denote the history of past treatment assignments. Write $\mathbf{d}_{-it}$ the leave-one-out vector treatment assignments, which omits d_{it}. Potential outcomes are

$$Y_{it}\big(d_{it},\, \mathbf{d}_{-it},\, H_{it}\big),$$

so that i's outcome may depend on her own treatment tuple, on the contemporaneous assignment of everyone else, and on the entire past assignment history. This allows for spillovers and general equilibrium effects, and dynamic effects.

Definition 1 (Scaling)**.** The policy *scales* if

$$Y_{it}\big((N^P, 0),\, \mathbf{d},\, H\big) \;=\; Y_{it}\big((N^P - 1,\, 1),\, \mathbf{d},\, H\big)$$

for any $N^P > 0$, any leave-one-out treatment vector $\mathbf{d}$, and any history H.

Definition 1 defines scaling as saying that implementing S or P for a given individual does not affect their outcomes, holding history and all other treatment assignments fixed. As such, this definition of scaling allows for scale-up spillovers and general equilibrium effects: these op rate through $\mathbf{d}$ and H, which are held fixed on both sides.

Scaling may appear as an untestable assumption, because it de-couples the assignment of S to individual i from the assignment of S to the rest of the population, whereas any real-world scale-up moves i's own exposure and everyone else's exposure together. Our method allows us to test scaling under the additional assumption of idempotency.

Definition 2 (Idempotency)**.** The policy is *idempotent* if

$$Y_{it}\big((N^P, 0),\, \mathbf{d},\, H\big) \;=\; Y_{it}\big((1, 0),\, \mathbf{d},\, H\big)$$

for any $N^P > 0$, any treatment vector $\mathbf{d}$, and any history H.

Idempotency means that repeated policy doses beyond the first leave outcomes unchanged.

The two-period design. Consider a setting with $t = 0, 1$ in which the scale-up S is implemented for all individuals at $t = 1$, while small randomized pilots of P are implemented at $t = 0$ and at $t = 1$. Own treatments are therefore

$$d_{i0} = \big(N^P_{i0},\, 0\big), \quad N^P_{i0} \in \{0, 1\}, \qquad\qquad d_{i1} = \big(N^P_{i1},\, 1\big), \quad N^P_{i1} \in \{0, 1\}.$$

The pilot at date t is run within a random sample $\mathcal{P}_t$ of the population, and N^P_{it} is randomly assigned within $\mathcal{P}_t$. For now we take $\mathcal{P}_0$ and $\mathcal{P}_1$ to be disjoint, in order to abstract from dynamic effects of the first pilot on the second; we return to this issue later. We assume there are no spillovers within either pilot:

Assumption 1 (Pilot SUTVA)**.** For $i \in \mathcal{P}_t$,

$$Y_{it}\big(d_{it}, \mathbf{d}_{-it}, H\big) = Y_{it}\big(d_{it}\big).$$

Because Y_{it} carries a time index, Assumption 1 still permits spillovers from the scale-up S at $t = 1$, including general equilibrium effects: these shift the level of $Y_{i1}(\cdot)$ as a function, rather than entering through $\mathbf{d}_{-i1}$. Dropping H means that the treatment assignment of the pilot in the previous period has no impact on the individuals from the $t = 1$ pilot sample.

Define the two estimands of interest. The *pure pilot effect*, for $i \in \mathcal{P}_0$, is

$$\tau_P \;\equiv\; \mathbb{E}\big[Y_{i0}(1, 0)\big] - \mathbb{E}\big[Y_{i0}(0, 0)\big],$$

and the *pilot-over-scale-up effect*, for $i \in \mathcal{P}_1$, is

$$\tau_{P|S} \;\equiv\; \mathbb{E}\big[Y_{i1}(1, 1)\big] - \mathbb{E}\big[Y_{i1}(0, 1)\big].$$

An RCT within $\mathcal{P}_0$ at $t = 0$ identifies τ_P; an RCT within $\mathcal{P}_1$ at $t = 1$ identifies $\tau_{P|S}$. We first examine the interpretation of the second experiment.

In general, $\tau_{P|S} = 0$ is a *joint* test of idempotency and scaling: under both assumptions

this treatment effect is zero. A non-zero value could therefore reflect either returns to repeated treatment (failure of idempotency) or a failure of scaling. Imposing one of the two assumptions isolates the other.

Proposition 1 (Test of scaling)**.** *Suppose the policy is idempotent (Definition 2) and the pilot satisfies SUTVA (Assumption 1). If the policy scales (Definition 1), then* $\tau_{P|S} = 0$*. Consequently, rejecting* $\tau_{P|S} = 0$ *rejects scaling.*

The short proof is in Appendix A.1, alongside an extension to non-idempotent policies with known returns to scale.

Remark 1. The test is one-sided in the usual sense: scaling implies $\tau_{P|S} = 0$, but $\tau_{P|S} = 0$ does not imply scaling. Offsetting failures of the two conditions cannot be ruled out.

2.3 Interpretation in an Activation Model

We now impose additional structure on the mechanism linking policies to outcomes, to give a more precise interpretation to τ_P and $\tau_{P|S}$. We continue to work with the two-period design of Section 2.2, $t = 0, 1$, with S active only at $t = 1$.

We assume both P and S operate by *activating* the same policy for individuals, that is, making the policy operative for that individual. Activation may capture multiple mechanisms such as policy awareness, salience, or the delivery of an object necessary for participants to benefit from the policy.

Let $A_{it} = \alpha_i(d_{it}) \in \{0, 1\}$ denote i's activation status, with $\alpha_i(0, 0) = 0$. Activation depends on i's own exposure only, and the map α_i is stable across periods. Let $\mathbf{A}_t = \{A_{jt}\}_j$ denote the period-t activation profile, with $\mathbf{A}_{-it}$ its leave-one-out version. We assume activation is a latent state unobserved by researchers.

We assume that the treatment assignment profile only affects outcomes through the profile of activation.

Assumption 2 (Exclusion restriction)**.**

$$Y_{it}\big(d_{it},\, \mathbf{d}_{-it},\, H_t\big) \;=\; Y_{it}^0(\mathbf{A}_{-it}) \;+\; \beta_{it}(\mathbf{A}_{-it})\, A_{it}.$$

Spillovers and general-equilibrium effects are otherwise unrestricted, entering through $Y_{it}^0(\cdot)$ and $\beta_{it}(\cdot)$. Assumption 1 lets us suppress the dependence of Y_{it}^0 and β_{it} on $\mathbf{A}_{-it}$ for $i \in \mathcal{P}_t$, with the time index carrying the environment: at $t = 0$ no one is exposed, at $t = 1$ everyone is exposed to S.

Write

$$A_i^P \equiv \alpha_i(1, 0), \qquad A_i^S \equiv \alpha_i(0, 1), \qquad A_i^{PS} \equiv \alpha_i(1, 1).$$

Assumption 3 (Activation monotonicity)**.** $A_i^{PS} \geq \max\left(A_i^P,\, A_i^S\right)$ for every i.

Assumption 3 requires activation to be monotone in each version separately: an additional offer does not de-activate the policy for anyone. It does not preclude $A_i^{PS} = 1$ and $A_i^P = A_i^S = 0$.

Assumption 4 (Time-stable treatment effects)**.** $\beta_{i0} = \beta_{i1} \equiv \beta_i$ for every i.

Partition individuals by their triple $\left(A_i^P, A_i^S, A_i^{PS}\right)$ that measures their activation profile. Under Assumption 3 only five cells are non-empty:

$$\underbrace{\mathcal{G}_P : (1,0,1)}_{\text{pilot-only}},\quad \underbrace{\mathcal{G}_C : (0,0,1)}_{\text{joint-activation}},\quad \underbrace{\mathcal{G}_B : (1,1,1)}_{\text{both}},\quad \underbrace{\mathcal{G}_S : (0,1,1)}_{\text{scale-up-only}},\quad \underbrace{\mathcal{G}_N : (0,0,0)}_{\text{never}}.$$

Proposition 2 (Decomposition)**.** *Under Assumptions 2–4,*

$$\tau_P = \mathbb{E}[\beta_i \, \mathbf{1}\{i \in \mathcal{G}_P \cup \mathcal{G}_B\}]\,, \tag{1}$$

$$\tau_{P|S} = \mathbb{E}[\beta_i \, \mathbf{1}\{i \in \mathcal{G}_P \cup \mathcal{G}_C\}]\,, \tag{2}$$

and hence

$$\tau_P - \tau_{P|S} = \mathbb{E}[\beta_i \, \mathbf{1}\{i \in \mathcal{G}_B\}] - \mathbb{E}[\beta_i \, \mathbf{1}\{i \in \mathcal{G}_C\}]\,. \tag{3}$$

The short proof is in Appendix A.1, together with a second proposition characterizing the relative magnitudes, under additional assumptions.

The result shows that $\tau_{P|S}$ is the average treatment effect of individuals who are only activated when they receive P, multiplied by their share of the population. There are two categories of such individuals, first, those for whom the policy is activated by P but not S, and those for whom neither P nor S alone activates, yet both together do.

The pre-scale-up contrast picks up $\mathcal{G}_B$, who are activated by the pilot at $t = 0$ but would have been activated by the at-scale policy anyway, instead of $\mathcal{G}_C$. The difference-in-differences nets the two against each other, and recovers the both-activated effect only when $\mathcal{G}_C$ is negligible.

Note that (1) and (2) hold separately without Assumption 4, with β_{i0} and β_{i1} in place of β_i; time-stability is needed only to subtract them.

3 Setting and Data Sources

This section describes Delhi's policy, the data we collected on women's travel, and the other data sets we use in analysis.

3.1 Delhi's Bus Policy Reform – the Pink Ticket Scheme

In late 2019, Delhi's government implemented a landmark reform of the public transit system, the "Pink Ticket" scheme, which has two main components. In June 2019, the Chief Minister of Delhi announced that all public transport would be made free for women. The government confirmed that public buses would be made free for all women in early August and budgeted for the reform in late August 2019, when the launch date was announced. The policy was ultimately implemented starting on October 29, 2019. This policy change came in the lead up to the January 2020 Delhi elections with push to make buses affordable and safer for women. As of 2025, the scheme is still operational and has been adopted by other states such as Punjab, Tamil Nadu, Karnataka and Andhra Pradesh.

Under the first component of the policy, women can travel for free on all public buses in the Delhi system.[5] Instead of paying for a bus ticket or using a bus pass, women receive a free pink ticket from the bus conductor when they board. During the period we study, each ticket included the message from the Chief Minister saying, "I wish that you and your family prosper. When women move forward, they take the country forward.".

In the second component of the policy, the government orchestrated a simultaneous massive expansion of bus marshals across city buses. These marshals are unarmed, uniformed officials—90 percent of whom are men—with limited formal training.[6] Their mandate is to promote safety, including deterring and responding to sexual harassment. Data we collected at bus stops across the city and that we describe below shows that bus marshal presence increased from fewer than 20 percent of the buses prior to the policy to over 90 percent after the policy.

3.2 Panel Data on Women's Travel Behavior

Sample of low-SES women in Delhi. Our primary sample consists of 746 low-income women in Delhi (referred to as "Delhi women" going forward). Our focus on low- and middle-income areas is based on the city-wide policy design that aimed to make public buses accessible to those who could not afford it otherwise (Express, 2019). We recruited this sample in person using a three-step approach. First, we selected 25 low- to middle-income wards within 15 km of the central business district (CBD) in Delhi, oversampling those with lower levels of education based on the 2011 Census, and those with high- or medium-density

[5]This includes 3,781 buses operated directly by the Delhi Transport Corporation (DTC), and 1,808 "cluster" buses serving the city of Delhi and some neighboring areas within the Delhi National Capital Region (NCR).

[6]Tiwari, A. (2019). "13,000 marshals on Delhi buses as free rides for women take off." The Indian Express, October 29.

based on manual classification using satellite images. Second, within these neighborhoods, we constructed 100m $\times$ 100m grids and restricted to those no more than 500 meters from a bus stop. We manually classified grid building density based on satellite images and retained those with medium- and high-density. We then randomly selected four grids to include in the sample. Third, within each selected grids, we randomly selected eight buildings and one household randomly within the building. (See Figure A.1.)

We used the following eligibility criteria. We excluded households that owned at least one car and adult women who were students (and therefore already had access to discounted bus travel). Given the focus of our analysis where we wanted to study women's travel behaviors after the free bus access was provided to them, we only included women who reported a minimum baseline level of mobility, defined as reporting to travel outside of their neighborhood at least once every week on average. We then randomly selected one adult woman in the household to enroll in the study. (We over-sampled young women below 30.)

Sample of low-SES men in Delhi. For a quarter of the sampled households, we also surveyed a randomly selected adult man, giving us a sample of 100 men (referred to as "Delhi men" henceforth).

Non-Delhi women sample. To assess the impact of the city-wide policy relative to women in other cities, we also recruited 337 women from the 15 largest Hindi-speaking Indian cities to track mobility outside Delhi (referred to as "non-Delhi women"). To do so, we posted ads on Facebook directed at women from these cities. Participants filled out an online baseline survey and were later contacted by phone for bi-weekly phone surveys.

Table A.1 presents summary statistics of the Delhi and non-Delhi women sample. Women in the Delhi sample are on average 35 years of age, and almost three-quarters of the sample has at most a middle-school education. Around a third of women reported work as their primary activity, while the rest report being housewives. More than half of the respondents reported that they need permission from someone else in the household before they leave their house. Before the free bus policy, more than half of women in Delhi made a trip the day before, and around a quarter used the bus at least once during their trips. Women in Delhi make around 60 percent of their trips alone and feel "very safe" (or the highest perceived safety as opposed to "somewhat safe," "somewhat unsafe" or "not at all safe") on around half of their trips. Women from other cities recruited on Facebook have higher education than the main sample and live in households more likely to own a car. They travel more but make fewer bus trips and report feeling the highest levels of perceived safety during three-fourths of their trips.

The baseline household survey for the main Delhi samples was conducted in September 2019 and collected demographic information, baseline typical travel behavior, and stated

preferences related to bus attributes. The online recruitment survey for women from other cities took place in late September and early October. Appendix Figure A.2 shows the timeline of data collection.

Panel Phone Surveys. We followed up with our three samples (Delhi women, Delhi men, non-Delhi women) using nine rounds of short phone surveys between October 2019 and January 2020. Delhi men and non-Delhi women were first included in the phone surveys in the third round, around mid-October. We conducted four weekly rounds before the policy launch on October 29^{th} and five bi-weekly rounds after the policy started. Each phone survey started with questions about travel over the past five days, overall and by bus. The survey then included a detailed travel diary for the day before. In the travel diary, surveyors asked respondents to report information for all trips the day before in chronological order. For each trip, they collected data on trip purpose, destination, departure time, overall trip duration, whether the respondent traveled alone or with someone, and additional information if they used the bus. The surveyor then elicited details by "legs" of the trip, including its mode, wait time (if any), travel duration, and feeling of safety.[7,8] Individual rounds occasionally included additional question modules. The average survey response rate was 58 percent per round for our main sample of Delhi women, and 91% of respondents were reached at least once (and 82% were reached at least once in the second half of the period).

