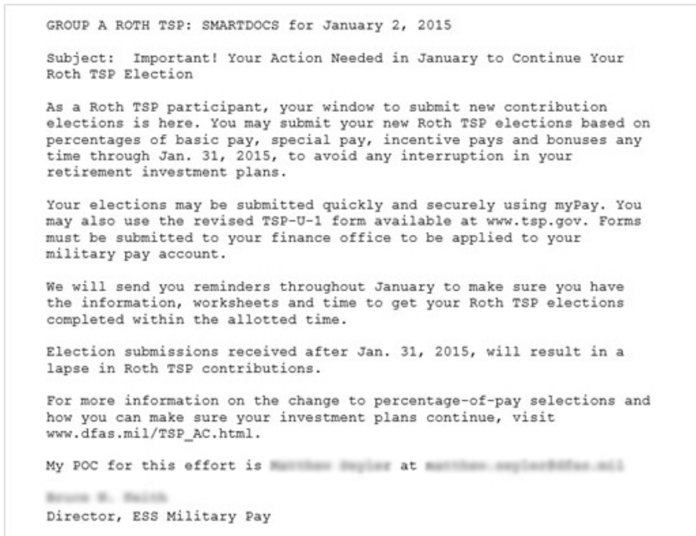
Figure 1: Example of nudges

(a) OES example: Control communication



GROUP A ROTH TSP: SMARTDOCS for January 2, 2015

Subject: Important! Your Action Needed in January to Continue Your Roth TSP Election

As a Roth TSP participant, your window to submit new contribution elections is here. You may submit your new Roth TSP elections based on percentages of basic pay, special pay, incentive pays and bonuses any time through Jan. 31, 2015, to avoid any interruption in your retirement investment plans.

Your elections may be submitted quickly and securely using myPay. You may also use the revised TSP-U-1 form available at www.tsp.gov. Forms must be submitted to your finance office to be applied to your military pay account.

We will send you reminders throughout January to make sure you have the information, worksheets and time to get your Roth TSP elections completed within the allotted time.

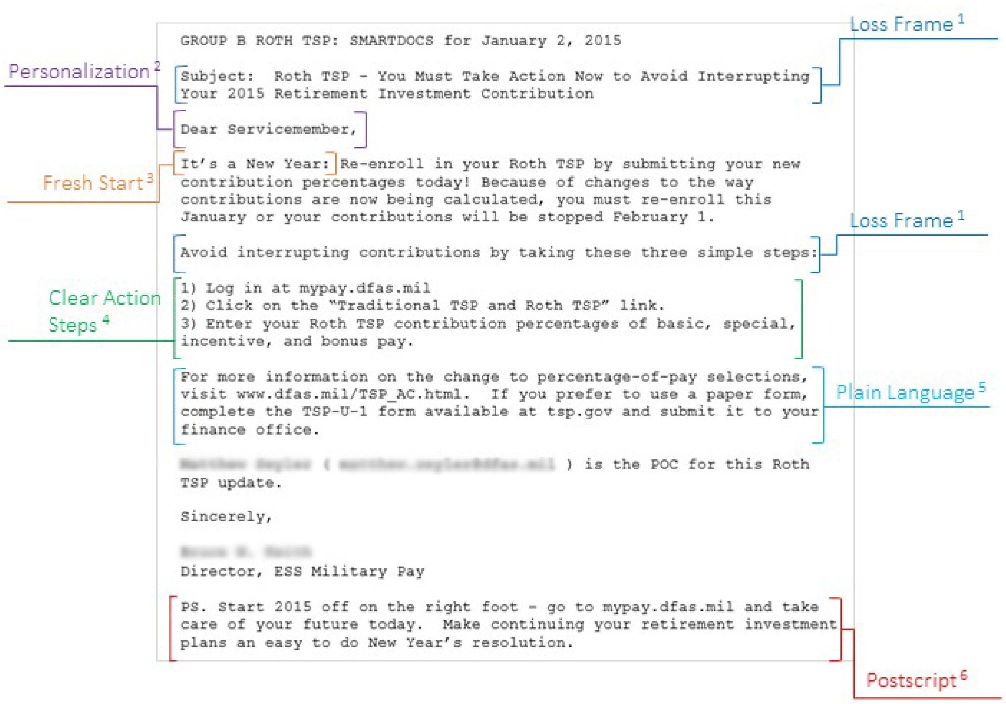
Election submissions received after Jan. 31, 2015, will result in a lapse in Roth TSP contributions.

For more information on the change to percentage-of-pay selections and how you can make sure your investment plans continue, visit www.dfas.mil/TSP_AC.html.

My POC for this effort is [illegible] at [illegible]

[illegible]
Director, ESS Military Pay

(b) OES example: Treatment communication



GROUP B ROTH TSP: SMARTDOCS for January 2, 2015

Subject: Roth TSP - You Must Take Action Now to Avoid Interrupting Your 2015 Retirement Investment Contribution

Dear Servicemember,

It’s a New Year: Re-enroll in your Roth TSP by submitting your new contribution percentages today! Because of changes to the way contributions are now being calculated, you must re-enroll this January or your contributions will be stopped February 1.

Avoid interrupting contributions by taking these three simple steps:

1) Log in at mypay.dfas.mil
2) Click on the “Traditional TSP and Roth TSP” link.
3) Enter your Roth TSP contribution percentages of basic, special, incentive, and bonus pay.

For more information on the change to percentage-of-pay selections, visit www.dfas.mil/TSP_AC.html. If you prefer to use a paper form, complete the TSP-U-1 form available at tsp.gov and submit it to your finance office.

[illegible] ([illegible]) is the POC for this Roth TSP update.

Sincerely,

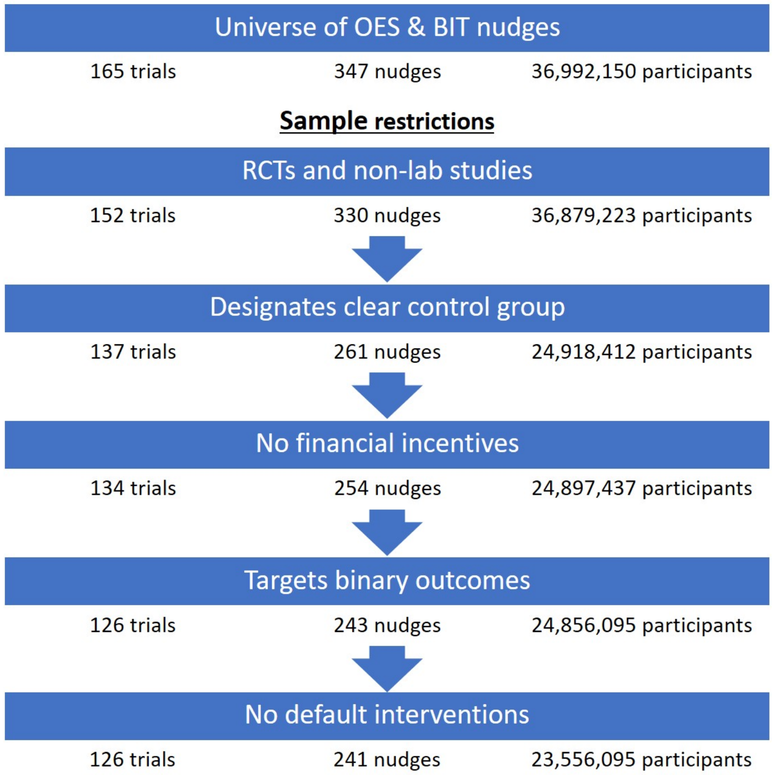
[illegible]
Director, ESS Military Pay

PS. Start 2015 off on the right foot - go to mypay.dfas.mil and take care of your future today. Make continuing your retirement investment plans an easy to do New Year’s resolution.

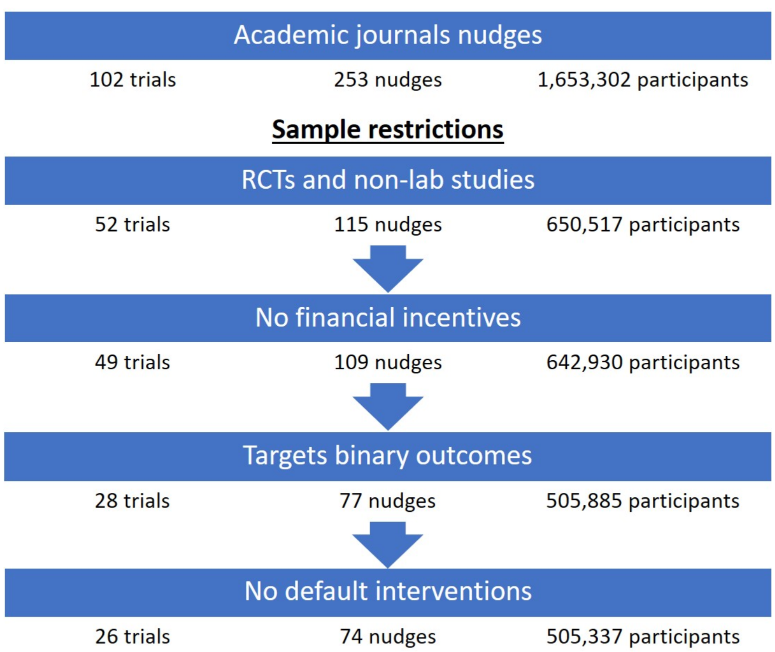
Figures 1a and 1b present an example of a nudge intervention from OES. This trial aims to increase service-member savings plan re-enrollment. The control group received the status-quo email (reproduced in Figure 1a), while the treatment group received a simplified, personalized reminder email with loss framing and clear action steps (reproduced in Figure 1b). The outcome in this trial is measured as savings plan re-enrollment rates.

Figure 2: Selection of nudge studies

(a) Selection among Nudge Units



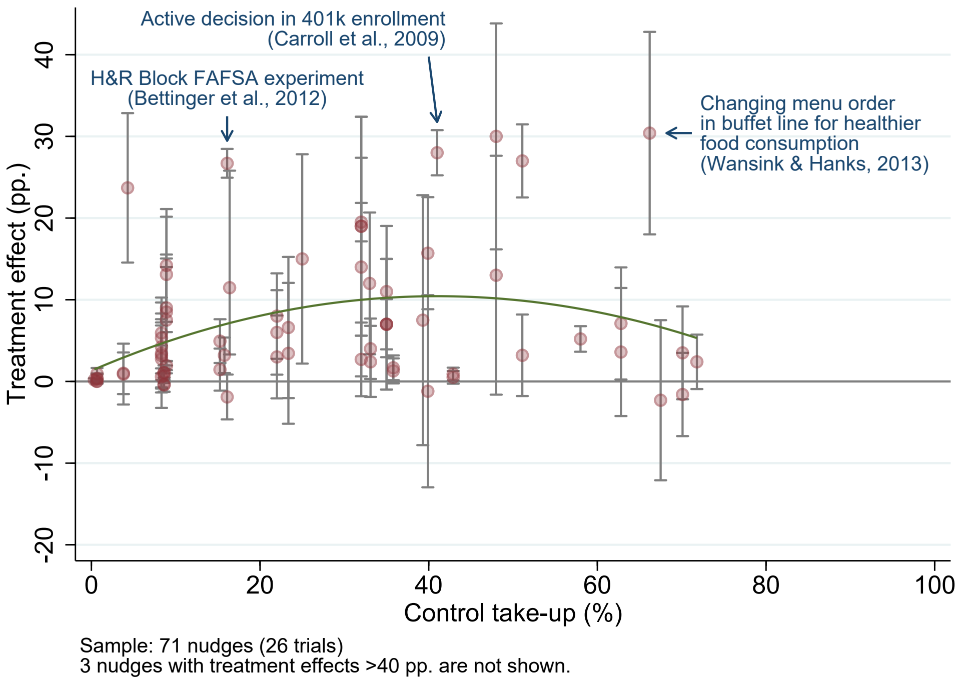
(b) Selection among Academic Journals



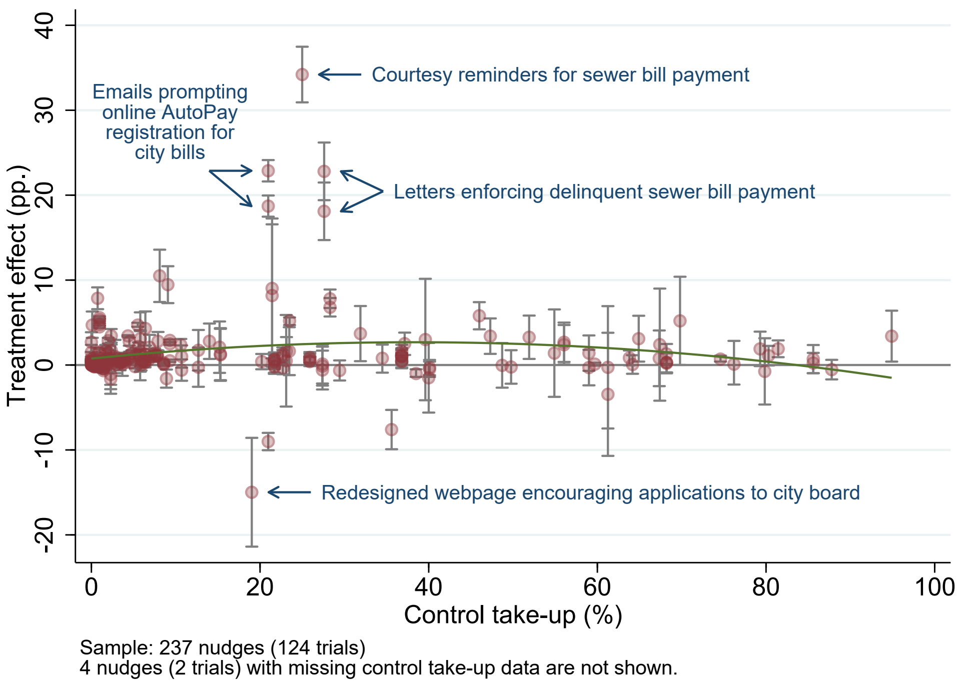
This figure shows the number of trials, treatments, and participants remaining after each sample restriction.

Figure 3: Nudge treatment effects

(a) Academic Journals sample



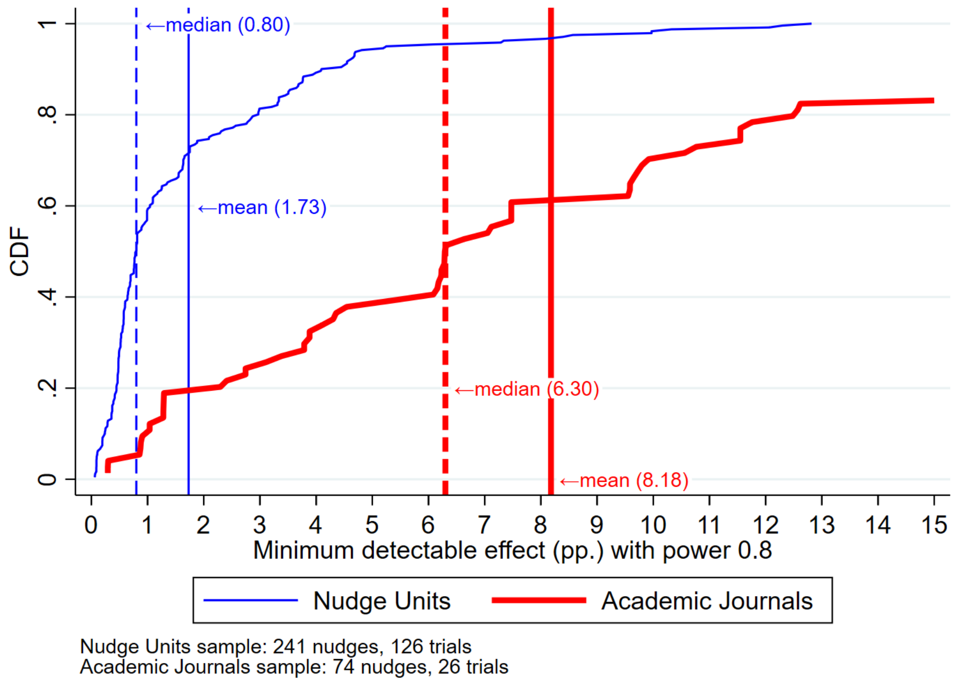
(b) Nudge Units sample



This figure plots the treatment effect relative to control group take-up for each nudge with the quadratic fit. Some of the outliers are labeled for context. Error bars show 95% confidence intervals.

