Analysis plan

Project name: Increasing competitiveness of small manufacturers via improved marketing

Project code: 2503

Date finalized: 7/23/2025

Project description

The goal of this evaluation is to improve the marketing capabilities of small manufacturing businesses who are set up to contract with the federal government and are participants in the SBA's contracting assistance programs. We are focusing on manufacturing businesses to further the SBA's Made in America Manufacturing Initiative by supporting small U.S. manufacturers.

To market themselves for government contracting opportunities, small businesses use capability statements. Capability statements are descriptions of the firm's core strengths and capabilities and are used to showcase and market the firm's services to potential federal buyers. They usually take the form of a 1-2 page PDF uploaded to various federal procurement platforms and are used in the market research conducted by federal procurement officials.

For this project, the SBA wants to understand how updating the design of capability statements (and proactively sharing capability statements) can help manufacturing firms compete for federal contracts. We are designing a template to create standardized capability statements, then testing whether the redesigned capability statements lead to better engagement from federal buyers. We will use the capability statements of (currently) 22 small manufacturing firms who have agreed to have the SBA create an updated version of their statement using the new template. We will then randomly assign procurement officials working in the same industries as those firms (our sample for this evaluation) to receive marketing emails about firms that include either an original capability statement or an updated capability statement created using the template. As our primary outcome, we will measure procurement officials' engagement with the statements by tracking whether they click a link in the email to view the statement or express their interest in the highlighted firm (the capability statement will be hosted on a web page that includes a link to a form for procurement officials to express interest in learning more about the firm).

In addition to that impact evaluation component, this study also has a descriptive component. We aim to learn about the extent to which procurement officials engage with marketing emails and whether it may be worthwhile for the SBA to encourage firms to conduct proactive outreach in this manner. To that end, we will report the percentage of officials who clicked on the links in their emails (and the percentage who filled out a form to contact the firm), both overall and in subgroups of our sample (e.g., officials working in manufacturing).

Preregistration details

This Analysis Plan will be posted on the OES website at [oes.gsa.gov](oes.gsa.gov) before outcome data are analyzed. We will hide it from public view until outcome data are received so that we do not risk revealing the details of the evaluation to recipients of the intervention emails, as this could impact their willingness to engage.

Hypotheses

There are two primary hypotheses:

H1: Procurement officials sent an email with an updated capability statement will be more likely to click the link in the email (to view the capability statement or express their interest in the firm) than procurement officials who receive an email with an original capability statement.

H2: Procurement officials sent an email with an updated capability statement will be more likely to provide their email addresses on an interest form than procurement officials who receive an email with an original capability statement.

Data and data structure

This section describes variables that will be analyzed, as well as changes that will be made to the raw data with respect to data structure and variables.

Data source(s):

Data collected for this study will fall into several categories, discussed more in Tables 1-3 below.

- Outcome measurement
  - We plan to track clicks on links embedded in the intervention emails. Each official sent an email will have their own unique link. Intervention emails will be sent to all procurement officials in our sample, varying only in their content.
  - When an official clicks on their link, they will be redirected to a form for them to fill out hosted on the OES website (discussed more below). But first, they will be routed through a cloud computing application we control that timestamps clicks. This will yield a record of click timestamps along with the unique ID of each link that was clicked (link click data). Those IDs will be randomly generated so that the records include no personally identifying information. A lookup table in our secure data environment will be needed to match those IDs to email addresses.
  - Records will be automatically stored in a cloud database. We will manually back-up records from it once a day (on weekdays) while the intervention is active. We currently anticipate recording click data for two weeks following the intervention emails (preliminary calculations from another ongoing procurement evaluation that involves tracking link clicks suggests that 1-2 weeks is appropriate).

    - We will also collect responses to the forms hosted on the OES website. These forms ask procurement officials to enter their email address if they would like to connect with a firm. This will yield a spreadsheet of procurement official emails by the firm whose form they filled out (interest form data).
- Procurement official and firm information
    - We have compiled datasets of federal contracting opportunities and the contact emails of procurement officials who posted those opportunities (opportunity data). Among other things, the opportunity data also includes information on the "place of performance" state of the procurement opportunity (if it has one) and all relevant industries listed (NAICS codes).
        - We leverage three different sources of contract opportunities data to identify procurement officials we could potentially contact for this study: (1) contact information from past solicitation notices posted on SAM.gov in fiscal years 2020 through 2023; (2) contact information from current solicitation, pre-solicitation, and sources sought notices active in SAM.gov on March 4th, 2025; and (3) contact information from select federal agencies' procurement forecasts circa late 2024.[1]
    - We will also have separate datasets providing a wide variety of information on the 22 firms whose capability statements will be used for the study, exported from their public registration data in SAM.gov. Among other things, this (firm data) includes information on where the firm is incorporated and what designations they have (e.g., veteran-owned, women-owned, etc.).
- Random assignment
    - For every official included in our sample and sent an email, we will have a dataset (assignment data) recording their block assignment (i.e., the strata in which they were randomly assigned to the updated or original capability statement), the number of opportunities they posted in our sources of contact information, the firm they were matched with, the source their contact information was gathered from, the agency they work in, and information on how they were matched to a specific firm.

Outcomes to be analyzed:

The primary outcome is whether the procurement official clicks on a link in the email (Form Clicks) to express their interest in being connected with the spotlighted firm. Specifically, in the email, we will include a call to action for officials to click a link to a form that lets the firm know that the official is interested in learning more about them. These links will be unique to each procurement official, and will lead to a landing page hosted by OES containing the interest form. However,

[1] The agencies are as follows: HHS, USDA, DOJ, DOE, Commerce, DOI, DOL, GSA, NRC, OPM, VA, and DOD (select offices with forecasts posted online). Some of these forecasts were downloaded from agency websites, others were downloaded from Acquistion.gov.

before officials reach that landing page, they will be routed through a Javascript application hosted on a cloud service that tracks each unique link click.

A secondary outcome is whether the procurement official goes a step further to fill out and submit the brief interest form so that we can share their information with the corresponding firm ***(Form Submissions)***. This is a "higher lift" behavior for the procurement official in the sense of requiring more effort than clicking a link, but it is also easier to measure.

