

# Exploratory Analysis of Contact Strategies for Annual Economic Surveys

---

Prepared for:

Stephen Cox, Melissa Cidade, Krysten Mesner, James Burton, Shelley Karlsson, Jason Bauer, and Lisa Donaldson.

Prepared by:

Katharyn Loweth, Coding it Forward Fellow

Hannah Wen, Coding it Forward Fellow

Economic Programs Directorate

U.S. Census Bureau

August 16, 2024

The authors of this report would like to acknowledge colleagues and mentors in the Economy-Wide Statistics Division (EWD) for their contributions and input on this project.

*Any opinions and conclusions expressed herein are those of the author(s) and do not reflect the views of the U.S. Census Bureau. The Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data (Project No. P-7529180, Disclosure Review Board (DRB) approval number: CBDRB-FY24-EWD001-006).*

## Contents

<b>Executive Summary.....</b>	<b>3</b>
<b>Review of the Communication Strategies for Annual Economic Surveys .....</b>	<b>4</b>
<i>Overview .....</i>	<i>4</i>
<i>Data &amp; Data Processing.....</i>	<i>4</i>
<i>Surveys and Data Sources.....</i>	<i>4</i>
<i>Creating a Relational Database from Collection Strategies Documentation .....</i>	<i>5</i>
<i>Respondent Engagement Metrics from Economic Datasets .....</i>	<i>9</i>
<i>Trends in Communication Outreach Efforts .....</i>	<i>10</i>
<i>Changes in Max Communication Attempts .....</i>	<i>10</i>
<i>Communication Outreach Efforts – Day of the Week Analysis.....</i>	<i>13</i>
<i>Communication Outreach Efforts and Respondent Engagement.....</i>	<i>15</i>
<i>Burn Rate and Check in Rate Over the Years.....</i>	<i>15</i>
<i>Burn Rate by Due Date and Burn Rate by Closeout Over the Years .....</i>	<i>16</i>
<i>Authentication Code Use in Relation to Survey Due Date .....</i>	<i>17</i>
<i>Communication and Burn Rate.....</i>	<i>19</i>
<i>Communication and Engagement Window.....</i>	<i>21</i>
<i>Due Date Reminder Timing.....</i>	<i>22</i>
<b>Limitations and Recommendations.....</b>	<b>24</b>
<b>Conclusion .....</b>	<b>25</b>

## Executive Summary

Communication strategies are central components of survey production. The emails, postal mailings, and phone calls of a communication strategy inform and remind potential survey respondents about the importance of survey completion. While these communications overall spur survey engagement, there is varying empirical evidence on the impact of these different communication methods and their particularities, such as sequencing and timing, on respondent engagement.

As part of the Coding it Forward 2024 Fellowship program, the Economic Directorate of the U.S. Census Bureau sponsored an investigation into the contact strategies of several annual legacy utilizing data from survey years 2015-2022. The goal of the analysis was to identify trends in communication strategies over time and to better understand the relationship between contact efforts and survey engagement. The recommendations from this study could then inform the future communication efforts of business surveys such as the Annual Integrated Economic Survey.

This summary report provides an overview of the communication strategy analysis performed as part of the Fellowship program. First, we discuss the data cleaning and transformation activities completed to utilize information from the communication strategies documentation as part of quantitative analysis. We then discuss the historical trends in communication strategies over time and the contact methods' different engagement windows. Based on our analysis, we identify limitations in the available data sources and provide recommendations for best practices in data handling and planning future communication strategies to better discern the causal impact of different contact methods on respondent engagement.

**Research Objectives:** The aim of this report is to analyze trends of annual economic legacy surveys that now comprise the Annual Integrated Economic Survey (AIES). It supports AIES by providing insights and recommendations on survey communication in relation to respondent engagement. However, these findings also are meant to aid the development for communication strategies for establishment surveys in general.

**Methodology:** For surveys ACES, ARTS, AWTS, SAS, and COS/ASM, information on the communication methods was collected from the collection strategies documents tracing back to survey year 2015. The research team inputted data into a compendium of Excel spreadsheets that would later act as a relational database. The team also assigned unique IDs to each communication point that were representative of certain characteristics, such as sequence. The outcome variable of interest is the authentication code burn date, which indicates when a respondent uses the code given to them to initiate the survey. This data was pulled from the U.S. Census Bureau's Standardized Economic Processing System (StEPS) and the Business Register.