Endline Phone Survey. We conducted an endline survey with the main Delhi sample between May and August 2020. Due to the COVID-19 crisis, the survey was conducted entirely remotely by phone, with a response rate of 67%.

3.3 Bus Stop Observation and Survey Data

We use two data sets collected in bus stops across Delhi and neighboring cities.

Bus Stop Observation Data. To measure female bus ridership, crowdedness in buses, and overall bus supply, we collect data at 50 bus stops across Delhi, both before and after the policy. To select data collection locations, we first obtained the universe of bus stops from Google Maps and web-scraped the routes for the 150 most frequent bus lines in Delhi. We restricted attention to bus stops within 1 km of wards with low or middle-income residential areas, and which have high-frequency bus lines. We sampled 50 bus stops at random from

[7]We detect recall bias effects in the initial questions over the past five days, with fewer trips reported further back in time. For this reason, for most of the analysis, we use the trip diary data for the previous day.

[8]We also collected data on sexual harassment, both experienced and observed, during the in-person baseline survey as well as during phone surveys. Despite strict privacy protocols, the data we collected features large, significant variation based on the mode of data collection (in-person or on the phone) as well as strong survey round effects. As a consequence and for most of the analysis, we do not report these results, instead focusing on self-reported feelings of safety.

this set, stratifying on latitude deciles. Figure A.3 shows the sampling frame of bus stops and the 50 selected bus stops.

Surveyors were stationed at bus stops, recording data on every bus that came by, giving us a final sample size of approximately 72,000 bus observations. Each surveyor worked six-hour shifts (either morning 8 AM – 2 PM or afternoon 2 PM – 8 PM), alternating sides of the road, and recorded the number of women passengers on board and the crowding level. Surveyors were randomly allocated to a daily bus stop rotation schedule, which was then repeated each week.

Bus Stop Survey Data from Baylis et al. We also use survey data from bus stops in Delhi and in adjacent National Capital Region (NCR) areas of Gurgaon and Noida, collected and shared with us by Patrick Baylis, Michael Greenstone, Kenneth Lee, and Harshil Sahai. Each day, surveyors visited bus stops within a cluster of bus stops in a pre-specified random order, and conducted 2-3 surveys of people standing in line at each bus stop. Surveyors conducted a short survey that includes data on demographics, as well as a question on how many times in the past week the respondent had used a bus.

4 Effects of the Pilot

We begin by describing the bus pass pilot experiment we ran immediately before the launch of the city-wide Free Bus policy. We examine key determinants of women's mobility on the demand-side that we studied as part of the pilot (1) women's latent demand for traveling, (2) potential of persistence in travel behavior driven by habit formation, (3) information frictions and lack of exposure to constrain travel behavior, and (4) household-level constraints restricting women's movement.

4.1 Bus Pass Experiment

As the pilot we designed a randomized intervention to provide women in our main sample with a free monthly bus pass after the city-wide policy was announced but before it was implemented. The bus passes made it possible for women to travel for free on any Delhi public bus and for any distance. The non-transferable passes included women's photo, name, and the price of the pass (800 INR or around 11 USD). Our team delivered the bus pass in person to each recipient (or to their family if she was not available) several weeks after baseline data collection and two- to four-weeks prior to the at-scale policy implementation. During that delivery, the surveyor explicitly informed participants that the study team "neither

encourages nor discourages" bus use or any other mode.[9] All bus passes were delivered less than a month before the policy launch date, so they expired after the launch of the policy. Figure A.2 shows the timeline detailing the sequence of data collection, pilot bus pass experiment, and at-scale policy implementation.

At baseline, we randomized whether we collected participant details and a photo for the bus pass application. Among this group, we randomized whether the participant was selected to receive the bus pass. This gives us three random treatment arms: the pure control group where participants were unaware of the bus passes, the "anticipation" control group where women were told they may be randomly selected to receive a bus pass (and eventually they were not selected), and the treatment group where participants received the bus pass. The randomization was stratified within each grid cell.[10,11] Our team successfully delivered bus passes to 60 percent of women in the treatment group, and we present intent-to-treat (ITT) results throughout.

Table A.2 reports balance results on outcomes collected at baseline. The treatment groups are generally well balanced, except that women assigned to the treatment group appear slightly younger and are 7 percent more likely to travel at baseline, both only statistically significant at 10 percent level. This potential imbalance does not drive our results. Our main empirical specification is a difference-in-difference. Table A.3 uses a difference-in-difference specification and shows that phone survey response rates (our measure of attrition) are balanced by treatment status.

4.2 Impact of the Bus Passes

We measure the effects of the pilot policy on travel around the city, travel by bus, travel alone and travel safety.

Our experimental design includes three randomly selected groups. The indicator $BusPass_i$ captures participants randomized to receive the bus pass, while $AnticipControl_i$ is switched on for participants informed that they might receive a bus pass and their photographs taken, who eventually did not receive one. The omitted category consists of "pure control" participants who were never told about a bus pass.[12]

[9] See Appendix A.2 for the complete scripts used at baseline and when bus passes were delivered.

[10] Example of grid cells is shown in Appendix Figure A.1.

[11] As each grid cell has only eight households, in analysis below, we include ward fixed effects (32 households were sampled per ward).

[12] In order to purchase bus passes for study participants, our team collected photographs and other details at baseline. Participants were informed that they may later receive a bus pass, which may lead to anticipatory behavior (Hussam et al., 2022).

We test this hypothesis by estimating the following equation:

$$Y_{idt} = \beta Post_t \times BusPass_i + \gamma Post_t \times AnticipControl_i + \psi_i + \phi_t + \omega_d + \varepsilon_{idt}, \qquad (4)$$

where i indexes a woman in our sample, surveyed on calendar date d and observed over weeks indexed by t relative to the week in which passes were delivered in her neighborhood.

The terms ψ_i, ϕ_t, and ω_d are fixed effects for person, week relative to treatment delivery, and calendar date, respectively. We cluster standard errors at the individual level. Let Y_{idt} denote the mobility outcome.

The coefficient of interest is β: the change in mobility among women assigned to receive a free bus pass relative to the pure control group, after bus pass delivery relative to before. Because treatment assignment was random, β identifies the causal impact of receiving a bus pass on mobility, as measured by the pilot (τ^{pre}.

We also report event study specifications that include separate coefficients for each week relative to treatment, omitting the coefficient on the week prior to treatment:

$$Y_{idt} = \sum_{\substack{-3 \leq t' \leq 11 \\ t' \neq -1}} \left(\beta_{t'} BusPass_i + \gamma_{t'} AnticipControl_i \right) + \psi_i + \phi_t + \omega_d + \varepsilon_{idt} \qquad (5)$$

Once the citywide policy made bus travel free for all women in Delhi, the bus passes in our intervention were rendered obsolete. We measurethe differential effects before versus after the citywide policy roll out by estimating two versions of equation 4. First, we use data before the pilot and only post periods that occur prior to the citywide policy roll out (pre-policy or short term effects). In a second specification, we use all the data before bus passes were delivered and only post periods after the citywide policy roll out (post-policy or long term effect).

In the pilot, we find that the bus pass intervention led to substantial increases in reported mobility, indicating strong latent demand for travel among this sample of low-income women. Table 1 presents the main intent-to-treat results on travel the day before the survey. Panel A shows pre-policy results, when women in the control group did not yet have access to free travel. We classify a trip as a bus trip if any leg of the journey uses a public bus. While the average respondent in the control group reported taking .25 bus trips the day before, women assigned to receive a free monthly bus pass make .157 more bus trips a day, an increase of 63 percent relative to the control mean in the post period (column 1). We next measure the impact on exploring more of the bus network, using the data we collected on the specific bus routes respondents used. In column 2 we measure the impact on the respondent using a new bus route that they have not reported in previous phone surveys, and in column 3 on

the cumulative number of unique bus routes reported to date. The coefficients are positive but insignificant.

We next measure whether the increase in bus use translates to more mobility overall. Column 4 reports the impact on the overall number of trips (irrespective of mode) yesterday. The coefficient is positive, indicating an increase of 14%, but not statistically significant. We find a large and marginally significant impact on the number of trips taken alone, which goes up by 35% from a base of 0.45 trips in the control group. This is suggestive evidence of higher mobility independence. Together, these results indicate that in the pilot policy, having access to free travel using the free bus pass spurred significantly more mobility, especially on buses, inconsistent with the hypothesis of low latent demand for travel.

We do not find evidence that this increase in public transit use and urban travel came at the expense of women's safety. Street harassment remains prevalent throughout Delhi, and at baseline, 23% percent of respondents reported experiencing harassment during travel the previous day.[13] Despite the increased travel and associated exposure, women in the treatment group do not report feeling less safe while taking buses (column 6 of Table 1), and we find a similar result using an unconditional travel safety outcome (not shown). Safety is measured as the fraction of trips taken where women report feeling "very safe." However, these results come with the important caveat that harassment experiences can be extremely difficult to accurately capture, and respondents may under-report how unsafe they felt, though it is unlikely that this response bias would vary differentially by treatment status.

We do not find any systematic evidence of substitution away from other modes (Table A.4). We classify a dominant mode for each trip as the highest ranked mode used, where modes are ranked in the following (increasing) order: walking, scooter, rickshaw, bus, metro, car. The results show a statistically significant increase in trips with bus as the dominant mode, and no change otherwise. In terms of the types of new trips induced by the bus pass intervention, in Table A.5 we split trips by its primary purpose: commuting, social, chores, accompanying someone else, or medical. We do not see any impact on any individual category, suggesting that the increase in mobility affected different types of trips for different participants. This reflects initial mobility patterns, where the trips women make are relatively split between all these categories.

Overall, we find evidence that the pilot policy of offering low-income women in Delhi a free bus pass strongly increases their bus usage, without evidence of reductions in other travel modes, and makes them more likely to travel alone.

Low underlying demand for travel by women has been found to contribute to the gender

[13]Harassment ranges from staring that makes women feel uncomfortable, to inappropriate comments or gestures to groping or grabbing, all the way to an assault.

disparity in urban mobility. For example, obligations related to household production such as childcare or a lack of opportunities for employment and leisure outside the home may reduce the value of travel even if constraints such as price or safety do not bind. Our results indicate that this sample of women have latent demand for travel outside of their neighborhood and that the bus pass intervention enabled them to travel more.

4.3 Potential for Misreporting in Self-Reported Data.

Since we primarily rely on self-reported data on mobility, differential reporting bias by treatment status presents one concern with the interpretation of these results. For instance, women assigned to the treatment group who received the free bus pass may feel pressure or experimenter demand to report bus trips that did not happen after receiving the bus pass, potentially inflating their mobility relative to those in the control group. Alternatively, participants in the control group may put in less effort responding to the survey, thus omitting real trips from the survey.

While we cannot definitively rule out this possibility, we present three pieces of evidence against misreporting bias explaining our findings. First, we explicitly discouraged experimenter demand effects by telling participants in the bus pass experiment that "we *[the team conducting the study]* are neither encouraging nor discouraging you to travel by bus or by any other mode," both during the baseline survey and when the bus pass was delivered to each participant or their family. (See Appendix A.2 for the full scripts.)

Second, we do not find heterogeneity by measured social-desirability (Dhar et al., 2022). If some respondents in the treatment arm perceive that the research team desired more bus travel, that may spur them to report more trips. At endline, we administer the Marlowe-Crowne social desirability scale to all respondents, which captures the extent to which an individual cares about projecting a positive image of themselves to others and responds in a socially expected way. Table A.7 shows that individual-level variation in social desirability does not predict treatment effects on bus trips, inconsistent with respondents' desire for pleasing the researchers driving the results.[14]

Finally, in a related exercise, we do not find heterogeneity by participant beliefs about the purpose of the study, measured at endline, using the question: "Do you think we (JPAL) would like the women in our study to ride the bus more?" Table A.8 reports the results and randomization inference test of the sharp null that the bus pass intervention did not affect the response to this question.

[14]We only measured social desirability at endline. This test is only valid under the sharp null that treatment status did not affect the generic social desirability status of respondents. We test for this in column 1 in Table A.7 and find a randomization inference p-value of 0.77.

While none of these pieces of evidence on their own can rule out biases in self-reported mobility influencing our results, taken together, they suggest randomly receiving a free bus pass did not just change reporting behavior but also increased women's travel.

5 Pilot Impact After the Citywide Policy

The results so far show that the bus pass intervention had immediate and important effects on women's mobility, before the citywide policy came into effect. It is important to note that the effects of the pilot may be due to making bus travel free (the price effect), or other demand-side channels, such as making bus travel more salient. By design, the pilot experiment kept supply-side factors, such as crowding, safety and female bus ridership, statistically similar across treatment groups.

We next analyze the impact of bus passes after buses become free for all women in Delhi. After this point, the treatment and control groups face the same (zero) price of taking the bus, and the two groups continue to face the same supply-side factors. This helps us separate the price effect from other possible effects of the pilot bus pass, and assess whether the bus passes delivered to a sample of women continued to impact mobility of these women over time.

We find highly persistent effects of the bus pass intervention on mobility, even after the bus passes become obsolete. Panel B in Table 1 reports the post-policy results that pool the time periods prior to the date of the bus pass delivery, and the time periods after October 29, when the city policy was launched. We find positive treatment effects, with a 55 percent increase in bus travel the previous day for those who received the free bus pass relative to women in the control group, despite both groups now having access to free bus travel. We can now detect a marginally significant increase of 26% more unique bus routes in the group that received bus passes. There is also a statistically significant increase of 27 percent in overall trips yesterday with over 41 percent of more solo trips by women in the treatment group. As with the short-term effects, the increased travel does not come at the cost of safety concerns for these women. Figure A.4 shows the impact by phone survey round. Although noisy, these estimates show divergence from the control group despite both receiving the same zero price after the launch of the citywide policy. Similar to the short-term, we find no substitution away from other modes (Table A.4, panel B), and in fact an increase in using walking as a mode, probably due to its complementarity with bus usage. We find some evidence that the increase in travel is concentrated among chores related to the household and medical trips, with both coefficients with p-values between 0.09 and 0.11 (Table A.5, panel B).

5.1 Rejection of at-scale policy subsuming pilot.

The strong evidence for persistent effects of the free bus passes delivered as part of the experiment allow us to reject a null hypothesis where $\tau^{\text{post}} = 0$. Under this null hypothesis, the launch of the citywide (at-scale) policy would erase any role for the (pilot) bus passes. But we do not detect any change in the magnitude of the treatment effect on the main mobility outcomes, nor any trend over this period. Our results indicate a failure of the citywide policy in the sense that it did not render the similar pilot policy obsolete.

These results have two implications. First, other aspects of the pilot bus pass intervention in addition to the fare reduction mattered. For example, bus passes—as physical objects—may have increased the salience of bus travel. It could also be that the mode of delivery of the bus pass, a personalized card that was delivered in-person to recipients, led to the price reduction associated with the bus pass being more effective.

Second, it is possible that women lack knowledge about bus routes, bus schedule or some other feature of travel by bus. And the initial increase in bus travel may have exposed participants to such information because of which they could make the most of the citywide policy once it was launched (Larcom et al., 2017). Second, using the bus for the first time might erase an initial aversion towards buses, the "exposure" effect documented in Dean et al. (2025). And finally, it could also be that the personalized bus pass made it possible for previously mobility-constrained women, for example because of familial constraints, to now travel and that initial exposure enabled them to continue using buses. We discuss each of these possible factors in greater detail below.

