When At-Scale Nudges Work and for Whom: A 10-Million-Person Field Experiment*

Ari Anisfeld[†]
University of Arkansas

Salman Khan[‡]
Harvard University

Dennis Kramer[§]
U.S. Department of Education

November 6, 2026

Abstract

Behavioral nudges have a mixed record: large campaigns, in particular, rarely show measurable impact at scale. We document one that works, and who it benefits. In a 9.91-million-student block-randomized experiment, a U.S. Department of Education FAFSA renewal email raised renewal by 0.98 percentage points (roughly 95,000 submissions) and college re-enrollment by 0.31 points (roughly 30,000 students) at near-zero marginal cost. The response is selective: enrollment effects are 77 percent of renewal effects for students previously full-time, 28 percent for the previously non-full-time, and approximately zero for the previously unenrolled. Paperwork moves everywhere; enrollment moves only among the already attached.

JEL Classification: C93, D91, I22, I23, I24

Keywords: Economics of Education, Financial Aid, Higher Education, Policy

*The views presented in this study are those of the authors alone and do not represent the position of the U.S. Department of Education or the Office of Federal Student Aid. Any errors are our own. We are grateful to Anjali Adukia, Desmond Ang, Christopher Avery, Peter Blair, Rajeev Darolia, Steven Durlauf, Susan Dynarski, Jordan Matsudaira, Tomás E. Monarrez, Richard Murnane, Bridget Terry Long, Carolyn Rose, Alex Spancake, Kevin Stange, Lesley Turner, and participants from AEFP and APPAM for helpful feedback. We are also grateful to the FSA Communications Team who permitted the experiment. This experiment was pre-registered through the AEA RCT Registry as AEARCTR-0011295 in May 2023 with an update in June 2023: https://doi.org/10.1257/rct.11295.

[†]Ari Anisfeld is an Assistant Professor at the University of Arkansas. Email: anisfeld@uark.edu.

[‡]Salman Khan is a recent PhD graduate of Harvard University and a Postdoctoral Fellow at the National Bureau of Economic Research (NBER). Email: skhan2@g.harvard.edu.

[§]Dennis Kramer is the Director of Data Policy and Evaluation at the Office of Federal Student Aid.

1 Introduction

Light-touch behavioral interventions conducted on a state-wide or national scale have yielded mixed results. Several at-scale campaigns find null effects: six preregistered experiments, sent variously by a state tax agency, a county welfare office, and a statewide nonprofit, varied the content, design, messenger, and mode of tax-credit outreach across roughly a million observations and changed neither filing nor claiming (Linos et al., 2022), and national campaigns encouraging students to apply to college, file the Free Application for Federal Student Aid (FAFSA) and persist in college, run by college-access organizations and state agencies, have produced precisely-estimated nulls (Bergman et al., 2019; Bird et al., 2021; Gurantz et al., 2021; Avery et al., 2021). Other nudge campaigns worked: simplified IRS notices raised tax-credit claiming among eligible non-claimants by four to nine points (Bhargava and Manoli, 2015), an information-only mailing raised SNAP enrollment by five points (Finkelstein and Notowidigdo, 2019), and letters about health-insurance coverage from a state marketplace and the IRS, together with text-message vaccine reminders from health systems, move take-up by roughly one to five points (Domurat et al., 2021; Goldin et al., 2021; Milkman et al., 2021). Across 126 trials reaching 23 million people, the average intervention run by two large U.S. nudge units raised take-up by 1.4 percentage points (DellaVigna and Linos, 2022).

In most of these settings, only the first-stage action is observed, so a campaign that moves paperwork cannot typically speak to the outcomes it seeks to improve. The take-up studies record submitting a form that is itself the benefit (Linos et al., 2022; Bhargava and Manoli, 2015; Finkelstein and Notowidigdo, 2019; Domurat et al., 2021), while education campaigns record enrollment but not the paperwork (Bergman et al., 2019; Bird et al., 2021; Gurantz et al., 2021), so a null cannot typically separate unmoved forms from unconverted ones. The closest study to our setting is a national text-message experiment urging college students to refile the FAFSA, which moved some students' filing earlier but did not raise overall filing or college persistence (Page et al., 2023).

We ask two questions. First, can a light-touch, government-run nudge move both a first-stage action and the welfare-relevant second-stage outcome on a national scale? Second, when a nudge moves paperwork, who converts that paperwork into the outcome that matters? In a 9.91-million-student block-randomized evaluation of the U.S. Department of Education's (DOE) annual FAFSA renewal email, the answer to the first is yes, at a marginal cost on the order of $0.40 per induced full-time college enrollment, and the answer to the second is sharply selective: the email raises renewal in every group we examine, but raises enrollment only among students already attached to college.

The FAFSA renewal cycle is a useful setting to explore these questions. First, scale: 9.9 million students is roughly two orders of magnitude larger than the at-scale samples in Bhargava and Manoli (2015); Finkelstein and Notowidigdo (2019); Domurat et al. (2021), so even the smallest subgroups we examine are precisely estimated. Second, both margins are observed in administrative data: the email targets a paperwork decision, renewing the FAFSA, while the welfare-relevant outcome is a separate decision, re-enrolling in college, and both are recorded for the same students in our data. Among the studies above, only Page et al. (2023) and the Texas school-district arm of Avery et al. (2021) record both FAFSA completion and later college enrollment for each student.

In this setting, roughly four in ten students who start college do not complete a degree within six years.[1] The returns to a degree are large, even for marginal students (Oreopoulos and Salvanes, 2011; Turner, 2016; Heckman et al., 2018; Ost et al., 2018). To maintain access to federal financial aid, and to access state and institutional aid that uses the FAFSA for means testing, students must annually resubmit the FAFSA. FAFSA non-renewal is common, even among students who likely would benefit from access to aid. Ten percent of Pell Grant recipients in good standing who return for their sophomore year fail to refile, and survey data show that non-renewal is negatively correlated with degree completion (Bird and Castleman, 2016). FAFSA non-renewal is a puzzle that poli-

[1]For the 2017 entry cohort, 62 percent of students graduated within six years, with substantial heterogeneity. https://nscresearchcenter.org/completing-college/

cymakers and advocates have sought to address with various interventions ranging from moderately intense counseling to low-cost nudges.

Our experiment covers 9.91 million students (9,716,427 treated and 198,270 in the holdout) who had filed a FAFSA for the 2022–23 academic year (AY2023) but had not yet renewed for 2023–24 (AY2024) as of March 28, 2023. The randomization was implemented with fidelity, the holdout received no other contact through the renewal window, and the sample is so large that even the smallest cells in our heterogeneity analysis (control $n > 30,000$) have standard errors below 0.3 percentage points. We can rule out an enrollment effect for the previously-unenrolled subgroup larger than 0.46 percentage points with 95% confidence.

We report three findings. First, in contrast to the at-scale nulls in this domain, the campaign caused a precisely-estimated increase in FAFSA renewal of 0.98 percentage points (s.e. 0.11), corresponding to roughly 95,000 additional submissions among the 9.7 million treated students. The point-in-time response is large: on the day a student receives an email, the probability of submitting FAFSA roughly doubles relative to the previous day. Second, the renewal effect translates into modest but precisely-estimated college re-enrollment effects in the aggregate: 0.31 percentage points (s.e. 0.10), or about 30,000 students. Combining standard per-send benchmarks (under $0.01 per email) with approximately 25,000 induced full-time enrollments yields a marginal cost of $0.40 per induced full-time enrollment, compared with roughly $20 to $55 per induced enrollment for mailed outreach in other benefit programs and $66 to $1,100 for intensive or targeted college-access programs.

Third, the aggregate ratio of 0.32 enrollees per submitter averages across populations with distinct responses. Read descriptively, the enrollment effect is 77% of the renewal effect for students who had been enrolled full-time in AY2023, 28% for the previously non-full-time, and approximately zero for students not enrolled in AY2023. The renewal effect itself, by contrast, is positive and economically meaningful in every subgroup. Section 4.4

cautions against a strict LATE interpretation, but either reading supports the qualitative pattern: the email moves paperwork in every group but moves enrollment only in groups already attached to college.

The paper makes two contributions to the at-scale behavioral interventions literature, with direct analogs in tax-credit take-up (Bhargava and Manoli, 2015), benefit enrollment (Finkelstein and Notowidigdo, 2019; Domurat et al., 2021), voter mobilization (Gerber and Green, 2000), and health-insurance outreach (Goldin et al., 2021). First, it documents an existence proof: a national, government-administered email campaign can move both renewal and enrollments at scale. Previous at-scale studies produced precisely-estimated null results (Bird et al., 2021; Gurantz et al., 2021). In our setting, we cannot causally identify the mechanism which drove the results. We do not observe variation across the source, message content, timing, or targeting of the email campaign that we can measure, and the comparison to prior at-scale nulls is suggestive rather than experimentally identified. We narrow down the candidate mechanisms in Section 5.3.

Second, by exploiting the block-randomized design and observing both the renewal margin and the enrollment margin, we report a descriptive cross-margin ratio. The cross-margin ratio is a natural summary statistic wherever an intervention moves a first-stage decision (apply, sign up, or file paperwork) while success is measured on a second-stage outcome (use, enrollment, or persistence), and our finding that the first margin moves nearly uniformly while the second moves selectively is a candidate stylized fact in each such setting, testable wherever both margins are observed. This pattern speaks to a nascent literature on the distributional consequences of nudges (Sunstein, 2022): an at-scale renewal nudge can produce a positive average enrollment effect while imposing burdens on a subgroup with little prospect of converting paperwork into enrollment. In this setting, we address this nascent literature and demonstrate that the paperwork is cheap enough that a few induced enrollments would justify the cost (Section 5.4).

2 Experiment and Policy Setting

Each year the U.S. Department of Education runs one or more communication campaigns to encourage students to renew their FAFSA application for the following year. The campaigns target students who have completed FAFSA for the current academic year, but have not yet renewed it for the upcoming academic year. This group includes students who use Title IV grants and loans in the current academic year and those who do not, including students who have no record of college enrollment in the current academic year.[2] To evaluate the campaign as it actually operates, we ran the 2023 campaign in the standard way, but withheld a roughly 2-percent holdout sample from all communications.

The DOE uses Adobe Campaign, an omnichannel marketing platform, to manage delivery. This platform can either randomize recipients automatically or use pre-randomized lists to send emails and text messages to students, tracking key metrics such as successful deliveries, opens, and clicks on embedded links. If the message fails to send, the system attempts delivery up to five times within the hour and then once again within five days. In the 2023 campaign, 98 percent of email sends resulted in a successful delivery. According to Department of Education staff, the exact timing of the campaign is determined by access to the servers and "warm-up" periods in which test emails are sent. In practice, the email system reads data in chunks and randomizes within these chunks such that the initial order of the data may influence which days communications go out. Overall, the experiment we conduct is reflective of a typical communication campaign run by the Department of Education. The 2023 communications campaign began on April 20, 2023, with 9,716,427 emails sent by May 20, 2023. For idiosyncratic reasons, most emails went out on May 13 or 16 (roughly 3 million each day).[3]

[2]DOE tracks student enrollment of students who have ever received Title IV aid, but will not typically have records for those who have never received aid, such as students with full-ride scholarships.

[3]In addition to the email campaign, the communications team sent reminder text messages to the 2.3 million students who opted in and were not in the holdout group. Each opted-in student received one SMS on one of four dates in early May, and 70 percent were successfully delivered. Because SMS receipt is nested within the treated arm, our estimates capture the combined effect of the email campaign and the accompanying SMS reminder for opted-in students. We do not have a flag within the control group for

Within the treated arm, the pre-registered design (AEARCTR-0011295) specified three message-content variants delivered to roughly equal-sized cells. A subsequent tone A/B test (neutral vs. "urgent" framing), introduced after preregistration in response to a priority change by the communications team, was overlaid onto the content cells. Roughly one in nine students received the original content messages as assigned and the remainder received the tone message. Appendix Figure A2 reproduces two representative renewal-email variants. Appendix J discusses why the within-treatment variation cannot be cleanly used to identify content effects. We do not exploit these tests in the main analysis since the operational implementation of the within-treatment assignment was nested inside the same chunked send schedule used for delivery rather than separately pre-randomized; balance checks on the realized cells indicate that segment assignment is correlated with covariates such as AGI and prior-year enrollment status, so a clean within-treatment comparison is not available. Importantly, our identification strategy relies only on the treatment-control split, which is unaffected by the within-treatment variation. We pool the different content variants and treat the campaign as a single intervention. While it is possible that imbalance in message content drives the heterogeneous result, we think it is unlikely given that the emails contain similar reminder messaging in the first half of the email and the main FAFSA renewal effects are similar across groups.