We believe that the click tracking method described above is the most reliable available way of tracking link clicks for this project, but it does raise technical concerns surrounding, for instance, whether the cloud application may fail due to some officials' security software. Our primary outcome would then undercount the true link click rate.[2] Adding an email to a contact form is asking more of the officials themselves, but it likely does not raise that same measurement concern.[3]

Raw data:

We will collect the following raw variables ourselves as a part of implementing this evaluation:

Table 1. Raw data collected

Source	Variables
Link Click Data Records stored in a cloud database, produced by a Javascript application deployed on another cloud service. There will be one row for each unique click of one of the links emailed as a part of this study.	Name: ***link_id*** Anonymized procurement official IDs - *Each procurement official contact email in our sample will have a different randomly assigned ID that is used only for link tracking.* - *These link IDs will be built into their unique links. When an official clicks on one of our links, the ID in that clicked link will be recorded.* - *These IDs can only be matched to emails using a cross-walk that is kept in our secure data environment*
	Name: ***link_time*** Link click timestamp (date/time) - The date/time of each unique click on a particular contact email's link
Interest Form Data	Name: ***interest_email***

[2] Nonetheless, of all available link click tracking methods for this project, in which we are emailing federal employees, we believe this method is likely to undercount the least. As emphasized by [Meyer and Mittag (2017)](), measurement error in a binary outcome in an OLS regression is *not necessarily* attenuating. Equation 6 of that study provides a method of estimating expected bias given assumptions about the false negative rate. We will consider this as a robustness check.

[3] There is the issue that officials could list a different email address in the form besides the one we emailed. If we cannot set up the form to auto-populate an "incoming email" field, and if there is a non-negligible number of emails in this category, we will consider (1) a fuzzy matching procedure and (2) the robustness check described in the footnote above.

When an official we have emailed clicks on our tracking link and goes further by adding their email to the interest form, it will be added to this database. There will be one row for each submitted email / participating firm combination.	A procurement official's provided email - *Submitted by the official themselves when filling out the form*
	Name: ***interest_firm*** The UEI for the participating firm in question - *Each procurement official is emailed only once, about a single firm.* - *In practice, officials will not all be linked to the same interest form. Each participating firm will have their own form (all hosted on the OES website).* - *This field lists the firm whose form the procurement official was linked to (auto-populated for all entries in a particular firm's form).*

Imported variables:

The procurement official and firm information referenced above will be imported from contracting opportunity data (SAM.gov or agency procurement forecasts) and firm registration data (SAM.gov). This includes the following variables. Note that these sources include additional idiosyncratic information we may have collected (e.g., projected opportunity price ranges, when available), but we focus here on the variables actually used in this evaluation.

Table 2. Imported variables

Source	**Variables**
Opportunity Data Combined dataset of contracting opportunities pulled from procurement forecasts, past SAM.gov open market "solicitation" notices (FY 2020-2023), and current SAM.gov "pre-solicitation" / "sources sought" open market notices (March 2025). Structured with one row for each: - Opportunity posting - Distinct procurement contact in an opportunity - Distinct NAICS code listed in an opportunity - Distinct place of performance listed in an opportunity In effect, many opportunities have multiple rows, though most opportunities do not vary on all three categories (multiple federal procurement contacts,	Name: ***row_state*** Place of performance - *One row for each place of performance provided in the opportunity.* - *Two letter state abbreviation. Based on cleaning unstructured text from various fields across different sources.* - *"NONE" if no state was provided, or if the opportunity lists the place of performance as national (or similar). For HHS forecasted opportunities only, performance location was imputed from contracting office location.* - *Another variable, **all_states**, concatenates all places of performance for a given opportunity*
	Name: ***opp_id*** Opportunity ID number - *Assigned during data cleaning.*

multiple NAICS codes, and multiple places of performance).	- *Unique within individual agencies for forecasted opportunities, or within sources for SAM opportunities.* - *See also the variable **solnum** (solicitation numbers for opportunities from SAM)*
	Name: ***source*** Procurement opportunity data source - *Forecasts, past SAM postings, or current SAM postings*
	Name: ***row_contact_id*** Federal procurement official ID - *One row for each procurement official contact listed in the opportunity (may be one or more)* - *Generated based on provided email addresses, assigning each unique email (across sources) its own numeric code.* - *Another variable, **all_contact_ids**, concatenates all IDs for a given opportunity*
	Name: ***row_contact_name*** Federal procurement official's name - *One row for each procurement official contact listed in the opportunity (may be one or more)* - *Inconsistently provided across sources, not always available* - *Another variable, **all_contact_names**, concatenates all names for a given opportunity (in the same order as **all_contact_IDs**)*
	Name: ***row_contact_email*** Federal procurement official's email - *One row for each procurement official contact listed in the opportunity (may be one or more)* - *Another variable, **all_contact_emails**, concatenates all emails for a given opportunity (in the same order as **all_contact_IDs**)*
	Name: ***row_contact_identities*** Federal procurement official's position - *One row for each procurement official contact listed in the opportunity (may be one or more)* - *Inconsistently provided across sources, not always available* - *Another variable, **all_contact_identities**,*

	concatenates all identities for a given opportunity (in the same order as ***all_contact_IDs****)*
	Name: ***row_naics_code*** Opportunity's industry/industries - *One row for each NAICS code listed in the opportunity (may be one or more)* - *Another variable,* ***all_naics****, concatenates all names for a given opportunity*
	Name: ***agency*** Posting agency's 3-letter abbreviation - *Sometimes provided directly, other times based on cleaning or manually reviewing unstructured text* - *See also, when provided, the varaibles* ***agency_code*** *(numeric agency indicators),* ***Department/Ind.Agency*** *(strings sometimes used to generate agency codes), and* ***AAC Code*** *(office codes)*
Firm data Firm registration data from SAM.gov; merged by UEI with data provided by SBA on the Business Opportunity Specialists (BOSs) working with each participating firm	Name: ***firm_name*** Firm's registered name in SAM
	Name: ***uei*** Firm's Unique Entity Identifier (UEI)
	Name: ***all_naics*** All NAICS (North American Industry Classification System) codes, i.e., industries, in which a firm is registered to work in SAM. - *Concatenated around "~"* - *Includes flags indicating whether the firm is a small business or not (per SBA classification rules) in each NAICS. See also the variable* ***naics_exception_string*** *for cases where the designation is more complicated.*
	Name: ***primary_naics*** The firm's "primary" NAICS in SAM
	Name: ***bos*** The firm's assigned Business Opportunity Specialist
	Name: ***state_of_incorporation***