Information We measured information about the bus network during the endline survey. Table A.6 first shows that response rates were statistically similar between the bus pass experiment group and the pure control (columns 1-2). The next two columns replicate our main result—a large, statistically precise, increase in daily overall number of trips outside the neighborhood—measured using a recall question for pre-COVID travel(The endline survey took place in mid-2020). We then report the impact on information about the Delhi bus network. We asked participants for all bus lines the respondent could identify and describe in terms of bus line number, a nearby bus stop, and the destination. Participants in the control group report 1.8 bus lines on average, and we cannot identify any effect of the bus pass experiment, with point estimates of -0.04 and 0.01 for our two specifications. We also do not detect any effect on the number of bus stops and bus line destinations. At the moment of the endline, bus pass recipients do not seem to have more information than those in the control group about Delhi bus lines.

Prior experience We check how the effects of the bus pass intervention vary with baseline bus usage.[15] Table 2 presents results on the impact of being randomly assigned to the treatment group separately by bus experience measured at baseline. To gain more power, we pool together all the phone survey data, including before and after the citywide policy launch. We classify respondents into three groups. First, we consider those women who—when listing their commonly frequented destinations—said that travel by bus was not an option. Among those who could travel by bus, we divide women into two additional groups: those who did use buses at least somewhat regularly and those who did not. We find small and insignificant impacts of treatment among those who initially could not have taken a bus to their desired location, which is a reassuring finding. Among those who did use the bus, we find effects of the bus pass among both groups, and we cannot reject equality. However, the "Low Bus at Baseline" group has much lower values of the all mobility outcomes in the control group. For example, for this group, the bus pass intervention increases the number of trips yesterday by 43%, compared to 13% for the "High Bus at Baseline." In other words, we find much larger *proportional* effects among participants with low bus use at baseline.

5.2 Household constraints

The household has been identified as potential source of women's mobility restrictions. In particular, male household members may monitor or control women's movement and travel (Jayachandran, 2021; Jayaraman and Sneha, 2026). To assess the scope for this constraint, we examine heterogeneity in the treatment effects of providing women with free bus passes by two baseline measures of their freedom to travel within their household: whether women need permission from someone else to leave the house and travelling alone. While the first proxy more directly targets the notion of within-household travel restrictions, the second is a revealed-preference measure.

Across both measures, we find strong evidence suggesting that the women facing lower intra-household restrictions on travel were also those able to capitalize on the free bus pass. Tables A.9 and A.10 show that the positive impact on the number of bus trips taken yesterday is entirely concentrated among women who do not need to obtain permission from another household member to travel, and among women who report some typical travel alone at baseline. For the impact on trips and trips made alone, we again see significant impacts of the experiment for women who do not need permission and those who sometimes travel alone, although we can no longer reject that there is no heterogeneity by these measures.

The specific pattern of results also help further allay concerns of misreporting of trips,

[15] Unlike Dean et al. (2025), we do not measure whether participants have *ever* used the bus in Delhi. In this sample, this seems likely to be the case for most of our sample.

as previously discussed. The finding of persistent effects on mobility, both for bus use and traveling alone, long after bus pass expired (Table 1), and these effects are concentrated among women for whom bus option exists for their baseline typical trips (Table 2), and among women with low household mobility constraints and who sometimes travel alone at baseline (Tables A.9 and A.10). In the endline survey, we find an effect on mobility just before COVID, but no effect on information about the bus network (Table A.6). This helps rule out an "effort" channel wherein the women who received the bus pass are willing to put in more effort to respond to our surveys, while the control group leaves out information, we would expect to see an effect on information.

Overall, these results suggest that while information and prior experience do not seem to be driving the persistent effects, household constraints play an important role.

6 Effect of the Citywide Free Bus Policy

We begin by studying the overall impact of the policy on women's mobility and then provide evidence to distinguish among the constraints. Because we combine evidence from multiple data sources, different results refer to different groups, such as low-income women in our survey, all women in Delhi, and low-education women.

6.1 The Overall Impact on Low-Income Women's Mobility

We first estimate the effect of the policy reform on bus usage and overall mobility among low-income women. We leverage the rich panel data on mobility we collected from phone surveys, excluding women assigned to the treatment group of the bus pass randomized intervention. We focus on the number of trips by bus, and the total number of trips, reported for the day before the survey, for three groups: our main sample of low-income Delhi women, women in other Hindi-speaking cities we recruited over Facebook as a comparison group (see section 3.2), and low-income Delhi men.

For low-income women in Delhi, who are intended policy beneficiaries, bus usage ebbs and flows, but we fail to detect a systematic increase in bus usage after the policy (Figure 1). The number of trips by bus hovers around 0.3 per day yesterday throughout the period for women in Delhi. Women in other cities have a significantly lower overall level of bus trips. Delhi men make more bus trips on average, and the variation over time does not appear to be correlated with that for Delhi women. Figure 2 repeats the same analysis using all trips rather than just bus trips. We fail to see an increase in travel for Delhi women after the policy launch. Mobility evolves very similarly for women from Delhi and from other cities.

Mobility for men in Delhi is higher and more stable.

Table 3 reports two types of analysis using the same outcome variables. First, we analyze pre-post differences for Delhi women (columns 1 and 4), which show significant decreases in bus and overall trips after the policy. The other columns use difference-in-differences specifications using as control women in other cities (columns 2 and 5) and men in Delhi (columns 3 and 6). When using women in other cities as control we find a negative coefficient and can reject an increase of 4 percentage points (or 10%) in bus use. The coefficient with Delhi men as control is positive but imprecise. For overall trips, we find a positive imprecise effect when comparing to women in other cities and a negative marginally significant effect when comparing to men. The assumption necessary for identifying the impact of the policy is that absent the policy, bus usage and overall trips for Delhi women would have evolved in parallel to mobility for the other groups. Based on the graphical evidence discussed above, none of the groups appears to be an excellent control group. Table A.11 repeats the analysis with a PPML specification, and using the fraction of bus trips as outcome.

Overall, in this analysis we lack a highly convincing comparison group for Delhi women, and our estimates are fairly imprecise. However, we fail to uncover any evidence that respondent in our sample of low-income women in Delhi reported making more bus trips or overall trips due to the policy. The fact we documented earlier that the positive mobility impact of the bus pass intervention persisted after the citywide policy is also consistent with no direct impact of the citywide policy (section 5). One interpretation of these results is that policy implementation details can be crucial, while another possibility is that general equilibrium effects of the citywide policy (such as crowding) reduced its effectiveness. We analyze this below.

Another possibility for our null results is that the citywide policy was not implemented in practice. This is not the case. In our survey sample, take-up of the free bus policy is very high. 89% of all Delhi women respondents who use the bus report taking advantage of the free pink bus ticket in the 6th round of the phone survey (mid-November), when we first asked the question. During the baseline survey in early September, 61% of respondents were already aware of the upcoming policy.[16]

6.2 The Heterogeneous Impact of the Free Buses Policy

Our results so far show null results of the policy among our survey sample of low-income women in Delhi. In this section, we show evidence that the policy led to an increase in female ridership *overall*, and that this effect is concentrated among women with at least 12 years of

[16]Moreover, we informed all phone survey respondents in the fourth round, just before the policy, of the upcoming policy.

education. We rely on two different datasets to triangulate and decompose the impact of the Delhi free bus policy on female ridership, and to measure impact on aggregate outcomes.

Bus Stop Observation Data. Data we collected at bus stops allows us to examine the impact of the policy on aggregate female ridership using a before-after design (see Section 3.3). Our key outcome of interest is the number of women passengers visible in a bus by the surveyor stationed at the bus stop. Our team of 11 surveyors spent 3,330 hours on 67 days observing buses across 50 different locations.

We document a sharp and steady rise in the number of women riding the bus after the policy launch (Figure 3). Relative to the two weeks before the policy, the number of women visible on the bus from the stop increases steadily from 5 per bus to around 6.9 per bus on average after the policy, a 37% increase (Table A.12). We do not detect any change in a qualitative indicator of crowding (panel B in Figure 3), and we find a small increase of 7% in bus supply (panel B in Figure A.5, and Table A.12).[17] Our results pass several plausibility checks. For example, the increase in female ridership is greater on air-conditioned (AC) buses, which were more expensive prior to the policy (panel A in Figure A.5). Initially, fewer women used the AC buses, while after the policy, usage is at par with non-AC buses.

Bus Stop Surveys. So far, we have shown a null effect of the policy among the sample of low-income women, and a positive effect for female bus ridership in general. To explore how these two results relate to each other, we use a separate data source to sheds more light on how the policy affected bus ridership by commuter characteristics. These data consist of surveys with bus passengers in bus stops across the greater Delhi region (see section 3.3).[18] Because these data capture passenger demographics, we have a better image of who travels by bus than in our bus stop observation data.

In Figure 4, we plot the reported number of bus trips in the past week by days relative to the policy launch, separately for women in Delhi with at least 12 years of education (high-education) and for those with less (low-education). High-education women travel slightly more at baseline, with a stable difference immediately prior to the policy (week 0 results include travel behavior during the past week, which includes days before the policy launch). High-education women have a large increase in bus usage around the policy launch. By contrast, there is no detectable change in bus usage for low-education women. These results

[17] This increase appears concentrated 1 and 2 weeks after the citywide policy launch, which coincided with an episode of driving restrictions (the "Odd-Even" policy) in Delhi.

[18] This data was collected as part of a study led by Patrick Baylis, Michael Greenstone, Kenneth Lee, and Harshil Sahai, and shared with us for this study. Each day, surveyors visited bus stops within a cluster of bus stops in a pre-specified random order, and conducted 2-3 surveys of people standing in line at each bus stop.

offer a potential explanation for our two previous sets of results: the policy led to an increase in ridership for more educated, richer women but not for low-education, low-income women.[19]

We now show that the increase in ridership was highly concentrated among high-education women in Delhi. The bus survey data includes observations for men, and in the adjacent cities of Noida and Gurgaon. Table 4 reports double- and triple-difference specifications. In the first two columns, we include women and men in Delhi, and compare ridership changes after the policy by gender, education, and their interaction, without and with bus stop fixed effects. (Interactions with "After" are also included but omitted for space, see Table A.13.) The increase in ridership after the policy loads onto the group of high-education women. The next two columns add data from outside Delhi, and add all interactions with Delhi. Once again, the increase in ridership loads onto women, who are high education, and from Delhi.

Overall, the results in this section show that the Delhi policy led to an increase in female bus ridership, and this was concentrated among high-education women. These results corroborate our earlier null result among our survey sample of low-income women, and offer a nuanced evaluation of the policy's impact.

6.3 Changes in Travel Safety

Safety is an important constraint to urban mobility for women in Delhi, and buses are typically perceived as less safe than other transport models (Borker, 2021). We now study the ridership impact of two dimensions of safety induced by the Delhi free bus policy: the increased presence of women on buses as a potential amenity (safety-in-numbers) and the presence of bus marshalls on buses.

6.3.1 The Role of Safety-in-Numbers for Low-Income Women

First, we study the "safety-in-numbers" hypothesis. The Delhi government hoped that its policy will have positive feedback effects due to increased female bus ridership. More female bus riders could make riding the bus safer for other women, potentially sparking a virtuous cycle.[20]

[19] This analysis has two limitations. First, these results are conditional on a woman's presence in the bus stop. Hence, they capture an intensive margin and a selection margin, but would not detect an extensive margin change in female bus ridership. Second, our survey sample was selected based on home neighborhood and household characteristics, while the bus stop survey data recorded education.

[20] Increased bus crowding might have the opposite effect. However, panel B in Figure 3 and Table A.12 show that we cannot detect any change in perceived by crowding based on assessments of the surveyors in bus stops.

We do not find evidence for this mechanism for women in our low-income sample. We find an increase in total female bus ridership that was anticipated, realized, and observed, yet this does not appear to have prompted any more bus travel among low-income women. In the previous section, we showed increases in female bus ridership, concentrated among educated women. While we have no evidence that Delhi women in our sample traveled more (by bus or in general) due to the policy, we find that those who did travel by bus observed an immediate and substantial increase in the number of *other* women passengers on buses (Figure 5). Conditional on taking the bus, this number jumps up from around 4 women passengers per bus before the policy to around 5.5 women passengers per bus after, an increase of around 35%. In other words, we can detect a clear first stage for the safety-in-numbers mechanism.

We have additional evidence that most women in our sample expected an increase in female bus ridership. In the fourth round of phone surveys—just before the policy launch—64 percent of our respondents believed that *other* women would travel more after the policy starts, while only 31 percent reported that they themselves would travel more.[21]

These results show that on buses used by women in the low-income sample, female bus ridership rose—an increase both anticipated and observed by women in our survey. Combined with the null effect on women's own bus use in this sample, this is consistent with the "safety-in-number" factor not being a binding mobility barrier for low-income women. Of course, our estimates may be imprecise, and it is possible that women respond to such changes over a longer time horizon or for even larger changes in female bus ridership than what we observe here.

6.3.2 The Impact of Bus Marshals on Aggregate Female Bus Ridership

The policy combined two main elements: making buses free for women and adding bus marshals to essentially all public buses in Delhi.

In order to separate the effects of these two components on overall female bus ridership, we leverage the fact that prior to the policy, bus marshals were concentrated on a certain type of buses and at certain intervals of the day. Prior to the policy, around half of A/C and non-A/C buses had marshals after 2:00 P.M., while there were no marshals before 2 P.M., and there were no marshals on cluster buses at any time of the day. This heterogeneity in the coverage of the bus marshal program is shown graphically in Figure A.7 and in column 1 in Table 5.

[21] We asked women to report their typical weekly bus use in the two preceding months and to forecast their bus use after the policy would come into effect. Then, we asked their beliefs about how other women in our study answered the same questions.

Immediately after the policy launch, bus coverage jumps to above 90% for essentially all times of the day and types of bus. We use the differential change in the presence of bus marshals to investigate whether women's bus ridership increased more for buses and time intervals where the increase in bus marshals was larger.

We perform two exercises. First, among non-cluster buses, we compare the policy impact on women's ridership before and after 2:00 P.M. (panel A, column 2 in Table 5). While the increase in bus marshals was 39 percentage points higher in the morning interval, the change in women's ridership is not higher in this period (the point estimate on Post $\times$ Before 2 P.M. is in fact negative and not statistically significant in column 2). We then combine these two results to perform an implied I.V. regression and find that the implied effect of going from 0% to 100% bus marshals has no statistically significant effect on women's ridership (column 3). The main limitation of this analysis is the possibility that the bus marshal expansion policy had spillover effects to other times of the day. To address this, we perform an analogous exercise among observations after 2:00 P.M., comparing the policy impact for cluster and non-cluster (A/C or non-A/C) buses. Cluster buses had a larger increase in bus marshals but no larger increase in women's ridership (panel B in Table 5). Our estimates are not very precise, yet for both analyses we can reject at 5% level an increase due to bus marshals around half of the size of the main effect of the policy.