Figure 4: Minimum detectable effects and forecasts

(a) Minimum detectable effect sizes



(b) Forecasts by background

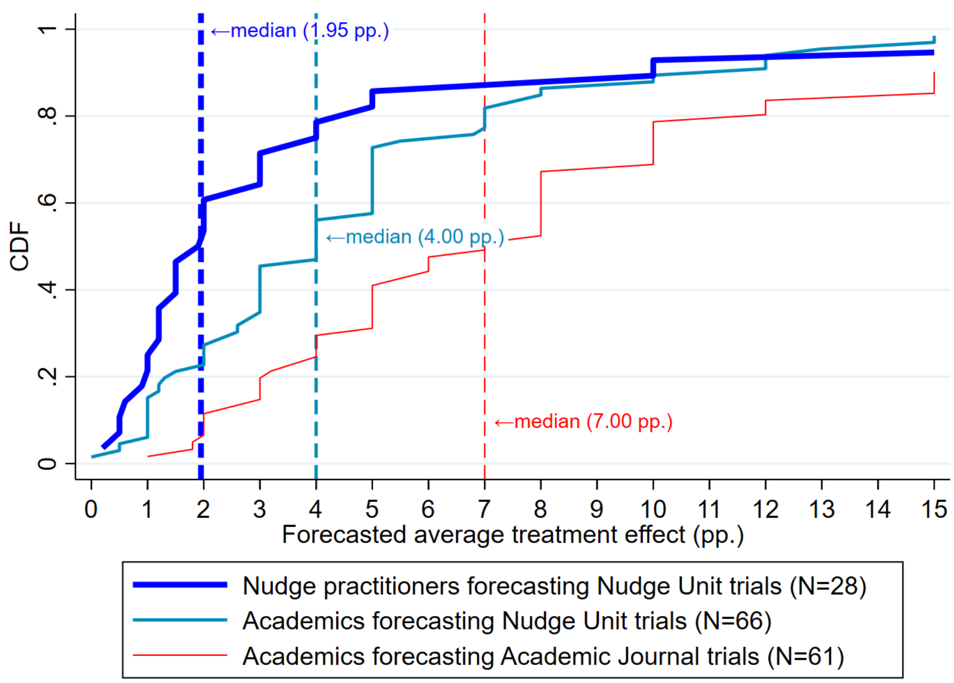
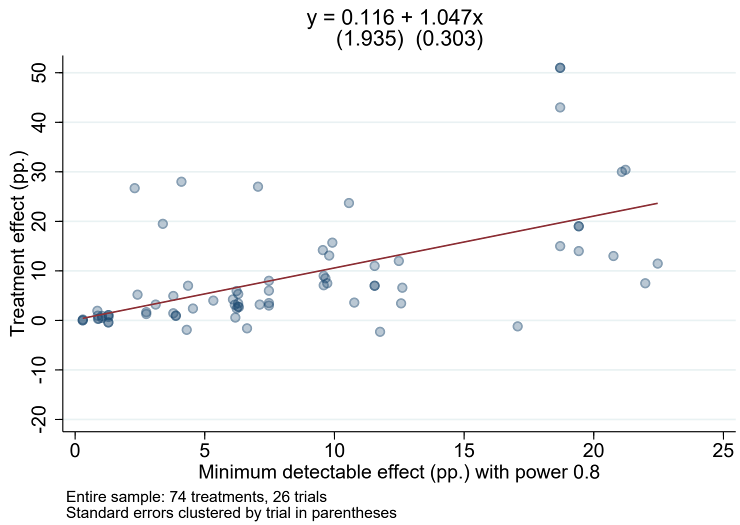


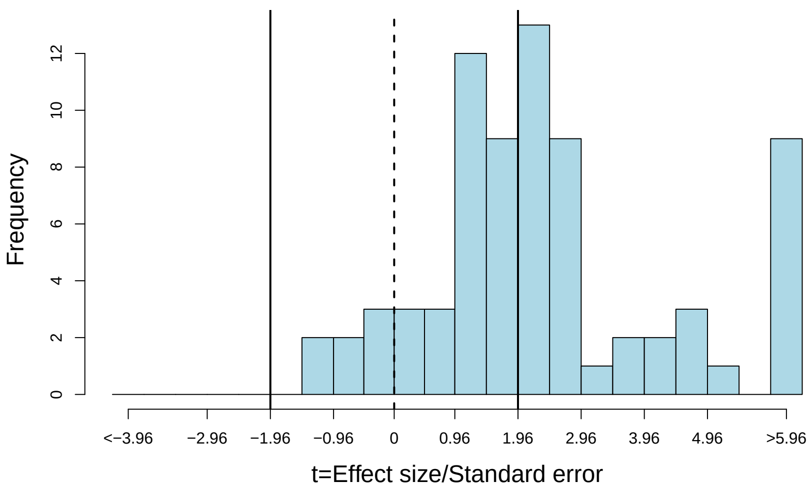
Figure 4a plots the CDF of the minimum detectable effects (MDE), or the size of the treatment effect that each treatment arm is powered to statistically detect 80% of the time given the control group take-up rate and the sample size. For 4 nudges (2 trials) in the Nudge Units sample that are missing control take-up data, the MDE is calculated assuming a conservative control group take-up of 50%. Control take-up is bounded below at 1% when calculating MDE.
Figure 4b shows various distributions of forecasts made by Nudge Unit practitioners and academics (university faculty and post-docs) on the treatment effect of nudges.

Figure 5: Publication bias tests: Academic Journals

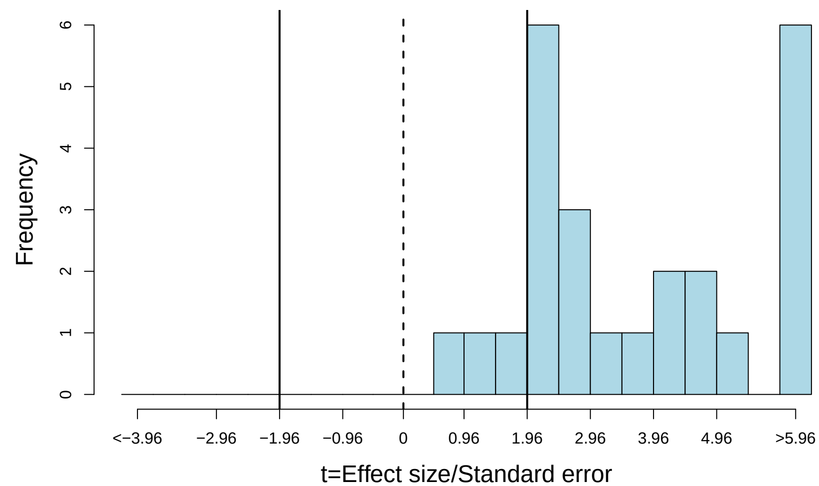
(a) Point estimate and minimum detectable effect



(b) t-stat distribution



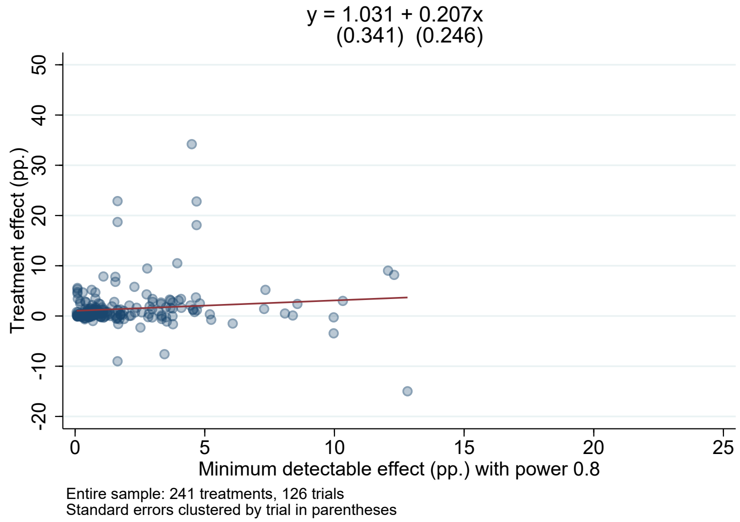
(c) Most significant nudges by trial



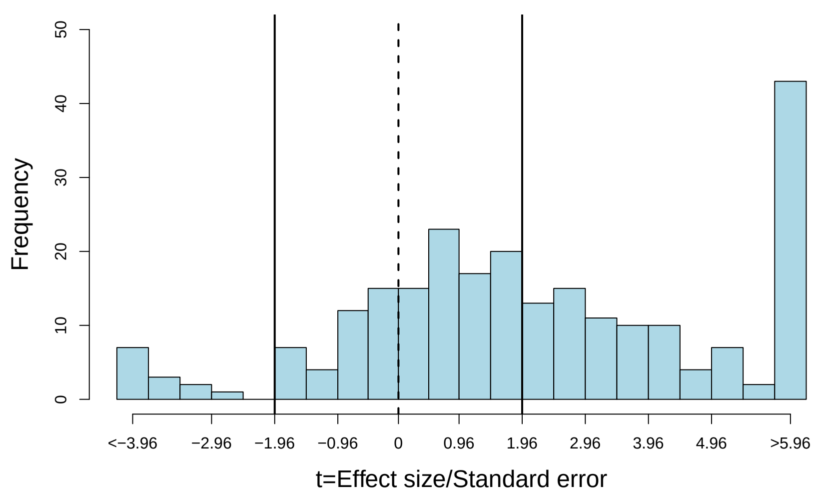
This panel displays tests for publication bias in the Academic Journals sample. Figure 5a plots the relationship between the minimum detectable effect and the treatment effect size. The estimated equation is the linear fit with standard errors clustered at the trial level. Figure 5b shows the distribution of t-statistics (i.e., treatment effect divided by standard error) for all nudges, and Figure 5c shows the distribution for only the max t-stat within each trial. Figure 5c excludes 1 trial in which the most significant treatment arm uses financial incentives.

Figure 6: Publication bias tests: Nudge Units

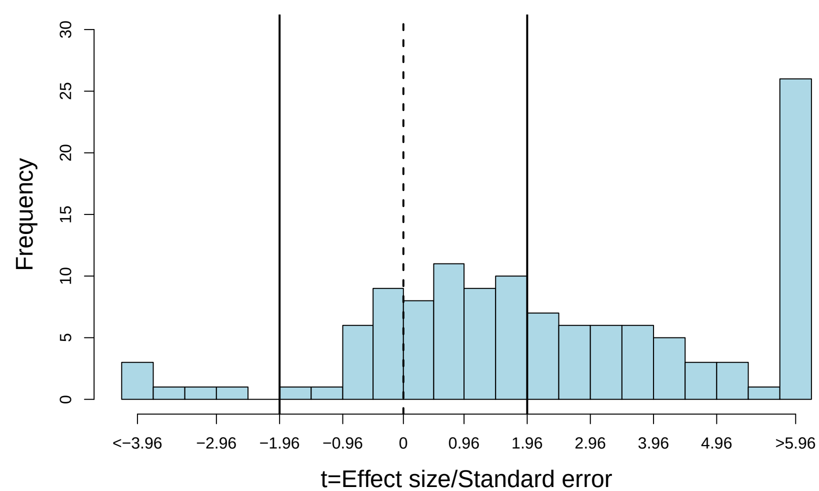
(a) Point estimate and minimum detectable effect



(b) t-stat distribution



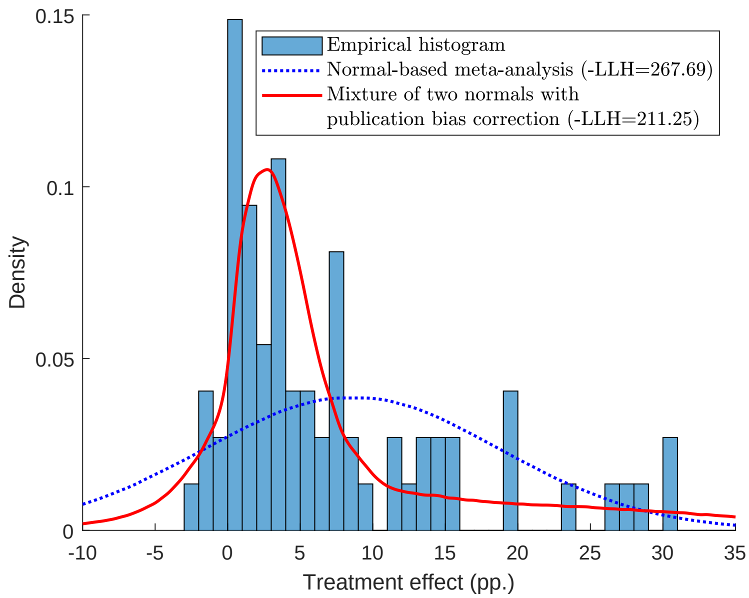
(c) Most significant nudges by trial



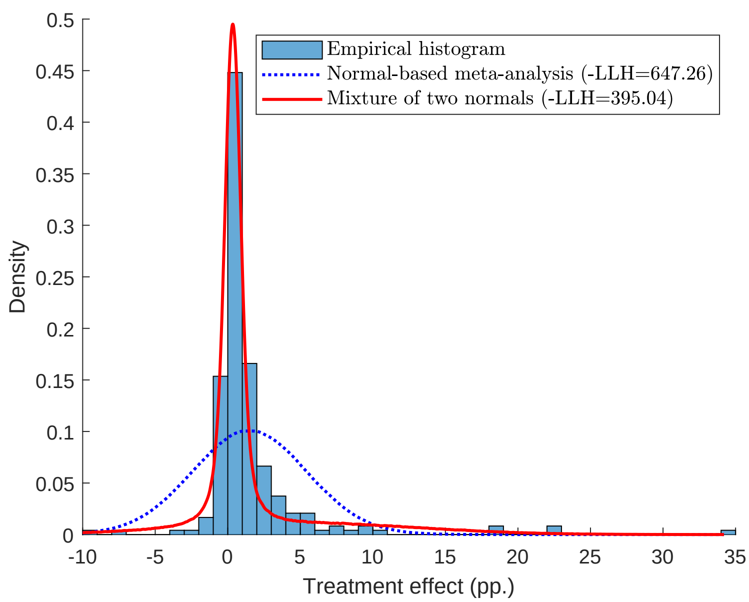
This panel displays tests for publication bias in the Nudge Units sample. Figure 6a plots the relationship between the minimum detectable effect and the treatment effect size. The estimated equation is the linear fit with standard errors clustered at the trial level. Figure 6b shows the distribution of t-statistics (i.e., treatment effect divided by standard error) for all nudges, and Figure 6c shows the distribution for only the max t-stat within each trial. Figure 6c excludes 2 trials in which the most significant treatment arm uses defaults/financial incentives.

Figure 7: Simulated densities from maximum likelihood and mixture of normals models

(a) Academic Journals



(b) Nudge Units



This figure plots the empirical histogram of observed nudge effects and compares the fit of a normal-based meta-analysis model (Panel A of Table V) to the fit of a mixture of two normals model (Panel B of Table V) for the Academic Journals sample in Figure 7a and for the Nudge Units sample in Figure 7b. 1 nudge in the Nudge Units sample with an effect less than -10 pp. and 3 nudges in the Academic Journals sample with effects greater than 35 pp. are not shown.

Table I: Comparison of nudge categories

	Nudge Units			Academic Journals		
	Freq. (%)	Nudges (Trials)	ATE (pp.)	Freq. (%)	Nudges (Trials)	ATE (pp.)
Date						
Early*	46.06	111 (49)	1.88	48.65	36 (14)	7.10
Recent*	53.94	130 (77)	0.97	51.35	38 (12)	10.18
Policy area						
Revenue & debt	29.05	70 (30)	2.43	17.57	13 (4)	3.60
Benefits & programs	22.41	54 (26)	0.89	10.81	8 (3)	14.15
Workforce & education	18.67	45 (24)	0.49	9.46	7 (2)	2.56
Health	12.45	30 (18)	0.73	28.38	21 (9)	8.98
Registration & regulation compliance	8.71	21 (16)	2.18	12.16	9 (2)	3.16
Community engagement	7.88	19 (10)	0.74	4.05	3 (2)	2.80
Environment	0.83	2 (2)	6.83	13.51	10 (3)	22.95
Consumer behavior	0	0 (0)	–	4.05	3 (1)	3.19
Medium of communication						
Email	39.83	96 (47)	1.09	12.16	9 (6)	3.75
Physical letter	29.88	72 (44)	2.41	16.22	12 (4)	1.67
Postcard	21.58	52 (22)	0.82	6.76	5 (1)	10.46
Website	2.90	7 (4)	-0.04	12.16	9 (3)	6.24
In person	0.83	2 (2)	3.05	28.38	21 (5)	14.82
Other	10.37	25 (15)	1.30	24.32	18 (9)	9.38
Control group receives:						
No communication	61.41	148 (66)	1.42	43.24	32 (9)	10.91
Some communication	38.59	93 (62)	1.34	56.76	42 (17)	6.99
Mechanism						
Simplification & information	58.51	141 (73)	1.19	5.41	4 (2)	16.34
Personal motivation	57.26	138 (76)	1.77	32.43	24 (9)	9.59
Reminders & planning prompts	31.54	76 (49)	2.54	35.14	26 (11)	5.02
Social cues	36.51	88 (58)	0.87	21.62	16 (7)	13.81
Framing & formatting	31.95	77 (47)	1.38	32.43	24 (8)	13.53
Choice design	6.22	15 (12)	7.01	20.27	15 (9)	8.85
Total	100	241 (126)	1.39	100	74 (26)	8.68

This table shows the number of nudges and trials in each category, and the average treatment effect within each category. Frequencies for *Medium* and *Mechanism* are not mutually exclusive and frequencies may not sum to 1.

**Early* refers to trials implemented between 2015-2016 for Nudge Units, and to papers published in 2014 or before for Academic Journals. *Recent* refers to trials and papers after these dates.

Table II: Comparison of trial features

	Academic Journals	Nudge Units			
	Mean [std. dev.]	Mean [std. dev.; p-value of difference from column 1]			
		All	BIT	OES	Academic-affiliated OES
	(1)	(2)	(3)	(4)	(5)
Academic faculty involvement	100%	19%	0%	50%	100%
Outcome features					
Control group take-up (%)	26.0	17.3	15.6	19.5	26.4
	[19.9]	[23.2; p=0.10]	[23.9; p=0.05]	[22.2; p=0.29]	[24.0; p=0.94]
Outcome time-frame (days)	68.7	60.2	38.6	101.7	141.5
	[91.7]	[74.5; p=0.59]	[38.0; p=0.11]	[104.9; p=0.25]	[110.9; p=0.04]
Trial design					
Mechanisms per treatment arm	1.5	2.2	2.0	2.5	2.3
	[0.7]	[1.0; p=0.00]	[1.0; p=0.00]	[0.9; p=0.00]	[0.9; p=0.00]
Treatment arms per trial	2.8	1.9	1.7	2.3	1.9
	[2.1]	[1.7; p=0.03]	[1.0; p=0.01]	[2.5; p=0.31]	[1.5; p=0.06]
Minimum detectable effect (pp.)	8.2	1.7	2.2	1.2	1.7
	[6.4]	[2.2; p=0.00]	[2.6; p=0.00]	[1.6; p=0.00]	[2.2; p=0.00]
Institutional constraints rating (1-5)	4.0	3.0	3.0	3.0	2.8
	[0.9]	[0.6; p=0.00]	[0.5; p=0.00]	[0.7; p=0.01]	[1.3; p=0.00]
Planning and implementention					
Total duration (months)	21.3	11.1	8.6	15.0	17.0
	[16.1]	[3.9; p=0.00]	[1.3; p=0.00]	[3.3; p=0.09]	[8.3; p=0.24]
Planning (including IRB)	6.6	4.6	4.0	5.6	5.1
	[6.1]	[2.3; p=0.17]	[1.1; p=0.06]	[3.4; p=0.61]	[2.5; p=0.28]
Intervention and data collection	6.7	4.5	3.4	6.2	6.5
	[7.1]	[2.0; p=0.16]	[1.2; p=0.03]	[1.8; p=0.77]	[2.3; p=0.91]
Data analysis and write-up	7.8	2.0	1.3	3.2	3.9
	[7.0]	[1.2; p=0.00]	[0.5; p=0.00]	[1.1; p=0.00]	[2.9; p=0.01]
Personnel full-time equivalent months	14.9	5.8	4.3	8.3	6.2
	[18.1]	[4.9; p=0.03]	[2.8; p=0.01]	[6.9; p=0.17]	[2.8; p=0.02]
Number of survey responses	25	13*	8*	5*	24
Number of trials	26	126	78	48	24

Data on the institutional constraints rating, duration, and personnel FTE months were collected from a survey of the researchers involved in the trials (see text and Appendix Section A.5 for details). Outcome duration is capped at 360 days, which only affects 1 trial in each of the Academic Journal and Nudge Unit samples. *In columns 2 to 4, the number of survey responses corresponds to the number of Nudge Unit staff members in leadership roles whom we surveyed.