<table>
<tr><td></td><td>The state in which the firm is incorporated
See also, where available, the variables mailing_address_state_or_province and physical_address_state_or_province</td></tr>
</table>

Transformations of variables and data structure:

Based on the data sources above, we will construct the following datasets for (1) random assignment and (2) final analysis.

Table 3. Constructed variables

<table>
<tr><th>Source</th><th>Variables</th></tr>
<tr><td rowspan="5">Assignment data

Variables we generate as a part of our firm matching (i.e., matching procurement officials to a participating firm) and random assignment processes. See the descriptions below.

Includes one row for each procurement contact email from our Opportunity Data that has posted at least one opportunity in a NAICS code (row_naics in the Opportunity Data) in which one of our participating firms works (all_naics in the Firm Data).</td><td>Name: uei

The firm with which an official is matched
See below for more detail on our matching procedure</td></tr>
<tr><td>Name: manufacturing_match

Manufacturing NAICS match
Binary indicator: whether this official was matched to a firm based on sharing a manufacturing NAICS code, i.e., a code between 31xxxx-33xxxx.
See also the manufacturing_naics variable, which provides the NAICS in question.</td></tr>
<tr><td>Name: contact_id

Unique contact IDs
row_contact_id from the Opportunity Data associated with this email</td></tr>
<tr><td>Name: contact_email

Unique contact email
row_contact_email from the Opportunity Data</td></tr>
<tr><td>Name: is_cluster

Are multiple emails possibly the same person?
In some cases, for the purposes of treatment assignment, we will “cluster” multiple emails that we need to tentatively treat as being the same person (see below).
We will list all emails in a cluster in the</td></tr>
</table>

	all_emails *variable. These emails will still have separate* ***contact_id****s, separate rows in the dataset, and separate* ***link_id****s.* - *We will also construct a binary indicator for these cases (****is_cluster****), and a count of the number of emails (****num_in_cluster****).*
	Name: ***match_group*** The strategy used to match emails to firms - *See below for more context. Note that this is computed at the "email cluster" level.* - *We will sort emails into groups based on the procedure used to match them to firms (priority order: manufacturing > location > firm primary NAICS). We expect larger treatment effects for firms that can be matched on location or primary NAICS, and we want to prioritize estimating a subgroup effect for manufacturing procurement.* - *This decision is not yet finalized. But it must be made before the emails go out, and thus will be made before outcome data could possibly be observed.*
	Name: ***block_id*** The block (strata) for random assignment - *The block within which a contact ("cluster") is randomly assigned to the updated capability statement.* - *Given the risk of some emailed officials no longer working in procurement (below), we need to maintain larger blocks. But, if feasible, we would like to block on several variables (priority order: match group > contact email source > firm UEI).* - *This decision is not yet finalized. But it must be made before the emails go out, and thus will be made before outcome data could possibly be observed.*
	Name: ***number_opps*** The observed number of opportunities - *Total number of opportunities in our gathered data that are associated with this* ***contact_id***
	Name: ***UpdatedCS*** Binary indicator for intervention assignment - *I.e., assignment to be emailed a templated*

	capability statement instead of an original capability statement - *Assigned at the “email cluster” level*
	Name: ***link_id*** The email’s ***link_id*** (see Link Click Data) - *When we have multiple emails in our data that we are tentatively treating as the same person, we will still email all of them (rather than guess which is monitored, e.g., in cases of possible typos)* - *Each email in a cluster will still have their own* ***link_id***
	Name: ***dep_of_defense*** Is this a Department of Defense opportunity? - *A binary indicator for contact emails from opportunities posted by offices within the Department of Defense.*
	Name: ***source*** The source we gathered a contact email from - *If any opportunities for a given email come from procurement forecasts, this will be “forecasts.”* - *If any come from current opportunities (as of March 4th, 2025), this will be “current.”* - *Otherwise, this will be “past.”*
Analysis Data The variables above, joined with our Link Click and Interest Form Data for constructing outcome measures. We will also incorporate any information we receive from our implementing partner on cases where an email is clearly no longer active (e.g., bounce backs).	Name: ***form_click*** Binary indicator for link clicks - *I.e., does this row’s* ***link_id*** *appear anywhere in the Link Click Data? If yes, =1, else =0.*
	Name: ***form_submit*** Binary indicator for interest form submission - *I.e., does this row’s* ***contact_email*** *appear anywhere in the Interest Form Data for their* ***uei****? If yes, =1, else =0.*
	Name: ***not_active*** Binary indicator for email bounce backs, etc. - *We aim to identify cases where an official has an outcome value of 0 (no link click) because they are no longer working in federal procurement rather than because they read the email and simply chose not to click on the link.* - *This issue is discussed more below.*

Sample construction:

Firm selection

We worked with three Business Opportunity Specialists (BOSs) at the SBA to choose partner firms for this study. These three BOSs are the regional SBA contacts (representing South Florida, Oklahoma, and Seattle) for firms participating in an SBA certification program. We preferred to do this through an existing certification program because it was important for us to find firms that already had relationships with the SBA. Among other criteria, we wanted to select firms who work in a diverse array of NAICS codes, and especially as many manufacturing NAICS codes as possible, while also finding interested partner firms who might benefit from their participation. All firms also needed to work in at least one manufacturing NAICS code according to their SAM registration. We found 22 firms interested in participating through an iterative process of discussion and outreach via the BOSs.