Overall, these results offer no evidence that bus marshals *per se* increased women's bus ridership after the policy launch. The main limitation of this analysis is the possibility that the bus marshal expansion policy had spillover effects to other times of the day or other types of buses. For example, if women's travel always combines a morning trip with an afternoon trip (e.g. for commuting), then the A.M. and P.M. decisions may be interlinked. Hence, we may not observe a *differential* effect, even if bus marshals did make one the time periods safer. In our survey sample, commuting trips only account for less than a third of all trips, which suggests that (at least among lower SES women in Delhi), not all trips are of this type. Our second design would be invalidated by spillovers across different types of buses, from cluster buses to non-cluster buses. While this is more difficult to motivate, we cannot rule it out definitively.[22]

[22]As a complement to these quasi-experimental results, in our main panel sample we designed and implemented an belief experiment leveraging the sudden expansion of bus marshals. The goal of the experiment was to randomly correct respondents' beliefs about the prevalence of bus marshals at their local bus stops. However, we find no evidence of changing beliefs about the prevalence of bus marshals. (We also do not find changes in beliefs about safety levels or number of female passengers or on actual travel behavior.) Hence, the results from this experiment are inconclusive.

7 Conclusion

This paper studies whether a successful small-scale intervention retains its effectiveness when a similar policy is implemented at scale. We develop a simple approach for idempotent policies: if the at-scale policy fully reproduces the benefit of the pilot, adding the pilot on top should have no additional effect. We apply this idea to Delhi's citywide policy making bus travel free for women.

Before the citywide rollout, a personalized free bus pass substantially increased bus use and mobility among low-income women. These effects persisted essentially unchanged after buses became free for all women, even though the experimental passes had become financially obsolete. In the language of our framework, the citywide policy did not subsume the pilot for this population. Complementary evidence also shows little increase in mobility among low-income women following the citywide policy.

The contrast between the pilot and the at-scale policy highlights the importance of implementation. Both eliminated fares, but the pilot was tangible, personalized, and individually delivered. Our results are therefore consistent with salience, personalization, or related features of delivery helping activate the policy for beneficiaries. More broadly, for idempotent policies, universal eligibility need not imply that a policy is effectively reaching its intended population. A randomized pilot layered onto a large-scale policy can provide a simple diagnostic of whether scaling has preserved the benefits of the smaller intervention.

The results also shed light on constraints to women's mobility. We find little evidence that the citywide expansion of bus marshals or an increase in female ridership meaningfully raised mobility among low-income women. By contrast, the personalized bus pass increased travel, particularly among women facing fewer household restrictions. Free transit policies can unlock substantial latent demand, but implementation shapes whether women's behavior is affected.

References

Al-Ubaydli, O., J. A. List, and D. L. Suskind (2017): "What Can We Learn from Experiments? Understanding the Threats to the Scalability of Experimental Results," *American Economic Review: Papers & Proceedings*, 107, 282–286.

Allcott, H. (2015): "Site Selection Bias in Program Evaluation," *Quarterly Journal of Economics*, 130, 1117–1165.

Baird, S., J. A. Bohren, C. McIntosh, and B. Özler (2018): "Optimal Design of Experiments in the Presence of Interference," *Review of Economics and Statistics*, 100, 844–860.

Banerjee, A., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukerji, M. Shotland, and M. Walton (2017): "From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application," *Journal of Economic Perspectives*, 31, 73–102.

Banerjee, A., R. Hanna, J. Kyle, B. A. Olken, and S. Sumarto (2018): "Tangible information and citizen empowerment: Identification cards and food subsidy programs in Indonesia," *Journal of Political Economy*, 126, 451–491.

Bold, T., M. Kimenyi, G. Mwabu, A. Ng'ang'a, and J. Sandefur (2018): "Experimental Evidence on Scaling up Education Reforms in Kenya," *Journal of Public Economics*, 168, 1–20.

Borker, G. (2021): "Safety First: Perceived Risk of Street Harassment and Educational Choices of Women," *World Bank Policy Research Working Paper 9731*.

Crépon, B., E. Duflo, M. Gurgand, R. Rathelot, and P. Zamora (2013): "Do Labor Market Policies Have Displacement Effects? Evidence from a Clustered Randomized Experiment," *Quarterly Journal of Economics*, 128, 531–580.

Davis, J. M. V., J. Guryan, K. Hallberg, and J. Ludwig (2017): "The Economics of Scale-Up," Working Paper 23925, National Bureau of Economic Research.

Dean, J. T., G. Kreindler, and O. Mbonu (2025): "The Effect of Exposure: Evidence from Spatial Choices in Nairobi," *Working paper*.

DellaVigna, S. and E. Linos (2022): "RCTs to Scale: Comprehensive Evidence from Two Nudge Units," *Econometrica*, 90, 81–116.

Dhar, D., T. Jain, and S. Jayachandran (2022): "Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India," *American economic review*, 112, 899–927.

Duflo, E. (2017): "The economist as plumber," *American Economic Review*, 107, 1–26.

Egger, D., J. Haushofer, E. Miguel, P. Niehaus, and M. Walker (2022): "General Equilibrium Effects of Cash Transfers: Experimental Evidence from Kenya," *Econometrica*, 90, 2603–2643.

Express, I. (2019): "Arvind Kejriwal announces free metro, bus rides for women in Delhi," The Indian Express.

Gerber, A. S. and D. P. Green (2000): "The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment," *American political science review*, 94, 653–663.

Hummel, D. and A. Maedche (2019): "How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies," *Journal of Behavioral and Experimental Economics*, 80, 47–58.

Hussam, R., A. Rabbani, G. Reggiani, and N. Rigol (2022): "Rational habit formation: experimental evidence from handwashing in India," *American Economic Journal: Applied Economics*, 14, 1–41.

Jamba, H., A. Devaraj, and C. Kanuri (2025): "Fare-free Bus Travel Scheme for Women: Lessons from Delhi," Working paper, World Resources Institute (WRI) India.

Jayachandran, S. (2021): "Social norms as a barrier to women's employment in developing countries," *IMF Economic Review*, 69, 576–595.

Jayaraman, R. and P. Sneha (2026): "In-Law Mortality and Women's Labor Supply in India," Working paper, July 24, 2026.

Larcom, S., F. Rauch, and T. Willems (2017): "The benefits of forced experimentation: Striking evidence from the London underground network," *The Quarterly Journal of Economics*, 132, 2019–2055.

List, J. A. (2022): *The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale*, New York: Currency.

——— (2024): "Optimally Generate Policy-Based Evidence Before Scaling," *Nature*, 626, 491–499.

Muralidharan, K. and P. Niehaus (2017): "Experimentation at scale," *Journal of Economic Perspectives*, 31, 103–124.

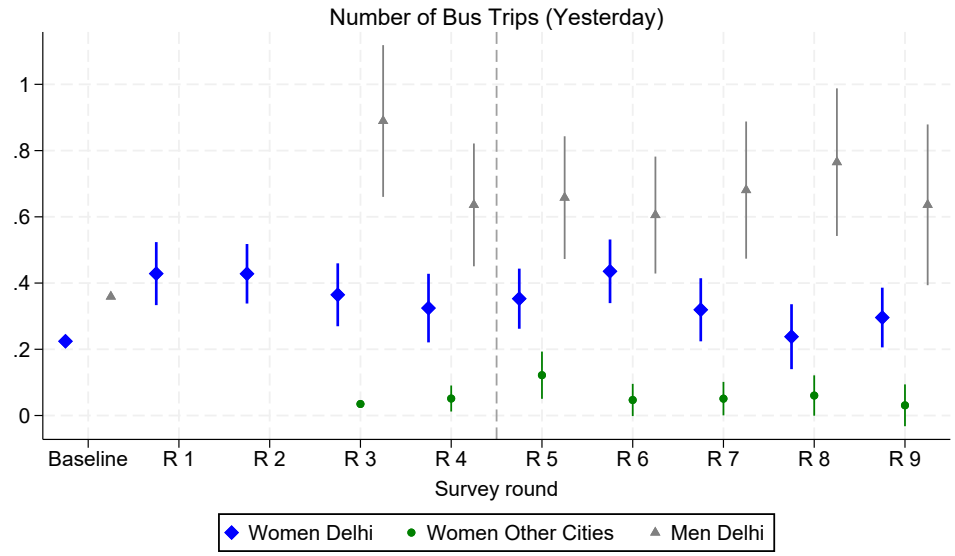
Muralidharan, K., P. Niehaus, and S. Sukhtankar (2016): "Building State Capacity: Evidence from Biometric Smartcards in India," *American Economic Review*, 106, 2895–2929.

Muralidharan, K. and A. Singh (2020): "Improving Public Sector Management at Scale? Experimental Evidence on School Governance in India," *NBER Working Paper 28129*.

Thaler, R. H. (2018): "From cashews to nudges: The evolution of behavioral economics," *American Economic Review*, 108, 1265–1287.

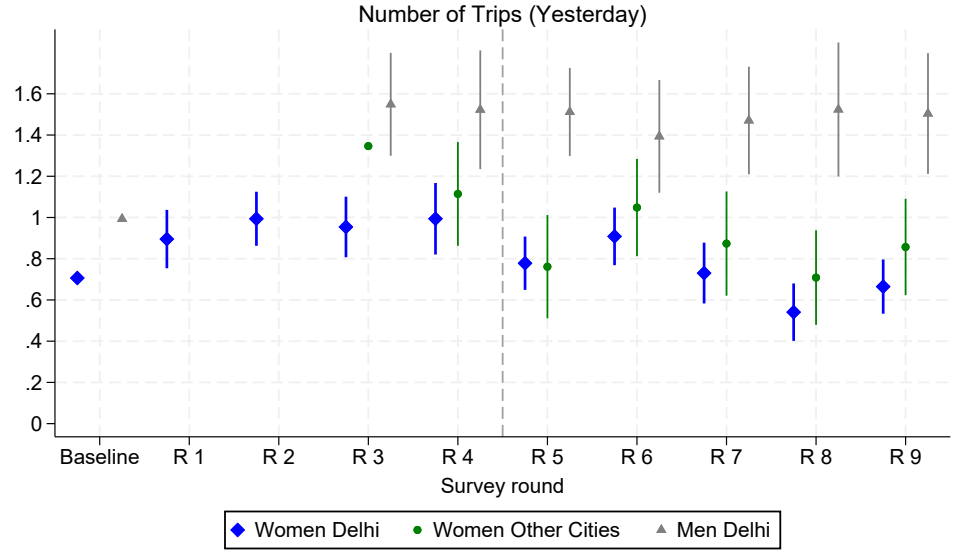
Vivalt, E. (2020): "How Much Can We Generalize from Impact Evaluations?" *Journal of the European Economic Association*, 18, 3045–3089.

Figure 1: (Lack of) Impacts of Free Bus Policy on Low-Income Women's Bus Usage



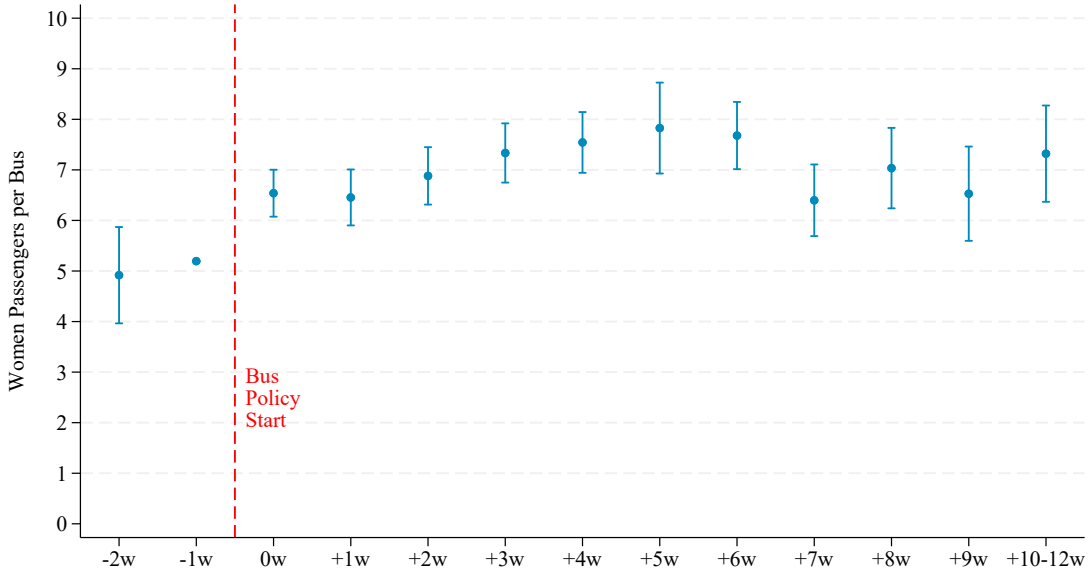
Note: Figure 1 shows results from three regressions of the number of trips by bus yesterday on indicators for the baseline survey and for phone survey round. The Delhi Free Bus policy was launched between the fourth and fifth survey rounds. All regressions include individual fixed effects. The figure reports results from three samples: the main sample of women in Delhi (the sample only includes the control group from the bus pass intervention), women in other cities, and men in Delhi. For each group, the reference time period is displayed without confidence intervals. Confidence intervals based on standard errors clustered at the individual level.

Figure 2: (Lack of) Impacts of Free Bus Policy on Low-Income Women's Mobility

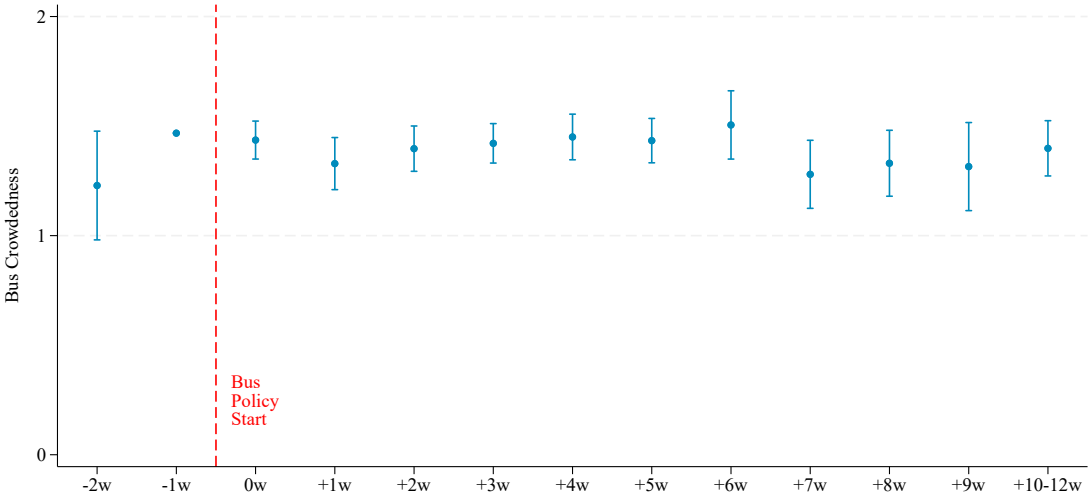


Note: Figure 2 repeats the analysis from Figure 1 using the overall number of trips yesterday.