Table III: Unweighted treatment effects

	Academic Journals	Nudge Units			
		All	BIT	OES	Academic-affiliated OES
	(1)	(2)	(3)	(4)	(5)
Average treatment effect (pp.)	8.682	1.390	1.698	1.023	0.978
	(2.467)	(0.304)	(0.528)	(0.206)	(0.408)
Nudges	74	241	131	110	45
Trials	26	126	78	48	24
Observations	505,337	23,556,095	2,008,289	21,547,806	8,923,186
Average control group take-up (%)	25.97	17.33	15.60	19.47	26.45
Distribution of treatment effects					
25th percentile	1.05	0.06	0.00	0.15	0.10
50th percentile	4.12	0.50	0.40	0.60	0.42
75th percentile	12.00	1.40	1.64	1.22	1.20

This table shows the average treatment effect of nudges. Standard errors clustered by trial are shown in parentheses. pp. refers to percentage point.

Table IV: Predicting nudge effect sizes

	Full sample				Academic-affiliated only		
Dep. Var.: Treatment effect (pp.)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	1.390	4.316	1.031	2.878	0.978	4.117	1.970
	(0.304)	(2.152)	(0.342)	(2.008)	(0.405)	(4.884)	(4.405)
Omitted group: Nudge Units							
Academic Journals	7.292	2.381	-0.915	0.030	7.704	6.122	-1.778
	(2.450)	(1.605)	(1.930)	(1.956)	(2.487)	(1.972)	(2.693)
Publication bias controls (Egger's test)							
Minimum detectable effect (MDE)			0.207	0.233			-0.084
			(0.247)	(0.273)			(0.168)
Academic Journals×MDE			0.840	0.342			1.076
			(0.386)	(0.375)			(0.372)
Nudge categories							
Policy area							
Benefits & programs		-0.266		-0.267			
		(1.006)		(0.927)			
Workforce & education		-2.319		-2.474			
		(1.003)		(0.940)			
Health		-0.876		-1.812			
		(1.555)		(1.469)			
Registrations & regulation compliance		-1.027		-1.014			
		(1.358)		(1.349)			
Community engagement		-1.625		-1.457			
		(1.595)		(1.289)			
Environment		9.287		5.491			
		(4.961)		(4.872)			
Consumer behavior		-10.959		-7.402			
		(3.670)		(3.578)			
Medium of communication							
Email		-1.883		-1.537			
		(1.429)		(1.392)			
Physical letter		-0.844		-0.308			
		(1.204)		(1.153)			
Postcard		0.125		-0.019			
		(1.514)		(1.360)			
Website		-2.236		-1.513			
		(3.180)		(2.745)			
In person		7.210		5.373			
		(3.146)		(3.417)			
Other		-0.438		-0.185			
		(1.727)		(1.678)			
Control group receives:							
Some communication		-1.223		-1.225			
		(0.953)		(0.892)			
Mechanism							
Simplification & information		0.878		0.872			
		(1.119)		(1.209)			
Personal motivation		-0.502		-0.330			
		(0.856)		(0.916)			
Reminders & planning prompts		0.349		0.789			
		(0.840)		(0.785)			
Social cues		0.040		0.233			
		(0.959)		(0.920)			
Framing & formatting		1.245		0.998			
		(0.934)		(0.912)			
Choice design		6.226		5.528			
		(2.356)		(2.315)			
Trial features							
Control take-up (%)		0.108		0.046			
		(0.059)		(0.056)			
Control take-up^2		-0.001		-0.001			
		(0.001)		(0.001)			
Log(outcome time-frame days)		-0.692		-0.309			
		(0.409)		(0.367)			
Ideal nudge implemented rating (1-5)						0.979	0.467
						(1.291)	(0.731)
Log(personnel FTE months)						0.671	0.902
						(0.857)	(0.711)
Log(planning & implementation months)						-2.721	-1.419
						(1.562)	(1.548)
Nudges	315	315	315	315	119	119	119
Trials	152	152	152	152	50	50	50
R-squared	0.18	0.46	0.38	0.49	0.14	0.22	0.45

This table shows OLS estimates with standard errors clustered by trial in parentheses. The MDE (minimum detectable effect) is calculated in pp. at power 0.8. Observations with missing data for outcome time-frame, control take-up result, trial duration, institutional constraints rating, or personnel FTE months are included with separate dummies.

Table V: Generalized meta-analysis models

	ATE (pp.)	$\hat{\gamma}$ (pub. bias)	Normal 1 $\hat{\bar{\beta}}_1$	Normal 1 $\hat{\tau}_{BT1}$	Normal 1 $\hat{\tau}_{WI1}$	Normal 2 $\hat{\bar{\beta}}_2$	Normal 2 $\hat{\tau}_{BT2}$	Normal 2 $\hat{\tau}_{WI2}$	$\hat{P}$(Normal 1)	-Log likelihood
Panel A. *Traditional parametric normal-based meta-analysis*										
Academic Journals	8.58	1 (fixed)	8.58	7.89	5.65	–	–	–	1 (fixed)	267.69
	(1.98)		(1.98)	(1.99)	(2.86)					
Nudge Units	1.50	1 (fixed)	1.50	3.04	2.38	–	–	–	1 (fixed)	647.26
	(0.34; p=0.00)		(0.34)	(1.24)	(1.20)					
Panel B. *Generalized mixture model with selective publication*										
Academic Journals	3.89	0.10	1.30	2.70	0.05	19.18	5.86	12.73	0.86	211.25
	(1.88)	(0.13)	(0.97)	(1.00)	(0.17)	(4.81)	(3.19)	(3.06)	(0.07)	
Nudge Units	1.38	1 (fixed)	0.35	0.41	0.23	5.09	4.64	6.40	0.78	395.04
	(0.33; p=0.19)		(0.10)	(0.12)	(0.09)	(1.72)	(3.53)	(3.41)	(0.06)	
Difference in observed ATE explained by publication bias: 66% (26%)										
Panel C. *Generalized mixture model with selective publication and heterogeneity based on observables*										
Parsimonious model of observables (see Column 3 of Table A.IXc):										
Difference in observed ATE explained by: publication bias 77% (19%), observable characteristics 21% (14%)										
Richer model of observables (see Column 4 of Table A.IXc):										
Difference in observed ATE explained by: publication bias 77% (19%), observable characteristics 20% (11%)										

This table shows the estimates from a traditional normal-based meta-analysis method in Panel A, and from generalized models with a mixture of normals in Panels B and C. Under the traditional normal-based meta-analysis assumptions, trial base effects β_i are drawn from a normal distribution centered at $\bar{\beta}$ with between-trial standard deviation τ_{BT}. Then, each treatment arm j within a trial i draws a base treatment effect $\beta_{ij} \sim N(\beta_i, \tau_{WI}^2)$, where τ_{WI} is the within-trial standard deviation. Each treatment arm also has some level of precision given by an independent standard error σ_{ij}. The observed treatment effect is $\hat{\beta}_{ij} \sim N(\beta_{ij}, \sigma_{ij}^2)$.

In Panel B, the mixture-of-two-normals model is a generalization of the normal-based meta-analysis, and allows trial base effects to be drawn from a second normal distribution. The model in Panel C adds a third normal, and also allows the probability of drawing effects from each normal to vary depending on observable trial characteristics (see Table A.IXc for details).

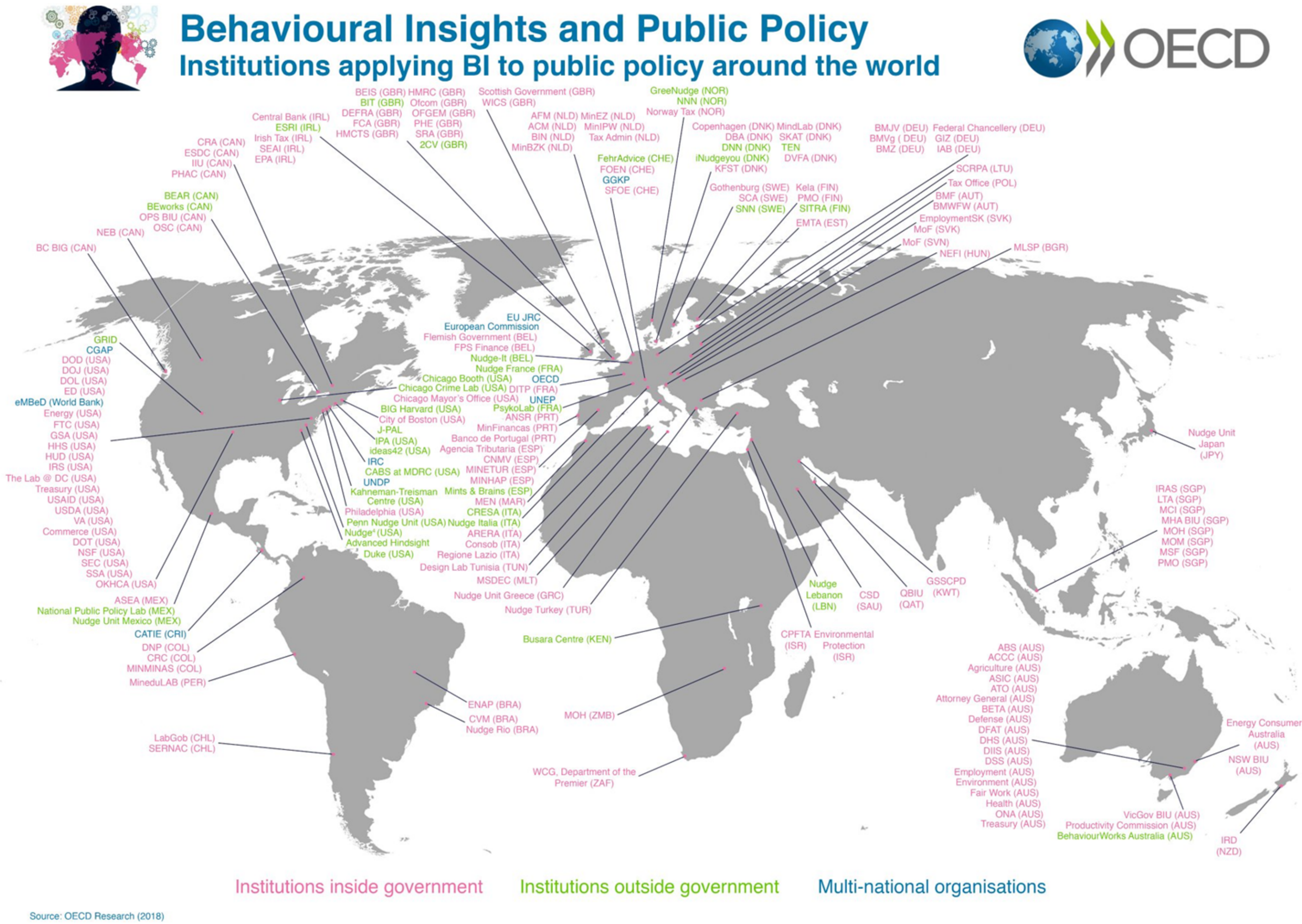
To capture the extent of selective publication, the probability of publication is allowed to differ depending on whether trial have at least one significant treatment arm. In particular, trials without any significant results at the 95% level are γ times as likely to be published as trials with significant results. Estimates are obtained using maximum likelihood. Bootstrap standard errors shown in parentheses. The p-value of the difference in the estimated average treatment effect (ATE) between the Academic Journals and Nudge Units samples is shown in the parentheses below the Nudge Unit ATE.

Table VI: Model counterfactuals

	Effect size distribution	Statistical power	Selective publication	Simulated ATE (pp.)
(1) Acad. J. as observed	Acad. J.	Acad. J.	Yes (as in Acad. J.)	7.33 (1.16)
Counterfactuals – Academic Journal effect sizes with:				
(2) High power	Acad. J.	Nudge Units	Yes (as in Acad. J.)	6.26 (1.11)
(3) No pub. bias	Acad. J.	Acad. J.	No (as in Nudge Units)	3.81 (0.77)
(4) High power & no pub. bias	Acad. J.	Nudge Units	No (as in Nudge Units)	3.78 (0.87)
Counterfactuals – Nudge Unit effect sizes with:				
(5) Low power & pub. bias	Nudge Units	Acad. J.	Yes (as in Acad. J.)	3.35 (0.69)
(6) Pub. bias	Nudge Units	Nudge Units	Yes (as in Acad. J.)	2.43 (0.57)
(7) Low power	Nudge Units	Acad. J.	No (as in Nudge Units)	1.39 (0.38)
(8) Nudge Units as observed	Nudge Units	Nudge Units	No (as in Nudge Units)	1.40 (0.38)

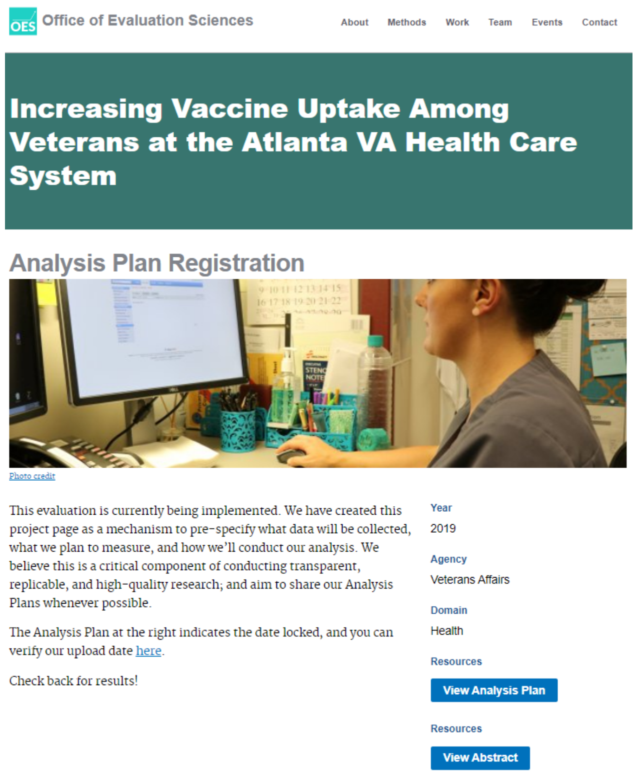
This table shows estimates for counterfactual simulated average treatment effects using the generalized model in Panel B of Table V. Each counterfactual exercise draws 1,000 samples of 152 simulated trials from the estimated mixture distribution for the sample of nudges indicated under "Effect size distribution". The number of experimental arms and their standard errors for these simulated trials are drawn with replacement from the sample listed under "Statistical power". Under selective publication, simulated trials without any positively significant treatment arms at the 95% level are "published" with probability $\hat{\gamma} = 0.1$ (as estimated in Panel B of Table V). Simulated trials with at least one positively significant treatment arm are published with probability 1. When selective publication is suppressed, all simulated trials are published. The "Simulated ATE (pp.)" column reports the average treatment effect in percentage points for all "published" treatment arms from the $1,000 \times 152 = 152,000$ simulated trials. The standard deviation of the observed ATE in the 1,000 simulated samples is reported in parentheses.

Figure A.1: Nudge Units around the world



This figure shows the various Nudge Units across the world.

Figure A.2a: Additional examples of nudges: OES website



OES Office of Evaluation Sciences

About Methods Work Team Events Contact

Increasing Vaccine Uptake Among Veterans at the Atlanta VA Health Care System

Analysis Plan Registration

Photo credit

This evaluation is currently being implemented. We have created this project page as a mechanism to pre-specify what data will be collected, what we plan to measure, and how we'll conduct our analysis. We believe this is a critical component of conducting transparent, replicable, and high-quality research; and aim to share our Analysis Plans whenever possible.

The Analysis Plan at the right indicates the date locked, and you can verify our upload date here.

Check back for results!

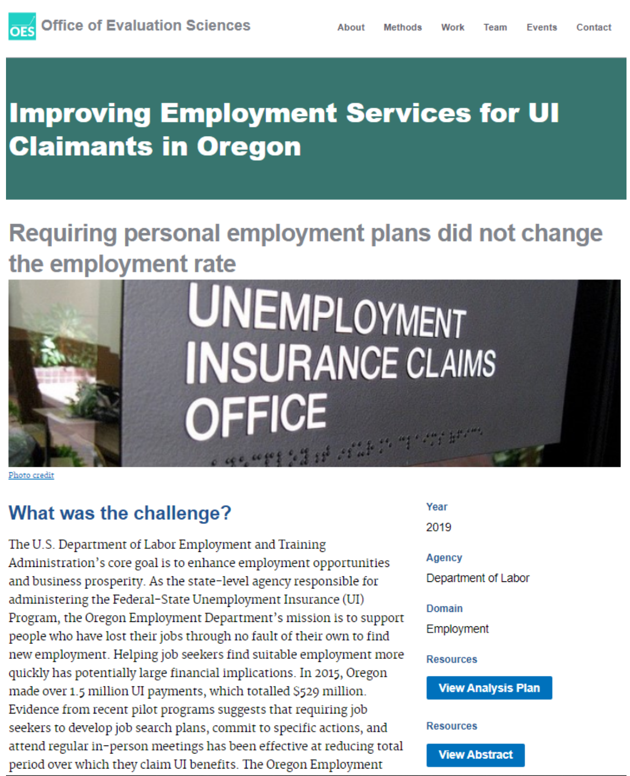
Year
2019

Agency
Veterans Affairs

Domain
Health

Resources
View Analysis Plan

Resources
View Abstract



OES Office of Evaluation Sciences

About Methods Work Team Events Contact

Improving Employment Services for UI Claimants in Oregon

Requiring personal employment plans did not change the employment rate

UNEMPLOYMENT INSURANCE CLAIMS OFFICE

Photo credit

What was the challenge?