Initial email selection

As outlined in Table 3, our analysis sample for this evaluation includes all procurement official contact emails from our Opportunity Data (Table 2) with at least one opportunity in a NAICS code where at least one of the participating firms works (Firm Data, Table 2). We do not plan to finalize our firm selection/matching/blocking until closer to intervention roll-out in case there are any late changes (e.g., firms that decide not to participate). However, based on a preliminary test list of firms, we anticipate identifying approximately 35,000-37,000 emails for inclusion in the study (approximately 20,000 of which would be opportunity postings in a manufacturing NAICS code). We emphasize that, because our outcome is only realized once our emails are sent, there is no risk of pending matching/blocking decisions being made while we have access to outcome data.

Email cleaning and data exclusion

Once we have an initial list of emails, the list will need to be cleaned. There are some cases of likely typos in the original procurement opportunities themselves (e.g., "john.smith@specific.agency.gov" and "jonhn.smith@specific.agency.gov") that we do not want to treat as separate emails. Similarly, there are also cases where a single person may work in two related offices or agencies (e.g., doing procurement for the Navy, and for the Defense Logistics Agency, on behalf of the Navy). We prefer a conservative approach that limits our likelihood of accidentally having the same person as both an intervention and control observation. We summarize our strategy here, and see Appendix A for a longer description:

1. Clean out email fields from the raw opportunity data that are not actually emails (e.g., "TBD", or names that do not include "@"). Additionally, clean obvious idiosyncratic typos such as ending the email with ".gvo" instead of ".gov."
2. Identify clusters of emails, ***within the same agency***, that are potentially a single person with multiple different email domains (the part after "@"), or two different emails with the same

domain that may in practice be managed by the same person (due to typos, or idiosyncratic reasons why a person may have multiple email accounts with similar names).

3. Some of those "clusters" could be incorrect and not actually the same person. However, we will treat them as such for the purposes of intervention assignment, assigning clusters of emails to intervention or control conditions instead of individual emails.
    a. This makes our tests more conservative, as assigning the intervention to fewer clusters of emails makes it harder to draw confident statistical conclusions, but will improve the accuracy of our estimates.
    b. In practice, an initial test of this procedure using ~20,000 emails only decreased the number of email clusters to about ~18,000.
4. All emails in a cluster will still be included separately in our intervention and analysis (incorporating step 1 cleaning). We will adjust our analyses for the clustered nature of intervention assignment.

Matching sampled officials to firms

All remaining email clusters will be matched to a single firm. One goal here is to match emails to firms in a way that increases potential engagement with our intervention email. Additionally, we want to match firms/emails as much as possible within the manufacturing industry, in support of estimating a subgroup effect for manufacturing procurement specifically.

We explain our planned strategy in more detail in Appendix B. Briefly, we will prioritize matching within the manufacturing industry first (emails associated with manufacturing opportunities in a NAICS in which one of our firm works), and then matching on location (based on the performance state of a contact email's opportunities, and a firm's state according to their SAM registry). If feasible, we will also match based on a firm's "primary" NAICS (match emails to firms whose primary NAICS matches one of that official's opportunities).

Ultimately, we will sort emails into "match groups" based on how they are being matched to a firm. We expect that officials matched to a firm based in their state, for instance, may be more likely to engage with the intervention email. One important constraint is that we prefer not to finalize the matching process until closer to intervention rollout, in case of any additional changes to our list of participating firms. Another constraint is that we need the match groups to remain relatively large, since they will be incorporated into our blocked randomization procedure (see the next section).

We plan to incorporate as many of the criteria listed above as we can (manufacturing industry, then location, then primary NAICS) while keeping the match groups relatively large in terms of the number of emails falling into each. If possible, we will sort emails into match groups based on every combination of the criteria we use (e.g., if matching based on manufacturing and location matches, this would yield four categories in total). But if this yields match groups that are too small, we will apply a "greedy" strategy following the priority order above (manufacturing matches first, then location matches among those without manufacturing matches, etc.).

We will use an email's assigned match group to narrow the range of candidate firms it could be matched with, and choose randomly among all that remain. Some firms in our sample will have more matches than others. We believe this is a worthwhile tradeoff in order to increase engagement with the email where possible, and to facilitate a subgroup analysis focusing on the manufacturing industry that is as efficient (i.e., well powered) as possible.

Random assignment procedure

We plan to randomly assign email clusters to the intervention (templated capability statement) or control (original capability statement) conditions within "blocks," or "strata," that share particular characteristics in common. We will then assign 50% of the email clusters within each block to the intervention.

We can think of each block as its own randomized sub-evaluation. When analyzing data from a block-randomized intervention, we are essentially estimating an overall treatment effect by pooling across separate within-block treatment effects. This can make treatment effect estimates more efficient (i.e., reducing the random noise that interferes with drawing confident statistical conclusions), particularly if we are interested in estimating treatment effects for subgroups of our sample (e.g., for procurement officials working in manufacturing) or if we expect treatment effect heterogeneity across our sample.[4]

As already discussed, we will not finalize the list of participating firms or match emails to firms until closer to intervention rollout. We will first prioritize assigning emails to blocks based on their match group. After this, if possible, within each match group we will assign firms to smaller blocks based on the source a procurement official's contact email is gathered from. Procurement officials sampled from procurement forecasts may be more likely to engage with our intervention email, or more likely to be interested in the updated capability statement, than an official sampled based on past SAM postings who rarely works in procurement anymore. After this, we may consider blocking based on the firm to which a procurement official is matched.

It can be important for both reducing bias and increasing efficiency to account for how blocked assignment occurred during analysis.[5] However, this is problematic when there is no intervention variation in blocks in the realized data (so it's not possible to estimate a within-block treatment effect). Some–or perhaps many–contact emails we have gathered will represent officials who are no longer working in federal procurement (see our Limitations). If we construct small blocks, and many emails in our sample cannot be delivered, adjusting for blocking anyway among observations that remain would amount to dropping all blocks without intervention variation from our "effective" analytical sample.