# Review of the Communication Strategies for Annual Economic Surveys

## Overview

Communication strategies are central components of survey production. The emails, postal mailings, and phone calls of a communication strategy inform and remind potential survey respondents about the importance of survey completion. While these communications overall spur survey engagement, there is varying empirical evidence on the impact of these different communication methods and their particularities, such as sequencing and timing, on respondent engagement.

The goal of this research is, in part, to track trends over time for the seven in-scope surveys. Based on the available data, we addressed the following exploratory questions:

- What is the rate of survey increase when mailings/emails are sent?
- What is the shelf life of engagement after mailings/emails are sent?
- Is there a sequence or timing of communication in relation to the survey due date that produces the highest proportion of authentication codes used?

Due to some limitations in the available data, we were unable to create a statistical model that captured the relationship between contact efforts and survey engagement at the business-level. From this research, we hope to provide recommendations for survey communications that will help others at the Census Bureau continue to explore the relationship between contact efforts and survey response.

## Data & Data Processing

### Surveys and Data Sources

The surveys included in this study are the seven legacy economic surveys that have now been combined into the Annual Integrated Economic Survey (AIES). These surveys are no longer in production.

#### **Annual Economic Surveys Initially Included in the Study:**

- SAS (Service Annual Survey)
- ARTS (Annual Retail Trade Survey)
- AWTS (Annual Wholesale Trade Survey)
- M3UFO\* (Manufacturers' Unfulfilled Orders Survey)
- ACES (Annual Capital Expenditures Survey)
- COS/ASM (Report of Organization/Annual Survey of Manufactures)

Because of the lack of information on M3UFO's communication strategies, M3UFO was not included in most of the analysis described in this report. Additionally, because of how the data was stored, COS/ASM are treated as one combined survey throughout this analysis.

The three key data sources included in the dataset are:

- 1. Collection Strategies Documentation:** From the collection strategies documentation, we were able to gather data on each communication used, such as date, method, and communication type, as part of the collection period.
- 2. Standardized Economic Processing System (StEPS):** From StEPS we pulled data on SAS, ARTS, AWTs, and ACES respondents' authentication code usage, specifically the date that the respondent used their authentication code to access the survey for the first time ("burning their code").
- 3. Business Register:** From the Business Register we pulled data on COS/ASM respondents' authentication code usage.

**Limitation:** Based on current stored data on the historical surveys, we do not have record of which businesses received each contact effort. Further, our measure of respondent engagement ("burning a code") only indicates when the business first accessed their survey. We do not have information on when/if completed the survey.

### Creating a Relational Database from Collection Strategies Documentation

We created a relational database framework to store the data from the collection strategies documentation. We took this approach because it would better support our subsequent quantitative analysis into the impact of different communication methods.

Our relational database consisted of the following tables:

- Survey
- Communication
- Email
- Mail
- Call
- Robocall

The **survey table** reports the basic information of each survey rollout for each year. Each observation, which represents a particular survey year, is given a unique ID (ex. ACES\_2017) which is recorded as the *survey ID* in other tables to keep track of communication methods.

- The *title* and *year* columns track the survey name and year.
- *open\_date*, *due\_date*, *soft\_close*, and *due\_date* track the important dates for each survey. Not every survey had a soft close, resulting in N/A values in the *soft\_close* column. For all columns, the date is formatted MM/DD/YYYY.

- *n\_participants* is the number of participants listed in the collection strategies documentation for each survey.
- *n\_subgroups* keeps track of the number of different groups of people who got either different types of emails/letters or who got emails/letters at different times.
- *feb\_birth*, *may\_birth*, *august\_birth* are binary variables which track whether certain birth groups were used for different mailout timelines.
- *email\_used*, *uaa\_used* and *uer\_used* are binary variables that track whether email communication was used for the survey, and whether Undeliverable as addressed (UAA) and Unexpired extension reminder (UER) messages were used as part of the communication strategy.
- *final\_br* is the burn rate (% of survey population who used their authentication code to access the survey) at survey closeout listed in the communication strategies document. This is the expected outcome variable for survey-level analysis.
- *early\_br* is the burn rate (% of survey population who used their authentication code to access the survey) by the due date listed in the collection strategies document and can be found in the Survey table.
- *final\_ci* is the final survey completion rate (% of survey population who completed the survey) listed in the communication strategies document.