Figure 3: Impacts of Free Bus Policy (Bus Stop Observations)



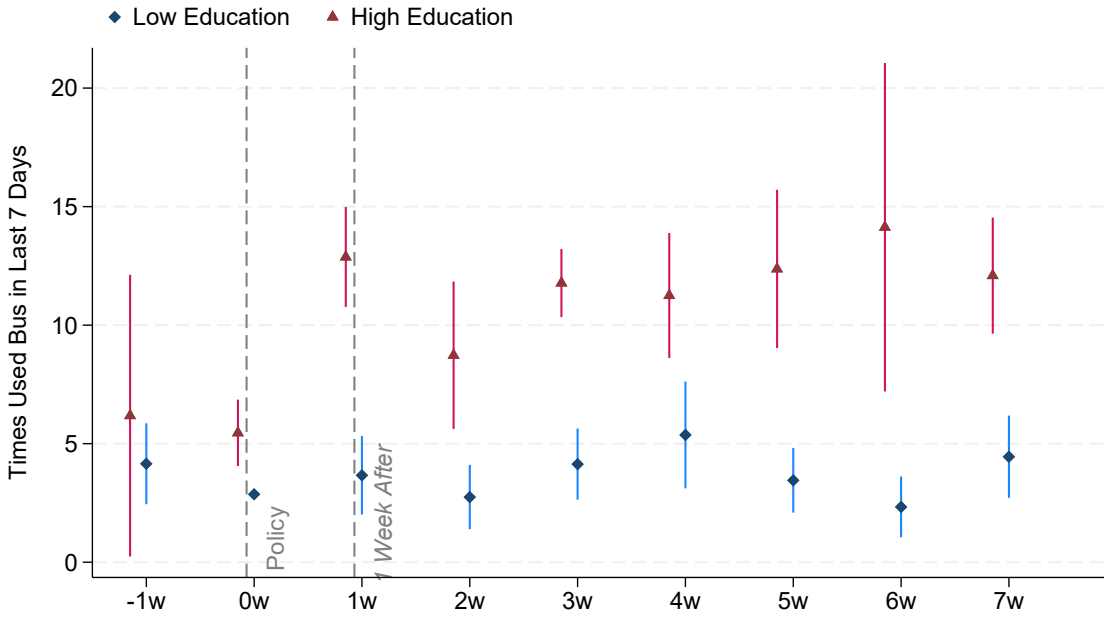
Panel (A) Impact on Female Passengers



Panel (B) Impact on Bus Crowdedness

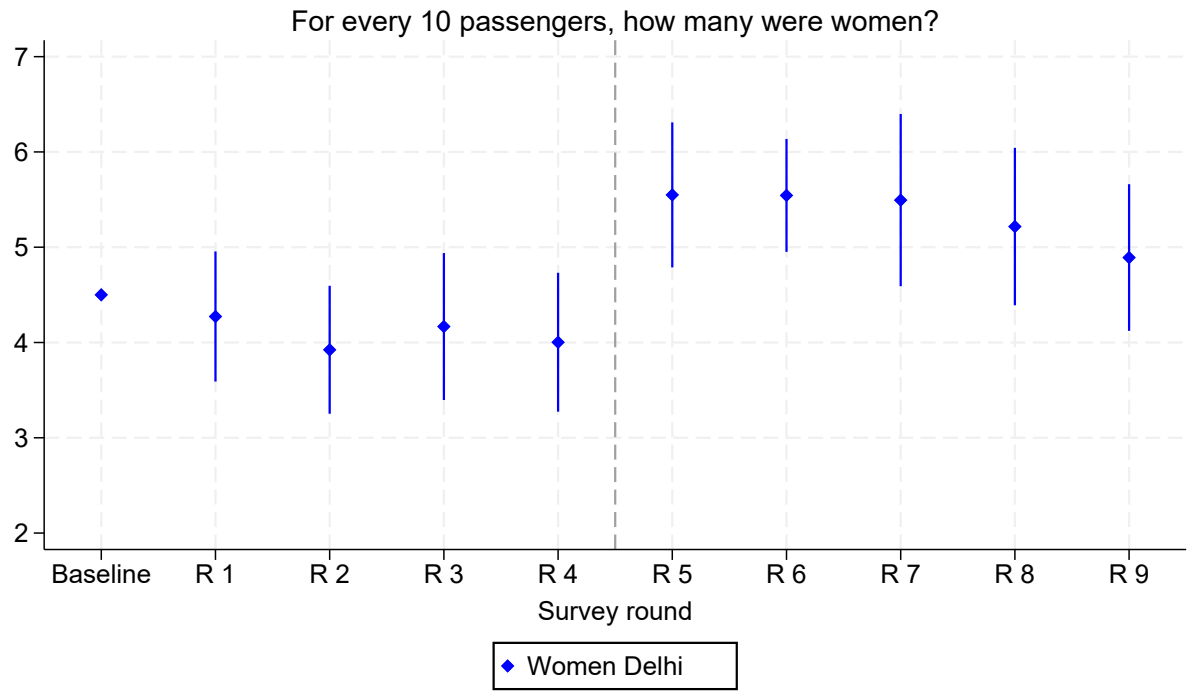
Note: Figure 3 shows the number of women visible in a bus from the bus stop (panel A) and bus crowdedness (panel B), by week relative to the policy roll-out, using the data collected at bus stops. The graphs plot week indicator coefficients in a specification with bus stop fixed effects. Bus crowdedness is categorized by surveyors on a sclae from 0 to 4, where 0 means "Many seats available," 1 means "Few seats available," 2 means "Standing room only," 3 means "Crowded standing room only," and 4 means "Crushed standing room only." Standard errors are clustered at the level of 50 bus stops.

Figure 4: Impacts of Free Bus Policy on Female Bus Usage by Education



Note: Figure 4 plots the coefficients from a regression of self-reported bus use over the past week, collected at bus stops across Delhi, on week indicators interacted with low- and high-education. The sample is women and bus stops in Delhi. High education is defined as 10th standard or more. Figure A.6 shows results from a specification that also includes bus stop fixed effects.

Figure 5: "Safety-in-Numbers" First Stage



Note: Figure 5 shows the experienced fraction of bus passengers who are women, by survey round. The Delhi Free Bus policy was launched between the fourth and fifth survey rounds. This data is reported by female respondents in the low-income Delhi sample for the bus rides they took. We run a regression on survey round indicators and respondent fixed effects. Standard errors are clustered at the respondent level.

Table 1: Impacts of Providing Free Bus Passes on Mobility

Panel A: Pre-Policy (Short Term)

	(1) Bus Trip	(2) New Bus Route	(3) Cum. Bus Routes	(4) Trip	(5) Trip Alone	(6) Safe | Bus
Treat × After	0.157**	0.054	0.113	0.129	0.159*	-0.083
	(0.062)	(0.051)	(0.072)	(0.093)	(0.085)	(0.176)
Anticipation control × After	0.044	0.016	0.046	0.106	0.031	-0.060
	(0.069)	(0.061)	(0.073)	(0.104)	(0.095)	(0.184)
Pure Control After Mean	0.25	0.11	0.53	0.91	0.45	0.47
Pure Control Respondents	164	164	164	164	164	25
Resp. Anticipation Control	146	146	146	146	146	30
Resp. Treated	316	316	316	316	316	69
Observations	2,276	2,276	2,276	2,276	2,276	337

Panel B: Pre-Treatment + Post-Policy (Long Term)

	(1) Bus Trip	(2) New Bus Route	(3) Cum. Bus Routes	(4) Trip	(5) Trip Alone	(6) Safe | Bus
Treat × After	0.154***	0.054	0.235*	0.190***	0.156***	0.017
	(0.050)	(0.043)	(0.130)	(0.071)	(0.060)	(0.100)
Anticipation control × After	0.020	0.031	0.080	0.117	0.038	-0.015
	(0.061)	(0.050)	(0.133)	(0.085)	(0.072)	(0.124)
Pure Control After Mean	0.28	0.09	0.91	0.70	0.38	0.57
Pure Control Respondents	170	170	170	170	170	31
Resp. Anticipation Control	152	152	152	152	152	43
Resp. Treated	342	342	342	342	342	97
Observations	3,593	3,593	3,593	3,593	3,593	598

Note: Table 1 shows the impact of receiving a bus pass on mobility. "Treat" denotes whether the respondent received a bus pass, and "Treat × After" indicates that the outcome is measured after the respondent receives the pass. "Anticipation control" indicates that the respondent was told they may be receive a bus pass but ultimately did not. Panel A shows impacts before the citywide policy launch. Panel B shows impacts excluding the period between bus pass distribution and citywide policy launch but including outcomes after the policy launch. Column (1) shows impacts on the number of trips on the day before the respondent was interviewed. Column (2) shows impacts on whether the respondent reports a bus route for the first time, conditional on riding the bus. Column (3) shows impacts on the cumulative number of bus routes taken so far. Column (4) shows impacts on the number of trips taken outside of the neighborhood yesterday, including all modes of transportation. Column (5) shows impacts on the number of trips taken *alone* outside of the neighborhood yesterday. Column (6) shows impacts on the share of trips the day before that the respondent perceived as safe conditional on riding the bus. Standard errors in parentheses are clustered by respondent. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Heterogeneous Impacts of Providing Free Bus Passes by Baseline Bus Use

	(1) Bus Trip	(2) Green Bus	(3) Red Bus	(4) Trip	(5) Trip Alone	(6) Safe | Bus
Treat × After × Bus Not Possible	0.07	0.05	0.02	0.09	0.13	-0.35
	(0.05)	(0.05)	(0.02)	(0.11)	(0.09)	(0.42)
Treat × After × Possible × Low Bus at Baseline	0.19***	0.19***	0.01	0.28**	0.12	-0.16
	(0.07)	(0.06)	(0.03)	(0.12)	(0.10)	(0.26)
Treat × After × Possible × High Bus at Baseline	0.16**	0.11	0.07**	0.12	0.16*	0.06
	(0.08)	(0.07)	(0.03)	(0.09)	(0.08)	(0.09)
Control Mean: Not Possible	0.08	0.06	0.02	0.59	0.28	0.62
Control Mean: Possible Low Bus	0.09	0.08	0.02	0.64	0.25	0.50
Control Mean: Possible High Bus	0.52	0.43	0.09	0.96	0.58	0.60
P-val Low = High	0.77	0.39	0.15	0.31	0.77	0.41
Respondents: Not Possible	134	134	134	134	134	14
Respondents: Possible Low Bus	183	183	183	183	183	24
Respondents: Possible High Bus	332	332	332	332	332	167
Observations	4,612	4,612	4,612	4,612	4,612	827

Note: Table 2 shows heterogeneous treatment effects of the bus pass experiment by baseline bus use. The sample includes all phone surveys rounds, both before and after the citywide policy launch. All specifications control for Anticipation control × Post interacted with the heterogeneity variables, but the coefficients are omitted for space. "Bus Not Possible" indicates that the respondent reported that it was not possible to travel by bus in their neighborhood. We measure baseline bus usage using an index equal to the sum of three indicators: whether the respondent uses the bus weekly based on the household roster, if the respondent reports using the bus in the past month, and if the respondent reports using a bus for at least one of its typical trips. We classify a respondent as "Low Bus at Baseline" if this index is 0 or 1, and "High Bus at Baseline" if the index is 2 or 3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Impacts of Free Bus Policy on Low-SES Women's Mobility

	# Bus Trips			# Trips		
	(1)	(2)	(3)	(4)	(5)	(6)
Post Policy	-0.06**			-0.25***		
	(0.03)			(0.04)		
Post Policy × Delhi Woman		-0.05	0.10		0.12	-0.21*
		(0.04)	(0.09)		(0.10)	(0.12)
Respondent FE	Yes	Yes	Yes	Yes	Yes	Yes
Sample: Delhi Women	Yes	Yes	Yes	Yes	Yes	Yes
Sample: Other Cities Women		Yes			Yes	
Sample: Delhi Men			Yes			Yes
Survey Rounds in Sample	1-9	3-9	3-9	1-9	3-9	3-9
Outcome Pre Mean	0.39	0.39	0.39	0.96	0.96	0.96
Observations	1,890	2,838	2,336	1,890	2,838	2,336

Note: Table 3 shows the pre-post and difference-in-differences impact of the Delhi citywide Free Buses policy on Delhi women's mobility, measured using the number of bus trips (columns 1-3) and the total number of trips (columns 4-6). Data from all nine rounds of phone surveys. All specifications include respondent fixed effects. Standard errors in parentheses are clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Impact of Free Bus Policy on Female Bus Usage

	Times Used Bus in Last 7 Days			
	(1)	(2)	(3)	(4)
After	1.8***	1.7***	0.4	-0.0
	(0.6)	(0.5)	(0.4)	(0.5)
After × Woman	-1.5	-2.8***	0.5	1.0*
	(0.9)	(0.9)	(0.6)	(0.5)
After × High-Ed	-1.3*	-0.7	-0.6	-0.5
	(0.8)	(1.0)	(0.6)	(0.7)
After × Woman × High-Ed	7.9***	5.1***	-0.1	-0.6
	(1.9)	(1.4)	(0.7)	(0.7)
After × Delhi			1.4**	1.7***
			(0.7)	(0.5)
After × Woman × Delhi			-2.0**	-3.9***
			(0.9)	(0.8)
After × Delhi × High-Ed			-0.7	-0.2
			(0.8)	(1.1)
After × Female × Delhi × High-Ed			8.1***	5.7***
			(1.8)	(1.2)
Bus Stop FE		Yes		Yes
Outcome Mean	7.76	7.77	5.67	5.68
Observations	1,803	1,800	2,970	2,957

Note: Table 4 reports double and triple difference results of the impact of the Delhi free bus policy by location, gender, and education level. This data was collected as part of a study led by Patrick Baylis, Michael Greenstone, Kenneth Lee, and Harshil Sahai. High education is defined as 10th standard or more. Columns 2 and 4 include fixed effects for each of the 421 bus stops visited during data collection. Specifications also include all direct effects, pairwise interactions, and all triple interactions in column 3 and 4. (Some coefficients are omitted here for space, see Table A.13.) Standard errors in parentheses clustered by date. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Impacts of Bus Marshals on Female Passengers

	Marshal	Women Passengers	
	(1)	(2)	(3)
Before 2 P.M.	-0.39***	-0.55	-1.36***
	(0.11)	(0.58)	(0.36)
Post	0.50***	2.97***	3.99**
	(0.11)	(0.56)	(1.65)
Post x Before 2 P.M.	0.44***	-0.89	
	(0.12)	(0.63)	
Bus Marshal			-2.03
			(1.87)
Specification	First Stage	Reduced Form	I.V.
Before 2 P.M. Mean	0.44	5.51	
Observations	42,893	43,009	42,872

Panel (A) Before vs after 2 P.M.