3 Data

Our sample is constructed from student-level records from the U.S. Department of Education's Data Warehouse. In particular, we rely on tables that contain records of FAFSA submission, aid receipt, and college enrollment in the academic year 2023-2024 and ear-

opt-in to SMS campaigns, so cannot evaluate the channels separately. Those who opted-in were more likely to come from lower-income households, to be female, and to be enrolled part-time rather than full-time. This group resubmits FAFSA at a lower rate than the rest of the population. Importantly, randomization ensures these do not bias the estimates but rather change the interpretation from an "e-mail" to a "business as usual communication campaign."

lier. It comprises the 9,914,697 students who had completed FAFSA for 2022-2023 and had not yet resubmitted FAFSA for 2023-2024 as of March 28, 2023. [4]

Student FAFSA applications for AY2023 provide baseline demographic information, including gender and whether the student was considered independent or dependent of a legal guardian. Enrollment information comes from the National Student Loan Data System (NSLDS). The NSLDS is the federal database that tracks information on all federal student loans and grants. Institutions and the Office of Federal Student Aid (FSA) work together to ensure that FSA has updated rosters that include currently enrolled students who have *ever* received Title IV aid. Key variables include enrollment status (e.g., full-time, part-time, withdrawn, graduated); an institution identifier, which we link to IPEDS and group as 4-year non-profit/public, 2-year non-profit/public and for-profit; and whether the student was awarded the Pell Grant in AY2023. The Pell Grant is the largest Title IV program that provides direct subsidies to students. Pell Grants are means-tested and so provide an indicator of being low income.

3.1 Outcomes (AY2024)

Our main outcomes are FAFSA submission, re-enrollment, full-time enrollment, and Title IV aid receipt for Academic Year 2024 (AY2024). These data are collected in the same manner as the baseline covariates. We also use the FAFSA submission date to construct dynamic outcomes: an indicator for submission by September 1, 2023 ("FAFSA by Fall") and the cumulative-by-week series used in Figure 1. Earlier submission of FAFSA is associated with higher state and institutional aid receipt, so the timing margin is of independent policy interest (McKinney and Novak, 2015).

One limitation of these data is that enrollment information is only available for students who have ever received Title IV aid. Hence, we do not observe enrollment status

[4]The pre-registered population (AEARCTR-0011295) differs by 146 records for which key NSLDS variables could not be retrieved at the data-pull date. The discrepancy is balanced across treatment and control.

for students who submit FAFSA in order to access state or institutional financial aid, but never receive grants or loans from the Federal government. We cannot distinguish between students who submit the FAFSA but never enroll and students who enroll but never use Title IV aid. Thus, randomization blocks labeled "not enrolled" include never-aided students who may or may not be enrolled. In total, never-aided students represent 52 percent of "not enrolled" students. Throughout the paper, we report results for four mutually exclusive AY2023-status groups that exhaust the sample: *Full-Time* (enrolled full-time in AY2023; holdout $n = 52{,}928$), *Non-Full-Time* (enrolled part-time, graduated, withdrawn, or other status; $n = 78{,}655$), *Not Enrolled* (no AY2023 enrollment record but previously Title-IV-aided; $n = 31{,}798$), and *Never-aided* (no Title IV history, so enrollment is unobserved; $n = 34{,}889$).

4 Empirical Strategy

4.1 Randomization

With the baseline characteristics outlined in Section 3, we formed 104 blocks:

$$\underbrace{\{\text{Male, Female+}\}}_{\text{Gender (2)}} \times \underbrace{\{\text{Dependent, Independent}\}}_{\text{Dependency Status (2)}} \times \underbrace{\{\text{Pell, Non-Pell}\}}_{\text{Pell Status (2)}}$$

$$\times \underbrace{\{\text{Not Enrolled}, \underbrace{\{\text{4-Year, 2-Year, For-Profit}\}}_{\text{Institution Type}} \times \underbrace{\{\text{Full-time, Part-time, Withdrawn, Other}\}}_{\text{Status}}\}}_{\text{Enrollment Type (13)}}$$

Within each block, we create an assignment vector to ensure the probability of treatment assignment is as close to 0.98 as possible. First, we calculate the number of complete groups of 100 within the block and, for each group, assign 98 treatment indicators and 2 control indicators. For any remaining units (those not fitting into groups of 100), we use

Bernoulli trials to determine the number assigned to treatment and control. Finally, we permute the vector to randomly assign individuals to either treatment or control. Appendix Table A2 shows balance is achieved on blocking variables and other baseline covariates.

4.2 Empirical Strategy

To estimate average treatment effects, we run the regression:

$$y_i = \tau_G D_i + \phi_g + \epsilon_i \tag{1}$$

y_i is an outcome for individual i. D_i is an indicator of being assigned an email. ϕ_g is a block fixed effect for block $g \in G$. τ_G is the treatment effect for students in group G which could be the full sample or one of the AY2023-status subgroups. τ_G is the object of interest reflecting the causal effect of being sent a reminder email. Because within-block treatment assignment probabilities are nearly constant, this is essentially a population-weighted average of within-block effects (Angrist, 1998). Standard errors are robust at the individual level. This is the appropriate level of aggregation for an experiment in which randomization occurs at the individual level within blocks.

The specification, primary outcomes (FAFSA renewal, AY24 enrollment, AY24 Title IV aid receipt), block design, and the use of multiple-testing correction were pre-registered (AEARCTR-0011295), but the AY24 full-time enrollment outcome, FAFSA-by-Fall timing outcome, and four-way AY2023-status decomposition (Full-Time / Non-Full-Time / Not Enrolled / Never-aided) are exploratory refinements of the pre-registered "enrollment" and "enrollment-status" subgroups.

Appendix Table A4 reports p-values adjusted via the sequentially-rejective procedure of Holm (1979) for all 25 outcome-by-subgroup ITT estimates in Table 1. The FAFSA-renewal effects in the pooled, Full-Time, Non-Full-Time, and Never-aided columns sur-

vive comfortably (adjusted $p < 0.001$); the pooled AY24 enrollment effect survives at the 5 percent level (Holm-adjusted $p \approx 0.047$); subgroup-specific enrollment effects in the Not Enrolled and Never-aided columns do not survive correction. We weight our discussion of individual cells accordingly.

4.3 A 2×2 Framework

Before introducing our estimands we sketch a conceptual framework to help interpret the results. Students face sequential decisions: *renew FAFSA* and *enroll in AY24*. First, the student decides whether to submit the FAFSA at a private cost c_p. After this decision, uncertainty about the net benefit of enrolling, including uncertainty about financial aid, resolves and the student decides whether to enroll. The net benefit includes the private returns of enrollment, financial aid, and costs such as re-engagement costs (c_e) for non-enrolled students. Submitting the FAFSA provides option value which allows the students to realize the value of FAFSA-dependent aid. The email reduces c_p but does not directly change other costs and benefits of enrollment.

The two decisions yield four student types, laid out as a 2×2 in Appendix Table A1: always-takers, who would renew and enroll without the email; students who would renew without the email but not enroll; and two kinds of compliers whose renewal the email induces, those who then enroll and those who only complete the paperwork. The email moves students across the renewal margin but does not change whether they would enroll conditional on renewing. The paperwork-only compliers create the policy-relevant tension: their renewals do not translate into enrollment, generating a paperwork burden without an enrollment benefit. Under this framework, the descriptive ratio r_G (defined below) would equal the share of compliers in group G who convert.

The framework yields two predictions we test below. (i) The renewal ITT τ_G^{renew} can be similar across groups when the density with option value below c_p is similar. (ii) The enrollment ITT τ_G^{enroll} varies with the share of compliers who convert, which is higher in

groups with low re-engagement cost c_e or less *ex-ante* uncertainty about the net benefit of re-enrollment (i.e., students already attached to college). We confirm both patterns in Section 5.3.

4.4 The Renewal-to-Enrollment Ratio

To summarize the relationship between the renewal margin and the enrollment margin we report the ratio of ITTs, $r_G \equiv \tau_G^{\text{enroll}} / \tau_G^{\text{renew}}$, in Table 1 Panel C. We interpret r_G descriptively, as the email-induced enrollment response per email-induced FAFSA renewer in group G. A formal LATE interpretation in the sense of Imbens and Angrist (1994) requires an exclusion restriction that the email moves AY24 enrollment *only* through FAFSA renewal. We are unable to test this directly: a salient government email about returning to college could in principle raise the perceived value of re-enrollment without operating through the renewed application. We therefore do not refer to r_G as a Wald ratio in the LATE sense, and we caution that cross-group differences in r_G confound two distinct sources of heterogeneity: differences in *who* the email induces to renew across groups, and differences in *what those induced renewers do next*. Both interpretations support the qualitative conclusion in Section 5.3, but quantitative statements like "77% of marginal renewers re-enroll" should be read with this caveat in mind.

4.5 Dynamic Effects

The timing of FAFSA submission matters as submitting earlier can help ensure access to state financial aid that may operate on a first-come, first-served basis. We get point-in-time treatment effect estimates by estimating the following equation:

$$y_{it} = \tau_t D_i \mathbb{1}\{\text{submission date} = t\} + \phi_g + \epsilon_i \tag{2}$$

This is nearly identical to Equation 1, except now the treatment assignment indicator is interacted with an indicator that equals 1 if the submission date equals t. The previous regressions produce daily treatment effects. We also care about the likelihood that a student will submit FAFSA before some date t. The following regression provides the *cumulative* impact of the communication campaign up to time t:

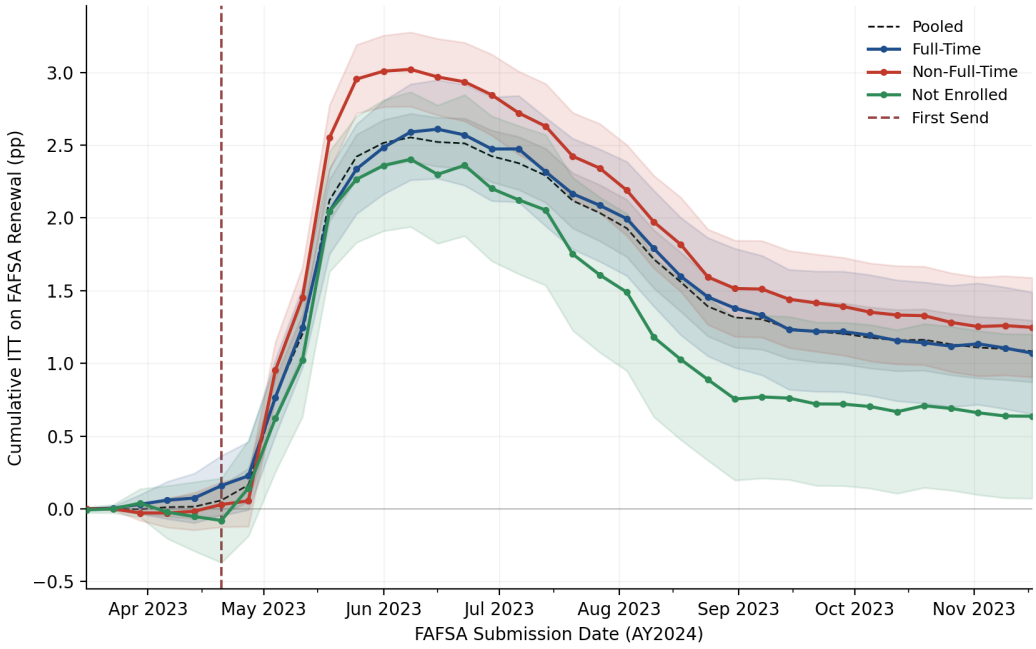
$$y_{it} = \tau_t D_i \mathbb{1}\{\text{submission date} \leq t\} + \phi_g + \epsilon_i \tag{3}$$

5 Results

5.1 Average Effects

In a 9.91-million-student block-randomized experiment, a single round of typical DOE FAFSA renewal emails caused 0.98 pp more renewals and 0.31 pp more re-enrollments. The re-enrollment effect sizes appear small but are statistically and economically significant. Most induced re-enrollments are full-time. The email moves paperwork uniformly across AY2023 status groups (renewal ITTs 0.51–1.15 pp) while moving enrollment only in the previously-enrolled subgroups (0.71 pp full-time; 0.32 pp non-full-time; statistically zero for the previously-unenrolled). Figure 1 shows the cumulative-by-status renewal trajectory.