The **communication table** logs the different forms of contact for each survey. Each observation is given a unique ID that represents a distinct moment of communication listed in the communication strategies document. The communication table ID is recorded as the *com\_ID* in other tables.

- *ID* is an 8-character ID created using the following logic:
  - characters 1-2 relate to the survey
  - characters 3-4 relate to the survey year
  - character 5 is a hyphen
  - character 6-7 indicate the sequence of the communication date in the communication strategies (01 = first communication date; 05 = fifth communication date). This does not represent a unique count of communication. We chose this approach because our outcome variable of interest is date-based and we do not know which communications businesses received.
  - character 8 indicates if more than 1 communication was launched on date, count of communication that occurs on that day. If observation is the only communication on listed day the value equals 0.
  - For example, AW15\_010 represents the sole communication that occurred on the first date of communication for 2015 AWTS.
- *date* column tracks the date communication was sent to applicable respondents. Date is in MM/DD/YYYY format.
- *survey ID* is pulled from the **survey table** and identifies the survey the communication is affiliated with
- *communication category* (*com\_cat*) is a 2-character code that indicates the sequence of the communication based on what is specifically indicated. For example, even if the second follow-up for email and mail occur at very different times in the sequence, they are both categorized as F2 because that is how the communication strategy labeled each communication. This helps us better understand the sequencing of communication across survey period.
  - Values include: IN=initial invitation message, DD = Due date reminder (DDR) message, F1 = first follow-up (FU) message, F2= second FU, F3 = third FU, F4 = fourth FU, F5 = fifth FU, F6 = sixth FU, F7 = seventh FU, F8 = eighth FU message, CL = closeout message, T1 =

first call FU, T2 = 2nd call FU, T3 = third call FU, T4 = fourth call FU, E1 = first Unexpired Extension Requests (UER) message, A1 = first undeliverable as addressed (UAA) message, E2 = second UER, A2 = second UAA, R1 = first robocall, R2 = second robocall, R3 = third robocall, and R4 = fourth robocall.

- *communication type* (com\_type) indicates whether the communication was a mailing, email, call, or robocall.
- *message type* is a broad categorization for the messaging based on how the communication is labeled in the communication strategy.
  - Value include: UAA = Undeliverable as Addressed, FU = follow up, initial = initial invitation, DDR = due date reminder, UER = Unexpired Extension Requests.
- *days past mailout* represents the number of days past the initial date a communication method is sent. The initial date was collected from the collection strategies document and can be found in the Survey table.
- *days\_last\_contact\_method* represents the number of days since the last communication method of the same communication type (com\_type) in the collection period for the survey year.
- *days\_last\_contact* represents the number of days since the last communication method in the collection period for the survey year.

The *com\_ID* is used in the **email, mail, call, and robocall tables** which log details of the specific communication respondents receive.

The **email table** keeps track of when and how many emails are sent out.

- *ID* is the unique 8-character ID used to track the emails in this table. Similar to com\_ID, it is created using the following logic:
  - characters 1-2 relate to the survey
  - characters 3-4 relate to the survey year
  - character 5 is a hyphen
  - character 6 is an "E" to signify this is an email communication
  - character 7-8 represent the count of the email communication for the survey year
  - For example, AC17\_E01 represents the first email sent during the 2017 ACES collection period.
- *survey\_ID* is the ID pulled from the **survey table** which links the observation to a certain survey year.
- *Com\_ID* is the ID pulled from the **communication table** which links the observation to a certain communication effort.
- *Send\_Date* is the date the email was sent (in YYYY-MM-DD format).
- *Day* is the day of the week calculated based on the send date (in numerical form).
- *n\_sent* is the number of emails sent to respondents.

The **mail table** keeps track of how and when the letters are sent out.

- *ID* is the unique 8-character ID used to track the mailings in this table. Similar to com\_ID, it is created using the following logic:
  - characters 1-2 relate to the survey
  - characters 3-4 relate to the survey year
  - character 5 is a hyphen
  - character 6 is an "M" to signify this is a mail communication
  - character 7-8 represent the count of the mail communication for the survey year

- For example, AC15\_M01 represents the first mailing sent during the 2015 ACES collection period.
- *survey\_ID* is the ID pulled from the **survey table** which links the observation to a certain survey year.
- *com\_ID* is the ID pulled from the **communication table** which links the observation to a certain communication effort.
- *mail\_class* represents the class of mail (FCR = first class regular, FCP = first class presort, PCR = priority class regular, PCP = priority class presort, S = standard) .
- *env\_type* indicates the type of envelope used to send the letter (R = regular, PS = pressure sealed).
- Binary variables *certified*, *letter*, *flyer*, and *supplemental* letter indicate whether the envelope was certified and the contents of the envelope.
- *n\_sent* represents the number of the communication that was sent out based on the actual numbers included in the communication strategies documents.