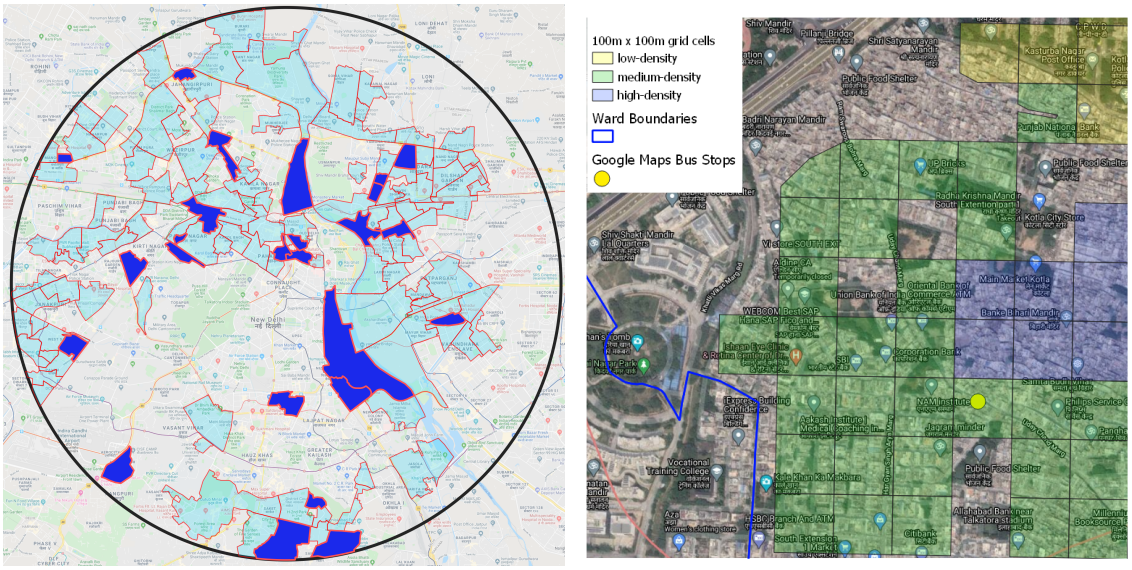
	Marshal	Women Passengers	
	(1)	(2)	(3)
Cluster Bus	-0.44***	1.46***	0.64**
	(0.11)	(0.40)	(0.24)
Post	0.50***	2.84***	3.82**
	(0.11)	(0.68)	(1.68)
Post x Cluster	0.37***	-0.69	
	(0.11)	(0.42)	
Bus Marshal			-1.93
			(1.65)
Specification	First Stage	Reduced Form	I.V.
Non-Cluster Mean	0.44	5.51	
Observations	34,359	34,442	34,342

Panel (B) Cluster vs non-cluster buses

Note: Table 5 shows difference-in-differences analysis of the impact of bus marshaps on female ridership, using bus stop observations data. Before the policy, bus marshals were active only in the afternoon and only on non-cluster buses. See Figure A.7 for the first stage. Standard errors are clustered at the level of 50 bus stops. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

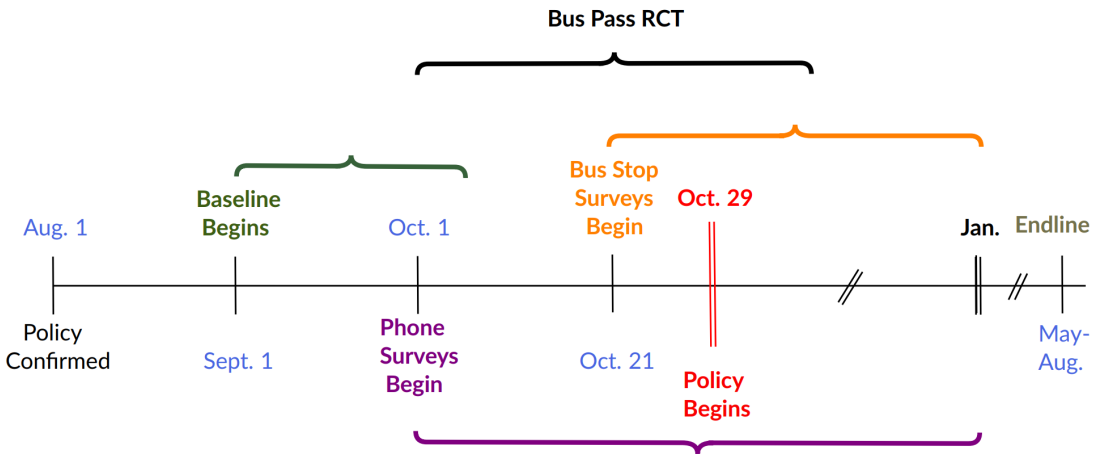
8 Additional Tables and Figures

Figure A.1: Sampling for the Main Study Sample



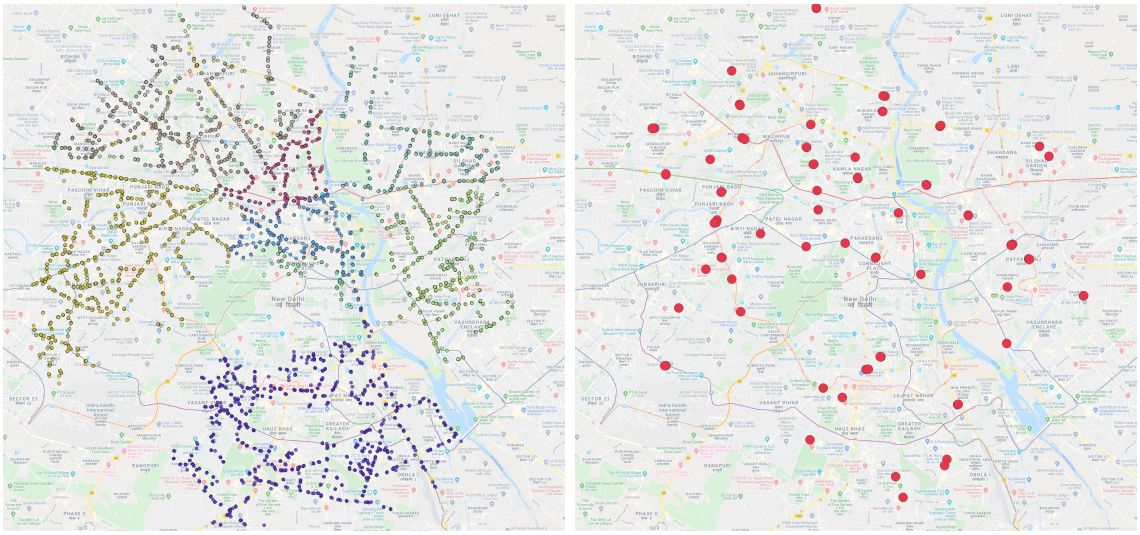
Note: Figure A.1 shows the 25 study wards (left) and grid cells (right). The sampling frame for wards consists of all Delhi wards within 15 km of the CBD that were manually categorized to contain high- or middle-density neighborhoods based on satellite imagery. Study wards were randomly selected over-weighting wards with low education in the 2011 census. Within these neighborhoods, we considered 100m × 100m grids no more than 500 meters from a bus stop. We manually classified grids and retained medium- and high-density grids based on satellite images, and randomly selected grids to include in the sample.

Figure A.2: Study Timeline



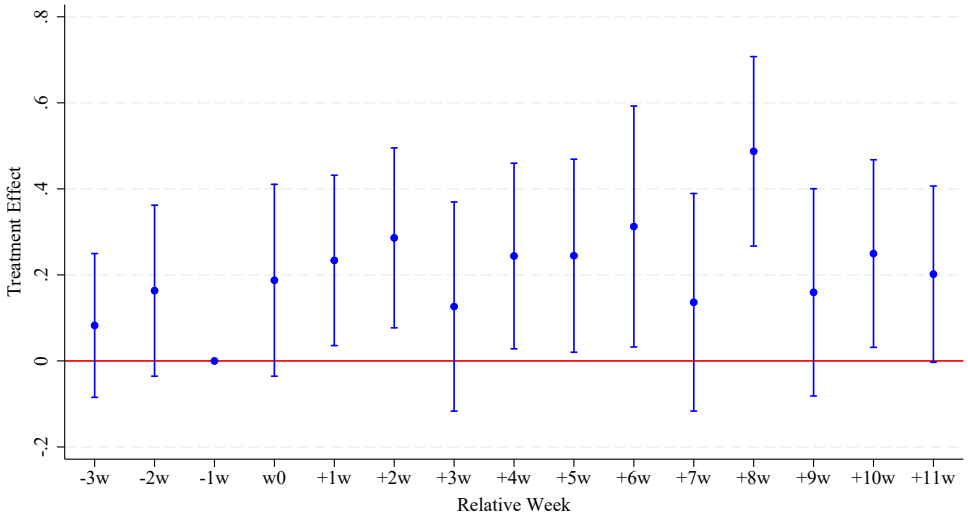
Note: Figure A.2 shows the sequence of policy announcements and launch, data collection exercises, and the bus pass pass experiment.

Figure A.3: Bus Stop Observation Data Collection Locations



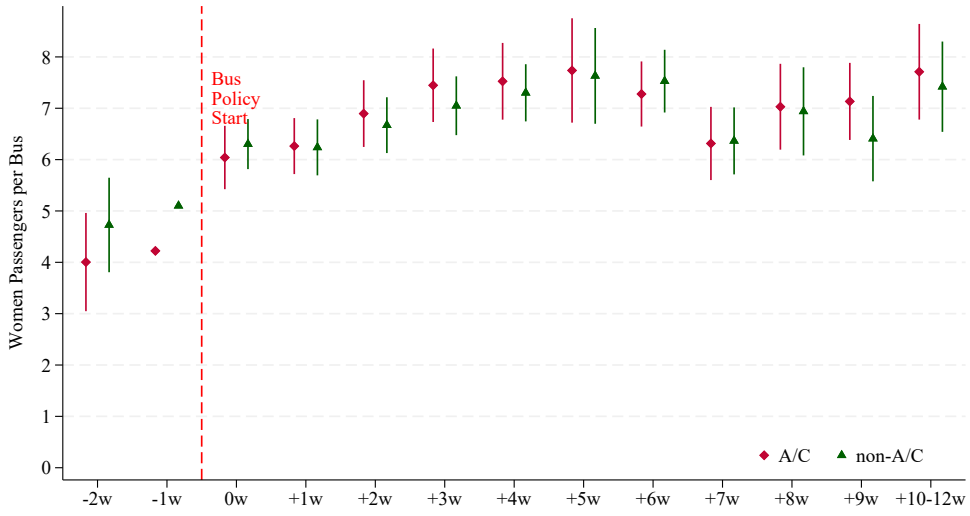
Note: Figure A.3 shows the sampling frame (left) and the 50 bus stops that were randomly selected stratified by latitude deciles (right). These were the locations of the bus stop observations.

Figure A.4: Impacts of Providing Free Bus Passes on Mobility

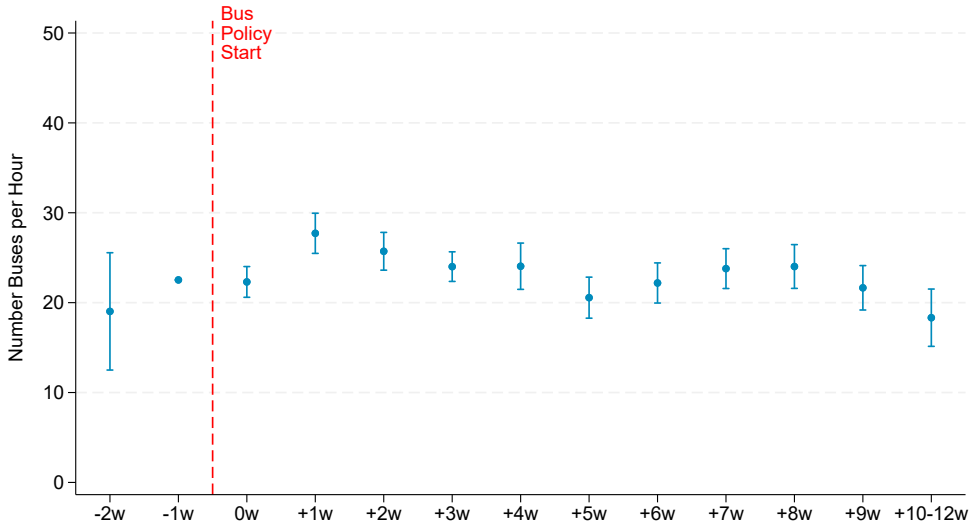


Note: Figure A.4 shows the impact of providing free bus passes over time on the total number of trips that included a bus, with treatment defined relative to the distribution of bus passes (or expected distribution for the control group).

Figure A.5: Impacts of Free Bus Policy (Bus Stop Observations)



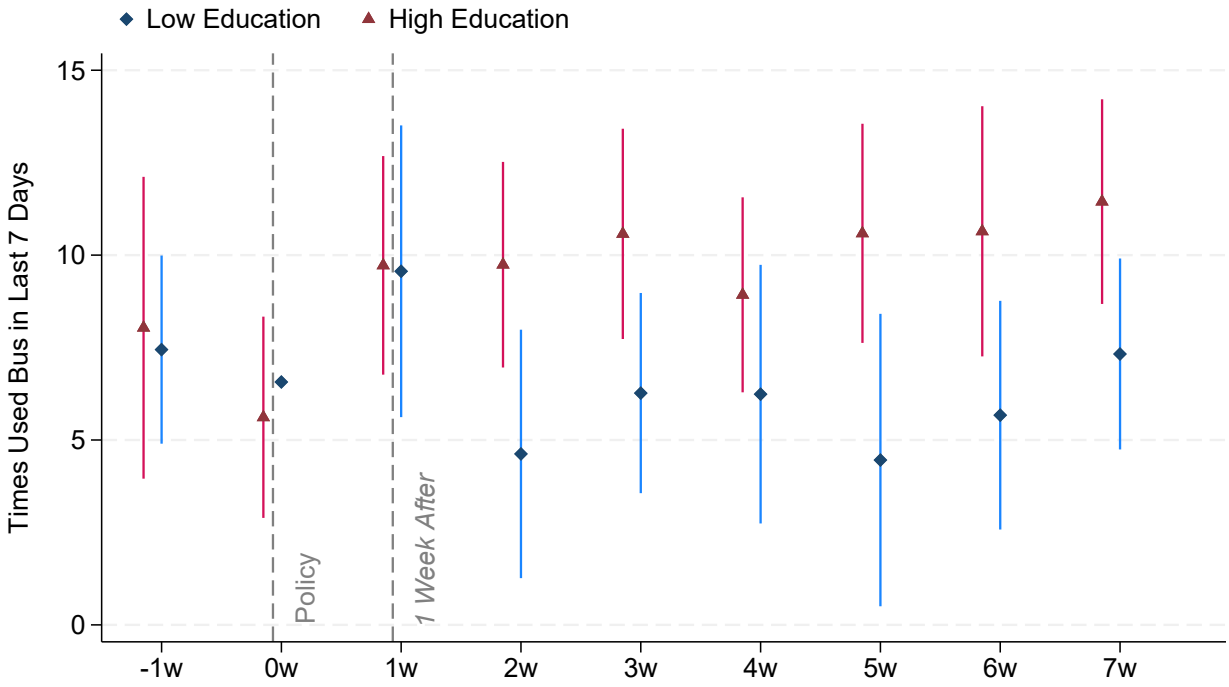
Panel (A) Impact on Female Passengers By Bus Type