The U.S. Department of Labor Employment and Training Administration's core goal is to enhance employment opportunities and business prosperity. As the state-level agency responsible for administering the Federal-State Unemployment Insurance (UI) Program, the Oregon Employment Department's mission is to support people who have lost their jobs through no fault of their own to find new employment. Helping job seekers find suitable employment more quickly has potentially large financial implications. In 2015, Oregon made over 1.5 million UI payments, which totalled $529 million. Evidence from recent pilot programs suggests that requiring job seekers to develop job search plans, commit to specific actions, and attend regular in-person meetings has been effective at reducing total period over which they claim UI benefits. The Oregon Employment

Year
2019

Agency
Department of Labor

Domain
Employment

Resources
View Analysis Plan

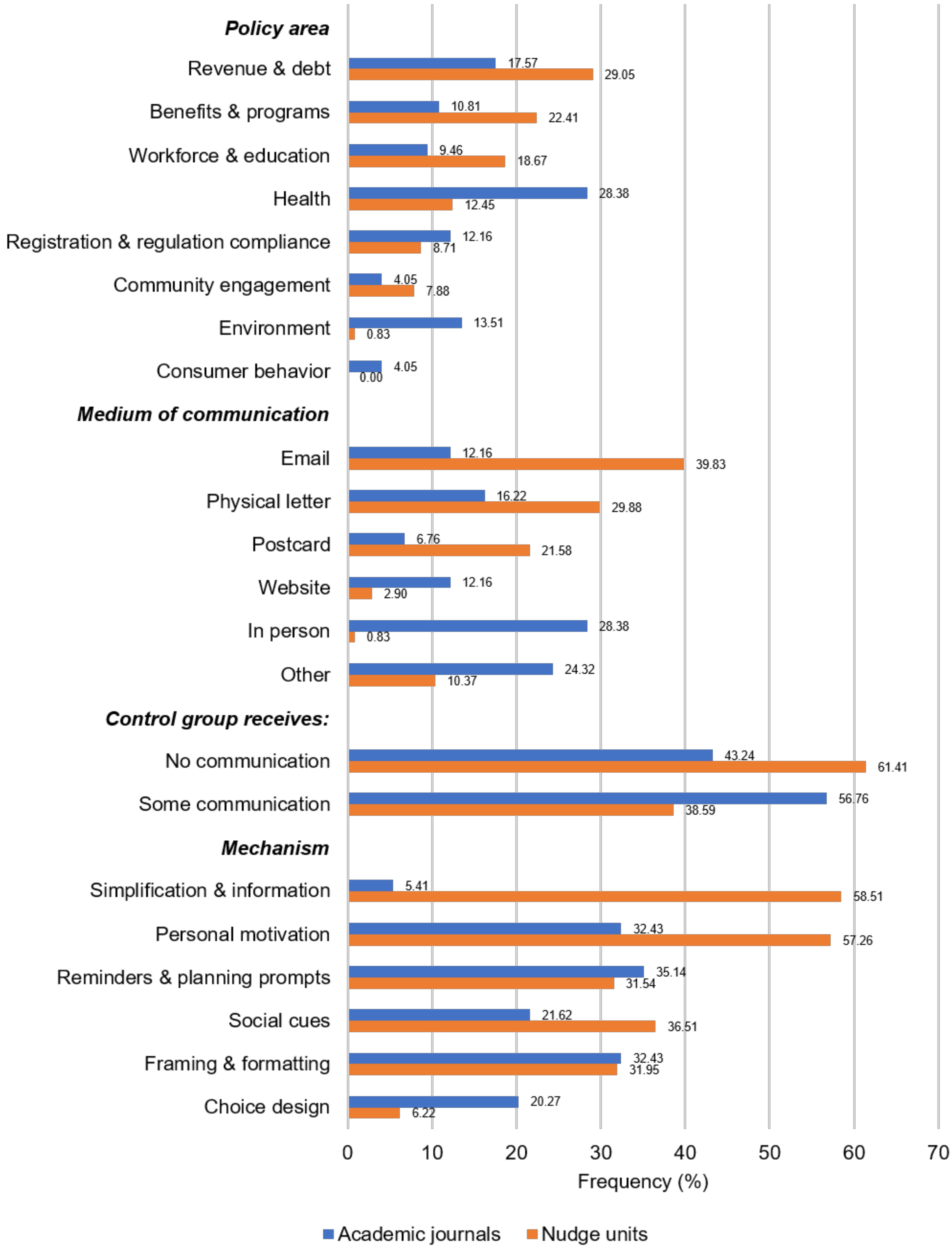
Resources
View Abstract

This figure shows screen captures directly from the Office of Evaluation Sciences website. The top page documents the analysis plan registration for an ongoing trial, whereas the bottom page presents the trial report from a concluded trial.

Figure A.2b: Additional examples of nudges: BIT-NA example

This figure presents an example of a nudge intervention run by BIT-NA. This trial encourages utilities customers to enroll in AutoPay and e-bill using bill inserts. The control group received the status quo utility bill that advertises e-bill and AutoPay on the back, while the treatment group received an additional insert with simplified graphics. The outcome in this trial is measured as AutoPay/e-bill enrollment rates.

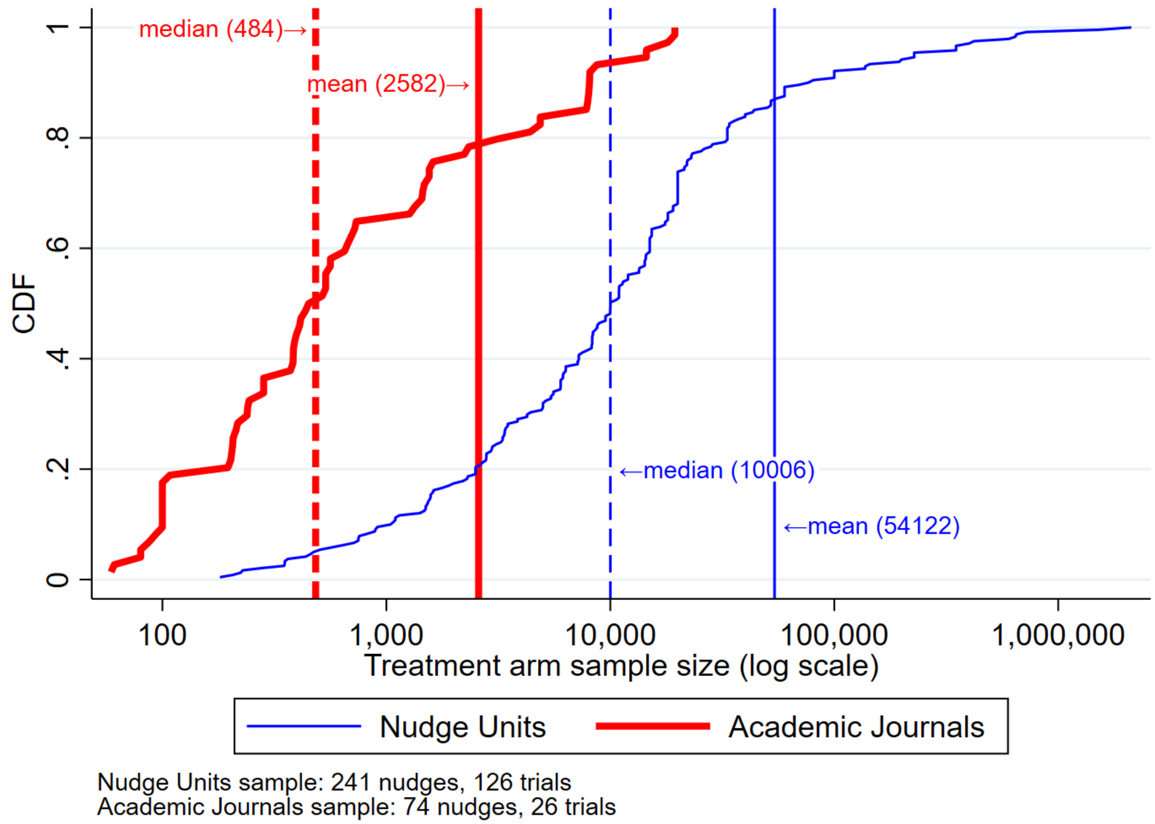
Figure A.3: Comparison of nudge categories



This figure shows the frequencies of nudges in category of characteristics. Categories for Medium and Mechanism are not mutually exclusive and frequencies may not sum to 1.

Figure A.4: Comparison of trial features between Nudge Units and Academic Journals

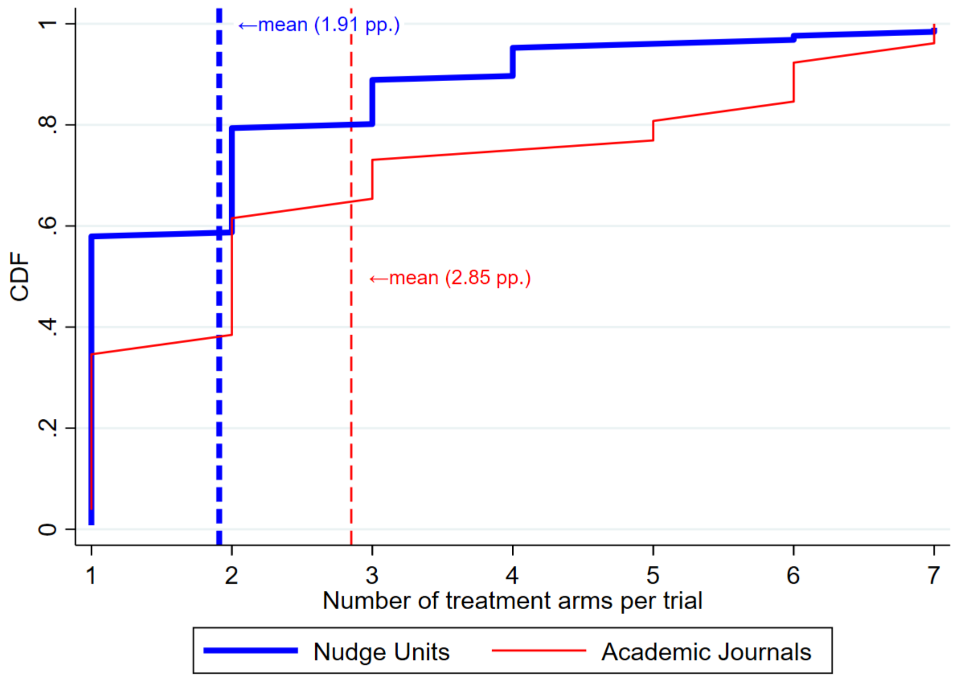
(a) Treatment arm sample size



This figure compares the distribution of nudge-by-nudge treatment arm sample sizes (i.e. excluding the control group sample size) between the Nudge Units and the Academic Journals samples.

Figure A.4: Comparison of trial features between Nudge Units and Academic Journals

(b) Treatment arms per trial



(c) Mechanisms per treatment arm

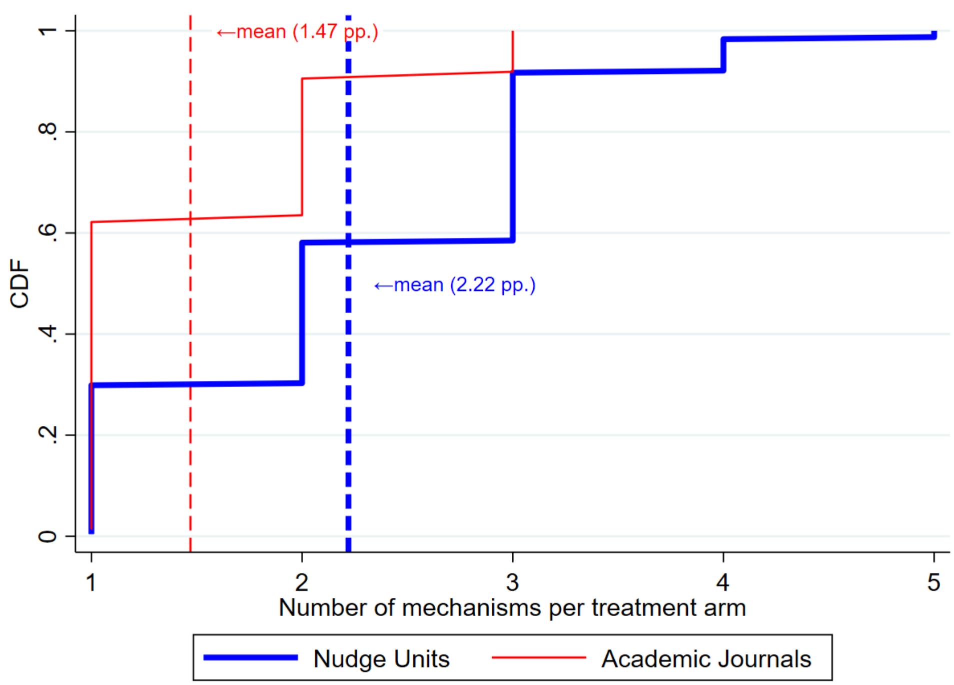
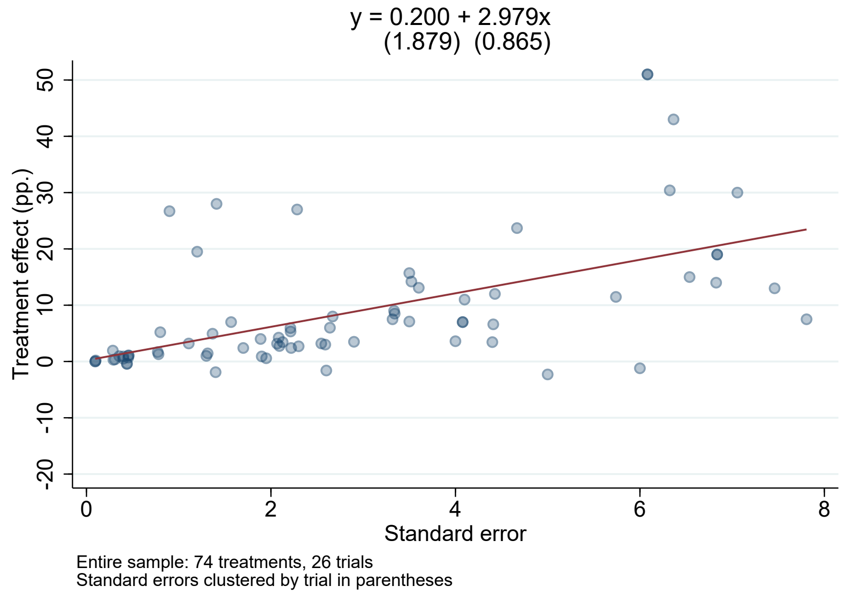
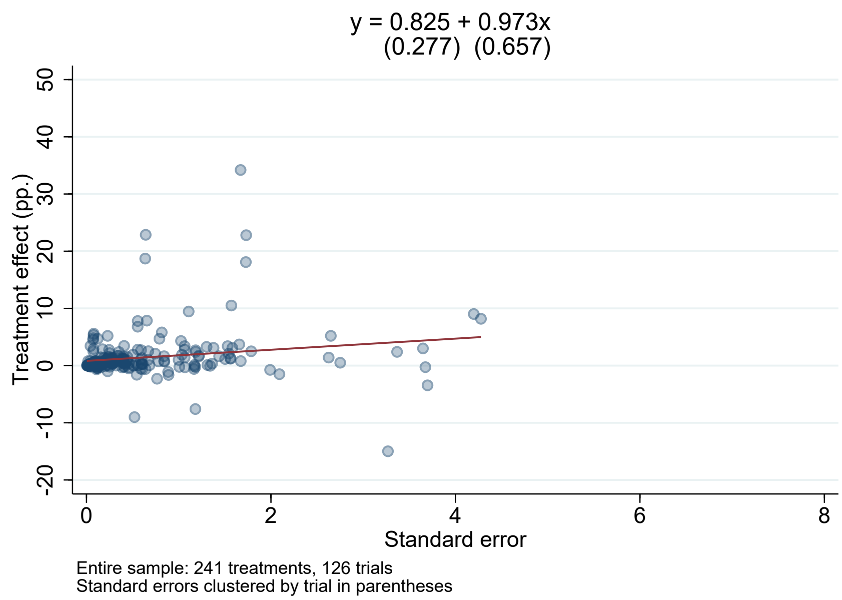


Figure A.5: Publication bias tests: Point estimate and standard error

(a) Academic Journals



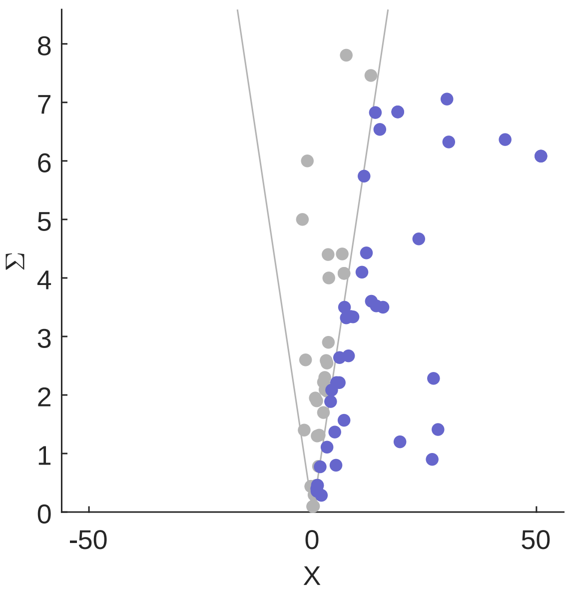
(b) Nudge Units



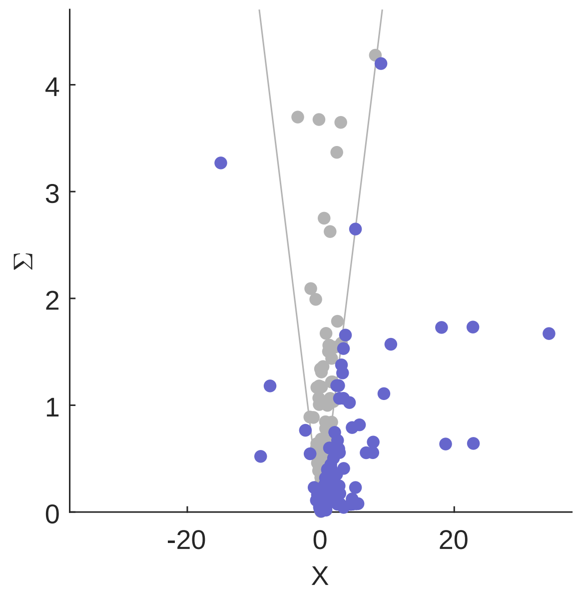
This figure plots the relationship between the standard error and the treatment effect for the Academic Journals sample (A.5a) and the Nudge Units sample (A.5b). The estimated equation is the linear fit with standard errors clustered at the trial level.