The **robocall table** keeps track of when and how many robocalls are used and the number of attempts. Notably, only COS/ASM, SAS, and ARTS utilize robocalls as part of their communication strategy.

- *ID* is the unique 8-character ID used to track the robocalls in this table. Similar to *com\_ID*, it is created using the following logic:
  - characters 1-2 relate to the survey
  - characters 3-4 relate to the survey year
  - character 5 is a hyphen
  - character 6 is an “R” to signify this is a robocall communication
  - character 7-8 represent the count of the robocall communication for the survey year
  - For example, CO21\_R01 represents the first robocall attempt during the 2021 COS/ASM collection period.
- *survey\_ID* is the ID pulled from the **survey table** which links the observation to a certain survey year.
- *com\_ID* is the ID pulled from the **communication table** which links the observation to a certain communication effort.
- *start\_date* is the start date of the robocall round.
- *end\_date* is the end date of the robocall round.
- *n\_calls* represents the number of cases in the robocall round recorded.
- *cases\_est* is a binary variable which records whether *n\_calls* is an estimated number or not.

The **call table** keeps track of when and how many live-agent telephone calls are used and the number of attempts.

- *ID* is the unique 8-character ID used to track the robocalls in this table. Compared to the other tables, the ID represents a call round rather than a specific call because this is how the method was described in the communication strategy documentation and each call round may have more than 1 call attempt to reach the respondent.
  - characters 1-2 relate to the survey
  - characters 3-4 relate to the survey year
  - character 5 is a hyphen
  - character 6-7 is an “TP” to signify this is a telephone phase.
  - character 8 represent the count of the telephone phase for the survey year.



- For example, AC15\_TP1 represents the first telephone phase during the 2015 ACES collection period.
- *com\_ID\_1* and *com\_ID\_2* represent the communication IDs which mark the beginning and ending of the telephone follow up call round, respectively. In the **communication table**, an observation would be made for the beginning and end dates of the call round since it is a continuous contact method.
- *survey\_ID* is the ID pulled from the **survey table** which links the observation to a certain survey year.
- *start\_date* is the start date of the telephone round recorded.
- *end\_date* is the end date of the telephone round recorded.
- *n\_calls* represents the number of calls in the telephone follow up round.
- *cases\_est* is a binary variable which records whether *n\_calls* is an estimated number or not.
- *n\_attempts* represents the number of call attempts in the telephone follow up round if it was provided.
- *max\_calls* represents the maximum number of call attempts each case was allowed if it was provided.

**Limitation:** Because we did not have access to call data for all surveys and periods, we ultimately did not analyze calls or robocalls as part of our analysis. Additionally, due to the lack of variation in the group of variables that describe mailing characteristics, we did not include these variables in our analysis.

### Respondent Engagement Metrics from Economic Datasets

As stated previously, we used data and variables pulled from the StEPS and the Business Register to better understand survey engagement. For the data that came from StEPS, the data was divided into different files by survey. Each datafile included information on all businesses that were a part of 2017-2022 collection periods. The key variables that are part of this analysis are described below:

- *STATP00* represents the statistical period the record is a part of.
- *\_ASTAT00* represents the status of the authentication code. U = used, E = expired, I = initial, and D = disabled.
- *AUTH\_USED* was a variable created by Stephen Cox that represented whether a code was used (U) or issued (I). *AUTH\_USED* = U could represent a code either being burned or expired. To determine if the code was burned, *AUTH\_USED* = U and *AUTHDATE\_BANK* is not null.
- *CKNDTE00* is the recorded date in the StEPS system when the code disposition changed. The date is formatted YYYY-MM-DD.
- *AUTHDATE\_BANK* is the data on record in the authentication code system when the code was used/disposition changed. This date should be similar to the *CKNDTE00* but there sometimes may be discrepancies based on when information was pulled. The date is formatted YYYY-MM-DD.

The data that came from the Business Register is formatted differently; compared to the StEPS dataset where only the final disposition of authentication code is recorded, in this dataset the history of the

authentication code usage is maintained across multiple records. For example, there is an observation that indicates the date an authentication code was issued and a second observation that indicates when the authentication code was used.