To mitigate this concern, we will be selective about the variables we choose to include in our blocking (and firm matching) strategies in practice. We will report our finalized strategy, including the number of emails in each block, in a technical appendix (posted publicly online) that accompanies our completed analysis.

[4] Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*.
[5] *Ibid*. See also: Aronow, P. M., & Samii, C. (2016). Does regression produce representative estimates of causal effects? *American Journal of Political Science*, 60(1), 250-267.

Outcome measurement

We expect that there may be some duplicates or extraneous link clicks that need to be excluded from our raw outcome data. We plan to handle this by measuring whether a link was ever clicked rather than measuring the number of times it was clicked. There may also be instances in our data of a link click that is not associated with a unique ID in our evaluation (e.g., if an official changes the link we send them in some way before following it). We will exclude clicks without IDs from our analysis since we will not be able to match them to a unit in the sample.

Treatment of missing data:

We do not anticipate having to handle any missing data. Since our outcome is whether individuals clicked on the link, any "missing" outcome data would be coded as not taking action for the outcome measure (see Limitations below). Similarly, all variables used in our construction of match groups and blocks can be observed for all emails without missingness, as can any variables listed in the tables above that we may potentially use as additional controls.

Treatment effect estimands under ineligibility:

While we expect to email approximately 35k-75k contact emails, we know that not all of these contacts will still be working in federal procurement at the time of our intervention (***ineligible***). And even those who are eligible may still choose to simply ignore the email or may be out of office at the time it is sent (***never-openers***). Only eligible ***always-openers*** can engage with their email (i.e., get Y=1). Everyone else will have Y=0 by default. In practice, we will likely be able to identify who some of these ineligible officials are (non-deliverable email messages, read receipts we may be able to implement, desk research on office closures, etc.). But we may not know who all of them are, and we may not know for sure how many are left in our sample (e.g., if they are no longer employed by the federal government but their email account hasn't been shut down yet).

We define a "true effect size" as the effect among eligible always-openers. For any given true effect size, the presence of ineligible emails and eligible never-openers decreases our estimated effect (relative to the true effect) and decreases the statistical power of our estimate (relative to the power we'd have without ineligibility or never-openers). In this sense, what we are estimating is an "Intention to Treat" (ITT) effect. As a part of our research design process, for a range of possible sample sizes and true effect sizes–*and for different assumptions about how much of our sample is ineligible*–we used simulations (coded manually in R) to estimate our power to detect SOME effect at a 95% confidence level (***Column 2 of Table 4***), and also the average ITT estimate we'd expect to see in that case (***Column 3 of Table 4***). For illustration, we provide select results from these simulations in Table 4 below. We focus only on the most extreme assumptions about ineligibility (5% and 50%) that we considered.

Table 4. Select simulated power/bias results

Simulation details:

- What is being estimated? - A comparison, on average, between officials sent an updated (templated) capability statement and officials sent an original capability statement.
- Outcome - A binary indicator with a 20% baseline rate. We expect a lower baseline in practice, but wanted to err towards a more conservative assumption for this simulation.[6]

Condition	**Smallest true effect at which we have at least 80% power to detect *some effect* at a 5% confidence level ("*minimum detectable estimand*")**	**Average observed effect for that true effect size**
n = 20k, ½ get control emails (½ an intervention email), 5% ineligibility, 5% are eligible but don't open emails	1.75 pp	1.6 pp
n = 20k, ½ get control emails (½ an intervention email), 50% ineligibility, 5% are eligible but don't open emails	2.5 pp	1.2 pp
n = 40k, ½ get control emails (½ an intervention email), 5% ineligibility, 5% are eligible but don't open emails	1.25 pp	1.15 pp
n = 40k, ½ get control emails (½ an intervention email), 50% ineligibility, 5% are eligible but don't open emails	1.75 pp	0.75 pp

If we included only 20k emails in our analysis, if ineligibility were minimal (5%), and if an additional 5 percentage points of officials were eligible but didn't open their emails, then we'd expect to have 80% power to detect some ITT effect if the true effect is at least 1.75 percentage points. In that case our expected ITT effect size would be about 1.6pp. In contrast, with much greater ineligibility at 50% of the 20k emails (and still 5pp of eligible officials simply not opening their emails), we'd expect to have 80% power to detect some ITT effect if the true effect were at least 2.5 percentage points. In that case our expected ITT effect size would be about 1.2pp.

In contrast, with our larger sample size of closer to 40k emails and a 20% baseline rate, we might expect to detect some ITT effect under true effects of 1.25pp (5% ineligible) or 1.75pp (50% ineligible). Respectively, we might then expect to observe ITT effects of 1.15pp or 0.75pp under those true effects.

[6] As the baseline rate approaches 0, all else equal, we might expect less variance in the outcome and thus more efficient treatment effect estimates. Based on a separate ongoing procurement evaluation in which we performed link click tracking, we might expect a real baseline outcome rate of about 4.5%. We will factor the observed baseline click rate into any supplementary simulations we perform during analysis.

Under a given realized sample size (once our list of participating firms is finalized closer to intervention roll-out), and given an observed baseline click rate, we can use simulations to develop more realistic expectations about how much our observed ITT effects may be underestimating the true effects among eligible always-openers (for different assumed ineligibility rates). Additionally, based on desk research and information incorporated into the not_active indicator from Table 3, we may also be able to develop reasonable bounds for the ineligibility rate. From there, it may be possible to estimate the treatment effects on eligible always-openers under different assumptions about the eligibility rate (e.g., by re-weighting observations with Y=0 so that our analytical sample corresponds to the assumed eligible sample size). We will consider both approaches to aid in our interpretation of our primary confirmatory results.

Descriptive statistics, tables, and graphs

We will report the number (and percentage) of procurement officials who clicked on an emailed link or filled out the interest form. We will also break this down by whether the firm and official were matched based on a manufacturing NAICS code, whether they were matched based on a shared location, and the source of the procurement official's contact email (procurement forecasts, current SAM listings, or historic SAM listings).