Figure A.5: Publication bias tests: Andrews-Kasy funnel plot

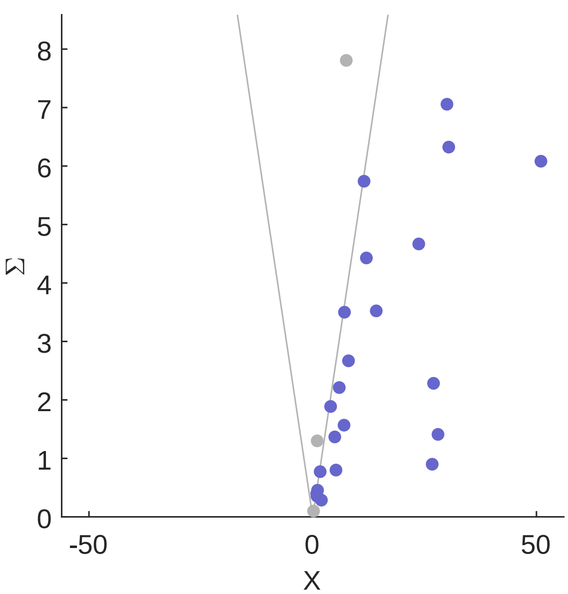
(c) Academic Journals: All nudges



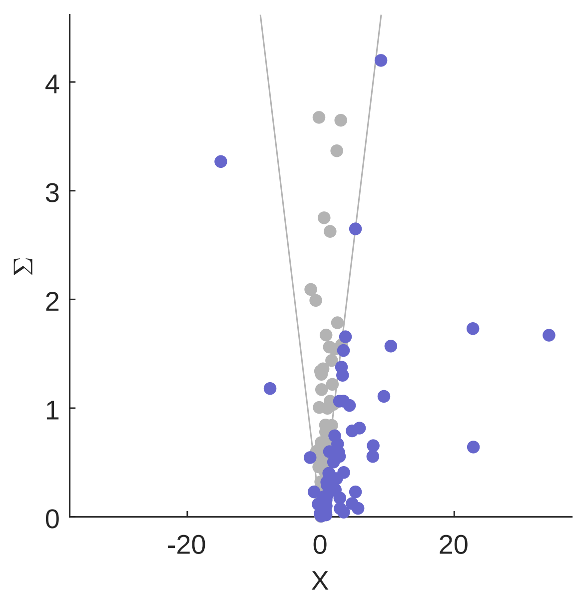
(e) Nudge Units: All nudges



(d) Academic Journals: Most significant nudges by trial



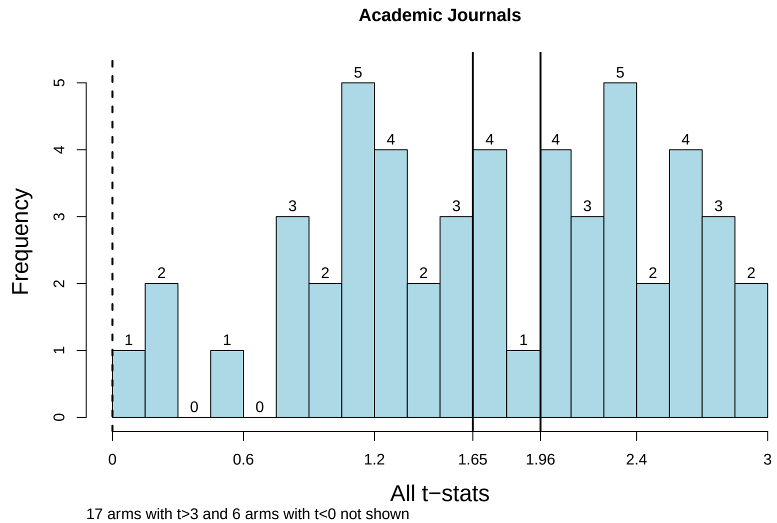
(f) Nudge Units: Most significant nudges by trial



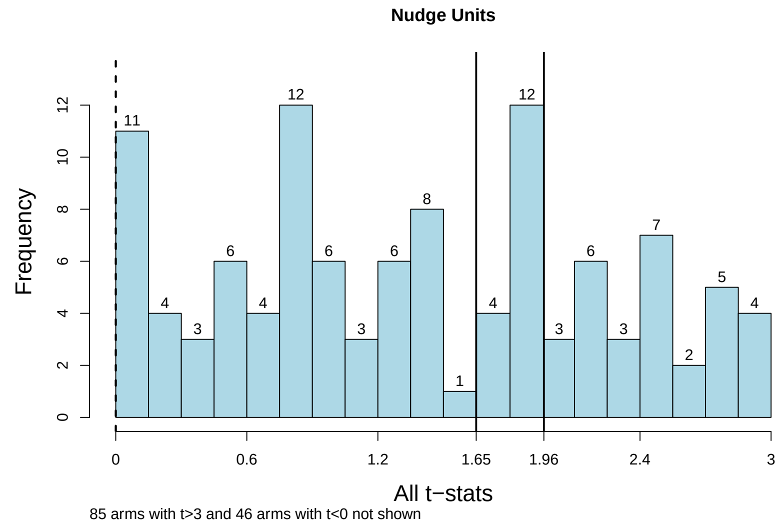
This figure presents funnel plots of the treatment effect (horizontal axis) against the standard error (vertical axis). Nudges within the two gray lines are insignificant at the 5% level (i.e., $|t| < 1.96$). Figures A.5c and A.5e show all the nudges in the samples, while A.5d and A.5f show only the nudges with the highest t-stat within each trial. 1 trial in the Academic Journals sample and 2 trials from the Nudge Units sample in which the most significant treatment uses defaults/financial incentives are excluded from A.5d and A.5f respectively.

Figure A.5: Publication bias tests: t-stat distribution (bin-width≈ 0.15)

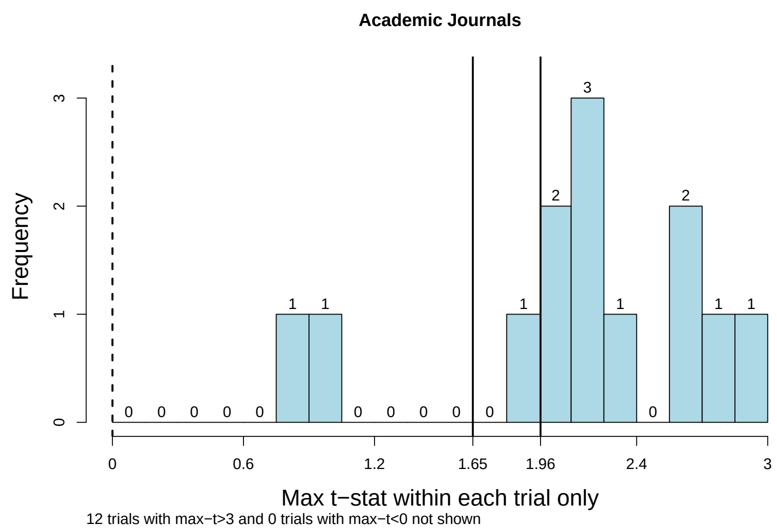
(g) Academic Journals: All nudges



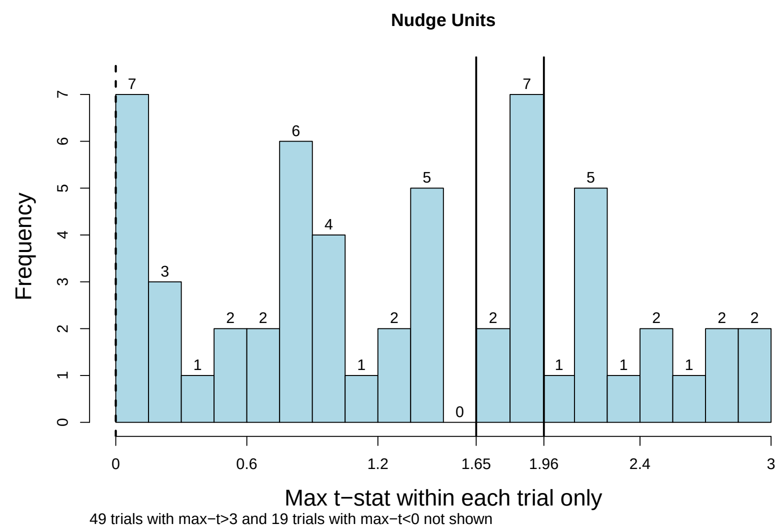
(i) Nudge Units: All nudges



(h) Academic Journals: Most significant nudges by trial



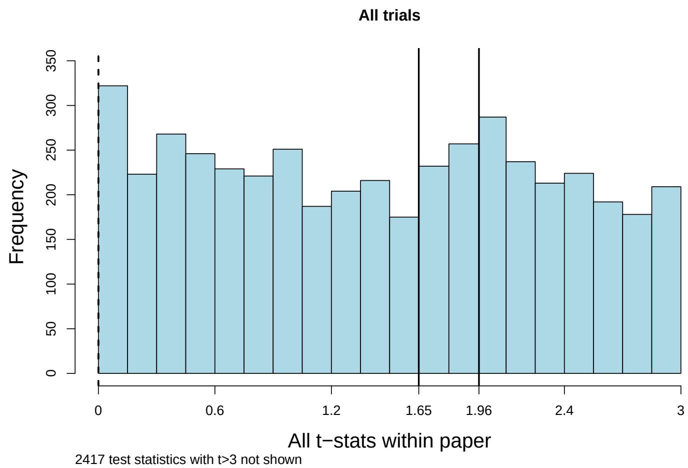
(j) Nudge Units: Most significant nudges by trial



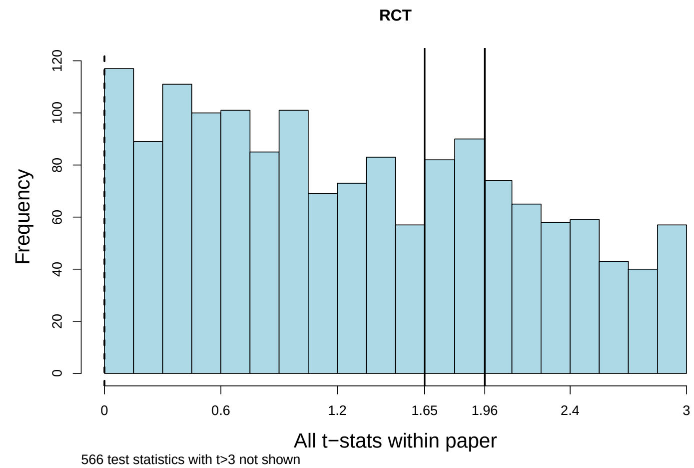
1 trial in the Academic Journals sample and 2 trials from the Nudge Units sample in which the most significant treatment uses defaults/financial incentives are excluded from A.5h and A.5j respectively.

Figure A.6: Distribution of t-stats from Brodeur, Cook, and Heyes (2020)

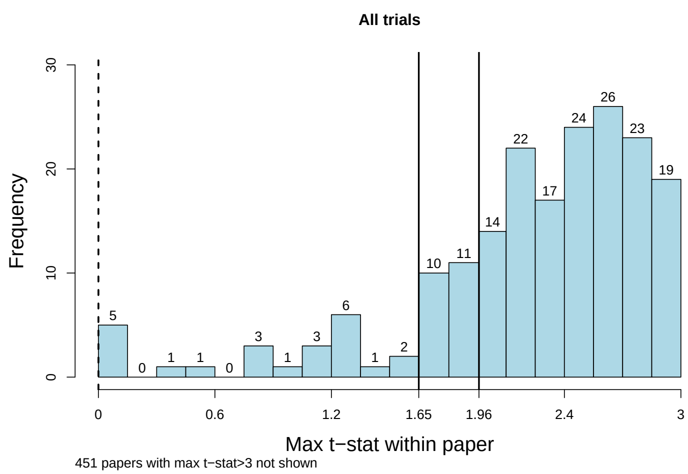
(a) All t-stats



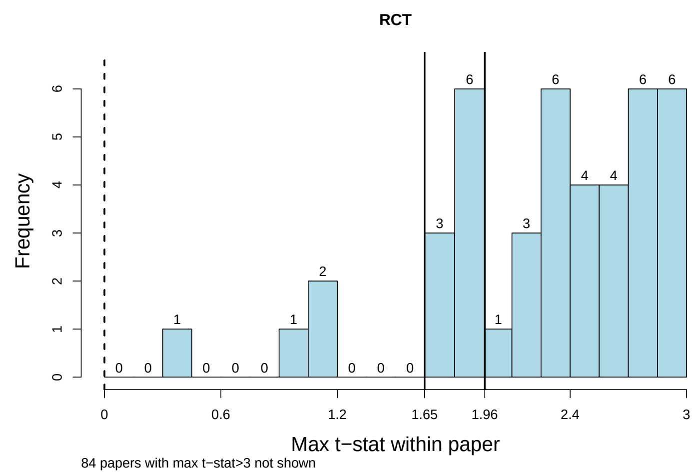
(c) RCT studies: All t-stats



(b) Max t-stat within each paper only



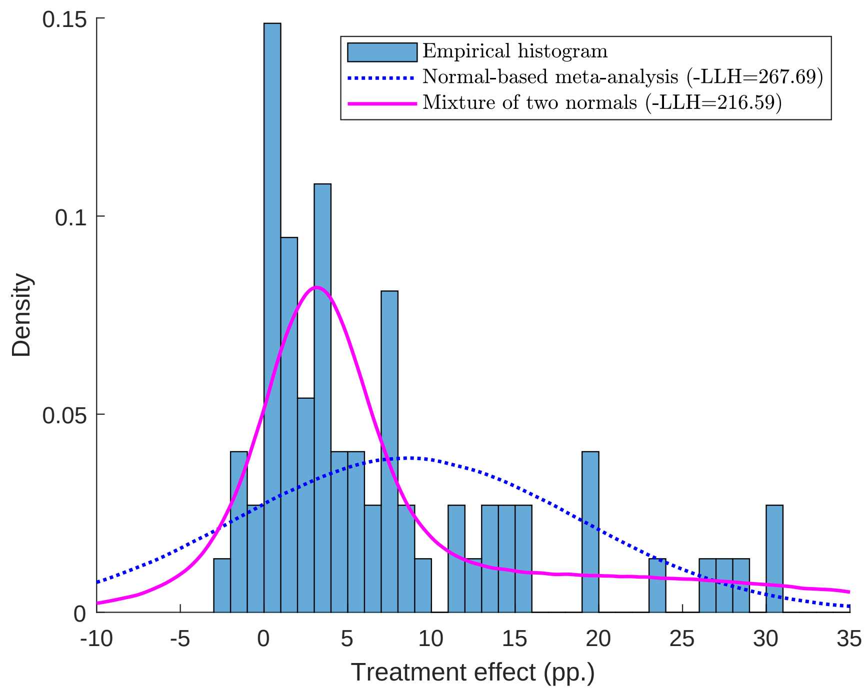
(d) RCT studies: Max t-stat within each paper only



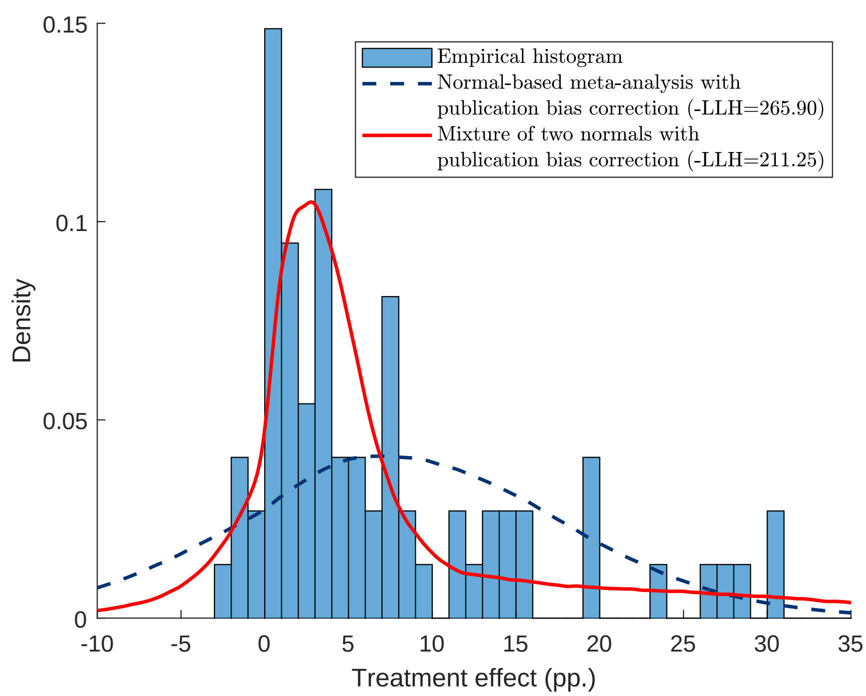
We thank Abel Brodeur for promptly sharing the data for this analysis. Brodeur et al. (2020) gather this data from the universe of papers published in the top 25 economics journals in 2015 and 2018. They categorize papers by empirical method (DID, IV, RCT, and RDD) and record the point estimate and standard error from the results in the main table of each article. Figure A.6a shows the distribution of all the t-stats from the main table of each paper for the entire sample of articles, while Figure A.6b shows the distribution of only the maximum t-stat within each paper. Figures A.6c and A.6d show the analog for the subsample of RCT papers.

Figure A.7: Academic Journals: Comparison of meta-analysis models

(a) Normal-based meta-analysis vs. mixture of two normals



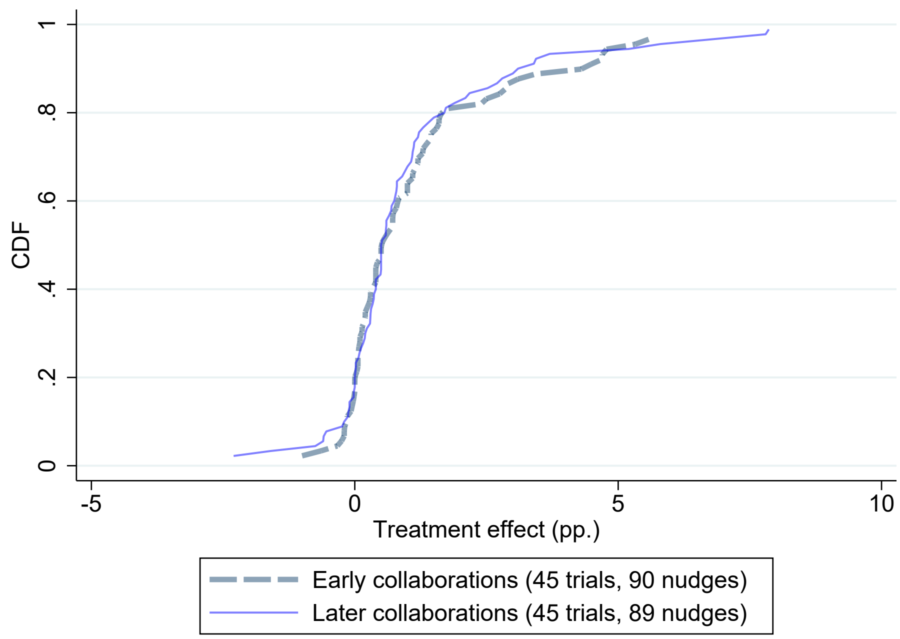
(b) With and without publication bias correction



This figure plots both empirical and simulated distributions of nudge effects and compares various meta-analysis specifications from Tables V and A.IXa. Figure A.7a compares the fit of a normal-based meta-analysis model and that of a mixture-of-two-normals model. A correction for publication bias is added to these two models in Figure A.7b. 3 nudges with effects greater than 35 pp. are not shown. The densities are kernel approximations from 1,000,000 simulated trials.