Figure 1: Cumulative ITT on FAFSA Renewal by AY2023 Enrollment Status



Notes: Cumulative ITT, in percentage points, plotted against the date by which a student had submitted; the text refers to weeks relative to the first send (Week 0 = April 20, 2023), which the dashed vertical line marks. Dashed black line: pooled effect; colored lines: sample split by AY2023 enrollment status. Each point comes from a separate cross-sectional regression of an indicator for "submitted by week t" on treatment with block fixed effects (Equation 3); pointwise ± 2 SE bands (heteroskedasticity-robust, rebuilt from block-level counts as described in Appendix Table A8) are shown but the series share the same individuals across t and should not be read as jointly independent. The cumulative effect peaks at roughly 2.5–3 percentage points around weeks 5–9 for previously-enrolled students and attenuates as the holdout catches up; for students not enrolled in AY2023, the cumulative effect peaks similarly but decays to roughly 0.6 percentage points by week 30.

Reading Figure 1, three features stand out. First, the pooled cumulative effect (dashed black line) rises rapidly to roughly 2.5 percentage points around week 7, then attenuates to 1.1 percentage points by week 30 as the holdout catches up. The treatment moves paperwork forward in the renewal window and leaves a persistent year-end gap. Second, the Non-Full-Time subgroup (red) exhibits the largest response throughout, peaking at 3.0 pp and remaining above the Full-Time subgroup (blue) at every horizon; this ordering

is consistent with the renewal-ITT magnitudes reported in Table 1 (1.15 pp Non-Full-Time vs. 0.92 pp Full-Time). Third, the Not-Enrolled subgroup (green) reaches a similar early peak of about 2.4 pp but decays substantially faster, closing to roughly 0.6 pp by week 30.

Table 1 reports the main results. In the full sample (column 1), the baseline FAFSA renewal rate in the holdout group was 37.2 percent. Receiving the campaign emails caused that rate to rise by 0.98 percentage points (s.e. 0.11), a 2.6 percent increase corresponding to roughly 95,000 additional FAFSA submissions across the 9.7 million treated students. With 9.9 million observations, the point estimate is more than nine standard errors from zero. The implied 95% confidence interval is $[0.76, 1.19]$ percentage points. The treatment also raised AY24 college enrollment by 0.31 percentage points (s.e. 0.10) and full-time enrollment by 0.26 percentage points (s.e. 0.09), corresponding to roughly 30,000 additional enrollments and 25,000 additional full-time enrollments. AY24 Title IV aid receipt rose by 0.23 percentage points (s.e. 0.10).

Table 1: Treatment Effects on FAFSA Renewal and AY24 Enrollment

	Full Sample	Full-Time	Non-Full-Time	Not Enrolled	Never-Aided
Panel A: Holdout-group outcome means					
Renew FAFSA	0.372	0.396	0.377	0.480	0.228
FAFSA by Fall	0.317	0.345	0.311	0.425	0.190
Enroll AY24	0.360	0.461	0.382	0.480	0.048†
Enroll Full-Time AY24	0.251	0.382	0.206	0.381	0.037†
Aid AY24	0.279	0.333	0.286	0.390	0.077
Panel B: ITT (percentage points)					
Renew FAFSA	0.98	0.92	1.15	0.51	1.15
	(0.11)	(0.21)	(0.17)	(0.28)	(0.23)
FAFSA by Fall	1.30	1.35	1.51	0.75	1.31
	(0.10)	(0.21)	(0.17)	(0.27)	(0.21)
Enroll AY24	0.31	0.71	0.32	-0.08	0.14†
	(0.10)	(0.21)	(0.17)	(0.27)	(0.12)
Enroll Full-Time AY24	0.26	0.52	0.29	0.04	0.10†
	(0.09)	(0.21)	(0.14)	(0.27)	(0.10)
Aid AY24	0.23	0.44	0.30	-0.18	0.20
	(0.10)	(0.20)	(0.16)	(0.27)	(0.14)
Panel C: Renewal-to-enrollment ratio ($\hat{\tau}^{enroll}$ / $\hat{\tau}^{renew}$)					
	0.32	0.77	0.28	—	—
N (holdout)	198,270	52,928	78,655	31,798	34,889

Notes: Panel A reports holdout-group means. Panel B reports ITT estimates from Equation 1 (outcome on a treatment indicator with block fixed effects), in percentage points, with heteroskedasticity-robust standard errors in parentheses. Panel C reports the descriptive ratio of the Enroll AY24 ITT to the Renew FAFSA ITT (Section 4.4); for "Not Enrolled" the enrollment ITT is statistically zero so the ratio is uninformative, and for "Never-aided" enrollment is not directly observed. *Subgroups.* Full-Time and Non-Full-Time refer to AY2023 enrollment status; Not Enrolled students were not enrolled in AY2023 but had received Title IV aid previously; Never-aided students never received Title IV aid. †Enrollment outcomes for Never-aided students are mechanically attenuated because NSLDS coverage is conditional on past Title IV receipt.

Engagement with the emails was high. Of the 9.7 million emails sent, 98 percent were successfully delivered, 58 percent were opened (or at least registered a read receipt), and 4.5 percent triggered a click on an embedded link. Engagement varied modestly across AY2023 enrollment groups: full-time students opened at the highest rate (61.4 percent), and the previously-unenrolled or never-aided opened at the lowest (54.2 percent). Appendix Table A3 reports the full disaggregation.

A useful framing for the magnitude is cost per induced outcome. The marginal delivery cost of an additional mass email at federal-government scale is well below $0.01 per send. Commercial enterprise platforms quote per-send costs in the $0.001–$0.005 range at high-volume tiers, and Granicus/GovDelivery is the federal-government industry standard.[5] The 9.7-million-recipient campaign therefore costs roughly $10,000 in marginal delivery cost. Against approximately 25,000 induced full-time enrollments, this implies a cost on the order of $0.40 per induced full-time enrollment, before accounting for the much larger returns to college (Oreopoulos and Salvanes, 2011). Like the comparison figures below, this counts program costs only. Students' own time, at 30 to 60 minutes per form, adds roughly 2 to 4 hours of paperwork per induced full-time enrollment ($14 to $113 at $7.25 to $30 an hour; Section 5.4).

This cost is roughly comparable to vaccine-reminder text campaigns (Milkman et al., 2021), which use the same near-zero marginal-cost delivery technology. It is one to two orders of magnitude below the per-induced-outcome costs of at-scale paper-mail outreach in SNAP (Finkelstein and Notowidigdo, 2019) and ACA marketplace settings (Domurat et al., 2021; Goldin et al., 2021), where per-piece mail costs of $0.25 to $1 limit how low the per-induced-outcome figure can go, and two to three orders of magnitude below intensive or targeted interventions in higher education (Bettinger et al., 2012; Dynarski et al., 2021). Appendix Table A10 reports the detailed per-recipient costs, effect sizes, and implied per-induced-outcome figures across these comparison studies. The email campaign

[5]Federal-agency platform pricing is not publicly disclosed. Our estimate is based on Mailchimp's published high-volume-tier pricing structure.

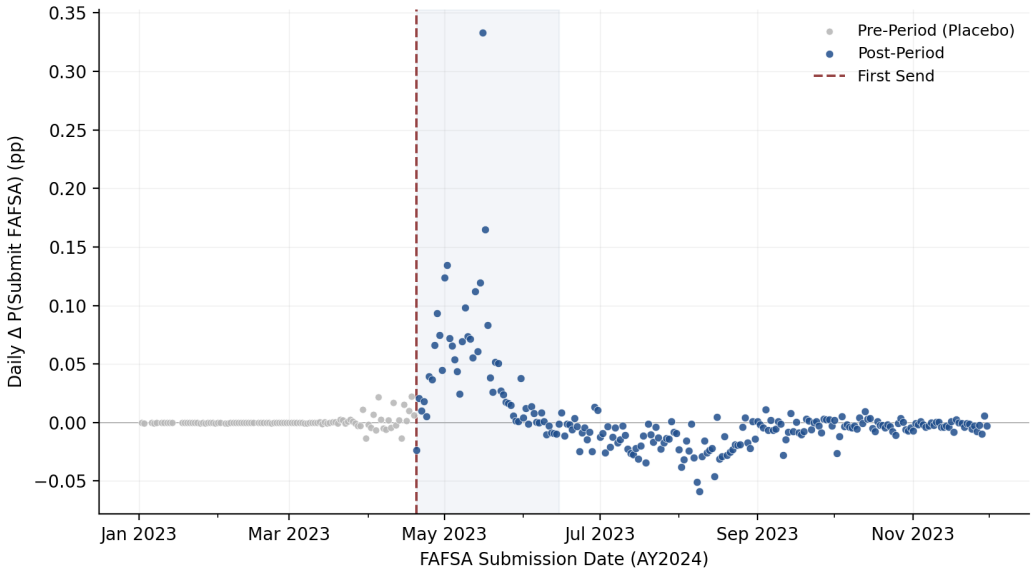
therefore sits in a distinct cost-effectiveness regime: it shares the near-zero marginal-cost structure of mass-text vaccine reminders while producing behavioral responses closer to those that have historically required more intensive or costly interventions.

The descriptive ratio of ITTs in the full sample, $r = 0.0031/0.0098 = 0.32$, indicates that the email-induced enrollment effect is roughly one third the email-induced renewal effect. As discussed in Section 4.4, we read this as a summary statistic and not as a LATE on enrollment. We return to its decomposition by AY2023 status in Section 5.3 after first establishing the timing of the response.

5.2 Dynamic Effects

Figure 2 plots the daily difference in the probability of submitting FAFSA for students who were sent reminder emails (treatment) compared to those who were not (control). Two features stand out. First, in the weeks between the data pull (March 28) and the first send (April 20), treatment and holdout students renew at nearly identical rates. This is a placebo check on the randomization: any pre-period divergence would call the design into question, and we find none. Second, immediately after April 20 the treatment rate spikes; on May 16, the highest-volume send day, the per-day probability of submitting FAFSA is more than twice the corresponding holdout rate.

Figure 2: Daily Difference in P(Submit FAFSA): Treatment vs. Holdout



Notes: Unadjusted daily difference in the probability of submitting FAFSA (treated minus holdout, in percentage points) by AY2024 FAFSA submission date. The dashed vertical line marks April 20, 2023, when the first email was sent. Pre-period (gray): treatment and holdout submit at indistinguishable rates. Post-period (blue): treatment submission spikes, with peaks on May 13 and May 16 (high-volume send days). Late June onward: daily differences turn negative as the holdout catches up.

Figure 1 shows the same response cumulatively. The campaign both moves submissions earlier in the renewal window, when state and institutional aid deadlines and first-come-first-served funds matter most (McKinney and Novak, 2015; Page et al., 2023), and produces a persistent year-end gap of 1.1 percentage points. The time path is broadly similar in shape across AY2023-status groups, but the not-enrolled series catches up much more completely: while the previously-enrolled groups retain a 1–1.5 percentage-point gap at week 30, the not-enrolled gap closes to roughly 0.6 points. Appendix Table A8 shows that this ranking is preserved at every fixed calendar cutoff from July 1 through November 15.