This descriptive data is valuable in its own right, as the SBA does not typically conduct this type of outreach on this scale. This evaluation will help us learn to what extent procurement officials engage with marketing emails and whether it may be worthwhile for the SBA to encourage firms to conduct proactive outreach in this manner. If feasible, we will also monitor replies to the intervention emails and report the number of replies we receive.

Statistical models and hypothesis tests

Statistical models:

To test **H1**, we will estimate the following model using OLS:

$$FormClick_i = \beta_0 + \beta_1 UpdatedCS_i + \delta_k X_{ik} + \varepsilon_i$$

where $FormClick_i$ represents the primary outcome of interest, a binary indicator variable for whether email i's assigned unique link was clicked. $UpdatedCS_i$ is an indicator variable equal to 0 if the email is in an email cluster randomly assigned to receive an original capability statement or 1 if the email cluster was randomly assigned to receive an updated capability statement. β_1 is our primary parameter of interest.

X_{ik} represents a vector of k pre-intervention control variables, primarily included to increase the precision of our treatment effect estimates. These control variables will be:

- Binary indicators for the block that email i is in (a binary indicator for each block except for one, which will instead be captured by the intercept).

- The number of emails in email i's cluster, ***num_in_cluster*** in Table 3.[7]
- The number of opportunities an email is associated with, ***number_opps*** in Table 3.

We include these controls additively in the expression above. But in practice, we will include them in our OLS model by applying what we call Lin (2013) adjustment: mean-centering each control within our analytical sample (including each separate fixed effect indicator), controlling for each mean-centered variable, and also controlling for the interaction between each mean-centered variable and our intervention indicator.[8] This approach helps further guard the precision of our treatment effect estimates.

To test ***H2***, we will fit an equivalent OLS model that replaces $FormClick_i$ with $FormSubmission_i$ (a binary indicator for manually adding their email address to the firm contact form). We will estimate both OLS regression models with CR2 standard errors at the email cluster level (***all_emails*** in Table 3).

Confirmatory analyses:

We will use the model outlined above to estimate treatment effects on $FormClick_i$ (***H1***) and $FormSubmission_i$ (***H2***). The β_1 parameters from each are our primary confirmatory treatment effect estimates. However, it is important to emphasize that these will represent "Intention to Treat" (ITT) effects. We are emailing contacts identified from procurement opportunities posted over the last few years, and we know that in practice not all of these officials will still be working in procurement now (i.e., they are no longer "eligible" for our study). Moreover, even among eligible officials, not everyone will open their email, and so they will not see our intervention as intended. Because our outcome measure represents engagement with the email, anyone who is ineligible or does not open it will by construction have an outcome of 0.

Our ITT estimates will underestimate the true effect among "eligible always-openers." Instead, they will speak most directly to the impacts of sharing updated capability statements in a marketing campaign targeting *potential* federal procurement contacts. However, as no centralized database of working procurement officials is readily available (to our knowledge), any such marketing campaign in the future would encounter similar logistical problems. We discuss this further in the Limitations section.

Finally, note that we are interested in one subgroup analysis that we consider to be confirmatory: estimating ITT effects for ***H1*** and ***H2*** among emails that were matched to a firm based on a manufacturing NAICS code (i.e., that official has posted an opportunity in a manufacturing NAICS in which the firm works). This corresponds to an ITT effect for a marketing campaign focused specifically on manufacturing procurement. Roughly, we anticipate that about half of our

---

[7] Following Aronow and Middleton (2015), as a robustness check we will consider an analysis that weights emails by their cluster size instead of controlling for cluster size. See: Middleton, J. A., & Aronow, P. M. (2015). Unbiased estimation of the average treatment effect in cluster-randomized experiments. *Statistics, Politics and Policy*, 6(1-2), 39-75.

[8] Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: *Reexamining Freedman's critique*. See also Gerber and Green (2012), referenced above, and: Miratrix, L. W., Sekhon, J. S., & Yu, B. (2013). Adjusting treatment effect estimates by post-stratification in randomized experiments. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 75(2), 369-396.

analytical sample will fall into this category. See the next section for more context on how we will perform that subgroup analysis.

Exploratory analysis:

Alternative model specification

Additive adjustment for block fixed effects, though common, can sometimes raise concerns about loss of precision or bias.[9] Nonetheless, we will check how dropping Lin (2013) adjustment from our estimation strategy above changes our ITT estimates for ***H1*** and ***H2***. Assuming results resemble our primary confirmatory tests in terms of substantive effect size and statistical significance, we will adopt simpler additive control specifications for all other exploratory tests and robustness checks that follow (depending on the subgroup analysis and our final blocking strategy, we may need to drop the block fixed effects or replace them with an alternative control).

Subgroup analysis

In addition to the confirmatory subgroup analysis above, we are interested in repeating our tests of ***H1*** and ***H2*** within the following groups as exploratory analyses:

- Contact emails sourced from procurement forecasts, current SAM postings, or past (historic) SAM postings.
    - Subgroup variable is source.
- Contact emails posting procurement opportunities for the Department of Defense.
    - Subgroup variable is dep_of_defense
- Contact emails that were or weren't matched to a firm that shares their location
    - Subgroup variable is location_match.
    - We will consider whether this subgroup analysis could be informative regardless of whether this variable is used in our email/firm matching in practice. Depending on how matching occurs, we may perform this analysis for a subset of our sample that could have been matched on this variable (e.g., if we use a "greedy" procedure).
- Contact emails that were or weren't matched to a firm whose primary NAICS matches one of their opportunities.
    - Subgroup variable is primary_naics_match.
    - We will consider whether this subgroup analysis could be informative regardless of whether this variable is used in our email/firm matching in practice. Depending on how matching occurs, we may perform this analysis for a subset of our sample that could have been matched on this variable (e.g., if we use a "greedy" procedure).