- *STATUS* represents the status of the authentication code. USED indicates the code has been burned, ISSUED indicates the code has been issued, and DISABLED indicates the code has been disabled and no longer in use.
- *DATE* is the data on record in the authentication code system when the code was used/disposition changed. The date is formatted YYYY-MM-DD.
- *YEAR* represents the statistical period/survey year, similar to STATP00 in the other dataset.

## Trends in Communication Outreach Efforts

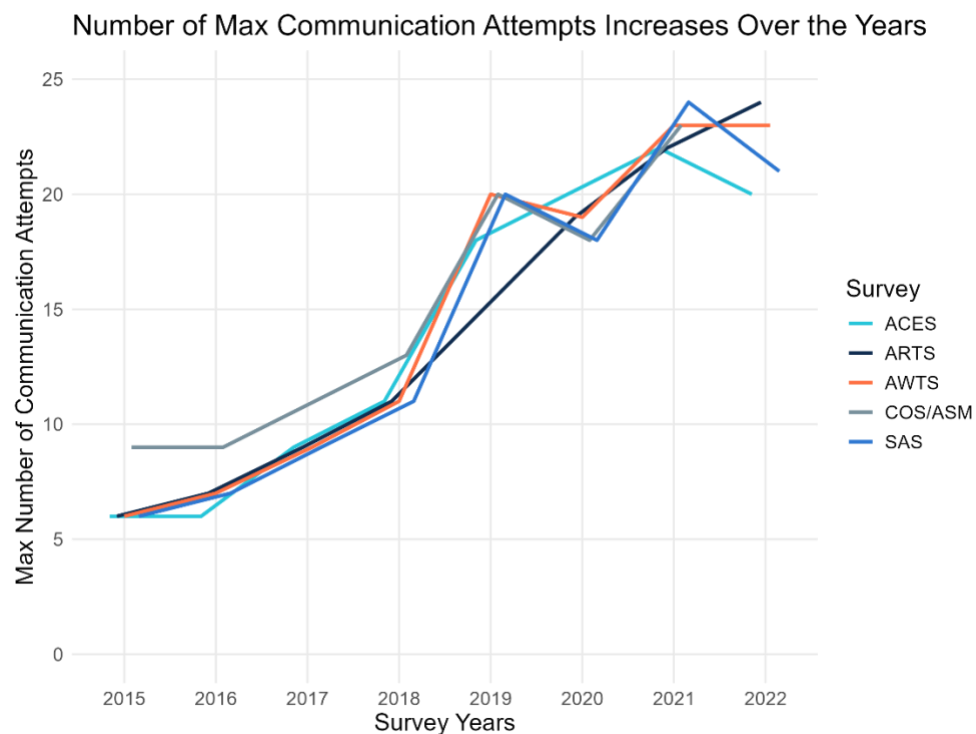
After creating the relational database, we reviewed the communication strategies in a few ways. First, we studied the changes in communication strategies over time. We investigated changes in the overall counts of communication attempts and for specific types of messages. Additionally, we looked at when these messages are sent, and how that may differ based on the communication method.

The maximum number of communication attempts refers to the maximum number of communications a respondent could have received. It was calculated by counting the number of unique com\_cat IDs and message types in the communication table for each survey year to differentiate contact points of different purposes and modes of communication.

## Changes in Max Communication Attempts

Across all surveys, there has been a significant increase in the number of communication attempts and methods used to reach survey respondents. As seen in Figure 1, the communication attempts have steadily increased since survey year 2015. There was also a sharp increase in max communication attempts in survey year 2019, which is likely a result of the Covid-19 pandemic.

Figure 1: Number of Maximum Communication Attempts for Survey Years 2015-2022



**Finding:** The maximum number of communication attempts triples over a 7-year period from survey years 2015-2022. In survey year 2015, most surveys had an average of 6 maximum communication attempts (COS/ASM had a maximum number of 9). In survey year 2022, all surveys had between 20 and 24 maximum communication attempts.

**Limitation:** This maximum number is based on if they received all the distinct communications listed in the communication strategies documentation. We are unable to determine the actual or average number of communications that respondents received.

**Recommendation:** None.

Table 1 breaks down how the Maximum Communication Attempts have changed by message type since survey year 2015. Between surveys years 2015 and 2022, there has been an increase in initial, due date reminder, and follow-up messaging and the introduction of new message types undeliverable as addressed (UAA), unexpired extension request (UER) and closeout messages.