Panel (B) Impact on Number of Buses per Hour

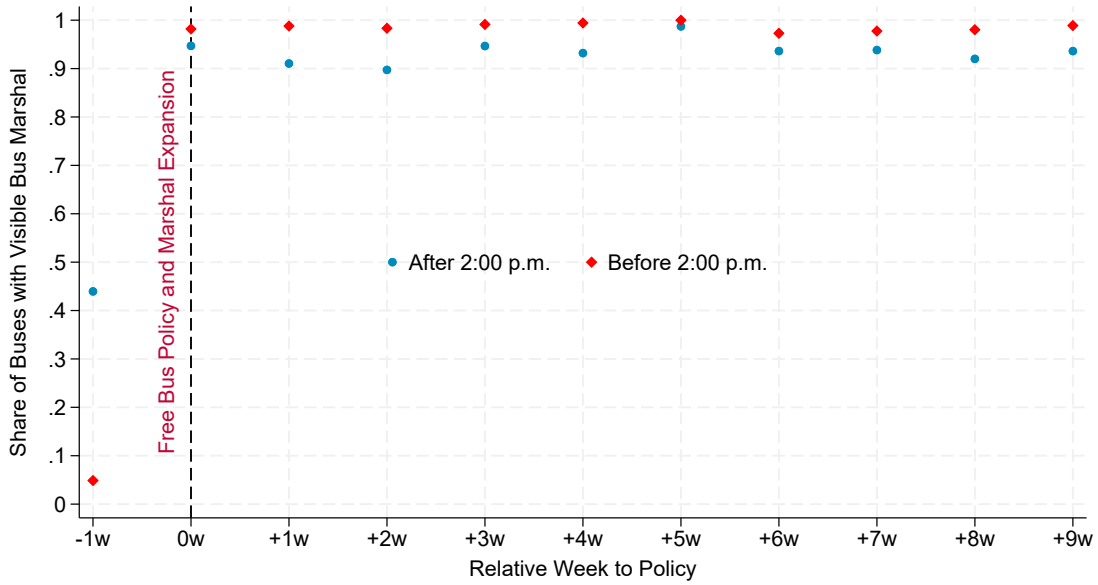
Note: Figure A.5 shows the number of women visible in a bus from the bus stop separately for green (non-AC) and red (AC) bus types (panel A) and the number of buses per hour (panel B), by week relative to the policy roll-out, using the data collected at bus stops. The graphs plot week indicator coefficients in a specification with bus stop fixed effects. Standard errors are clustered at the level of 50 bus stops.

Figure A.6: Impacts of Free Bus Policy on Female Bus Usage by Education

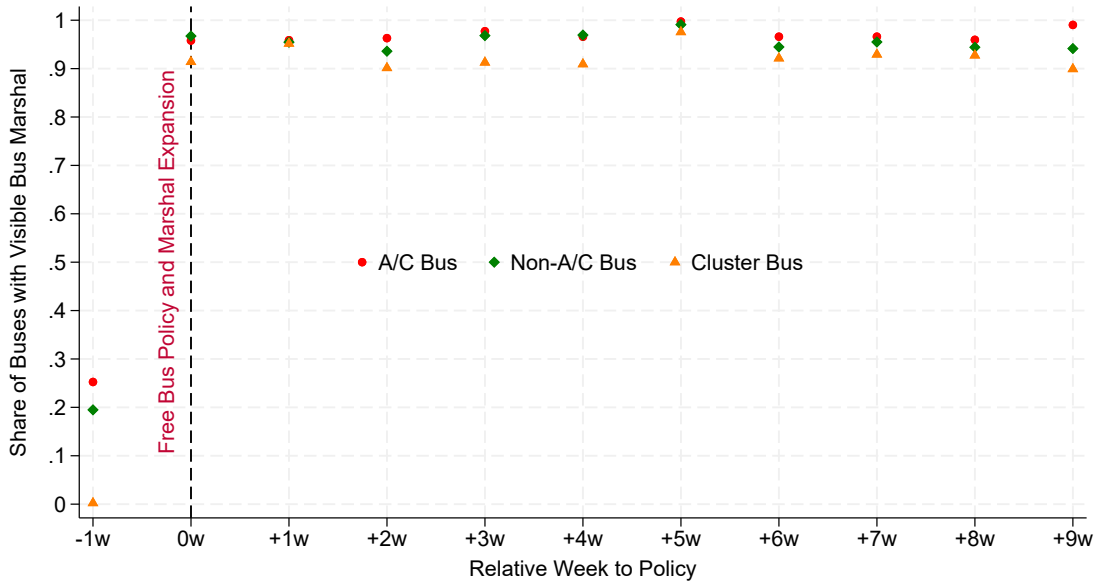


Note: Version of Figure 4 from a specification that also includes bus stop fixed effects.

Figure A.7: Bus Marshals By Time of Day and Bus Type (Bus Stop Observations)



Panel (A) Before and After 2 PM



Panel (B) Cluster vs Other Bus Types

Note: Figure A.7 shows whether a bus marshal is visible in a bus from the bus stop separately by time of day (panel A) and bus type (panel B), by week relative to the policy roll-out, using the data collected at bus stops. The graphs plot week indicator coefficients in a specification with bus stop fixed effects. Standard errors are clustered at the level of 50 bus stops.

Table A.1: Summary Statistics

	Sample:	Delhi Women	non-Delhi Women
		Mean [SD]	Mean [SD]
(1)	Age	35.4	30.0
		[11.2]	[9.7]
(2)	Middle-school or less	0.72	0.23
(3)	Primary activity: student	0.00	0.23
(4)	Primary activity: work	0.31	0.36
(5)	Primary activity: housewife	0.67	0.31
(6)	Need Permission to Travel	0.60	
(7)	Made trip yesterday	0.55	0.63
(8)	Made bus trip yesterday	0.23	0.06
(9)	% trips alone	0.57	0.46
(10)	% trips very safe	0.49	0.72
	Number of respondents	747	257

Note: Table A.1 presents summary statistics. The Delhi sample includes all women in the main panel survey. The non-Delhi sample includes all women recruited via Facebook and followed via repeated phone surveys from other Indian cities. The first six rows use data from the baseline survey (in-person for the Delhi survey, online for the non-Delhi survey), while the other rows use phone survey data before the policy.

Table A.2: Balance Tests Among Experimental Sample

		Treatment	(S.E.)	Control Anticipation	(S.E.)	Obs.	Control Pure Mean
(1)	Age	-1.92*	(1.02)	-1.76	(1.13)	746	36.88
(2)	SES index	0.02	(0.09)	0.12	(0.10)	707	-0.03
(3)	Own smartphone	-0.01	(0.04)	-0.01	(0.05)	747	0.55
(4)	Stay at home	-0.04	(0.04)	-0.02	(0.05)	746	0.69
(5)	Household size	0.03	(0.15)	0.16	(0.18)	747	4.95
(6)	Children under 5	0.06	(0.04)	0.04	(0.05)	747	0.22
(7)	Travel daily	0.07*	(0.04)	0.06	(0.04)	746	0.62
(8)	Number of trips	-0.01	(0.06)	-0.03	(0.07)	747	0.73
(9)	Any bus use	0.06	(0.05)	0.02	(0.05)	747	0.21
(10)	DTC bus use	0.05	(0.05)	0.02	(0.05)	747	0.20
(11)	Metro use	-0.01	(0.02)	-0.01	(0.03)	747	0.05

Note: Table A.2 shows shows experimental balance checks. Each row reports coefficients from a regression with the row header as outcome on indicators for the treatment group and for the anticipation control group, relative to the pure control group. SES index is an index of the respondent's socio-econonmic status created using principal component analysis. The index is based on the number of refrigerators, televisions, beds, bathrooms owned by the respondent's household and whether the respondent owns their house, a computer, a bicycle, a smartphone and if they have a shared bath. All regressions include ward fixed effects. Robust standard errors are shown in parentheses.

Table A.3: Attrition by Treatment Status over Time

	Short-term (pre policy)		Long-term (post policy)		Endline	
	(1)	(2)	(3)	(4)	(5)	(6)
Treated (× After)	0.01	0.00	0.02	-0.01	0.08**	0.04
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)
Anticipation Control (× After)		-0.03		-0.06		-0.09*
		(0.04)		(0.04)		(0.05)
Specification	DD	DD	DD	DD	T vs C	T vs C
Control mean	0.54	0.54	0.60	0.60	0.63	0.63
Observations	2988	2988	4844	4844	747	747
R^2	0.57	0.57	0.52	0.52	0.08	0.08

Note: Table A.3 shows response rates (our measure of attrition) to phone surveys by time period, and separately for the endline survey. Column 1-4 use a differences-in-difference specification based on equation (4).

Table A.4: Impacts of Providing Free Bus Passes on Travel Mode

Panel A: Pre-Policy (Short Term)

	(1) Walking	(2) Rickshaws	(3) DTC Bus	(4) Metro	(5) Scooter	(6) Car	(7) Green Bus	(8) Red Bus
Treat × After	0.134	0.009	0.157**	0.029	0.011	0.003	0.115**	0.047*
	(0.092)	(0.062)	(0.062)	(0.030)	(0.021)	(0.018)	(0.058)	(0.026)
Anticipation control × After	0.030	0.011	0.044	0.012	0.035	0.047	0.067	-0.022
	(0.104)	(0.073)	(0.069)	(0.033)	(0.026)	(0.031)	(0.065)	(0.026)
Pure Control After Mean	0.82	0.29	0.25	0.03	0.03	0.02	0.20	0.04
Pure Control Respondents	164	164	164	164	164	164	164	164
Resp. Anticipation Control	146	146	146	146	146	146	146	146
Resp. Treated	316	316	316	316	316	316	316	316
Observations	2,276	2,276	2,276	2,276	2,276	2,276	2,276	2,276

Panel B: Pre-Treatment + Post-Policy (Long Term)

	(1) Walking	(2) Rickshaws	(3) DTC Bus	(4) Metro	(5) Scooter	(6) Car	(7) Green Bus	(8) Red Bus
Treat × After	0.132*	0.053	0.154***	0.001	0.013	0.003	0.114**	0.044**
	(0.070)	(0.045)	(0.050)	(0.020)	(0.019)	(0.019)	(0.047)	(0.022)
Anticipation control × After	0.084	0.082	0.020	-0.015	0.003	-0.007	-0.002	0.026
	(0.085)	(0.055)	(0.061)	(0.023)	(0.019)	(0.021)	(0.058)	(0.032)
Pure Control After Mean	0.63	0.20	0.28	0.03	0.03	0.02	0.23	0.06
Pure Control Respondents	170	170	170	170	170	170	170	170
Resp. Anticipation Control	152	152	152	152	152	152	152	152
Resp. Treated	342	342	342	342	342	342	342	342
Observations	3,593	3,593	3,593	3,593	3,593	3,593	3,593	3,593

Note: Table A.4 shows the impact of the bus pass intervention on the number of trips with a given dominant mode, including one of two types of buses: Green (non-AC) buses and Red (AC) buses. A trip's dominant mode is the highest ranked mode used at least once on the trip, where modes are ranked in increasing order as follows: walking, scooter, rickshaw, bus, metro and car. Standard errors are clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Impacts of Providing Free Bus Passes on Travel Purpose

Panel A: Pre-Policy (Short Term)

	(1) Commute	(2) Social	(3) Chores	(4) Accompany	(5) Medical
Treat × After	0.018	0.008	0.074	-0.043	0.002
	(0.055)	(0.044)	(0.065)	(0.042)	(0.048)
Anticipation control × After	0.014	0.119**	-0.100	-0.009	0.003
	(0.065)	(0.057)	(0.071)	(0.047)	(0.059)
Pure Control After Mean	0.25	0.11	0.26	0.13	0.13
Pure Control Respondents	164	164	164	164	164
Resp. Anticipation Control	146	146	146	146	146
Resp. Treated	316	316	316	316	316
Observations	2,276	2,276	2,276	2,276	2,276

Panel B: Pre-Treatment + Post-Policy (Long Term)

	(1) Commute	(2) Social	(3) Chores	(4) Accompany	(5) Medical
Treat × After	-0.018	-0.005	0.063	0.032	0.063*
	(0.044)	(0.038)	(0.039)	(0.029)	(0.037)
Anticipation control × After	-0.019	0.041	0.011	0.030	-0.026
	(0.050)	(0.047)	(0.050)	(0.031)	(0.044)
Pure Control After Mean	0.26	0.14	0.12	0.05	0.10
Pure Control Respondents	170	170	170	170	170
Resp. Anticipation Control	152	152	152	152	152
Resp. Treated	342	342	342	342	342
Observations	3,593	3,593	3,593	3,593	3,593

Note: Table A.5 shows the impact of the bus pass intervention on the number of trips with a given trip purpose. Standard errors are clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Impacts of Providing Free Bus Passes on Bus Knowledge

	Reached Endline		Typical Trips		N Bus Lines		N Bus Stops		N Destinations	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treated	0.04	0.04	0.40***	0.35***	-0.04	0.01	-0.01	0.03	0.03	0.08
	(0.04)	(0.04)	(0.11)	(0.11)	(0.21)	(0.21)	(0.14)	(0.14)	(0.18)	(0.19)
Anticipation Control	-0.09*	-0.08*	0.19	0.15	-0.29	-0.23	-0.14	-0.08	-0.15	-0.09
	(0.05)	(0.05)	(0.14)	(0.14)	(0.25)	(0.26)	(0.16)	(0.16)	(0.22)	(0.22)
Controls		Yes		Yes		Yes		Yes		Yes
Control Mean	0.67	0.68	0.73	0.73	1.76	1.76	1.15	1.15	1.47	1.47
Observations	747	746	483	483	462	462	462	462	462	462

Note: Table A.6 shows the impact of the bus pass experiment on outcomes measured in the endline survey, which was conducted by phone in mid 2020. Even columns include controls measured at baseline for age and an indicator for traveling outside the neighborhood daily, the two variables with weak imbalance in Table A.2. The outcome in the first two columns is an indicator for responding to the endline survey. All other columns are conditional on responding to the endline survey. The next two columns measure a recall question about typical number of daily trips pre-COVID. The next three pairs of columns measure the number of bus lines the respondent could identify and describe in terms of bus line number, a nearby bus stop, and the destination.