Figure A.8: Within-collaboration Nudge Unit effects

(a) Nudge Unit treatment effects in early vs. later collaborations with the same agency/city



(b) Success of first collaboration and number of total collaborations with the same agency/city

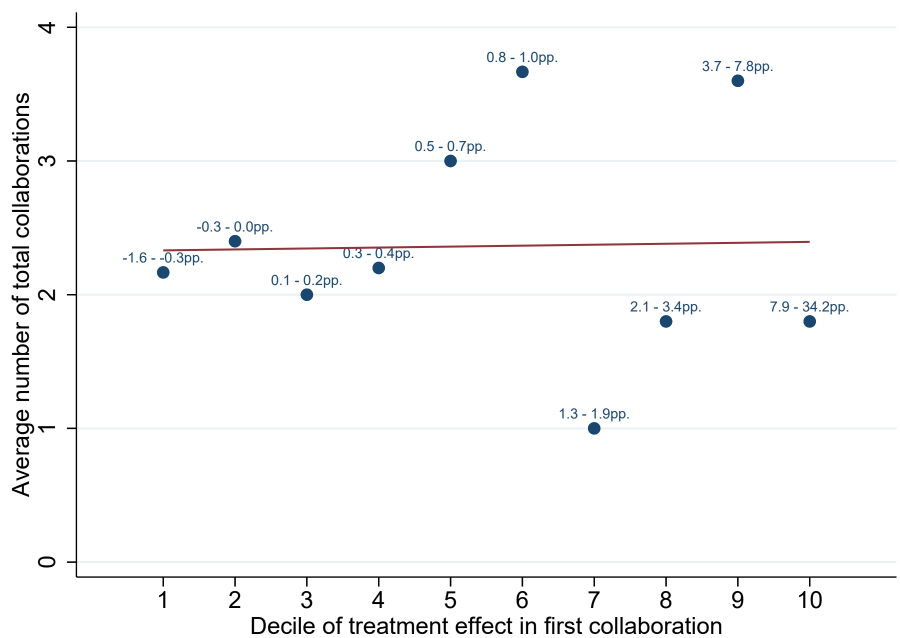
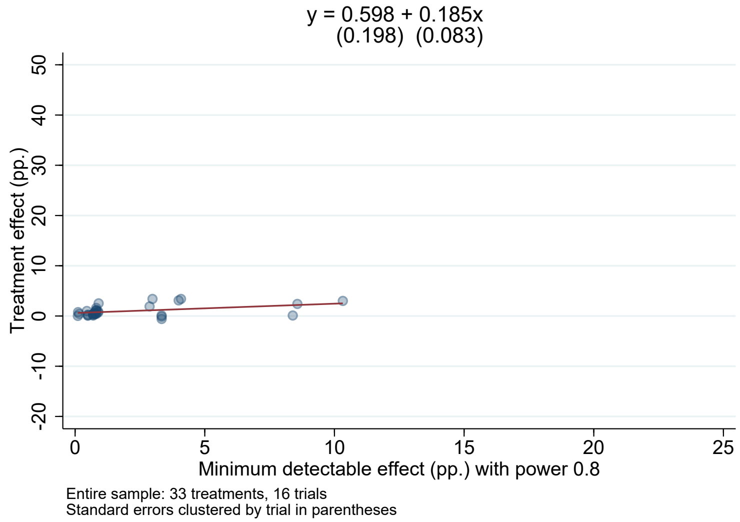


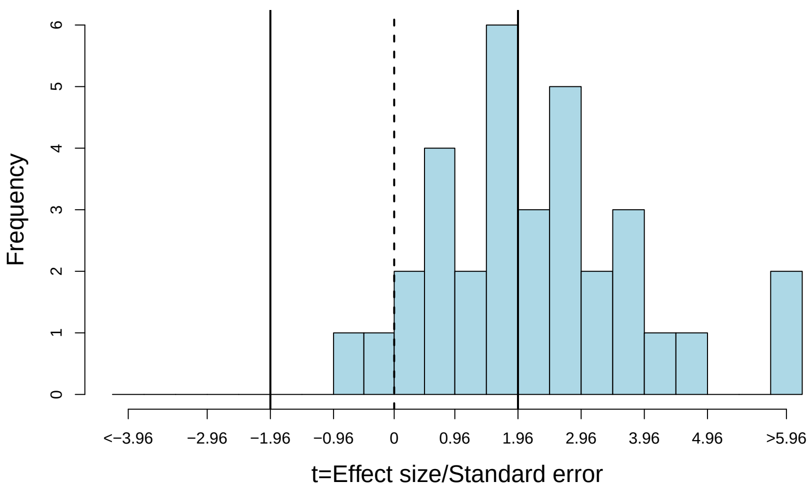
Figure A.8a compares the CDF of the treatment effects in percentage points between the first half of trials ("early") in a series of collaborations with the same government agency or city and the latter half of trials in the same series of collaborations ("latter"). Trials that were one-time collaborations with an agency or city are not included. When there is an odd number of trials in a collaboration, the median trial is not included. Figure A.8b categorizes the first trials in each series of collaborations with a partnering government agency or city (which may be one-time) into deciles based on the treatment effect of their most effective arm. This figure shows the average total number of collaborations for each decile. The labels for each point reports the range of treatment effect sizes in each decile.

Figure A.9: Publication bias tests: Published Nudge Unit trials

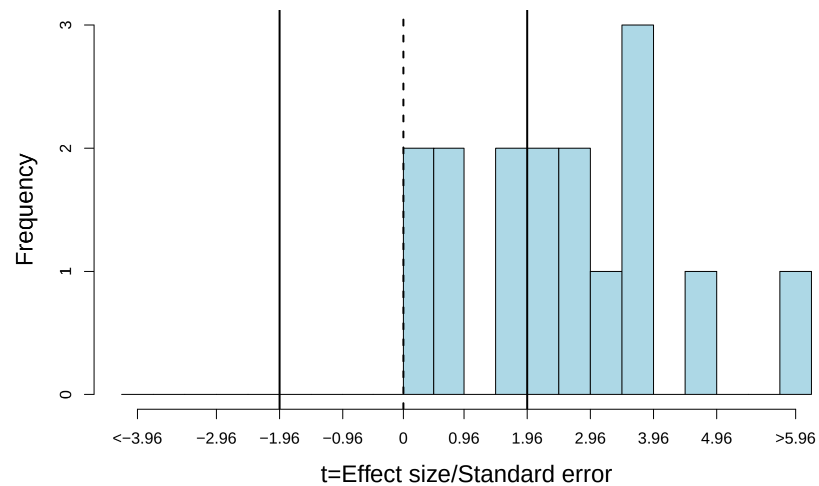
(a) Point estimate and minimum detectable effect



(b) t-stat distribution



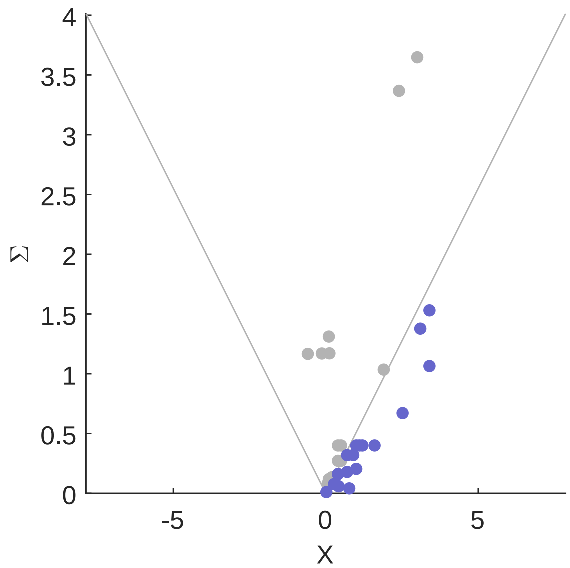
(c) Most significant nudges by trial



This panel displays tests for publication bias in the Published Nudge Units sample. Figure A.9a plots the relationship between the minimum detectable effect and the treatment effect size. The estimated equation is the linear fit with standard errors clustered at the trial level. Figure A.9b shows the distribution of t-statistics (i.e., treatment effect divided by standard error) for all nudges, and Figure A.9c shows the distribution for only the max t-stat within each trial.

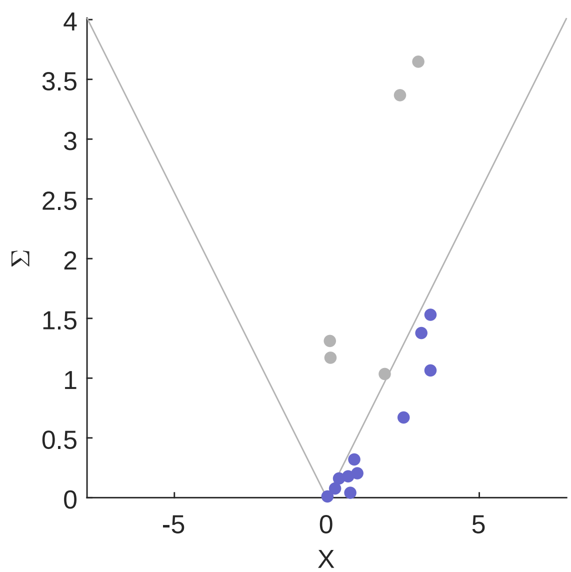
Figure A.9: Publication bias tests: Published Nudge Unit trials

(d) Andrews-Kasy funnel plot



(e) Andrews-Kasy funnel plot: Most significant treatments

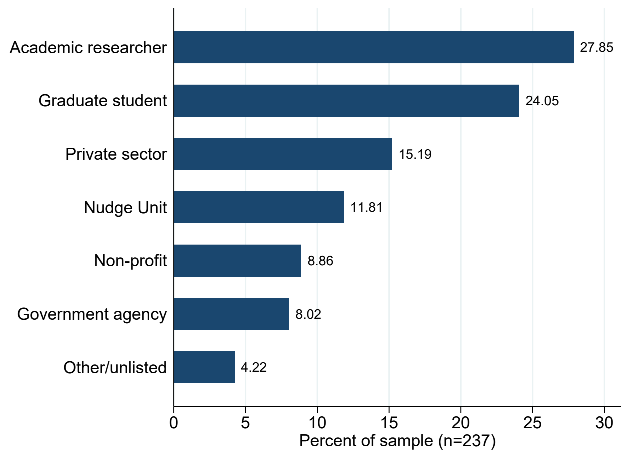
.



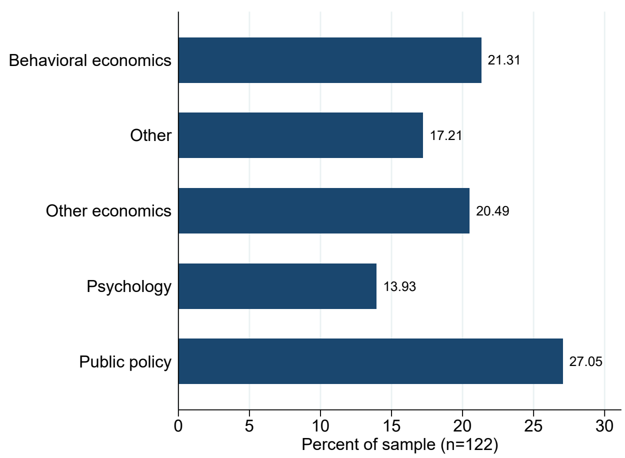
This figure plots the treatment effects (horizontal axis) against the standard errors (vertical axis). Nudges within the two gray lines are insignificant at the 5% level (i.e., $t < 1.96$). Figure A.9d shows all the nudges in the Published Nudge Units sample, while A.9e shows only the nudges with the highest t-stat within their trial.

Figure A.10: Characteristics of forecasters

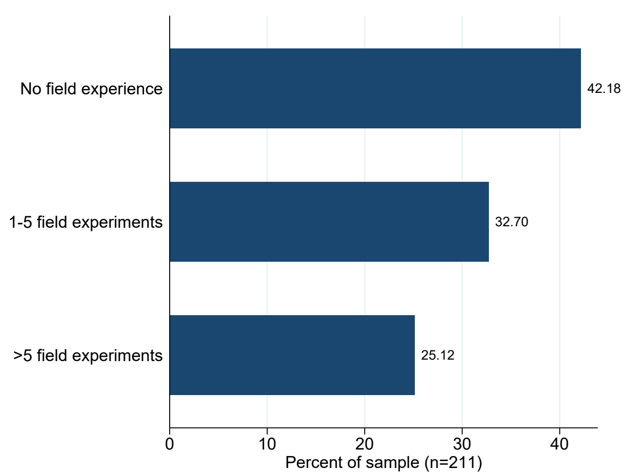
(a) By affiliation



(b) By academic background



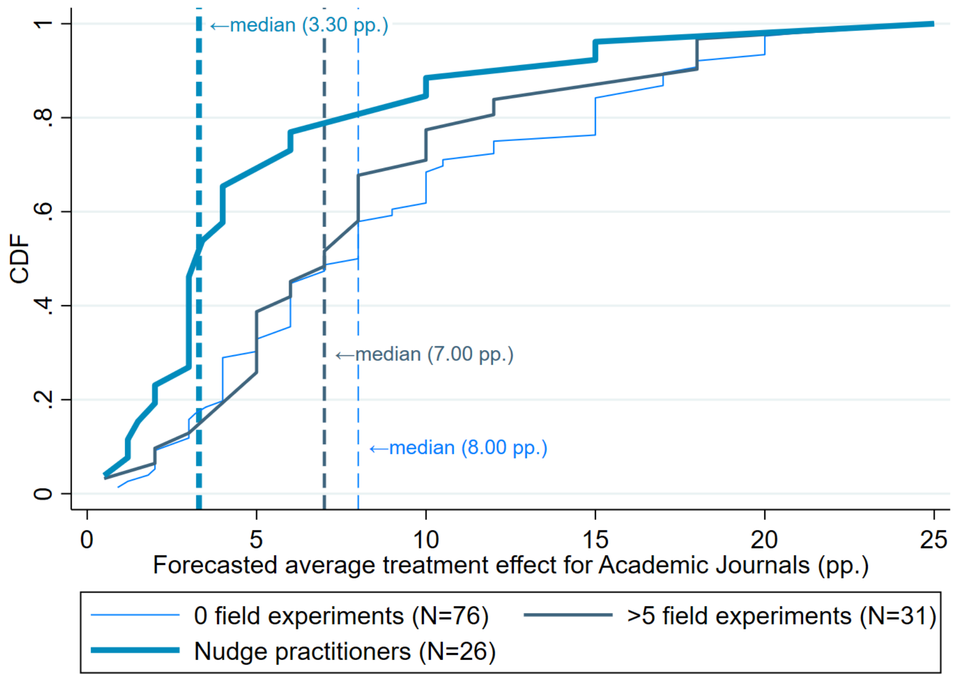
(c) By experience



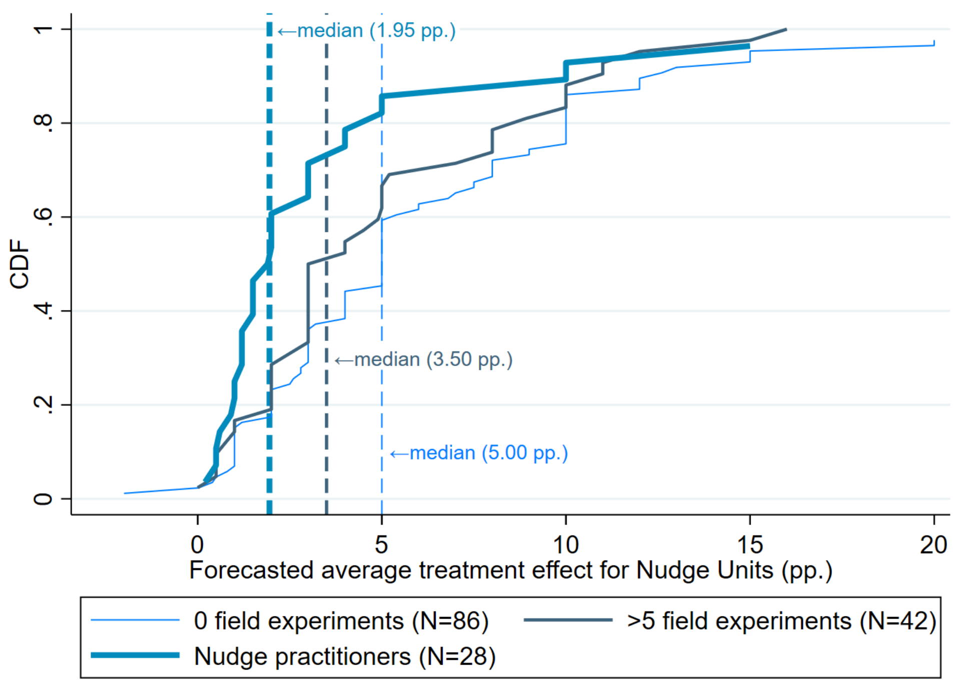
This figure shows the characteristics of the forecasters along several dimensions. Figure A.10a categorizes forecasters by their professional affiliation, A.10b by their academic background (if they are university faculty/(under)graduate students), and A.10c by their experience in conducting field experiments.