Across the message types, the number of potential follow-up messages have increased the most both in the methods used and rounds of the communication. The initial and due date reminder messaging increase from 1 to 2 is a result of the use of both mailing and email contact methods. An initial and due date reminder mail/email is only sent one time. The UAA and UER messaging, which are only delivered through one contact method (mail and email, respectively), may be sent at two different times. In contrast, follow-up communication may consist of mailings, emails, telephone calls, and robocalls, and

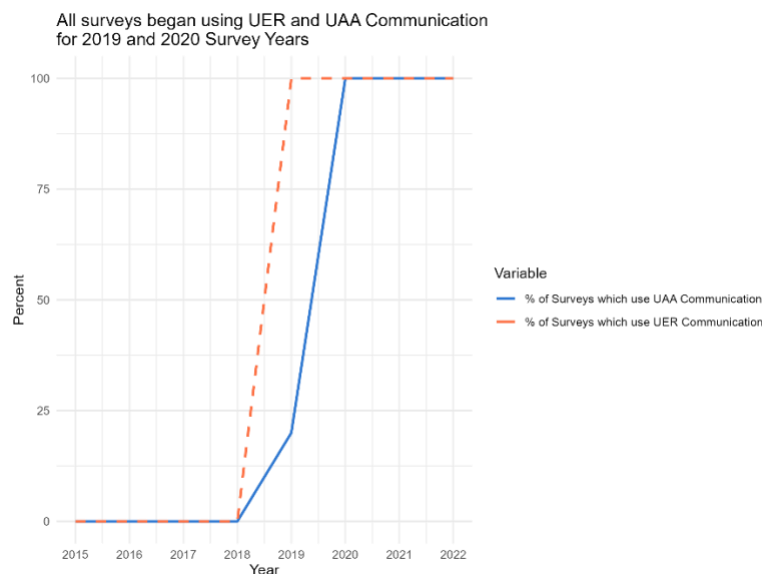
each method may be used multiple times. In more recent survey years, as many as 8 follow-up emails may be sent to a potential respondent.

Table 1: Maximum Number of Contact Attempts by Message Type for survey years 2015 and 2022/2021.

<b>Maximum Number of Contact Attempts, by Message Type</b>										
	<b><u>SY 2015</u></b>					<b><u>SY 2021/2022*</u></b>				
Message Type	<b>Count</b>					<b>Count</b>				
	<b>ACES</b>	<b>ARTS</b>	<b>AWTS</b>	<b>SAS</b>	<b>COS/ASM</b>	<b>ACES</b>	<b>ARTS</b>	<b>AWTS</b>	<b>SAS</b>	<b>COS/ASM</b>
Initial	1	1	1	1	1	2	2	2	2	2
Due Date Reminder	1	1	1	1	1	2	2	2	2	2
Follow-up	4	4	4	4	7	11	15	14	12	16
Unavailable as Addressed	0	0	0	0	0	2	2	2	2	0
Unexpired Extension Request	0	0	0	0	0	2	2	2	2	2
Closeout	0	0	0	0	0	1	1	1	1	0
Total	6	6	6	6	9	20	24	23	21	22

While each survey had their own production timelines and strategies, as seen in Figure 2 and Figure 3 below, there are similarities in when the surveys implemented significant changes in their communication strategies.

Figure 2: Percentage of Surveys using Unexpired Extension Request (UER) and Undeliverable As Addressed (UAA) Specific Communication Over the Years