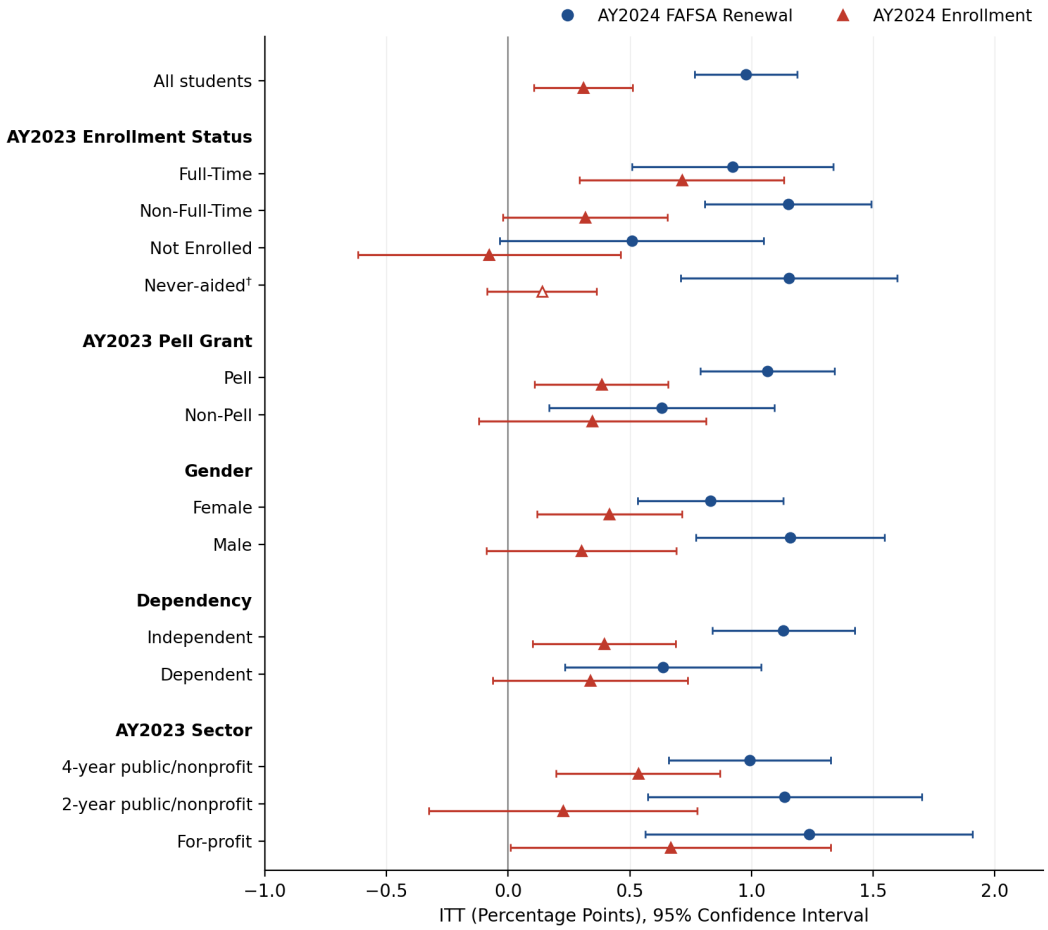
5.3 Who Responds?

The aggregate ratio $r = 0.32$ averages across populations with distinct responses. Columns 2–5 of Table 1 report ITTs for the four AY2023-status subgroups. The renewal effect is positive in every subgroup but varies in magnitude: 0.92 percentage points (Full-Time), 1.15 (Non-Full-Time), 0.51 (Not Enrolled), and 1.15 (Never-aided). The enrollment effect, where observable, is heterogeneous: 0.71 percentage points (s.e. 0.21) for the previously full-time, 0.32 (s.e. 0.17) for the previously non-full-time, and an estimate indistinguishable from zero (-0.08, s.e. 0.27; the 95% confidence interval rules out effects above 0.46 percentage points) for the previously unenrolled.[6]

The descriptive ratios r_G are 0.77 for the previously full-time, 0.28 for the non-full-time, and approximately zero for the previously unenrolled. As emphasized in Section 4.4, these are summary statistics rather than LATE estimates. In any case, the email moves paperwork in every group but moves enrollment only in groups already attached to college. For the previously-unenrolled, the campaign produces a renewed FAFSA without a detectable enrollment. Figure 3 summarizes the two-margin pattern across every subgroup we examine: the renewal ITT is positive and of similar magnitude in all of them, whereas the enrollment ITT is positive only for the previously-enrolled groups.

[6]Enrollment is not directly observed for Never-aided students (Section 3); we omit a renewal-to-enrollment ratio for that column in Table 1 Panel C.

Figure 3: ITTs on FAFSA Renewal and AY24 Enrollment by Subgroup

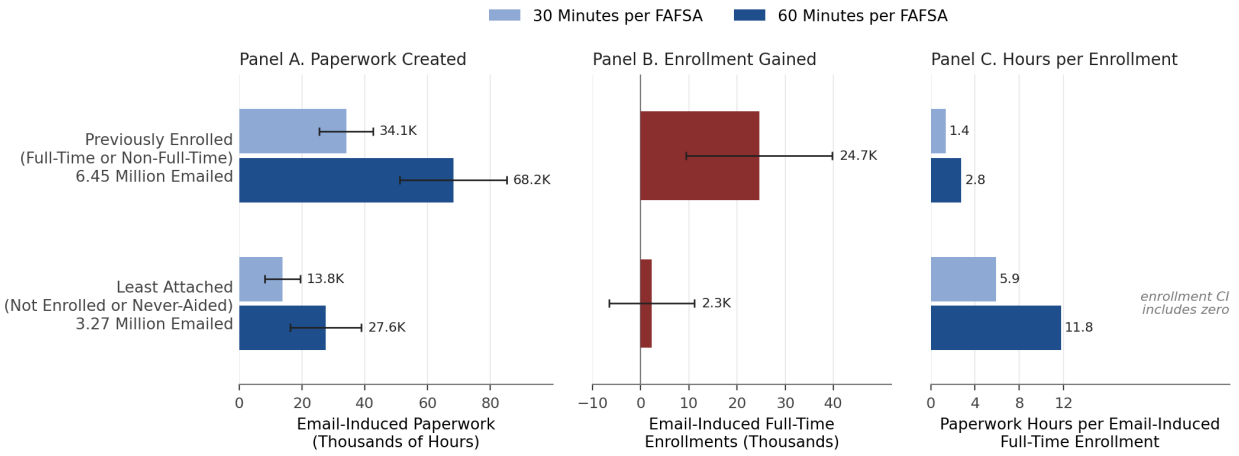


Notes: Point estimates and 95% confidence intervals (heteroskedasticity-robust) of the ITT on FAFSA renewal (circles) and AY24 enrollment (triangles), in percentage points. The first five rows are the full sample and the four AY2023-status groups of Table 1. †Enrollment is not directly observed for Never-aided students, so is observed only for those who receive Title IV aid in AY24, which the email itself raises, and is not comparable to the other rows (hollow marker). The remaining rows are the pre-registered Pell, gender, dependency, and AY2023-sector splits, aggregated from the 104 block-level estimates with block-fixed-effects weights; they exclude never-aided students (Appendix Table A5), and the sector rows are restricted to students enrolled in AY2023.

5.4 Paperwork and Targeting

Figure 4 converts the subgroup ITTs in Table 1 into totals for two groups: the previously enrolled (AY2023 Full-Time and Non-Full-Time), who are the most attached to college, and the least attached to college (AY2023 Not Enrolled and Never-aided). At 30 to 60 minutes per form, the email created 34,000 to 68,000 hours of paperwork among the previously enrolled and 14,000 to 28,000 among the least attached. The previously enrolled gained 24,700 full-time enrollments; the least attached gained 2,300. Paperwork per induced full-time enrollment is therefore 1.4 to 2.8 hours for the previously enrolled and 5.9 to 11.8 hours for the least attached, a ratio of two totals rather than a cost borne by any one student. Even the higher ratio is cheap: valuing student time at $7.25 to $30 an hour and a renewal without enrollment at zero, emailing the least attached pays for itself whenever an induced full-time enrollment is worth more than $50 to $360.

Figure 4: Email-Induced Paperwork and Full-Time Enrollment by Attachment to College



Notes: Previously enrolled: AY2023 Full-Time and Non-Full-Time students. Least attached: AY2023 Not Enrolled and Never-aided students. Induced counts are the subgroup ITTs in Table 1 multiplied by the treated subgroup size, summed across the two subgroups in each row. Panel A multiplies induced renewals by 30 or 60 minutes per form; the Department's Paperwork Reduction Act estimate for the 2023–24 FAFSA, about 40 minutes (OMB No. 1845-0001: 22.8 million burden hours across 34.0 million responses). Panels A and B show 95% confidence intervals from the robust standard errors in Table 1, treating the two subgroups in each row as independent. Panel C divides induced paperwork hours by induced full-time enrollments and reports point estimates only; it is a ratio of two totals, not a per-student cost. Enrollment for Never-aided students is observed only if they receive Title IV aid in AY24, which the email itself raises; they account for 1,700 of the 2,300 least-attached induced full-time enrollments.

Section 6 returns to the question of *why* the campaign works where prior at-scale nudges have failed. The design does not separate the four candidate channels (source, targeting, content, timing) experimentally, but the data focus the discussion. The daily-spike pattern in Figure 2 (and Appendix Figure A1 by status) is consistent with an attention/salience channel, though the design does not rule out other short-latency mechanisms. The cross-status enrollment heterogeneity suggests that a pure-information mechanism is unlikely to be the full story: every group receives the same email and the same baseline aid information, so what varies across groups must be the post-renewal enroll-

ment decision, not the renewal decision itself. And the contrast with Bird et al. (2021), whose at-scale statewide campaign also targeted current college students like our renewers, narrows the targeting explanation, leaving *source* and *content* as the most plausible remaining candidates. Definitive identification awaits a campaign that cross-randomizes the message source. One plausible candidate is a messenger or authority effect: as Appendix Figure A2 shows, the emails prominently identify "Federal Student Aid" as the sender, which may lead recipients to give them more consideration than a message from an unfamiliar organization. Evidence from EITC outreach cautions against over-weighting this channel, however: Linos et al. (2022) vary the messenger (a state tax agency versus a nonprofit, with different logos, signatures, and return addresses) across roughly 96,000 households and find no effect on take-up. A parallel line of inquiry, pursued in a related setting (renewal reminders to community-college students) by Athey et al. (2025), uses machine-learning tools (causal forests) to characterize the full distribution of individual-level effects and to derive optimal targeting policies. Their finding that intermediate-baseline students yield the largest per-nudge welfare gain is consistent with the intermediate ranking of our Non-Full-Time subgroup and points toward a natural extension of the analysis in this paper.

6 Conclusion

Each year the U.S. Department of Education attempts to remind students to resubmit the Free Application for Federal Student Aid (FAFSA). The goal is to nudge students to file FAFSA early so that they maintain access to means-tested financial aid and can continue to receive financial aid towards their college journey. In a 9.91-million-student block-randomized experiment, a typical U.S. Department of Education FAFSA renewal email campaign caused a 2.6 percent increase in renewal (about 95,000 additional submissions) and a 0.9 percent increase in college re-enrollment (about 30,000 additional enrollments),

at a marginal cost on the order of $0.40 per induced full-time enrollment (Section 5.1). The campaign moved students rapidly: on the highest-volume send day, the probability of submitting FAFSA was more than twice the holdout rate.

These positive effects stand in apparent tension with the precisely-estimated null effects from prior at-scale nudge experiments (Bird et al., 2021; Gurantz et al., 2021). We are cautious in attributing this difference to any single mechanism. The campaigns differ from one another along several dimensions: source (the federal aid administrator vs. external partners), targeting (FAFSA-renewal-eligible students vs. broader populations), message content, and timing. Our design does not vary any of these due to restrictions placed by the Office of Federal Student Aid, but Section 5.3 bounds the candidates using the within-campaign data. The targeting channel is partially ruled out by overlap with the sample in Bird et al. (2021), the daily-spike pattern is consistent with attention/salience, and the cross-status enrollment heterogeneity suggests that a pure-information story is unlikely to be the full explanation. Source (the federal aid administrator) and content (the specific message tone and information) remain the most plausible candidates. The federal-administrator source is the dimension along which this campaign differs most sharply from the partner-administered nulls, so it is the leading candidate, but the null messenger effects in Linos et al. (2022) warrant caution.

The aggregate enrollment effect masks sharp heterogeneity. Read descriptively, the email-induced enrollment effect is roughly 77 percent of the email-induced renewal effect for the previously-full-time enrolled, falls to 28 percent for the previously non-full-time, and is essentially zero for those not enrolled the year before. The framework of Section 4.3 interprets this as variation in the share of email-induced compliers who would convert renewal to enrollment ($v - c_e > 0$): low for the previously-unenrolled and high for the previously-attached. Either reading, however, supports the qualitative conclusion: only students already attached to college appear to convert a renewal nudge into a re-enrollment. This parallels evidence that nudge-induced take-up is selective: the marginal

enrollees induced by SNAP and ACA marketplace outreach differ systematically from other enrollees (Finkelstein and Notowidigdo, 2019; Domurat et al., 2021), and the mortality gains from IRS coverage outreach are concentrated among middle-aged adults (Goldin et al., 2021). The contribution of our setting is the simultaneous observation of both margins in administrative data, which permits the cross-status decomposition we report.

Two policy implications follow. First, the federal government's FAFSA renewal email campaigns are inexpensive enough and effective enough that the marginal cost of an additional round is small relative to the marginal benefit. The campaign produces full-time enrollment at a marginal cost on the order of $0.40 per induced student, in the same near-zero marginal-cost regime as vaccine-text reminders (Milkman et al., 2021), one to two orders of magnitude below at-scale paper-mail outreach in SNAP and ACA settings (Finkelstein and Notowidigdo, 2019; Domurat et al., 2021; Goldin et al., 2021), and two to three orders of magnitude below the per-induced-enrollment costs of intensive or targeted interventions in higher education (Bettinger et al., 2012; Dynarski et al., 2021). Defaulting to ongoing operation across years and renewal cycles is well-justified by the evidence here.