Figure A.11: Findings vs. expert forecasts

(a) Forecasts for Academic Journals by forecaster experience



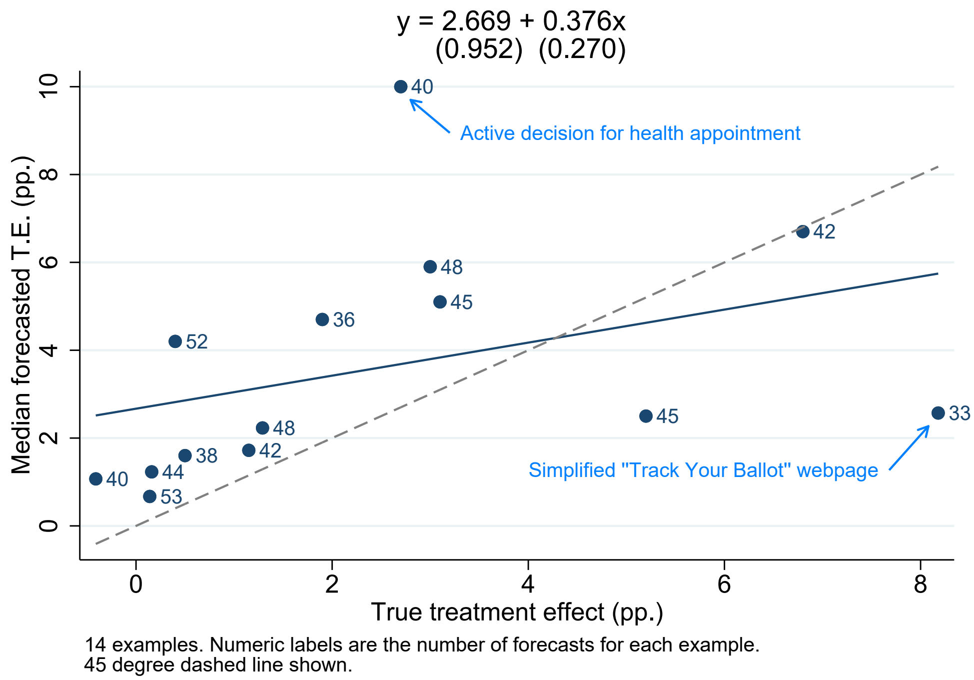
(b) Forecasts for Nudge Units by forecaster experience



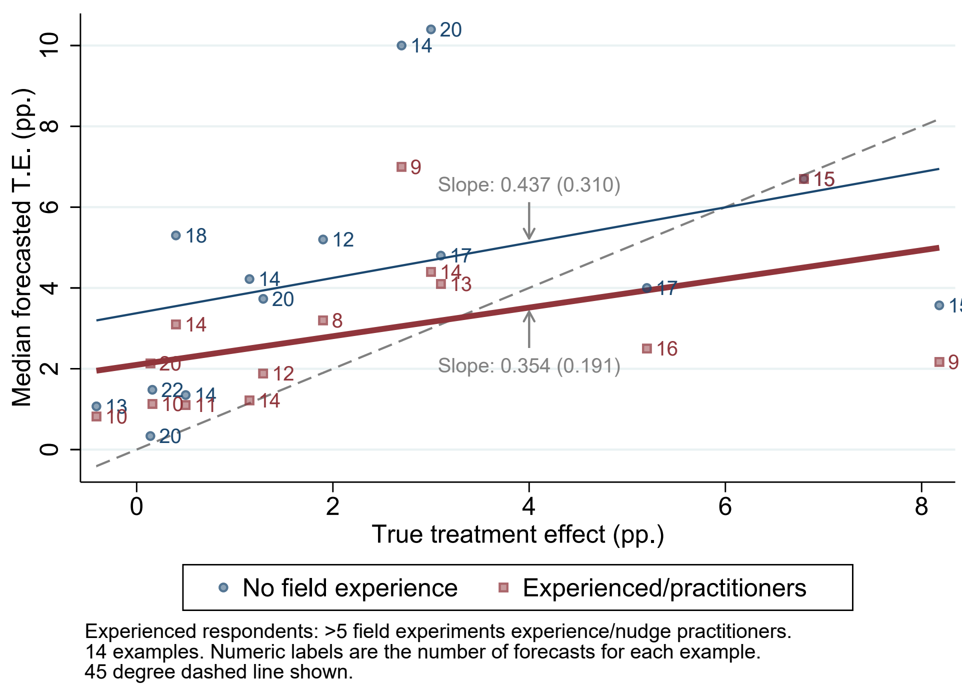
Figures A.11a and A.11b show the distributions of forecasts for treatment effects in the Academic Journals and Nudge Units samples respectively, separated by the forecasters' experience in running field experiments.

Figure A.12: Example-by-example forecasts

(a) All respondents



(b) Forecasts by forecaster experience



This figure plots the median forecasted treatment effect for each of the 14 examples shown on the forecast survey against the true treatment effect. Figure A.12a presents forecasts from all the respondents, and A.12b splits the forecasts by experience.

Table A.Ia: List of published papers in the Nudge Units sample

Published papers featuring OES trials

1. Anteneh et al. 2020. "Appraising praise: experimental evidence on positive framing and demand for health services." *Applied Economimcs Letters*. Cited by 0 (Insignificant)
2. Benartzi et al. 2017. "Should Governments Invest More in Nudging?" *Psychological Science*, 28(8): 1041-1055. Cited by 281
3. Bowers et al. 2017. "Challenges to Replication and Iteration in Field Experiments: Evidence from Two Direct Mail Shots." *American Economic Review, Papers and Proceedings*, 107(5): 462-65. Cited by 0
4. Castleman and Page. 2017. "Parental influences on postsecondary decision-making: Evidence from a text messaging experiment." *Educational Evaluation and Policy Analysis*, 39(2): 361-77. Cited by 26
5. Chen et al. forthcoming. "The Effect of Postcard Reminders on Vaccinations Among the Elderly: A Block-Randomized Experiment." *Behavioural Public Policy*. Cited by 0
6. Guyton et al. 2017. "Reminders and Recidivism: Using Administrative Data to Characterize Nonfilers and Conduct EITC Outreach." *American Economic Review, Papers & Proceedings*, 107(5): 471-75. Cited by 8
7. Leight and Safran. 2019. "Increasing immunization compliance among schools and day care centers: Evidence from a randomized controlled trial." *Journal of Behavioral Public Administration*, 2(2). Cited by 2 (Insignificant)
8. Leight and Wilson. 2019. "Framing Flexible Spending Accounts: A Large-Scale Field Experiment on Communicating the Return on Medical Savings Accounts." *Health Economics*, 29(2): 195-208. Cited by 0 (Insignificant)
9. Kramer and Cooper. 2020. Paper based on trial "Using Proactive Communication to Increase College Enrollment for Post-9/11 GI Bill Beneficiaries", R&R at *Education Finance and Policy*.
10. Sacarny, Barnett, and Le. 2018. "Effect of Peer Comparison Letters for High-Volume Primary Care Prescribers of Quetiapine in Older and Disabled Adults." *JAMA Psychiatry*, 75(10): 1003-1011. Cited by 21
11. Yokum et al. 2018. "Letters designed with behavioural science increase influenza vaccination in Medicare beneficiaries." *Nature Human Behaviour*, 2: 743-749. Cited by 5

Published papers featuring BIT-NA trials

1. Linos. 2017. "More Than Public Service: A Field Experiment on Job Advertisements and Diversity in the Police." *Journal of Public Administration Research and Theory*, 28(1): 67-85. Cited by 25
2. Linos, Ruffini, and Wilcoxen. 2019. "Belonging Affirmation Reduces Employee Burnout and Resignations in Front Line Workers." Working paper. Cited by 0
3. Linos, Quan, and Kirkman. 2020. "Nudging Early Reduces Administrative Burden: Three Field Experiments to Improve Code Enforcement." *Journal of Policy Analysis and Management*, 39(1): 243-265.
   (covers 3 trials) Cited by 0 (2/3 trials are insignificant)

Table A.Ib: List of papers in the Academic Journals sample

1. Altmann and Traxler. 2014. "Nudges at the Dentist." *European Economic Review*, 11(3): 634-660. Cited by 69
2. Apesteguia, Funk, and Iriberri. 2013. "Promoting Rule Compliance in Daily-Life: Evidence from a Randomized Field Experiment in the Public Libraries of Barcelona." *European Economic Review*, 63(1): 66-72. Cited by 36
3. Bartke, Friedl, Gelhaar, and Reh. 2016. "Social Comparison Nudges—Guessing the Norm Increases Charitable Giving." *Economics Letters*, 67: 8-13. Cited by 16
4. Bettinger and Baker. 2011. "The Effects of Student Coaching in College: An Evaluation of a Randomized Experiment in Student Mentoring." *Educ. Eval. & Policy Analysis*, 33: 433-461. Cited by 31
5. Bettinger, Long, Oreopoulos, and Sanbonmatsu. 2012. "The Role of Application Assistance and Information in College Decisions: Results from the H & R Block FAFSA Experiment." *Quarterly Journal of Economics*, 8(10): e77055. Cited by 780

6. Carroll, Choi, Laibson, Madrian, and Metrick. 2009. "Optimal Defaults and Active Decisions." *Quarterly Journal of Economics*, 53(5): 829-846. Cited by 581

7. Castleman and Page. 2015. "Summer Nudging: Can Personalized Text Messages and Peer Mentor." *Journal of Economic Behavior and Organization*, 16(1): 15-22. Cited by 273

8. Chapman et al.. 2010. "Opting in Vs. Opting out of Influenza Vaccination." *Journal of the American Medical Association*, 76: 89-97. Cited by 135

9. Cohen et al.. 2015. "Effects of Choice Architecture and Chef-Enhanced Meals on the Selection and Consumption of Healthier School Foods: A Randomized Clinical Trial." *JAMA Pediatrics*, 124(4): 1639-1674. Cited by 77

10. Damgaard and Gravert. 2016. "The Hidden Costs of Nudging: Experimental Evidence from Reminders in Fundraising." *Journal of Public Economics*, 121(556): F476-F493. Cited by 66 (Insignificant)

11. Fellner, Sausgruber, and Traxler. 2013. "Testing Enforcement Strategies in the Field: Appeal, Moral Information, Social Information." *Journal of the European Economic Association*, 108(26): 10415-10420. Cited by 285

12. Gallus. 2016. "Fostering Public Good Contributions with Symbolic Awards: A Large-Scale Natural Field Experiment at Wikipedia." *Management Science*, 115: 144-160. Cited by 68

13. Goswami and Urminsky. 2016. "When Should the Ask Be a Nudge? The Effect of Default Amounts on Charitable Donations." *Journal of Marketing Research*, 60(573): e137-43. Cited by 57

14. Holt, Thorogood, Griffiths, Munday, Friede, and Stables. 2010. "Automated electronic reminders to facilitate primary cardiovascular disease prevention: randomised controlled trial." *British Journal of General Practice*, 152: 73-75. Cited by 35

15. Kristensson, Wästlund, and Söderlund. 2017. "Influencing Consumers to Choose Environment Friendly Offerings: Evidence from Field Experiments." *Journal of Business Research*, 304(1): 43-44. Cited by 22

16. Lehmann, Chapman, Franssen, Kok, and Ruiter. 2016. "Changing the default to promote influenza vaccination among health care workers." *Vaccine*, 36(1): 3-19. Cited by 22

17. Löfgren, Martinsson, Hennlock, and Sterner. 2012. "Are Experienced People Affected by a Pre-Set Default Option—Results from a Field Experiment." *Journal of Env. Econ. & Mgmt.*, 64: 266-284. Cited by 69 (Insignificant)

18. Luoto, Levine, Albert, and Luby. 2014. "Nudging to Use: Achieving Safe Water Behaviors in Kenya and Bangladesh." *Journal of Development Economics*, 63(12): 3999-4446. Cited by 30

19. Malone, and Lusk. 2017. "The Excessive Choice Effect Meets the Market: A Field Experiment on Craft Beer Choice." *Journal of Behav. & Exp. Econ.*, 129: 42-44. Cited by 13

20. Miesler, Scherrer, Seiler, and Bearth. 2017. "Informational Nudges As An Effective Approach in Raising Awareness among Young Adults about the Risk of Future Disability." *Journal of Consumer Behavior*, 169(5): 431-437. Cited by 7

21. Milkman, Beshears, Choi, Laibson, and Madrian. 2011. "Using Implementation Intentions Prompts to Enhance Influenza Vaccination Rates." *PNAS*, 34(11): 1389-92. Cited by 297

22. Nickerson, and Rogers. 2010. "Do You Have a Voting Plan? Implementation Intentions, Voter Turnout, and Organic Plan Making." *Psychological Science*, 127(3): 1205-1242. Cited by 243

23. Rodriguez-Priego, Van Bavel, and Monteleone. 2016. "The Disconnection Between Privacy Notices and Information Disclosure: An Online Experiment." *Economia Politica*, 21(2): 194-199. Cited by 4

24. Rommela, Vera Buttmannb, Georg Liebig, Stephanie Schönwetter, and Valeria Svart-Gröger. 2015. "Motivation Crowding Theory and Pro-Environmental Behavior: Experimental Evidence." *Economics Letters*, 157: 15-26. Cited by 14

25. Stutzer, Goette, and Zehnder. 2011. "Active Decisions and Prosocial Behaviour: A Field Experiment on Blood Donation." *Economic Journal*, 72: 19-38. Cited by 65 (Insignificant)

26. Wansink and Hanks. 2013. "Slim by Design: Serving Healthy Foods First in Buffet Lines Improves Overall Meal Selection." *PLoS ONE*, 110: 13-21. Cited by 93

Citations are updated as of March 5, 2020. The "(Insignificant)" label applies to papers that have no nudge treatment arms with a *t*-stat above 1.96.

Table A.II: Comparison of nudge categories

	Nudge Units			Academic Journals		
	Freq. (%)	Control take-up (%)	Trial-level N	Freq. (%)	Control take-up (%)	Trial-level N
Date						
Early*	46.06	14.01	194,229	48.65	25.34	24,208
Recent*	53.94	20.06	142,634	51.35	26.58	5,518
Policy area						
Revenue & debt	29.05	11.90	151,075	17.57	10.98	23,380
Benefits & programs	22.41	17.37	381,021	10.81	27.66	4,312
Workforce & education	18.67	14.39	134,726	9.46	66.16	3,950
Health	12.45	19.48	85,164	28.38	24.57	4,854
Registration & regulation compliance	8.71	45.41	7,981	12.16	14.42	8,917
Community engagement	7.88	8.77	196,286	4.05	40.27	135,912
Environment	0.83	23.37	9,478	13.51	28.20	419
Consumer behavior	0	–	0	4.05	15.43	7,253
Medium of communication						
Email	39.83	13.03	205,076	12.16	21.06	17,962
Physical letter	29.88	26.05	184,759	16.22	13.17	14,911
Postcard	21.58	15.39	122,838	6.76	8.90	1,227
Website	2.90	9.85	22,822	12.16	10.83	2,492
In person	0.83	27.50	4,242	28.38	35.40	2,299
Other	10.37	22.20	120,825	24.32	38.28	26,304
Control group receives:						
No communication	61.41	15.14	230,798	43.24	29.51	25,709
Some communication	38.59	20.78	84,493	56.76	23.28	8,149
Mechanism						
Simplification & information	58.51	17.23	217,529	5.41	24.08	4,057
Personal motivation	57.26	15.91	208,042	32.43	30.97	4,347
Reminders & planning prompts	31.54	27.13	160,849	35.14	25.17	26,246
Social cues	36.51	17.55	98,317	21.62	31.11	8,230
Framing & formatting	31.95	12.74	205,766	32.43	23.78	1,614
Choice design	6.22	14.05	334,554	20.27	23.60	2,723
Total	100	17.33	23,556,095 (sum)	100	25.97	505,337 (sum)

This table shows the frequency of nudges in each category, and the average control group take-up and trial-level N within each category. Frequencies for *Medium* and *Mechanism* are not mutually exclusive and frequencies may not sum to 1.

**Early* refers to trials implemented between 2015-2016 for Nudge Units, and to papers published in 2014 or before for Academic Journals. *Recent* refers to trials and papers after these dates.

Table A.IIIa: Unweighted treatment effects in log odds ratio

	Academic Journals	Nudge Units			
		All	BIT	OES	Academic-affiliated OES
	(1)	(2)	(3)	(4)	(5)
Average treatment effect (log odds ratio)	0.499	0.273	0.257	0.292	0.339
	(0.110)	(0.0671)	(0.0717)	(0.120)	(0.265)
Nudges	74	229	123	106	44
Trials	26	121	75	46	23
Observations	505,337	23,370,543	1,913,572	21,456,971	8,919,795
Average control group take-up (%)	25.97	17.94	16.62	19.47	26.45
Distribution of treatment effects					
25th percentile	0.12	0.02	0.00	0.02	0.01
50th percentile	0.32	0.10	0.12	0.08	0.04
75th percentile	0.69	0.34	0.49	0.23	0.17

This table shows the average treatment effect of nudges. Standard errors clustered by trial are shown in parentheses.

Table A.IIIb: Unweighted treatment effects for Published Nudge Unit trials

	Percentage points	Log odds ratio
	(1)	(2)
Average treatment effect	0.970	0.202
	(0.234)	(0.0981)
Nudges	33	33
Trials	16	16
Observations	2,136,014	2,136,014
Average control group take-up (%)	31.93	31.93
Distribution of treatment effects		
25th percentile	0.20	0.02
50th percentile	0.50	0.05
75th percentile	1.20	0.14

This table shows the average treatment effect of nudges. Standard errors clustered by trial are shown in parentheses.

Table A.IVa: Categorization of treatment effects

	Academic Journals		Nudge Units	
	Nudges	Freq. (%)	Nudges	Freq. (%)
Significant & positive	40	54.05	116	48.13
Insignificant & positive	28	37.84	79	32.78
Insignificant & negative	6	8.11	33	13.69
Significant & negative	0	0	13	5.39
Total	74	100	241	100

Significance is determined at the 95% level.

Table A.IVb: Robustness checks

	Academic Journals	Nudge Units	Published Nudge Units
	(1)	(2)	(3)
Average treatment effect (pp.)	8.68	1.39	0.97
	(2.47)	(0.30)	(0.23)
Panel A. *ATE including:*			
Defaults	9.57	1.46	0.97
	(2.60)	(0.31)	(0.23)
Most policy relevant	6.47	1.55	1.00
	(1.73)	(0.47)	(0.24)
Low cost interventions	–	1.35	1.18
		(0.36)	(0.67)
Panel B. *ATE weighted by:*			
Citations	7.89	–	0.76
	(2.01)		(0.14)
asinh(citations)	8.25	–	0.92
	(2.19)		(0.19)
Nudges	74	241	33
Trials	26	126	16
Observations	505,337	23,556,095	2,136,014

This table shows the average treatment effects including default nudges, only the outcomes in the top half of policy relevance, or only nudges with low cost interventions, and weighting treatment effects by citations. Standard errors clustered by trial are shown in parentheses. The Nudge Units sample has 2 nudges (from 1 trial) that use defaults on 1.3 million participants and have treatment effects in pp. (standard errors) of 9.4 (0.15) and 11.2 (0.15). The Academic Journals sample has 3 nudges (from 3 trials) that use defaults on 548 participants and have treatment effects in pp. (standard errors) of -0.1 (3.6), 3.9 (7.78), and 91 (2.87). Policy relevance is determined by priority scores in response to the question: *How much of a priority is this outcome to its policy area?* Seven undergraduates reported their scores for each trial outcome on a 3-point scale (1-Low, 2-Medium, 3-High). The most policy relevant nudges are defined as those in the top half of average priority scores. For the Academic Journals outcomes, the Cronbach's alpha for the scoring is 0.83, and for the Nudge Units, 0.62. 65 percent of Nudge Unit trials are considered low cost interventions, which are either email communications or cases in which the control group was receiving a status quo communication. Citations are updated as of March 5, 2020. Trials with zero citations are assigned a citation count of 1 in the weighting analysis. See Tables A1a and A1b for the list of published trials and their citation counts.