We will estimate the following model for our subgroup analyses (focusing on ***H1*** for illustration):

[9] Aronow, P. M., & Samii, C. (2016). Does regression produce representative estimates of causal effects? *American Journal of Political Science*, 60(1), 250-267. Gerber, A. S., & Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*.

$$FormClick_i = \beta_0 + \beta_1 UpdatedCS_i + \beta_2 Subgroup_i + \beta_3 UpdatedCS_i * Subgroup_i + \delta_k X_{ik} + \varepsilon_i$$

In these subgroup analyses we ***will not*** employ Lin (2013) adjustment–we will use standard additive adjustment for controls, as in the subgroup analysis above. We are interested in two quantities in these subgroup analyses: first, the estimated ITT effect in each subgroup of interest (based on a linear hypothesis test of $\beta_1 + \beta_3$); and second, whether the ITT effect in this subgroup differs to a statistically significant degree (based on β_3).

Additional robustness checks

As discussed in footnotes above, we will check the robustness of our primary, confirmatory tests to (1) weighting based on email cluster size instead of controlling for it. We will also consider (2) estimating the bias we might expect due to measurement error in our primary outcomes (based on realistic assumptions about the false negative rate attributable to link tracking errors).

Inference criteria, including any adjustments for multiple comparisons:

We will classify findings as statistically significant if they yield estimated two-tailed p-values of less than 0.05 (i.e., we evaluate statistical significance at a standard 95% confidence level). If any of our confirmatory ITT estimates are statistically insignificant, we will use an equivalence test to determine whether we can support a claim of "no meaningful ITT effect" with 95% confidence.[10] For these equivalence tests, we tentatively define a region of -0.3 percentage points to +0.3 percentage points as representing a degree of change that is not meaningfully different from 0 (our "equivalence region").

Our primary confirmatory tests of ***H1*** and ***H2*** are related in the sense that, if we see supportive results for either outcome, we would conclude that sharing the updated capability statement increases engagement with the marketing email. For this reason, we need to adjust for multiple testing when drawing conclusions across outcomes.[11] However, results for these tests are likely to be correlated, and so standard corrections like Bonferroni adjustment may be overly conservative (i.e., penalizing multiple testing more than is necessary).

To address those concerns, when drawing conclusions across outcomes, we plan to perform an omnibus test using a stacked regression.[12] This requires creating a new dataset where each observation in our original dataset has two rows, one for each outcome (with an additional binary indicator for a row's outcome measure). We can then fit a modified version of our primary confirmatory model, interacting each right hand side term from the original model with the binary outcome indicator (CR2 errors still at the email cluster level). A linear hypothesis test for the joint significance of treatment effects across each outcome will yield a single p-value (side-stepping multiple testing concerns). The stacked regression model should yield the same ITT estimates for

[10] Rainey, C. (2014). Arguing for a negligible effect. *American Journal of Political Science*, 58(4), 1083-1091.
[11] Rubin, M. (2024). Inconsistent multiple testing corrections: The fallacy of using family-based error rates to make inferences about individual hypotheses. *Methods in Psychology*, 100140.
[12] Oberfichtner, M., & Tauchmann, H. (2021). Stacked linear regression analysis to facilitate testing of hypotheses across OLS regressions. *The Stata Journal*, 21(2), 411-429. Equivalently, we could use a "seemingly unrelated estimation" (SUE) approach, which that study discusses as well. SUE also allows for a joint test of intervention effects across outcomes in two different regressions. If we encounter any unexpected issues in implementing this procedure, we will default to using [randomization-inference based simulations](#) to account for multiple testing concerns.

each outcome as the separate models.[13] The primary difference is simply that it allows for testing for an overall ITT effect across OLS regressions on both outcomes at once.

Multiple testing adjustment may be necessary when drawing conclusions across outcomes (either within our overall sample, or within our confirmatory manufacturing subgroup analysis). But we do not believe it is necessary to adjust for multiple testing across these different groups of observations (the overall sample vs. the manufacturing sample). It is possible that we would see supportive results in one sample but not another. Moreover, while we are interested in each sample independently, we do not have any particular conclusions we plan to draw based on the patterns we see across both.

Appendices

Appendix A. Email cleaning procedure

1. Load the Opportunity Data for all emails that will tentatively be included in our sample. Do some initial basic email cleaning, yielding ***cleaned email step 1***.
    a. Ensure that emails from the opportunity listings are not blank, or "TBD," etc.
    b. Using a regular expression (regex) command, verify that they are structured like emails: "[text]@[text].[text]".
2. Among step 1 cleaned emails, identify cases where the same email "base" (the portion before "@") has different domain names (the portion after "@"). Only do this among opportunity postings from the same agency.
    a. Review unique combinations of domains that this yields manually, and correct cases where this is due to obvious typos.
        i. E.g., ".gvo" should clearly be ".gov".
        ii. Apply similar corrections to all emails in the sample.
    b. Check that remaining combinations of domains appear plausible.
        i. E.g., the US Army has used the domain first.last@mail.mil in the past, but is migrating to first.last@army.mil. However, not everyone has fully switched over. Similarly, officials may often procure for both the Navy and DLA (on behalf of the Navy), or both the Air Force and Space Force.
    c. Within each matching-base-in-agency combination, randomly choose one of the emails to use as ***cleaned email step 2***.
3. Among step 2 cleaned emails, define a maximum string difference between two email "bases" before we can no longer consider them potential matches. Identify all "bases" with the same domain (and agency) that are within this cutoff from each other.

[13] If the outcome indicator =1 for ***form_submission*** rows, then the coefficient on the un-interacted ***UpdatedCS*** indicator represents the ITT effect on clicks, and that plus the interaction term represents the ITT effect on form submissions.

a. Specifically, we calculate pairwise Levenshtein distances between all email bases (name part before "@") from the same domain/agency.

b. We set a maximum string distance of 2. I.e., any two bases that differ from each other by 1 or 2 edits are flagged as potential matches. In an initial test, this yielded about 1000 potential match pairs (out of ~20,000 emails).

c. To be overly conservative, we assume all of these pairs must be potentially treated as the same person. Within each "person," we choose whichever email has the most opportunities as the "real" one. This yields ***cleaned email step 3***.