Second, the paperwork cost alone does not justify dropping the least-attached students. Sunstein (2022) argues that nudges should be judged by their distributional effects as well as their average effects, and that reminders may help inattentive people most. Our reminder reached every group but raised enrollment only among students already attached to college, so its enrollment gains are concentrated while its paperwork cost is widespread. However, the paperwork cost is cheap: excluding the least attached would spare 14,000 to 28,000 hours (Figure 4), yet emailing them pays for itself whenever an induced full-time enrollment is worth more than a few hundred dollars. If the least attached are also less advantaged, the attention to distributive justice that Sunstein (2022) calls for would only strengthen the case for including them.

References

Angrist, Joshua D. 1998. "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants." *Econometrica* 66 (2): 249–288. 10.2307/2998558.

Athey, Susan, Niall Keleher, and Jann Spiess. 2025. "Machine Learning Who to Nudge: Causal vs. Predictive Targeting in a Field Experiment on Student Financial Aid Renewal." *Journal of Econometrics* 249 105945. 10.1016/j.jeconom.2024.105945.

Avery, Christopher, Benjamin L. Castleman, Michael Hurwitz, Bridget Terry Long, and Lindsay C. Page. 2021. "Digital messaging to improve college enrollment and success." *Economics of Education Review* 84 102170. 10.1016/j.econedurev.2021.102170.

Bergman, Peter, Jeffrey T. Denning, and Dayanand Manoli. 2019. "Is Information Enough? The Effect of Information about Education Tax Benefits on Student Outcomes." *Journal of Policy Analysis and Management* 38 (3): 706–731. 10.1002/pam.22131, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pam.22131.

Bettinger, Eric P., Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. 2012. "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment." *The Quarterly Journal of Economics* 127 (3): 1205–1242. 10.1093/qje/qjs017.

Bhargava, Saurabh, and Dayanand Manoli. 2015. "Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment." *American Economic Review* 105 (11): 3489–3529. 10.1257/aer.20121493.

Bird, Kelli A., Benjamin L. Castleman, Jeffrey T. Denning, Joshua Goodman, Cait Lamberton, and Kelly Ochs Rosinger. 2021. "Nudging at scale: Experimental evidence from FAFSA completion campaigns." *Journal of Economic Behavior & Organization* 183 105–128. 10.1016/j.jebo.2020.12.022.

Bird, Kelli, and Benjamin L. Castleman. 2016. "Here Today, Gone Tomorrow? Investigating Rates and Patterns of Financial Aid Renewal Among College Freshmen." *Research in Higher Education* 57 (4): 395–422. 10.1007/s11162-015-9390-y.

Cochran, William G. 1954. "The Combination of Estimates from Different Experiments." *Biometrics* 10 (1): 101–129. 10.2307/3001666.

DellaVigna, Stefano, and Elizabeth Linos. 2022. "RCTs to Scale: Comprehensive Evidence From Two Nudge Units." *Econometrica* 90 (1): 81–116. 10.3982/ECTA18709, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA18709.

Domurat, Richard, Isaac Menashe, and Wesley Yin. 2021. "The Role of Behavioral Frictions in Health Insurance Marketplace Enrollment and Risk: Evidence from a Field Experiment." *American Economic Review* 111 (5): 1549–1574. 10.1257/aer.20190823.

Dynarski, Susan, C. J. Libassi, Katherine Michelmore, and Stephanie Owen. 2021. "Closing the Gap: The Effect of Reducing Complexity and Uncertainty in College Pricing on the Choices of Low-Income Students." *American Economic Review* 111 (6): 1721–1756. 10.1257/aer.20200451.

Finkelstein, Amy, and Matthew J. Notowidigdo. 2019. "Take-Up and Targeting: Experimental Evidence from SNAP." *The Quarterly Journal of Economics* 134 (3): 1505–1556. 10.1093/qje/qjz013.

Gerber, Alan S., and Donald P. Green. 2000. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment." *American Political Science Review* 94 (3): 653–663. 10.2307/2585837.

Goldin, Jacob, Ithai Z. Lurie, and Janet McCubbin. 2021. "Health Insurance and Mortality: Experimental Evidence from Taxpayer Outreach." *The Quarterly Journal of Economics* 136 (1): 1–49. 10.1093/qje/qjaa029.

Gurantz, Oded, Jessica Howell, Michael Hurwitz, Cassandra Larson, Matea Pender, and Brooke White. 2021. "A National-Level Informational Experiment to Promote Enrollment in Selective Colleges." *Journal of Policy Analysis and Management* 40 (2): 453–479. 10.1002/pam.22262, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pam.22262.

Heckman, James J., John Eric Humphries, and Gregory Veramendi. 2018. "Returns to Education: The Causal Effects of Education on Earnings, Health, and Smoking." *Journal of Political Economy* 126 S197–S246. 10.1086/698760, Publisher: The University of Chicago Press.

Higgins, Julian P. T., and Simon G. Thompson. 2002. "Quantifying Heterogeneity in a Meta-Analysis." *Statistics in Medicine* 21 (11): 1539–1558. 10.1002/sim.1186.

Holm, Sture. 1979. "A Simple Sequentially Rejective Multiple Test Procedure." *Scandinavian Journal of Statistics* 6 (2): 65–70.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2): 467–475. 10.2307/2951620.

Linos, Elizabeth, Allen Prohofsky, Aparna Ramesh, Jesse Rothstein, and Matthew Unrath. 2022. "Can Nudges Increase Take-Up of the EITC? Evidence from Multiple Field Experiments." *American Economic Journal: Economic Policy* 14 (4): 432–452. 10.1257/pol.20200603.

McKinney, Lyle, and Heather Novak. 2015. "FAFSA Filing among First-Year College Students: Who Files on Time, Who Doesn't, and Why Does It Matter?" *Research in Higher Education* 56 (1): 1–28. 10.1007/s11162-014-9340-0.

Milkman, Katherine L., Mitesh S. Patel, Linnea Gandhi et al. 2021. "A Megastudy of Text-Based Nudges Encouraging Patients to Get Vaccinated at an Upcoming Doctor's

Appointment." *Proceedings of the National Academy of Sciences* 118 (20): e2101165118. 10.1073/pnas.2101165118.

Oreopoulos, Philip, and Kjell G. Salvanes. 2011. "Priceless: The Nonpecuniary Benefits of Schooling." *Journal of Economic Perspectives* 25 (1): 159–184. 10.1257/jep.25.1.159.

Ost, Ben, Weixiang Pan, and Douglas Webber. 2018. "The Returns to College Persistence for Marginal Students: Regression Discontinuity Evidence from University Dismissal Policies." *Journal of Labor Economics* 36 (3): 779–805. 10.1086/696204, Publisher: The University of Chicago Press.

Page, Lindsay C., Bruce I. Sacerdote, Sara Goldrick-Rab, and Benjamin L. Castleman. 2023. "Financial Aid Nudges: A National Experiment With Informational Interventions." *Educational Evaluation and Policy Analysis* 45 (2): 195–219. 10.3102/01623737221111403.

Sunstein, Cass R. 2022. "The distributional effects of nudges." *Nature Human Behaviour* 6 (1): 9–10. 10.1038/s41562-021-01236-z, Number: 1 Publisher: Nature Publishing Group.

Turner, Lesley J. 2016. "The returns to higher education for marginal students: Evidence from Colorado Welfare recipients." *Economics of Education Review* 51 169–184. 10.1016/j.econedurev.2015.09.005.

A The 2×2 Framework

Table A1 lays out the four student types described in Section 4.3. The subsequent sections provide a more detailed decision process.

Table A1: Framework: A 2×2 of FAFSA Renewers and College Re-enrollers

	Would Enroll if Renewed $(v - c_e + a > 0)$	**Would Not Enroll if Renewed** $(v - c_e + a \leq 0)$
Would Renew Without Email $(OV > c_p)$	"Always-takers" No Email Effect	"Pre-FAFSA only" No Email Effect
Would Not Renew Without Email $(OV \leq c_p$, but $OV > c_p^{\text{email}})$	**Marginal: Converts** Enroll because of email	**Marginal: Paperwork Only** Paperwork Cost, No Gain

Notes: Stylized 2×2 of student types defined by (i) whether they would renew FAFSA absent the email and (ii) whether they would enroll conditional on renewal. The email lowers the paperwork friction c_p and moves marginal students from the top row to the bottom row. Among bottom-row compliers, those in the left column convert renewal to enrollment; those in the right column do not. The empirical question is what share of compliers in each AY2023-status group lies in each bottom-row column.

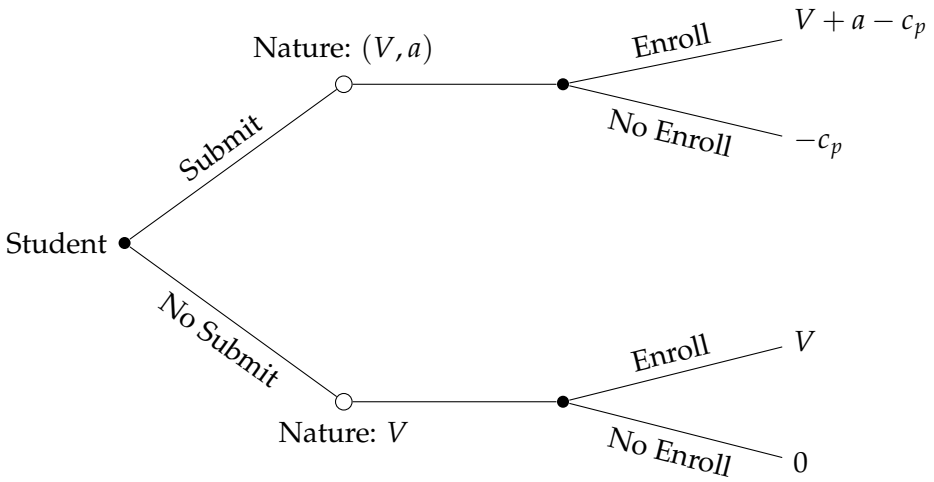
For ease of exposition, the framework excludes always-takers who would enroll with or without renewing FAFSA, but are nudged to submit FAFSA.

A.1 Setup and Timing

A student decides whether to submit the FAFSA, $s \in \{0, 1\}$, at cost c_p. After this decision, uncertainty about the value of enrolling and about aid resolves:

- $V_i = u_i - c_i^e$ is the realized net benefit of enrolling: the continuation value u_i minus the cost of enrolling c_i^e;
- $a_i \geq 0$ is realized financial aid, which offsets enrollment costs and is available only if the student submitted.

The student then chooses whether to enroll, with outside option normalized to 0.



Filled nodes are the student's decisions; hollow nodes are moves by Nature.

A.2 Enrollment and Submission Decisions

Enrollment. Given realizations, a student who submitted enrolls if $V + a > 0$; a student who did not submit enrolls if $V > 0$.

Submission. Students form subjective beliefs $\tilde{F}$ over (V, a); in particular, they may not precisely understand how a is realized. The student submits if

$$\tilde{\mathbb{E}}\big[\max(V + a,\, 0)\big] - c_p \;>\; \tilde{\mathbb{E}}\big[\max(V,\, 0)\big],$$

that is, if

$$c_p < OV \equiv \tilde{\mathbb{E}}\big[\max(V + a,\, 0) - \max(V,\, 0)\big].$$

Submitting provides an *option*: the right to enroll with aid, exercised only after uncertainty resolves. The email lowers the paperwork friction c_p and moves marginal students from the top row to the bottom row.

B Balance and Holdout-Group Outcome Means

Table A2 reports balance on baseline covariates (Panel A) and holdout-group means for the five AY24 outcomes used throughout the paper (Panel B). Differences in Panel A are absolute treatment-minus-holdout differences with heteroskedasticity-robust standard errors. Balance is achieved across all observable covariates.