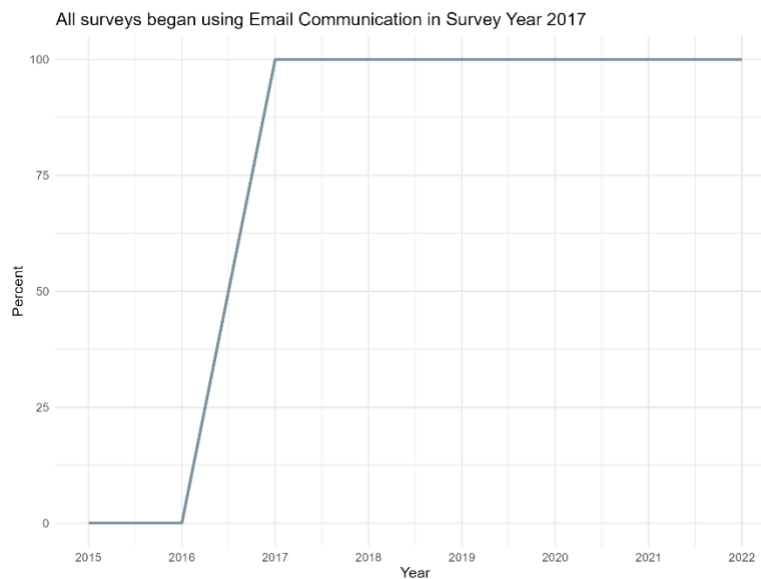


**Finding:** All surveys start using UER-specific communication during the 2019 collection period and UAA-specific communication during the 2020 collection period.

UAA-specific communication is explicitly sent through mail and UER-specific communication is explicitly sent through email. UER communication began one year after email was introduced to contact methods in survey year 2017.

**Recommendation:** None.

*Figure 3: Percentage of Surveys using Email Communication Over the Years*



**Finding:** All surveys start using emails as a communication method in the 2017 collection period.

**Recommendation:** None.

### Communication Outreach Efforts – Day of the Week Analysis

As part of the analysis, we reviewed the weekdays that mailings and emails were sent to respondents. Compared to the distinct counts in Table 1, this analysis includes all mail and email observations.

#### *Email.*

As seen in the Table 2 below, survey email communications were most frequently sent on Tuesdays. Across all surveys and collection periods, forty percent of the communications were sent on Tuesday.

More than eighty percent of the email communication was sent between Tuesday and Thursday, and few messages were sent at the beginning and end of the work week.

Reviewing the different message types, nearly all the message types were most sent on Tuesdays. However, initial messages were overwhelmingly sent on Thursdays. This reflects the strategic timing of certain communication.

Table 2: Count of Email Communications by Message Type and Day of Week

<b>Count of Email Communications by Message Type and Day of Week (total count=229)*</b>										
<b>Message Type</b>	<b>Monday</b>		<b>Tuesday</b>		<b>Wednesday</b>		<b>Thursday</b>		<b>Friday</b>	
	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>
Initial	2	11.8%	1	5.9%	1	5.9%	<b>13</b>	<b>76.4%</b>	0	0%
Due Date Reminder	3	9.7%	<b>17</b>	<b>54.8%</b>	5	16.1%	3	9.7%	3	9.7%
Follow-up	21	14.7%	<b>61</b>	<b>42.7%</b>	28	19.5%	32	22.4%	1	0.7%
Unexpired Extension Request (UER)	2	5.8%	<b>11</b>	<b>32.4%</b>	9	26.5%	<b>11</b>	<b>32.4%</b>	1	2.9%
Closeout	0	0%	<b>2</b>	<b>50%</b>	<b>2</b>	<b>50%</b>	0	0%	0	0%
Total	28	12.2%	<b>92</b>	<b>40.2%</b>	45	19.6%	59	25.8%	5	2.2%

\*Unavailable as Addressed messages were only sent by mail. Therefore, it is not included in the above table.

\*Bolded numbers indicate the highest number/percentage for that message type. Percentage calculated using row totals.

## Mail.

In contrast to the email communication, mail communication was most frequently sent on Thursdays. As seen in Table 3, forty-four percent of all mailing were sent on Thursday. Thursday is also the most popular day for all message types. The next most common days are Monday and Tuesday. In contrast to email communication, less than ten percent of the mailings are sent out on Wednesday.

Table 3: Count of Mail Communications by Message Type and Day of Week

<b>Count of Mail Communications by Message Type and Day of Week (total count = 326)*</b>										
<b>Message Type</b>	<b>Monday</b>		<b>Tuesday</b>		<b>Wednesday</b>		<b>Thursday</b>		<b>Friday</b>	
	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>	<b>N</b>	<b>%</b>
Initial	6	6.6%	14	15.6%	8	8.9%	<b>51</b>	<b>56.7%</b>	11	12.2%
Due Date Reminder	3	5.7%	12	22.5%	2	3.8%	<b>32</b>	<b>60.4%</b>	4	7.6%
Follow-up	34	21.9%	26	16.8%	15	9.7%	<b>54</b>	<b>34.8%</b>	26	16.8%

Unavailable as Addressed (UAA)	8	28.6%	6	21.4%	5	17.9%	<b>9</b>	<b>32.1%</b>	0	0%
Total	51	15.6%	58	17.8%	30	9.2%	<b>146</b>	<b>44.8%</b>	41	12.6%

\*Unexpired Extension Request and Closeout messages were only sent via email. Therefore, they are not represented in this table.