Table A.7: Heterogeneous Treatment Effects by Social Desirability

	Social Desirability (z-score)	Bus Trip		Trip		Trip Alone	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treated	-0.03						
	(0.12)						
Anticipation Control	0.01						
	(0.14)						
Treat × After		0.16***	0.16***	0.17**	0.17**	0.16**	0.16**
		(0.05)	(0.05)	(0.07)	(0.07)	(0.07)	(0.07)
Anticipation Control × After		-0.03	-0.03	0.03	0.03	-0.00	-0.01
		(0.06)	(0.06)	(0.09)	(0.09)	(0.08)	(0.08)
Treat × After × SD z-score			0.01		0.02		0.04
			(0.05)		(0.07)		(0.07)
A. C. × After × SD z-score			0.03		0.02		-0.09
			(0.06)		(0.09)		(0.09)
RI p-value Treated	0.77						
Control After Mean	-0.01	0.33	0.33	0.82	0.82	0.45	0.45
Respondents Control		180	180	180	180	180	180
Respondents Treated		222	222	222	222	222	222
Observations	406	3,090	3,090	3,090	3,090	3,090	3,090

Note: Table A.7 shows heterogeneous treatment effects of the bus pass experiment by the Marlow-Crowe index of social desirability from (Dhar et al., 2022), measured at endline. Column 1 estimates the causal impact of the bus pass intervention on this index, and reports the randomization inference p-value based on 1,000 runs. Columns 3, 5, and 7 interact the treatment times post indicators with the z-score of the social desirability measure. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: Heterogeneous Treatment Effects by Perceived Desirability of Bus Use

	Bus Use Perceived Desirable	Bus Trip		Trip		Trip Alone	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treated	0.03						
	(0.06)						
Anticipation Control	0.00						
	(0.08)						
Treat × After		0.19***	0.21***	0.23***	0.24**	0.20***	0.16*
		(0.06)	(0.08)	(0.08)	(0.11)	(0.07)	(0.09)
Anticipation Control × After		0.09	0.04	0.09	0.03	0.08	0.05
		(0.07)	(0.10)	(0.10)	(0.12)	(0.08)	(0.12)
Treat × After × Perceived Desirability			-0.04		-0.06		0.08
			(0.12)		(0.16)		(0.14)
A. C. × After × Perceived Desirability			0.12		0.13		0.08
			(0.13)		(0.19)		(0.16)
RI p-value Treated	0.61						
Control After Mean	0.45	0.32	0.32	0.79	0.79	0.41	0.41
Respondents Control		177	177	177	177	177	177
Respondents Treated		171	171	171	171	171	171
Observations	352	2,538	2,538	2,538	2,538	2,538	2,538

Note: Table A.8 shows heterogeneous treatment effects of the bus pass experiment by whether respondents answered yes to the endline survey question "Do you think we (JPAL) would like the women in our study to ride the bus more?" Column 1 estimates the causal impact of the bus pass intervention on this outcome, and reports the randomization inference p-value based on 1,000 runs. Columns 3, 5, and 7 interact the treatment times post indicators with this outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: Heterogeneous Treatment Effects by Need Permission to Leave Status

	(1) Bus Trip	(2) Green Bus	(3) Red Bus	(4) Trip	(5) Trip Alone	(6) Safe | Bus
T × A × Needs permission	0.08	0.07	0.02	0.12	0.17**	0.02
	(0.05)	(0.05)	(0.02)	(0.09)	(0.07)	(0.12)
T × A × Does not	0.27***	0.20***	0.07**	0.25**	0.12	0.07
	(0.08)	(0.07)	(0.03)	(0.10)	(0.09)	(0.14)
Control Mean: Needs permission	0.29	0.24	0.05	0.75	0.36	0.59
Control Mean: Does not	0.34	0.28	0.07	0.86	0.51	0.60
P-val for heterogeneity	0.05	0.15	0.20	0.35	0.62	0.75
Respondents: Needs permission	408	408	408	408	408	115
Respondents: Does not	274	274	274	274	274	96
Observations	4,617	4,617	4,617	4,617	4,617	826

Note: Table A.9 shows heterogeneous treatment effects of the bus pass experiment by whether the respondent reported at baseline that she needs permission to leave the house. The sample includes all phone surveys rounds, both before and after the citywide policy launch. All specifications control for Anticipation control × Post interacted with the heterogeneity variable, but the coefficients are omitted for space. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.10: Impacts of Providing Free Bus Passes by Baseline Travel Alone

	(1) Bus Trip	(2) Green Bus	(3) Red Bus	(4) Trip	(5) Trip Alone	(6) Safe | Bus
Treat × After × Never alone	-0.01	-0.03	0.03	0.13	0.10	0.10
	(0.07)	(0.06)	(0.02)	(0.12)	(0.08)	(0.26)
Treat × After × Sometimes alone	0.23***	0.18***	0.05*	0.19**	0.17**	0.01
	(0.06)	(0.06)	(0.03)	(0.08)	(0.07)	(0.11)
Control Mean: Never alone	0.24	0.21	0.03	0.71	0.28	0.54
Control Mean: Sometimes alone	0.34	0.28	0.07	0.84	0.49	0.61
P-val for heterogeneity	0.01	0.01	0.52	0.65	0.50	0.74
Respondents: Never alone	217	217	217	217	217	51
Respondents: Sometimes alone	463	463	463	463	463	160
Observations	4,612	4,612	4,612	4,612	4,612	826

Note: Table A.10 shows heterogeneous treatment effects of the bus pass experiment by whether the respondent reported at baseline that she sometimes made typical trips by herself. The sample includes all phone surveys rounds, both before and after the citywide policy launch. All specifications control for Anticipation control × Post interacted with the heterogeneity variable, but the coefficients are omitted for space. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.11: Impacts of Free Bus Policy on Low-SES Women's Mobility

	Log Bus Trips			Log Trips			Fraction Bus Trips		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Post Policy	-0.18**			-0.30***			0.03		
	(0.08)			(0.05)			(0.03)		
Post Policy × Delhi Woman		-0.53	0.08		0.06	-0.30***		-0.04	0.07
		(0.47)	(0.16)		(0.10)	(0.10)		(0.04)	(0.04)
Specification	PPML	PPML	PPML	PPML	PPML	PPML	OLS	OLS	OLS
Respondent FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample: Delhi Women	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sample: Other Cities Women		Yes			Yes			Yes	
Sample: Delhi Men			Yes			Yes			Yes
Survey Rounds in Sample	1-9	3-9	3-9	1-9	3-9	3-9	0-9	0, 3-9	0, 3-9
Outcome Pre Mean	0.73	0.74	0.74	1.04	1.04	1.04	0.39	0.39	0.39
Observations	1,030	1,108	1,271	1,746	2,616	2,154	1,789	2,727	2,554

Note: Table A.11 replicates the analysis from Table A.11 using additional outcomes and specifications. Column 1-6 fit PPML models, while column 7-9 use OLS with the outcome being the *fraction* of trips by bus (defined as zero if no trips are reported). All specifications include respondent fixed effects. Standard errors in parentheses are clustered at the individual level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.12: Impacts of Free Bus Policy (Bus Stop Observations)

	(1)	(2)	(3)	(4)
	Frac with Bus Marshals	Number Buses	Bus Crowdedness	Women Passengers Per Bus
After Citywide Policy	0.77***	1.78**	-0.04	1.84***
	(0.02)	(0.75)	(0.04)	(0.18)
Pre-Policy Mean	0.16	26.24	1.41	5.06
Observations	3,310	3,310	3,310	3,310
Bus Stop FE	Yes	Yes	Yes	Yes

Note: Table A.12 shows results from a regression of the column outcome on an indicator for after the citywide policy, including bus stop fixed effects. Data is averaged at the bus stop times hour of the day level. The outcomes are the share of buses with a visible bus marshal, the number of buses per hour, the average crowdedness score (see notes for Figure 3), and the average number of visible women passengers per bus. Standard errors are clustered at the level of 50 bus stops.

Table A.13: Impact of Free Bus Policy on Female Bus Usage (All Coefficinets)

	Times Used Bus in Last 7 Days			
	(1)	(2)	(3)	(4)
After	1.8***	1.7***	0.4	-0.0
	(0.6)	(0.5)	(0.4)	(0.5)
Woman	-1.2	1.4*	-0.5	-1.6***
	(0.8)	(0.8)	(0.4)	(0.4)
High-Ed	3.6***	1.6*	1.1*	0.6
	(0.6)	(0.8)	(0.5)	(0.6)
After × Woman	-1.5	-2.8***	0.5	1.0*
	(0.9)	(0.9)	(0.6)	(0.5)
After × High-Ed	-1.3*	-0.7	-0.6	-0.5
	(0.8)	(1.0)	(0.6)	(0.7)
Woman × High-Ed	-2.2	-1.8	0.3	0.7
	(1.7)	(1.1)	(0.5)	(0.6)
After × Woman × High-Ed	7.9***	5.1***	-0.1	-0.6
	(1.9)	(1.4)	(0.7)	(0.7)
Delhi			2.8***	0.0
			(0.4)	(.)
After × Delhi			1.4**	1.7***
			(0.7)	(0.5)
Woman × Delhi			-0.7	3.0***
			(0.6)	(0.6)
High-Ed × Delhi			2.5***	1.0
			(0.6)	(1.0)
After × Woman × Delhi			-2.0**	-3.9***
			(0.9)	(0.8)
After × Delhi × High-Ed			-0.7	-0.2
			(0.8)	(1.1)
Female × Delhi × High-Ed			-2.6	-2.5***
			(1.6)	(0.9)
After × Female × Delhi × High-Ed			8.1***	5.7***
			(1.8)	(1.2)
Bus Stop FE		Yes		Yes
Outcome Mean	7.76	7.77	5.67	5.68
Observations	1,803	1,800	2,970	2,957

Note: This table displays all coefficients from the specification in Table 4. Standard errors in parentheses clustered by date. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A Appendix

A.1 Econometric Framework: Additional Results

Proof of Proposition 1. We omit the dependence of potential outcomes on $\mathbf{d}$ and H by Assumption 1. Applying scaling, idempotency, and scaling again gives

$$Y_{i1}\big((0,1)\big) =^{\text{(Def 1)}} Y_{i1}\big((1,0)\big) =^{\text{(Def 2)}} Y_{i1}\big((2,0)\big) =^{\text{(Def 1)}} Y_{i1}\big((1,1)\big).$$

The left-most and right-most potential outcomes above coincide for each individual and hence in expectation. Under Assumption 1 and random assignment of N_{i1}^P within $\mathcal{P}_1$, the difference in means across the two arms identifies $\tau_{P|S}$, which is therefore zero. □

Remark 2. This result can be generalized to policies with known returns to repeated treatment, i.e. $\mathbb{E}Y_{it}((2,0),\mathbf{d},H) = \lambda EY_{it}((1,0),\mathbf{d},H)$ for known λ. In this case, under the null of scaling, the impact of the additional pilot ($\tau_{P|S}$) should be a known multiple of the pure pilot impact (τ_P). With and additional assumption of time-stable impacts, both the pure pilot effect and λ may be measured before the roll-out of the scaled up policy, using a randomized experiment with three treatment arms offering up to two "doses" of the pilot policy: $d_0 = (0,0)$, $d_1 = (1,0)$ and $d_2 = (2,0)$.

Activation

Proposition 2. Fix $i \in \mathcal{P}_0$. By Assumptions 2–2 and Assumption 1, $Y_{i0}(1,0) = Y_{i0}^0 + \beta_{i0}A_i^P$ and $Y_{i0}(0,0) = Y_{i0}^0$, using $\alpha_i(0,0) = 0$. The level cancels, so $\tau_P = \mathbb{E}\big[\beta_{i0}A_i^P\big]$; since $A_i^P = 1$ exactly on $\mathcal{G}_P \cup \mathcal{G}_B$, Assumption 4 gives (1).

Fix $i \in \mathcal{P}_1$. Similarly $\tau_{P|S} = \mathbb{E}\big[\beta_{i1}\big(A_i^{PS} - A_i^S\big)\big]$. By Assumption 3 the difference lies in $\{0,1\}$ and equals $\mathbf{1}\{A_i^S = 0,\ A_i^{PS} = 1\}$, the indicator of $\mathcal{G}_P \cup \mathcal{G}_C$; Assumption 4 gives (2), and (3) follows by subtraction. Random assignment of N_{it}^P within $\mathcal{P}_t$ identifies each expectation by the difference in arm means. □

Signing the contrasts

Assumption 5 (Non-negative effects)**.** $\beta_i \geq 0$ for all i.

Assumption 5 is strong—it requires that activation harms no one—and we invoke it only to sign and bound the objects in Proposition 2.

Proposition 3 (Bounds and sharp subsumption)**.** *Under Assumptions 2–5:*

(i) $\tau_{P|S} \geq \mathbb{E}\big[\beta_i \mathbf{1}\{\mathcal{G}_P\}\big] \geq 0$, *with* $\tau_{P|S} = 0$ *if and only if* $\beta_i = 0$ *for almost every* $i \in \mathcal{G}_P \cup \mathcal{G}_C$*: the at-scale policy subsumes the pilot.*

(ii) $\tau_{P|S} > \tau_P$ *implies* $\Pr(\mathcal{G}_C) > 0$*: a post contrast exceeding the pre contrast is evidence that the two versions are complements in activation.*

(iii) If $\Pr(\mathcal{G}_C) = 0$*, then* $0 \leq \tau_{P|S} \leq \tau_P$*, so* $\tau_{P|S}/\tau_P \in [0, 1]$ *whenever* $\tau_P > 0$*; and* $\tau_P - \tau_{P|S} = 0$ *if and only if* $\beta_i = 0$ *for almost every* $i \in \mathcal{G}_B$*.*

Proof. (i) By (2), $\tau_{P|S}$ is the expectation of a non-negative random variable, bounded below by its restriction to $\mathcal{G}_P$ and vanishing iff the integrand is zero a.e. (ii) By (3), $\tau_P - \tau_{P|S} < 0$ requires $\mathbb{E}[\beta_i \mathbf{1}\{\mathcal{G}_C\}] > \mathbb{E}[\beta_i \mathbf{1}\{\mathcal{G}_B\}] \geq 0$. (iii) With $\Pr(\mathcal{G}_C) = 0$, (2) and (3) are both expectations of non-negative random variables. □

Interpretation. When $\mathcal{G}_C$ is negligible, the ratio $\tau_{P|S}/\tau_P \in [0, 1]$ is the share of the pilot's measured impact that the at-scale policy fails to reproduce: at one extreme the scale-up reproduces the pilot in full, at the other it leaves the pilot's impact undented, because everyone the pilot activates and benefits is someone the at-scale policy does not activate. Part (ii) shows this reading is not imposed by assumption—the unit interval is a testable restriction, and a ratio above one is a signature of joint activation the design can detect.

A.2 Bus Pass Experiment

Bus pass experiment script During the baseline survey, surveyors used the following script to introduce the bus pass experiment:

> *In this study, we aim to understand the constraints to traveling by public transportation in Delhi. That is why we will be offering some of our study participants a free bus pass, to understand if price is such as constraint.*
>
> *If you are selected to receive the free bus pass, it will be mailed or delivered to you at home in the next weeks. We would like to collect from you now the information necessary to apply for the bus pass on your behalf, including a photo of you. If you are selected, we plan to submit this information on the Delhi government's website for purchasing bus passes online (http://www.dtcpass.delhi.gov.in/). The photograph we take today will be permanently erased no later than 3 months from now.*
>
> *Remember that we are neither encouraging nor discouraging you to travel by bus or by any other mode. You should decide whether and how to travel taking into account all relevant factors, such as, in no particular order: travel duration,*

reliability, convenience, personal safety, price, other factors that are important to you. etc. In particular, if you do not plan to travel by bus, you do not need to use the bus pass provided.

Remember: the photograph and this information will ONLY be used to obtain the bus pass and deleted afterward. We will never share this data with anyone else. We are a reputed research organization.

At the time of delivery, the surveyor used the following script:

Remember that we are neither encouraging nor discouraging you to travel by bus or by any other mode. You should decide whether and how to travel taking into account all relevant factors, such as, in no particular order: travel duration, reliability, convenience, personal safety, price, other factors that are important to you. etc. In particular, if you do not plan to travel by bus, you do not need to use the bus pass provided.