Table A2: Balance and Holdout-Group Outcome Means

	Holdout Group		Treatment Difference	
	Mean	SD	|Diff|	p value
Panel A: Baseline covariate balance				
Prop. Female	.61	(.49)	$< 2 \times 10^{-5}$	.99
Prop. Pell AY23	.605	(.49)	$< 3 \times 10^{-5}$	.98
Not Enrolled AY23	.336	(.47)	$< 1 \times 10^{-6}$	>.99
Enrolled Full-Time AY23	.267	(.44)	$< 1 \times 10^{-5}$	>.99
Age	27.99	(9.56)	0.024	.28
Expected Family Cont. AY23	14,778	(44,618)	135	.18
Panel B: AY24 outcome means in the holdout group				
Renew FAFSA	.372	(.483)	—	—
FAFSA by Fall	.317	(.465)	—	—
Enroll AY24	.360	(.480)	—	—
Enroll Full-Time AY24	.251	(.434)	—	—
Aid AY24	.279	(.448)	—	—
Sample size	**Holdout: 198,270**		**Treatment: 9,716,427**	

Notes: Panel A shows balance on baseline characteristics. Differences are absolute treatment-minus-holdout differences from regressions of each covariate on the treatment indicator, without block fixed effects, with heteroskedasticity-robust standard errors, with block fixed effects the age and EFC differences have $p = 0.15$. Panel B shows holdout-group means and standard deviations for the five AY24 outcomes used in Table 1.

C Email Engagement Statistics

Table A3 reports email delivery, open, and click rates for the treated arm overall and by AY2023 enrollment status. These statistics motivate the inline summary in Section 5.1 and are descriptive of campaign reach.

Table A3: Email Engagement Statistics

	Total	Not Enrolled	Non-Full-Time	Full-Time
	Mean	Mean	Mean	Mean
Success Rate	.979	.969	.984	.984
Open Rate	.584	.542	.597	.614
Click Rate	.045	.043	.050	.039
Sample Size (millions)	9.72	3.27	3.85	2.59

Notes: Shows proportion of emails sent successfully, opened, and clicked as a fraction of the total (sub)sample. Information disaggregated by AY23 enrollment status; sample sizes are treated students, and the Not Enrolled column includes never-aided students.

D Family-Wise Error Correction

Table A4 reports p-values adjusted using the sequentially-rejective procedure of Holm (1979) for the family of 25 outcome-by-subgroup ITT estimates shown in Table 1 Panel B. Reading down the columns: the FAFSA-renewal and FAFSA-by-Fall effects in the pooled, Full-Time, Non-Full-Time, and Never-aided columns survive correction with adjusted $p < 0.001$. The pooled AY24 enrollment effect survives at the 5% level after correction (adjusted $p \approx 0.047$), and the Full-Time AY24 enrollment effect remains significant at the 5% level (adjusted $p \approx 0.015$). Several subgroup-specific cells, in particular all enrollment-related cells in the Not Enrolled columns, do not survive correction. The qualitative conclusion of Section 5.3, that enrollment effects are concentrated among the previously-enrolled, is therefore robust to family-wise correction in its directional form: the Full-Time enrollment effect remains significant, the Not Enrolled enrollment effect does not, and we are not in a position to claim a precise zero in the Not Enrolled subgroup but only a non-rejection of the null.

Table A4: Holm-Adjusted p-Values for Table 1 ITT Estimates

	Full Sample	Full-Time	Non-Full-Time	Not Enrolled	Never-aided
Panel A: Unadjusted p-values					
Any FAFSA	<0.001	<0.001	<0.001	0.067	<0.001
FAFSA by Fall	<0.001	<0.001	<0.001	0.006	<0.001
Enroll	0.003	<0.001	0.066	0.78	0.23
Enroll FT	0.006	0.012	0.046	0.88	0.32
Aid in AY24	0.023	0.030	0.070	0.50	0.18
Panel B: Holm-adjusted p-values (family of 25 hypotheses)					
Any FAFSA	<0.001	<0.001	<0.001	0.59	<0.001
FAFSA by Fall	<0.001	<0.001	<0.001	0.089	<0.001
Enroll	0.047	0.015	0.59	1.00	1.00
Enroll FT	0.085	0.15	0.46	1.00	1.00
Aid in AY24	0.27	0.33	0.59	1.00	1.00

Notes: Panel A reports the raw p-values associated with the ITT estimates in Table 1 Panel B, computed from heteroskedasticity-robust standard errors. Panel B reports Holm-adjusted p-values for the family of all 25 outcome-by-subgroup hypotheses. Cells with adjusted $p < 0.05$ remain statistically significant after correction; cells with adjusted $p > 0.10$ should be read as not separately distinguishable from zero under the family-wise correction.

E Heterogeneity Along the Pre-Registered Dimensions

The pre-analysis plan specified heterogeneity by Pell status, gender, dependency status, and institution type in addition to AY2023 enrollment status. Table A5 reports subgroup ITTs along each of these dimensions, built from the 104 block-level estimates (the table notes describe the construction; aggregating the same estimates by AY2023 status reproduces the corresponding columns of Table 1 to within 0.01 percentage points). Three patterns stand out. First, the renewal response is larger among Pell recipients (1.07 vs. 0.63 pp), male students (1.16 vs. 0.83 pp), and independent students (1.13 vs. 0.64 pp), although only the dependency gap approaches conventional significance for renewal ($p = 0.05$) and is significant for FAFSA-by-Fall ($p < 0.01$); the gender gap is significant only for aid receipt (-0.50 pp, $p = 0.04$), echoing Table A7. Second, the enrollment response is statistically indistinguishable across every demographic dimension: the Pell, gender, and dependency differences in the AY24 enrollment ITT are all within 0.11 pp of zero, with standard errors of roughly 0.25 pp. Third, the one sector pattern that emerges is consistent with the attachment story of Section 5.3: students at two-year institutions show a renewal response as large as any other sector (1.14 pp) but essentially no full-time enrollment response (0.01 pp, versus 0.50 pp at four-year institutions; difference $p = 0.08$). Taken together, the pre-registered demographic dimensions shift how strongly students respond on the paperwork margin but not whether renewal converts to enrollment. AY2023 enrollment status remains the dimension along which enrollment effects differ.

Table A5: Treatment Effects Along the Pre-Registered Heterogeneity Dimensions

	Renew FAFSA	FAFSA by Fall	Enroll AY24	Enroll FT AY24	Aid AY24	Enroll/Renew	N (millions)
Panel A: Pell Grant receipt (AY2023)							
Pell	1.07***	1.41***	0.38***	0.34***	0.32**	0.36	5.99
	(0.14)	(0.14)	(0.14)	(0.12)	(0.13)		
Non-Pell	0.63***	1.04***	0.35	0.27	0.07	0.55	2.18
	(0.24)	(0.23)	(0.24)	(0.23)	(0.23)		
Difference (Pell − Non-Pell)	0.43	0.38	0.04	0.07	0.25		
	(0.27)	(0.27)	(0.28)	(0.26)	(0.26)		
Panel B: Gender							
Female	0.83***	1.20***	0.42***	0.33**	0.07	0.50	5.15
	(0.15)	(0.15)	(0.15)	(0.14)	(0.15)		
Male	1.16***	1.50***	0.30	0.31*	0.57***	0.26	3.02
	(0.20)	(0.19)	(0.20)	(0.19)	(0.19)		
Difference (Female − Male)	-0.33	-0.30	0.11	0.02	-0.50**		
	(0.25)	(0.24)	(0.25)	(0.23)	(0.24)		
Panel C: Dependency status							
Independent	1.13***	1.56***	0.39***	0.38***	0.33**	0.35	5.22
	(0.15)	(0.14)	(0.15)	(0.13)	(0.14)		
Dependent	0.64***	0.89***	0.34*	0.22	0.11	0.53	2.95
	(0.21)	(0.20)	(0.20)	(0.20)	(0.20)		
Difference (Independent − Dependent)	0.50*	0.67***	0.06	0.16	0.22		
	(0.25)	(0.25)	(0.25)	(0.24)	(0.25)		
Panel D: AY2023 institution sector (students enrolled in AY2023)							
4-year public/nonprofit	0.99***	1.39***	0.53***	0.50***	0.35**	0.54	4.17
	(0.17)	(0.16)	(0.17)	(0.16)	(0.16)		
2-year public/nonprofit	1.14***	1.47***	0.23	0.01	0.28	—	1.48
	(0.29)	(0.28)	(0.28)	(0.23)	(0.27)		
For-profit	1.24***	1.69***	0.67**	0.51*	0.51	0.54	0.93
	(0.34)	(0.32)	(0.34)	(0.30)	(0.32)		
Difference (2-year − 4-year)	0.14	0.07	-0.31	-0.49*	-0.07		
	(0.33)	(0.32)	(0.33)	(0.28)	(0.31)		
Difference (For-profit − 4-year)	0.24	0.30	0.13	0.02	0.16		
	(0.38)	(0.36)	(0.38)	(0.34)	(0.36)		

Notes: ITT estimates in percentage points, with standard errors in parentheses, for the subgroups defined by the four pre-registered heterogeneity dimensions (AEARCTR-0011295). Estimates are built from the 104 block-level treatment–holdout differences that also underlie Figure **??** and Tables A6–A7; those regressions are estimated on students with any prior Title IV award, so never-aided students (whose enrollment is unobserved) are excluded throughout and N sums to the "Drop Never-aided" sample. Each subgroup estimate is the weighted average of the block differences in that subgroup with weights $n_g p_g(1-p_g)$, which is the block-fixed-effects estimator of Equation 1 restricted to the subgroup; aggregating by AY2023 status this way reproduces the Full-Time, Non-Full-Time, and Not Enrolled columns of Table 1 to within rounding. Standard errors treat blocks as independent samples, $\sqrt{\sum_g w_g^2 \, \widehat{\mathrm{se}}_g^2} / \sum_g w_g$, using heteroskedasticity-robust within-block standard errors. "Difference" rows report the gap between two subgroups with its standard error and a two-sided normal test. Panel D is restricted to students enrolled in AY2023, for whom an institution sector is observed; "For-profit" is the proprietary-sector category of the block design. The Enroll/Renew column is the descriptive ratio of Section 4.4, omitted where the enrollment ITT is within one standard error of zero. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

F Block-Level Heterogeneity

To assess whether the dispersion in block-level ITT estimates across the 104 randomization blocks reflects genuine heterogeneity rather than sampling variation, Table A6 reports Cochran's Q statistic (Cochran, 1954) and the corresponding I^2 (Higgins and Thompson, 2002) for the four observed outcomes. Cochran's Q is the inverse-variance-weighted sum of squared deviations of block-level estimates from the precision-weighted mean and is distributed χ^2_{K-1} under a constant-effects null. I^2 is the share of variation across blocks attributable to true heterogeneity rather than sampling. The block-level regressions are estimated on students with any prior Title IV award; never-aided students, whose enrollment is unobserved, are excluded from all block-level analyses in this appendix. The $\hat{\beta}$ column reports the inverse-variance-weighted mean of the block estimates; it differs from the pooled ITTs in Table 1 both for this reason and because the block-fixed-effects regression weights blocks by $n_g p_g(1 - p_g)$ rather than by the inverse of the estimated block variance.

Three patterns emerge. First, for AY24 full-time enrollment, $I^2 \approx 0\%$ ($p \approx 0.55$): we cannot reject a constant-effects null, so the apparent dispersion in block-level full-time ITTs is consistent with sampling noise alone. Second, for FAFSA renewal and AY24 enrollment, $I^2 \approx 16$–17% ($p \approx 0.08$–0.10): the evidence for heterogeneity across blocks is suggestive but not statistically conventional. Third, AY24 aid receipt shows clear heterogeneity ($I^2 \approx 43\%$, $p < 0.001$). This last pattern, substantial across-block variation in the aid-receipt effect, is consistent with institutional differences in how Title IV awards convert from a renewed FAFSA, and is a natural target for future heterogeneity work.