Table A.V: Targeted power in MDE calculations from AEA registry trials

	Number of trials
(1) All trials in AEA registry as of March 2020	3379
(2) Trials registered prior to intervention start date	1315
(2a) Trials with non-empty MDE field	555
(2b) Trials specifying targeted power level for MDE calculation	267
(2c) Trials using a target power level of 0.8 for MDE calculation	240

The trials included in this table were scrapped from the AEA RCT Registry in March 2020. The registry contains an optional field titled "Minimum detectable effect size for main outcomes (accounting for sample design and clustering)". We use the responses in this field to compile data on targeted power levels in minimum detectable effect size (MDE) calculations for trials that were registered prior to the start of their intervention. Row (2a) includes trials that (i) stated a MDE without specifying the target power level, (ii) referred to a separate document without stating the MDE and its target power level in the MDE field, or (iii) calculated the power based on an expected effect size (instead of calculating the minimum detectable effect size based on a target power level); these trials are excluded in rows (2b) and (2c).

Table A.VIa: Heterogeneity in effects by nudge categories: Academic Journals

Dep. Var.: Treatment effect (pp.)	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)	OLS (6)	OLS (7)	OLS (8)	Lasso (9)
Minimum detectable effect (pp.)	1.047							-0.820	0.554
	(0.303)							(0.457)	
Control take-up %		0.706						1.077	
		(0.289)						(0.332)	
Control take-up %^{2}		-0.009						-0.011	
		(0.004)						(0.006)	
Log(outcome time-frame days)		-1.676						-3.543	
		(0.945)						(1.432)	
Date									
Recent (published after 2014)			3.086					0.295	
			(4.760)					(3.302)	
Policy area									
Benefits & programs				10.547				6.892	
				(5.170)				(6.455)	
Workforce & education				-1.046				-11.559	
				(3.483)				(11.008)	
Health				5.379				-1.754	
				(3.885)				(6.904)	
Registrations & regulation compliance				-0.447				-22.885	
				(3.482)				(8.069)	
Community engagement				-0.803				-20.176	
				(4.039)				(9.863)	
Environment				19.351				1.318	2.474
				(7.723)				(8.461)	
Consumer behavior				-0.409				-23.615	
				(3.436)				(10.004)	
Medium of communication									
Email					-5.629			9.886	
					(3.683)			(5.623)	
Physical letter					-7.710			-1.022	
					(3.253)			(4.866)	
Postcard					1.078			19.467	
					(3.124)			(7.729)	
Website					-3.144			10.777	
					(4.307)			(11.767)	
In person					5.442			3.703	
					(5.331)			(6.083)	
Control group receives:									
Some communication						-3.920		-5.335	
						(5.319)		(4.553)	
Mechanism									
Simplification & information							14.333	13.567	
							(4.649)	(5.847)	
Personal motivation							0.288	1.571	
							(3.984)	(4.114)	
Reminders & planning prompts							0.286	2.870	
							(3.183)	(4.388)	
Social cues							9.382	9.953	
							(6.724)	(4.640)	
Framing & formatting							8.999	8.429	
							(4.496)	(4.363)	
Choice design							3.766	10.424	
							(4.183)	(6.037)	
Constant	0.116	3.721	7.098	3.603	9.382	10.907	2.003	1.106	3.819
	(1.935)	(4.566)	(1.638)	(3.436)	(3.124)	(5.047)	(3.679)	(7.969)	
Nudges	74	74	74	74	74	74	74	74	74
Trials	26	26	26	26	26	26	26	26	26
Observations	505,337	505,337	505,337	505,337	505,337	505,337	505,337	505,337	505,337
R-squared	0.34	0.24	0.02	0.35	0.17	0.03	0.23	0.72	
Avg. control take-up	25.97	25.97	25.97	25.97	25.97	25.97	25.97	25.97	25.97

Standard errors clustered by trial are shown in parentheses. The minimum detectable effect (MDE) is calculated in pp. at power 0.8. The penalty parameter in the linear lasso model is selected with cross-validation.

Table A.VIb: Heterogeneity in effects by nudge categories: Nudge Units

Dep. Var.: Treatment effect (pp.)	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)	OLS (6)	OLS (7)	OLS (8)	Lasso (9)
Minimum detectable effect (pp.)	0.207							0.225	0.105
	(0.246)							(0.253)	
Control take-up %		0.089						-0.002	
		(0.059)						(0.049)	
Control take-up %^{2}		-0.001						-0.000	
		(0.001)						(0.001)	
Log(outcome time-frame days)		0.268						0.259	0.099
		(0.268)						(0.326)	
Date									
Recent (2017-)			-0.904					-0.130	-0.026
			(0.640)					(0.644)	
Policy area									
Benefits & programs				-1.541				-1.230	-0.128
				(1.004)				(0.740)	
Workforce & education				-1.935				-1.206	-0.209
				(0.935)				(0.833)	
Health				-1.700				-2.750	-0.585
				(0.968)				(1.258)	
Registrations & regulation compliance				-0.251				-0.720	
				(1.233)				(1.463)	
Community engagement				-1.685				-1.538	
				(1.537)				(1.186)	
Environment				4.404				4.878	3.361
				(1.180)				(1.876)	
Medium of communication									
Email					-0.309			-1.036	-0.048
					(0.659)			(0.801)	
Physical letter					1.144			1.039	0.883
					(0.807)			(0.728)	
Postcard					-0.765			-0.361	
					(0.665)			(0.722)	
Website					-1.408			-0.193	
					(3.376)			(2.844)	
In person					1.266			1.263	
					(1.550)			(2.809)	
Control group receives:									
Some communication						-0.080		-0.281	
						(0.630)		(0.588)	
Mechanism									
Simplification & information							-0.220	-0.774	
							(0.483)	(0.683)	
Personal motivation							0.860	0.953	0.546
							(0.515)	(0.550)	
Reminders & planning prompts							1.347	1.092	0.753
							(0.632)	(0.590)	
Social cues							-0.341	-0.457	-0.107
							(0.457)	(0.611)	
Framing & formatting							0.007	-0.424	
							(0.586)	(0.673)	
Choice design							5.615	4.858	4.943
							(3.030)	(2.609)	
Constant	1.031	-0.002	1.878	2.426	1.367	1.421	0.375	1.242	-0.001
	(0.341)	(0.819)	(0.530)	(0.919)	(0.567)	(0.378)	(0.562)	(1.716)	
Nudges	241	241	241	241	241	241	241	241	241
Trials	126	126	126	126	126	126	126	126	126
Observations	23,556,095	23,556,095	23,556,095	23,556,095	23,556,095	23,556,095	23,556,095	23,556,095	23,556,095
R-squared	0.01	0.04	0.01	0.06	0.03	0.00	0.17	0.26	
Avg. control take-up	17.33	17.33	17.33	17.33	17.33	17.33	17.33	17.33	17.33

Standard errors clustered by trial are shown in parentheses. The minimum detectable effect (MDE) is calculated in pp. at power 0.8. The penalty parameter in the linear lasso model is selected with cross-validation. The 4 nudges (2 trials) missing control take-up data are dummied out when including control take-up in the regression.

Table A.VII: Weighted decomposition between Nudge Units and Academic Journals

Dep. Var.: Treatment effect (pp.)	(1) Egger's test	(2)	(3)	(4)
Academic Journals	-0.282	1.676	3.902	-0.054
	(0.100)	(1.314)	(1.712)	(0.763)
Standard error (SE)	4.237			
	(1.116)			
Academic Journals×SE	-0.816			
	(1.292)			
Constant	0.044	1.107	1.597	1.174
	(0.041)	(0.393)	(0.368)	(0.365)
Nudges	315	315	311	311
Trials	152	152	150	150
R-squared	0.112	0.021	0.078	0.000
Weighted by $1/SE^2$	✓			
Weighted by 1/MDE		✓		✓
Weighted by P-score from nudge categories			✓	✓

Standard errors clustered by trial are shown in parentheses. The coefficient on Academic Journals sample is the estimated average difference in percentage point (pp.) treatment effects between the Academic Journals and Nudge Units samples. MDE (minimum detectable effect) is calculated in pp. at power 0.8. P-score is the propensity score of being in the Academic Journals sample using predicted probabilities from a logit regression that includes the same nudge category controls as in Column 2 of Table IV.

Table A.VIII: Traditional meta-analysis models (without correction for selective publication)

	True study-level effects distributional assumption	Academic Journals		Nudge Units		Published Nudge Units	
		(1) ATE (pp.)	(2) $\hat{\tau}$	(3) ATE (pp.)	(4) $\hat{\tau}$	(5) ATE (pp.)	(6) $\hat{\tau}$
Unweighted	None	8.68	–	1.39	–	0.97	–
		(2.47)		(0.30)		(0.23)	
Maximum Likelihood	Normal	7.86	9.68	1.32	3.50	0.55	0.34
		(2.11)		(0.27)		(0.14)	
Empirical Bayes	Normal	7.95	10.40	1.33	3.71	0.62	0.49
		(2.15)		(0.27)		(0.14)	
DerSimonian-Laird	None	5.41	2.53	0.95	0.63	0.57	0.38
		(1.42)		(0.17)		(0.14)	
Card, Kluve, and Weber (2018)	None	1.90	–	1.26	–	0.82	–
		(0.96)		(0.25)		(0.18)	
Fixed effect	Degenerate	2.40	0.00	1.22	0.00	0.71	0.00
		(1.09)		(0.38)		(0.16)	

This table shows the average treatment effects using various meta-analysis methods. Standard errors clustered by trial are shown in parentheses. $\hat{\tau}$ is the estimated standard deviation in between-study true effect sizes. Following Card, Kluve, and Weber (2018), we winsorize weights from their method at the 10th and 90th percentiles. Mantel-Haenszel weights are used for the fixed-effect model.

Table A.IXa: Generalized meta-analysis models: Additional specifications

	ATE (pp.)	$\hat{\gamma}$ (pub. bias)	Normal 1 $\hat{\bar{\beta}}_1$	$\hat{\tau}_{BT1}$	$\hat{\tau}_{WI1}$	$\hat{P}$(N1)	Normal 2 $\hat{\bar{\beta}}_2$	$\hat{\tau}_{BT2}$	$\hat{\tau}_{WI2}$	$\hat{P}$(N2)	Normal 3 $\hat{\bar{\beta}}_3$	$\hat{\tau}_{BT3}$	$\hat{\tau}_{WI3}$	$\hat{P}$(N3)	-Log likelihood
Panel A. *Traditional parametric normal-based meta-analysis*															
Academic Journals	5.19	0.25	5.19	9.00	5.47	1	–	–	–	–	–	–	–	–	265.90
	(3.84)	(0.32)	(3.84)	(2.58)	(2.74)										
Published Nudge Units	0.68	1 (fixed)	0.68	0.45	0.14	1	–	–	–	–	–	–	–	–	31.66
	(0.36)		(0.36)	(0.30)	(0.07)										
Published Nudge Units	0.35	0.07	0.35	0.42	0.13	1	–	–	–	–	–	–	–	–	26.15
	(0.23)	(0.08)	(0.23)	(0.19)	(0.05)										
Panel B. *Mixture of two normals meta-analysis*															
Academic Journals	8.50	1 (fixed)	3.09	2.48	0.05	0.69	20.43	5.44	12.41	0.31	–	–	–	–	216.59
	(1.97)		(1.04)	(0.78)	(0.20)	(0.11)	(4.68)	(2.78)	(2.46)	(0.11)					
Published Nudge Units	1.07	1 (fixed)	0.47	0.29	0.13	0.74	2.74	0.00	0.00	0.26	–	–	–	–	28.69
	(0.36)		(0.15)	(0.11)	(0.06)	(0.15)	(0.57)	(0.01)	(0.02)	(0.15)					
Published Nudge Units	0.36	0.07	0.09	0.08	0.04	0.59	0.75	0.11	0.17	0.41	–	–	–	–	23.96
	(0.17)	(0.09)	(0.12)	(0.11)	(0.03)	(0.19)	(0.37)	(0.27)	(0.09)	(0.19)					
Panel C. *Mixture of three normals meta-analysis*															
Academic Journals	3.23	0.07	0.26	0.17	0.03	0.59	3.11	2.88	0.01	0.30	19.21	5.91	12.80	0.11	205.68
	(1.48)	(0.08)	(0.34)	(0.14)	(0.13)	(0.15)	(1.59)	(1.40)	(0.25)	(0.14)	(5.10)	(2.89)	(2.28)	(0.06)	
Nudge Units	1.48	1 (fixed)	0.21	0.28	0.10	0.61	2.34	1.83	0.66	0.32	8.54	0.00	13.21	0.07	355.33
	(0.34)		(0.07)	(0.08)	(0.03)	(0.09)	(0.64)	(0.55)	(0.19)	(0.07)	(3.97)	(5.96)	(5.11)	(0.04)	

This table shows additional results from generalized normal-based meta-analysis model in Table V. Under the normal-based meta-analysis assumptions in Panel A, trial base effects β_i are drawn from a normal distribution centered at $\bar{\beta}$ with between-trial standard deviation τ_{BT}. Then, each treatment arm j within a trial i draws a base treatment effect $\beta_{ij} \sim N(\beta_i, \tau^2_{WI})$, where τ_{WI} is the within-trial standard deviation. Each treatment arm also has some level of precision given by an independent standard error σ_{ij}. The observed treatment effect is $\hat{\beta}_{ij} \sim N(\beta_{ij}, \sigma^2_{ij})$. In Panel B, the mixture of normals model is a generalization of the normal-based meta-analysis, and allows trial base effects to be drawn from a second normal distribution (and a third, in Panel C). $\hat{P}$(N) is the estimated proportion of effects drawn from each normal distribution. To capture the extent of selective publication, the probability of publication is allowed to differ depending on whether trials have at least one significant treatment arm. In particular, trials without any significant results at the 95% level are γ times as likely to be published as trials with significant results. Estimates are obtained using maximum likelihood, and bootstrap standard errors are shown in parentheses.

Table A.IXb: Mixture of three normals with stacked data

	Sample			Parameters of normals		
	Academic Journals	Nudge Units		Mean	Between-trial SD	Within-trial SD
P(Normal 1)	0.49 (0.13)	0.63 (0.07)	Normal 1	0.22 (0.07)	0.28 (0.08)	0.10 (0.04)
P(Normal 2)	0.38 (0.08)	0.30 (0.06)	Normal 2	2.58 (0.59)	2.11 (0.49)	0.66 (0.20)
P(Normal 3)	0.12 (0.08)	0.07 (0.03)	Normal 3	13.34 (4.36)	7.80 (4.48)	12.95 (1.96)
ATE (pp.)	2.75 (1.24)	1.82 (0.28)				
Pub. bias	0.07 (0.06)	1 (fixed)				
-Log likelihood	208.08	356.55				

This table shows the joint estimation of the mixture of three normals meta-analysis combining both the Academic Journals and Nudge Units samples of nudges. (Panel C of Table A.IXa presents the results when the model is estimated separately for the two samples.) The mean, between-trial variance, and within-trial variance of each of the three normal distributions are assumed to be the same for both samples of nudges, and the two samples only differ in the probability of drawing a trial from each of the normals. The probabilities of drawing from the three normals are modeled using ordinal probit assumptions (see notes in Table A.IXc for details). The results in this table correspond to Column 2 in Table A.IXc. Standard errors from 50 bootstrapped samples are shown in parentheses.

Table A.IXc: Generalized mixture model with selective publication and heterogeneity based on observables

	(1)	(2)	(3)	(4)
Academic Journals	1.22	0.34	-0.04	-0.01
	(0.26)	(0.35)	(0.38)	(0.46)
In-person			1.48	1.48
			(0.63)	(0.58)
Email				-0.08
				(0.31)
Control receives communication			0.06	0.00
			(0.25)	(0.26)
Workforce & education			-0.56	-0.51
			(0.30)	(0.39)
Consumer behavior				-0.72
				(0.89)
Choice design			0.91	0.94
			(0.58)	(0.60)
Framing & formatting				0.40
				(0.32)
θ_1	0.33	0.33	0.34	0.43
	(0.19)	(0.20)	(0.24)	(0.28)
θ_2	1.58	1.50	1.65	1.76
	(0.28)	(0.24)	(0.39)	(0.33)
γ	1 (fixed)	0.07	0.08	0.08
	–	(0.06)	(0.06)	(0.07)
Academic Journals				
ATE at $\bar{X}_{AJ}$ (pp.)	6.67	2.75	3.05	3.05
	(1.93)	(1.24)	(1.44)	(1.41)
ATE at $\bar{X}_{NU}$ (pp.)			1.53	1.56
			(0.74)	(0.87)
Nudge Units				
ATE at $\bar{X}_{NU}$ (pp.)	1.88	1.82	1.61	1.58
	(0.39)	(0.28)	(0.30)	(0.25)
ATE at $\bar{X}_{AJ}$ (pp.)			3.20	3.09
			(1.02)	(0.97)
-Log likelihood	573.50	564.63	558.48	557.25
Nudges	315	315	315	315
Trials	152	152	152	152

This table shows results from the mixture of three normals meta-analysis on a stacked data set combining both Academic Journal and Nudge Unit samples of nudges. The parameters of each of the three normals (mean, between-trial variance, and within-trial variance) are held constant between both samples. The two samples of nudges differ in the probability of drawing a trial from each of the three normals. These probabilities are estimated under an ordinal probit model. Specifically, the probability that a trial i draws its effect size from the first (lowest) normal is $P(X_i'\eta + \varepsilon < \theta_1)$, where X_i is a $k \times 1$ vector of trial characteristics, such as being in the Academic Journal sample. η is a $k \times 1$ vector of coefficients, and the error ε follows a standard normal distribution. The probability that a trial i draws its effect size from the second (middle) normal is $P(\theta_1 \leq X_i'\eta + \varepsilon < \theta_2)$, and the probability of drawing from the third (highest) normal is $P(\theta_3 \leq X_i'\eta + \varepsilon)$. The thresholds θ_1, θ_2 and the coefficient vector η are jointly estimated. This table shows the estimated coefficients for observable trial and treatment features (e.g., delivering the intervention via email). Observables that vary at the treatment level are included by taking the within-trial average. For tractability, Column 3 includes only the most significant (i.e., with the highest t-stat) medium, policy area, and mechanism as estimated in Column 4 of Table IV and the indicator for whether the control group receives any communication. Column 4 allows for more observables and includes the *two* most significant groups from each category. The table also shows the thresholds in the ordinal probit, and γ, the probability that a trial with no significant results is published relative to a trial with at least one significant result. Below these estimates, the table shows the average treatment effect (ATE) for the two samples separately. For each sample, the ATE is calculated twice, first holding X_i at the average levels within its own sample, and then at the average levels within the other sample (except the indicator for being in the Academic Journals sample). Standard errors from at least 40 bootstrap samples are reported in parentheses.