4. At the end of this process, we can use the step 3 cleaned email to identify unique contacts (i.e., "clusters") for the purposes of treatment assignment.

    a. In practice, in a preliminary test, this procedure yielded a sample size of email clusters that was only about 2,000 smaller than our original test sample.

    b. The assumption made in step 3c is incorrect, but it is incorrect in a way that only decreases the number of email clusters over which we can assign treatment (and so shrinks our "effective sample size" for the purposes of treatment effect estimation more than is needed). To avoid sending the same person treatment and control emails as much as possible, we prefer to err towards making it harder for us to draw firm conclusions from this analysis.

5. We preserve the original emails as well (those that remain after step 1). We will email all of them in practice, and all of them will be included as separate observations in our analysis (with appropriate adjustments to account for email clustering).

    a. We will clean these emails again separately before the intervention begins (applying corrections similar to those in 2a in cases of obvious typos, like ".gvo" instead of ".gov").

Appendix B. Matching procedure

1. For each email cluster, identify all NAICS codes in which any constituent email has posted an opportunity. Identify all participating firms that work in at least one of those NAICS codes. These are the candidate firms for this email cluster.

    a. In a preliminary test of this procedure, only approximately 6% of emails had a single candidate, making it necessary for us to choose one in most cases.

2. Divide observations into match groups.

    a. As our list of participating firms is still tentative, this process is not finalized. We need the match groups to remain large (as they will be incorporated into our blocking strategy), but also incorporate additional information that could boost email engagement, where possible.

    b. We will prioritize the following when creating match groups:

    i. First, matching emails to a firm that works in a manufacturing NAICS for which that email has posted at least one contract opportunity.
    ii. Second, matching emails to a firm that is located in a state that matches the place of performance of at least one of that email's opportunities.
    iii. Third, matching emails to a firm that has listed one of that email's NAICS codes (across posted opportunities) as its "primary" NAICS.
  c. If possible, we will create match groups as mutually exclusive categories. If not, we will apply a "greedy" grouping strategy (match on manufacturing first, then match remaining emails on location, etc.).
    i. In our preliminary test, there was little overlap between, e.g., primary NAICS and location matches, suggesting a greedy strategy is more realistic.
  d. The match group chosen determines which of an email's remaining candidates are excluded or retained as an option.
    i. E.g., if they are in the manufacturing group, all candidates that don't share a manufacturing NAICS with the email are dropped
3. Randomly choose a firm among each email's remaining candidates (which is determined based on their match group).
  a. If an email cannot be matched based on a manufacturing NAICS, location, or primary NAICS, their firm will just be chosen randomly among all of their candidates

Appendix C. Expanding on the ineligibility issue

A non-standard issue we have to confront when thinking about bias and power for this research design is ***sample ineligibility***: not all of the contact emails we have gathered will correspond to officials who are still working in procurement at the time of our intervention. Ineligibility may vary across the sources we use to identify contact emails: contacts listed in procurement forecasts are most likely to still be eligible for our study, while contacts listed in past SAM solicitations are least likely to still be working in procurement. We cannot guarantee that we will know who all of these ineligible officials are. It is possible that we can identify some of them: (1) some emails will bounce back if the account is officially closed, (2) we can find information online about offices that have been closed, and (3) we may be able to employ tools like "read receipts."

Given that all of our treatment arms, including the control group, are receiving an email, we can think of ineligibility as one source of ***noncompliance***: officials we intended to open an email and see the treatment/control may not in practice. Another more typical source of noncompliance is people who are eligible and see the email but simply ignore it (***never-openers*** instead of ***always-openers***). We plan to use the same subject line for all intervention emails, which rules out the possibility of "compliers" or "defiers" who open conditional on treatment status.

Anyone who is ineligible or a never-opener will receive Y=0 for our outcome measure: Y=1 is only possible for eligible always-openers. In other words, Y=0 has several possible interpretations for procurement officials who we are not able to rule out using (1)-(3) above: it may represent an official choosing not to engage with the email, or it may represent an official being a never-opener, or it may represent the official no longer working in federal procurement. Both ineligibility and the presence of never-openers will lead to downward bias in our ability to estimate an average treatment effect among eligible always-openers, and will also decrease our statistical power.

To see why, imagine first (***case 1***) that we had perfect read receipts for every intervention email, and so we classified Y=NA (missing) for every official who did not open it. By virtue of having the same subject line across arms, being an always-opener is a pre-intervention characteristic (we can control for it or subset data for analysis based on it). And by virtue of random assignment, the rate of always openers should be the same on expectation across arms. With perfect knowledge of which officials are still working and which opened the email, we could simply restrict our analysis to the officials who opened their email and estimate an unbiased average treatment effect among always-openers (with statistical power then based on the number of always-openers instead of our larger sample size).

As an example, assume that we sent 1000 emails across only two arms (500 each), with half of our sample (500) opening the email (eligible and not never-openers). Assume also that we know who all of them are (assigning Y=NA in these cases). If 10% of those who opened the email engaged with the email in the control group (25 of 250), while 16% did in the treatment group (40 of 250), then that yields an estimated effect of 6pp among those with non-NA outcome values.

Instead, imagine now (***case 2***) that all the NAs are replaced with 0: we have no read receipts for any official, and only know that they didn't engage with the email, but not why. We'd then be adding 250 observations with Y=0 to both the treatment and control group. This means the control group now has 25 of 500 (5%) engaging with the email, vs 40 of 500 (8%) in the treatment group, yielding an estimated effect size of 3pp (rather than the true effect size of 6pp among always-openers). We are not restricting our sample now, but we still lose statistical power, this time because the observed effect within the sample is lower.[14]

This evaluation will likely fall somewhere in between ***case 1*** and ***case 2*** (knowing who some ineligible officials or never-openers are, but not all). However, as discussed in the Limitations section, it may be feasible to partially identify the effect among eligible always-openers under a range of realistic assumptions about the eligibility rate. If possible, we will perform such an analysis as an exploratory check on our findings.

[14] This is similar to the reason why, in randomized trials with more typical kinds of non-compliance, power for estimating ITT effects and LATEs is often similar. The ITT estimate is downwardly biased, which decreases power, but the standard error of the LATE is higher. Power tends to net out about the same.