Table A7 asks a complementary question: which observable block characteristics predict the dispersion in block-level treatment effects? For each outcome, we run a single weighted-least-squares regression of the block ITT on indicators for AY2023 Pell receipt, gender, dependency, school sector, and AY2023 enrollment status, using inverse-variance weights and heteroskedasticity-robust standard errors. Three findings stand out. First,

Table A6: Block-Level Heterogeneity Tests

Outcome	$\hat{\beta}$ (pp)	Cochran's Q	df	p	I^2
FAFSA renewal	0.98	123.5	103	0.083	17%
AY24 enrollment	0.39	122.2	103	0.096	16%
AY24 full-time enrollment	0.33	100.7	103	0.546	0%
AY24 aid receipt	0.34	180.0	103	<0.001	43%

Notes: For each outcome, we test whether the block-level ITT estimates from the 104 randomization blocks reject a constant-effects null. $\hat{\beta}$ is the inverse-variance-weighted mean of the block estimates (in percentage points). Cochran's Q is the inverse-variance-weighted sum of squared deviations from $\hat{\beta}$ and is distributed χ^2_{K-1} under the constant-effects null. $I^2 = \max(0, (Q-(K-1))/Q)$ is the share of variation across blocks attributable to true heterogeneity rather than sampling variation.

the most consistent predictor is the female-block indicator: female blocks have systematically smaller ITTs on both FAFSA renewal (-0.44 pp, $p < 0.05$) and AY24 aid receipt (-0.70 pp, $p < 0.01$), but no detectable effect on enrollment (0.08 pp, s.e. 0.25). Second, the AY2023-Not-Enrolled indicator predicts smaller renewal and aid-receipt effects (-0.73 and -1.36 pp, both $p < 0.10$), consistent with the cross-status pattern documented in Table 1. Third, no other predictor: Pell, dependency, or school sector is statistically distinguishable from zero. The block-level heterogeneity that Cochran's Q detects is therefore concentrated along the gender and enrollment-status dimensions; sector and Pell-status differences contribute little after conditioning on these. We caution that these are correlations across the 104 randomization blocks, not causal effects of demographic characteristics on treatment response.

Table A7: Predictors of Cross-Block Heterogeneity in Treatment Effects

	FAFSA Renewal	AY24 Enrollment	AY24 Aid Receipt
Intercept	1.23**	0.57	1.66*
	(0.54)	(0.55)	(1.00)
Pell (AY23)	0.23	-0.01	-0.15
	(0.25)	(0.29)	(0.35)
Female	-0.44**	0.08	-0.70***
	(0.22)	(0.25)	(0.26)
Independent	0.32	-0.03	-0.05
	(0.27)	(0.28)	(0.36)
2-year sector	-0.20	-0.41	-0.69
	(0.41)	(0.44)	(0.57)
4-year sector	-0.23	-0.23	-0.66
	(0.31)	(0.39)	(0.56)
Full-Time AY23	-0.21	0.35	-0.14
	(0.25)	(0.30)	(0.35)
Not Enrolled AY23	-0.73*	-0.67	-1.36*
	(0.43)	(0.51)	(0.76)
N (blocks)	104	104	104

Notes: Each column reports a single weighted-least-squares regression at the block level. The dependent variable is the block's ITT estimate from Equation 1 for the indicated outcome (in percentage points). Predictors are block characteristics: AY2023 Pell status, gender, dependency, school sector, and AY2023 enrollment status. Sector indicators (2-year, 4-year) are relative to private as the omitted category. Enrollment-status indicators (Full-Time AY23, Not Enrolled AY23) are relative to Non-Full-Time as the omitted category. Weights are inverse variances of the block-level estimates. Standard errors (in parentheses) are heteroskedasticity-robust. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

G Cumulative ITT at Fixed Calendar Cutoffs

Table A8 reports cumulative-ITT estimates at five calendar cutoffs of operational interest, motivated by state and institutional FAFSA-deadline policies. The pooled and by-status patterns are stable across reasonable cutoff choices: the pooled effect is roughly 2.4 percentage points by July 1, declining to 1.9 by August 1 and 1.1 by mid-November as the holdout catches up. The Not Enrolled subgroup catches up most completely (gap of 0.6 pp by November 15), the Non-Full-Time subgroup least (1.3 pp). The qualitative ranking documented in Figure 1 is preserved at each horizon.

Table A8: Cumulative ITT on FAFSA Renewal at Fixed Calendar Cutoffs

Cutoff (calendar)	Pooled	Full-Time AY23	Non-Full-Time AY23	Not Enrolled AY23
July 1 (week 11)	2.38 (0.09)	2.47 (0.18)	2.72 (0.14)	2.12 (0.25)
August 1 (week 15)	1.93 (0.10)	1.99 (0.20)	2.19 (0.16)	1.49 (0.27)
Sept. 1 (week 19)	1.32 (0.10)	1.38 (0.20)	1.51 (0.17)	0.76 (0.28)
Oct. 1 (week 23)	1.20 (0.10)	1.22 (0.21)	1.39 (0.17)	0.72 (0.28)
Nov. 15 (week 30)	1.08 (0.11)	1.07 (0.21)	1.25 (0.17)	0.64 (0.28)

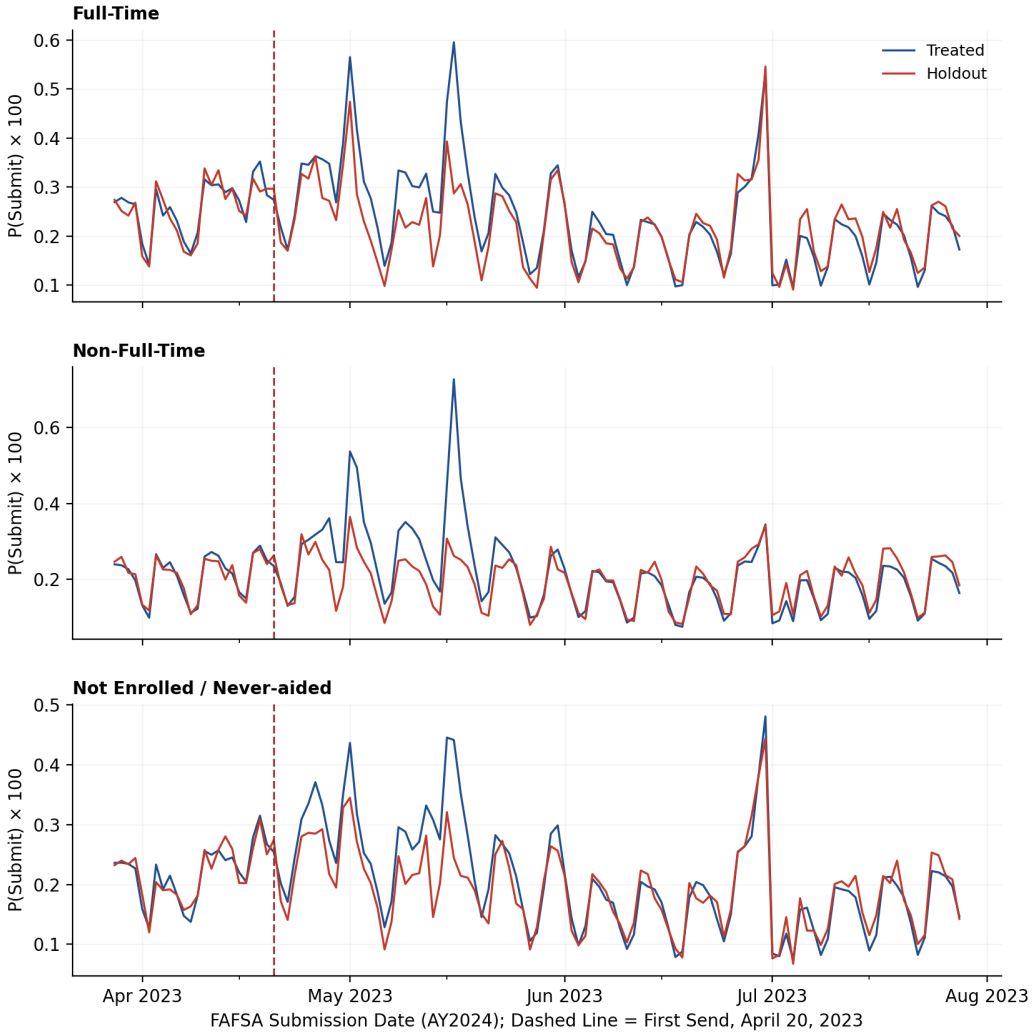
Notes: Reports cumulative ITT estimates on the indicator "submitted FAFSA on or before week t" from Equation 3 (block fixed effects, individual-level robust SEs in parentheses), in percentage points. Week 0 = April 20, 2023. Each calendar label is approximate; the exact week-to-date mapping is week t = April 20 + $7t$ days. Estimates by AY2023 status come from regressions restricted to each subgroup; "Never-aided" students are not directly observed in NSLDS enrollment and are excluded from this table.

H Daily Renewal Response by AY2023 Status

Figure A1 disaggregates the pooled daily-spike pattern in Figure 2 by AY2023 enrollment status. Two patterns emerge. First, all three panels show a clear treatment spike on May 13 and May 16–the two highest-volume send days–with the treated daily probability reaching 0.4–0.7 percent while the holdout series shows no corresponding send-day spike. Both series spike together at the start of July, independent of treatment status. The timing pattern documented in the pooled figure is therefore not driven by any single AY2023-status group; the email moves students rapidly in every subgroup for which we observe daily submissions.

Second, the magnitude of the spike differs by group. The Non-Full-Time and Not-Enrolled panels exhibit the sharpest treated-versus-holdout gap on send days, while the Full-Time panel shows the highest baseline submission rate but a more compressed treated-holdout difference. This is consistent with the cumulative pattern in Figure 1: groups with lower baseline renewal probabilities show larger *relative* spikes when nudged, but the holdout catches up faster in the Not-Enrolled / Never-Aided combined group, eroding the cumulative gap by year-end.

Figure A1: Daily Probability of Submitting FAFSA, Treated vs. Holdout by AY2023 Enrollment Status



Notes: Each panel plots the daily share of the population (per 100) submitting FAFSA on each calendar date for the treated arm (blue) and the holdout (red). The dashed vertical line marks April 20, 2023 (first send). Because the randomization-block identifier alone cannot distinguish students who never received Title IV aid from students who were not enrolled in AY2023 but had been Title-IV-aided previously, the bottom panel combines both groups.

I Pre-Registration Crosswalk

The experiment was pre-registered at the AEA RCT Registry on May 31, 2023 (last updated June 6, 2023), as AEARCTR-0011295 (Khan, Anisfeld, and Kramer, 2023), prior to the receipt of any AY24 outcome data. The registry lists a planned intervention window of April 11–14, 2023. Operational warm-up, managed by the US Department of Education, delayed the first send to April 20, with the bulk of sends on May 13 and May 16 (Section 2). Table A9 maps each element of the pre-analysis plan to its location in the paper, distinguishing pre-specified analyses from exploratory ones.

Table A9: Pre-registration Crosswalk

Pre-registered Element	Location in Paper	Status
Block-randomized design over 104 strata defined by gender, dependency, Pell, institution type, and AY2023 enrollment status	§4; balance in Appendix Table A2	Pre-specified
Sample: ≈10M FAFSA-renewal-eligible students; control ≈ 2%	§3; n = 9,914,697 analytic	Pre-specified
Primary outcome: FAFSA renewal	Table 1 ("Renew FAFSA"); Fig. 1	Pre-specified
Primary outcome: AY24 re-enrollment	Table 1 ("Enroll AY24")	Pre-specified
Primary outcome: AY24 Title IV aid receipt	Table 1 ("Aid AY24")	Pre-specified
Heterogeneity by Pell status, gender, dependency, institution type, and enrollment status	§5.3; Fig. 3; Appendix Table A5 (subgroup ITTs); Appendix Table A7 and Fig. **??**	Pre-specified
Multiple-testing correction across primary outcomes and subgroups	Appendix Table A4 (Holm); §4	Pre-specified
Three pre-registered message-content variants (pooled in the main analysis)	§2; not exploited individually in main analysis	Pre-specified (operational implementation prevented clean estimation; see Appendix J)
AY24 full-time enrollment as a separate outcome	Table 1 ("Enroll Full-Time AY24")	Exploratory
FAFSA-by-Fall timing outcome (cumulative by Sept. 1)	Table 1 ("FAFSA by Fall"); Fig. 1	Exploratory
Four-way AY2023-status decomposition (separate "Never-aided" stratum)	§3; Table 1 columns 2–5	Exploratory
Renewal-to-enrollment ratio r_G (descriptive)	§4.4; Table 1 Panel C	Exploratory
2×2 conceptual framework	§4.3; Appendix Table A1	Exploratory
Cumulative-by-week and daily dynamics	§5.2; Figs. 1, 2	Exploratory
Block-by-block heterogeneity analysis (104 blocks)	§5.3; Fig. **??**; Appendix F	Exploratory
Cost-effectiveness back-of-envelope (∼$0.40 / induced FT enrollment)	§5.1; §6	Exploratory
Tone A/B (neutral vs. urgent)	§2	Added post-registration; not analyzed for content effects

Notes: The first block lists elements specified in the original pre-analysis plan. The second block lists elements added during analysis that respond to the design opportunity (e.g., the descriptive ratio, the cumulative dynamics) or to data realities (e.g., the Never-aided stratum, made necessary because NSLDS enrollment is unobserved for that group). The tone A/B was implemented after registration and is documented but not used for inference. All main-text claims about renewal, enrollment, and aid effects rest on the pre-specified analysis.

J Email Designs and Within-Treatment Variation

Email Designs

Figure A2 reproduces the proofs of the two most commonly sent variants of the renewal email delivered. Both communicate the same core message—that the Federal Pell Grant and Title IV aid do not automatically renew, that students must resubmit FAFSA for AY24, and that many state and institutional aid programs use first-come-first-served deadlines. They differ in tone and emphasis. Pell eligible students receive messaging that replaces "FAFSA form" with "Pell Grant". The pre-registered design also included two additional content arms not displayed here. These are very similar to Figure A2 (a), except they include more information about the call to action in the "above the fold" blue region[7] and have links that provide information about loan forgiveness programs (PSLF) or income-driven repayment plans. We treat the content arms as a single pooled treatment in the main analysis.

Content Effects

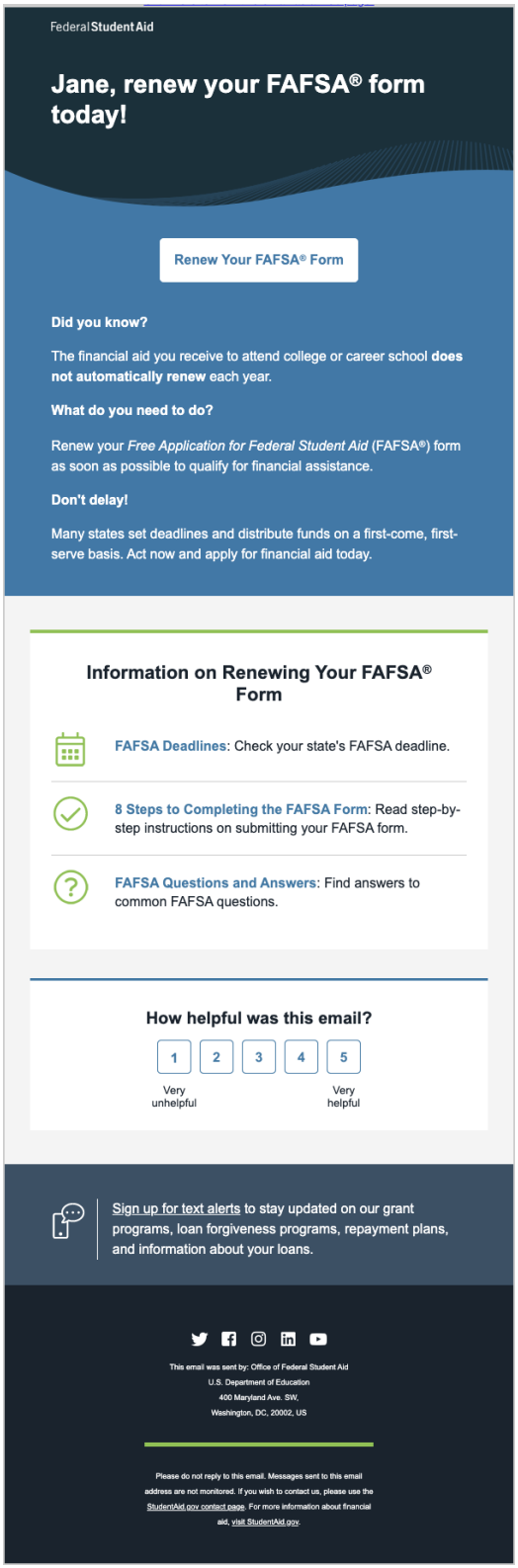
We do not identify separate content effects because of implementation issues. The initial design assigned individuals to the messaging variants; however, due to our implementation partner's needs, they introduced the A/B test described above. They decided to keep the treatment assignment for one in nine of the students in our initial 3-arm test, creating an effective 5-arm experiment. Unfortunately, while respecting our pre-assigned treatment labels, there was a failure of randomization in how the system chose which emails to enter into the A/B test. Balance checks on the realized composition show non-trivial imbalance on baseline covariates that were balanced across the treatment-control split (e.g., AGI and AY2023 enrollment-status indicators). A simple OLS regression of

[7]The first and third text blocks are identical. In the second, the emails ask "How do you renew your FAFSA form?" and include an additional sentence about renewing FAFSA if you plan on attending college or career school between July 1, 2023, and June 30, 2024.

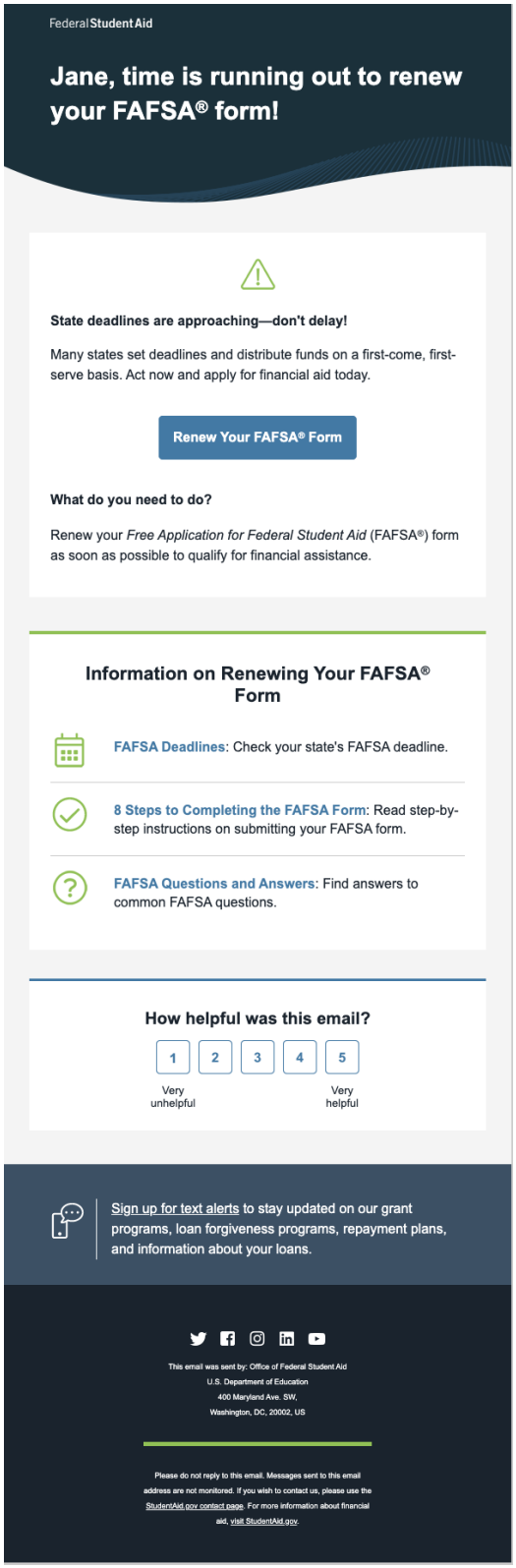
message-cell indicators on baseline covariates rejects the null of independent assignment.

Accordingly, we treat the 2023 campaign as a single intervention in the main analysis and do not draw conclusions about the relative efficacy of tone or content variants. We feel that the main treatment-control split still provides clear evidence about the efficacy of email campaigns. Namely, all the renewal effects are similar across groups. A small non-random sample of students received messaging about income driven repayment or public service loan forgiveness which may influence enrollment, which we cannot rule out. However, this represents less than 8 percent of the sample and so the message would need to be very effective to drive the heterogeneity results. We hope that future at-scale campaigns will follow cross-randomized variation in source and content so that these mechanism questions can be addressed.

Figure A2: Email Designs



(a) Neutral tone ("Jane, renew your FAFSA form today!").



(b) Urgent tone ("Jane, time is running out to renew your FAFSA form!").

K Cost-Effectiveness Across Behavioral Interventions

Table A10 situates the FAFSA email campaign's cost-effectiveness alongside other behavioral interventions in adjacent settings cited in the introduction. The table is intended to discipline the qualitative "orders-of-magnitude" comparisons in Sections 5.1 and 6 rather than to claim precision: cost figures from comparison studies are paper-reported where available and otherwise reconstructed from reported per-recipient costs and effect sizes. Bhargava and Manoli (2015), compare notice designs among tax filers who all received a notice and do not report costs, so no per-induced-claim cost can be computed. For Finkelstein and Notowidigdo (2019), the contrast between information-only and assistance arms within the same study illustrates the intensive-vs-at-scale gap directly. Milkman et al. (2021) report that the two text messages cost less than a dime per patient. The null entries for Bird et al. (2021) and Gurantz et al. (2021) reflect precisely-estimated null treatment effects, so the implied per-induced-outcome cost is undefined.

Three patterns emerge from the table. First, at-scale paper-mail outreach in SNAP and ACA settings produces per-induced-outcome costs of roughly $20 to $55, bounded below by per-piece mail costs of $0.25 to $1. Second, mass digital delivery (text or email) lowers the per-recipient cost by another one to two orders of magnitude, putting the FAFSA email campaign and the vaccine-text megastudy in the same low-marginal-cost regime. Third, intensive or targeted interventions in higher education (Bettinger et al., 2012; Dynarski et al., 2021) produce much larger per-recipient effects at per-induced-enrollment costs of roughly $66 to $1,100, two to three orders of magnitude above the email campaign; door-to-door canvassing (Gerber and Green, 2000), at roughly $8 per induced vote, sits between the two regimes.

Table A10: Cost-effectiveness comparison: at-scale and intensive behavioral interventions

Study	Setting	Scale	Cost / recipient	Effect on primary outcome	Implied $ / induced outcome
Panel A: This paper					
Anisfeld, Khan & Kramer (this paper)	FAFSA renewal email (federal source)	At-scale (n = 9.91M)	\$0.001–\$0.005 / email (marg.)	+0.98 pp on renewal; +0.26 pp on FT enrollment	\$0.39-\$1.93 / FT enrollment
Panel B: At-scale outreach (mailed letters, mass digital)					
Bhargava and Manoli (2015)	EITC IRS reminder notice	At-scale ($n \approx$ 35K)	N/A	+4 to +9pp	N/A
Finkelstein and Notowidigdo (2019)	SNAP info-only mailing	At-scale ($n \approx$ 32K)	∼\$1 / recipient	+4.7 pp on enrollment	∼\$20 / induced enrollment
Domurat et al. (2021)	ACA Covered California letter	At-scale ($n \approx$ 87K)	∼\$0.69 / letter	+1.3 pp on enrollment	∼\$53 / induced enrollment
Goldin et al. (2021)	IRS letter to mandate-payers (ACA)	At-scale ($n \approx$ 4.5M)	∼\$0.49 / letter	+0.69 pp on any coverage	∼\$36 / induced enrollment
Milkman et al. (2021)	Vaccine text megastudy	At-scale ($n \approx$ 47K)	∼\$<0.10 / patient	+2.1 pp avg	∼\$2-\$5 / induced vaccination
Bird et al. (2021); Gurantz et al. (2021)	National FAFSA nudges (nonprofit and state partners)	At-scale	Low (mass digital)	∼0 (precise null)	N/A
Panel C: Intensive, targeted, or local interventions					
Finkelstein and Notowidigdo (2019)	SNAP application assistance	Intensive ($n \approx$ 11K)	∼\$7 / recipient	+11.8 pp on enrollment	∼\$60 / induced enrollment
Gerber and Green (2000)	Door-to-door voter canvassing	Local ($n \approx$ 5.8K)	∼\$1.50 / contact	+8.7 pp on turnout	∼\$8 / induced vote
Bettinger et al. (2012)	H&R Block FAFSA + counseling	Local & intensive ($n \approx$ 17K)	∼\$88 / recipient	+16 pp FAFSA filing; +8 pp dependent enrollment	∼\$1,100 / induced enrollment
Dynarski et al. (2021)	HAIL Michigan (targeted outreach)	Targeted at-scale ($n \approx$ 3.9K)	∼\$10 / recipient	+15 pp on U-M enrollment	∼\$66 / induced enrollment

Notes: Cost-per-recipient figures are paper-reported where the original study reports them and otherwise approximated from standard federal direct-mail rates (∼\$1 / piece) or platform-pricing benchmarks for mass digital delivery. Cost-per-induced-outcome figures are paper-reported where available, and otherwise computed as cost-per-recipient ÷ percentage-point effect on the primary outcome, expressed in dollars per induced enrollee/claim/vote/vaccination. Numbers should be treated as "in the right order of magnitude" rather than precise: actual federal procurement costs and intervention-implementation budgets are not always reported.