\*Bolded numbers indicate the highest number/proportion for that message type. Percentage calculated using row totals.

## Communication Outreach Efforts and Respondent Engagement

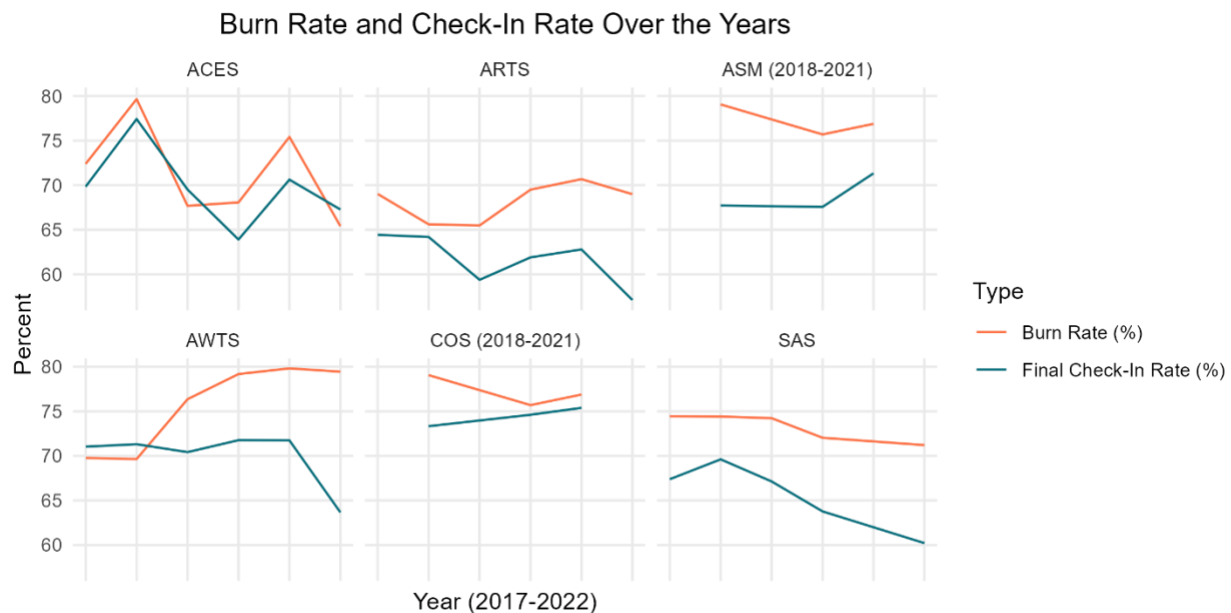
This section highlights the authentication code burn rate and respondent engagement over the duration of the collection period in relation to survey and the number of communication attempts.

The burn rate was calculated by dividing the total number of authentication codes used by the total number of authentication codes issued for a survey year. These burn rates were then linked to each survey in the **Survey table**. We calculated burn rate at two different times in survey collection. Burn rate by due date refers to the number of authentication codes burned by the survey due date (approximately 2 months after survey open date) divided by the number of authentication codes issued. The burn rate by closeout date is the burn rate for the entire collection period (which ranges from 7-12 months). We chose this approach to better understand how survey engagement changes over the collection period.

### Burn Rate and Check in Rate Over the Years

As stated previously, burn rate only represents the percentage of sample that have used their authentication code to begin the survey and does not represent those who complete the survey. The completion rate for these surveys is called the check-in rate. In our dataset we can discern the daily authentication code usage but only know the final check-in rate for the survey. Thus, in this section authentication code burn rate will be our primary outcome of interest. However, it is important to view the relationship between check-in rate and burn rate to see how these two measures of respondent engagement compare. As seen in the figure below, the relationship between burn rate at closeout and final check-in rate varies significantly across survey and collection period.

Figure 4: Burn and check-in rate for each survey over the years.



**Finding:** Surveys ARTS, AWTS, and SAS have an increasing gap between the number of people burning their codes and the reported final check-in rate in recent years. Surveys COS and ASM have a decreasing gap between the burn and check rate.

**Limitation:** Some records of authentication codes being burned may be missing and that cases where respondent data was collected over the phone may result in respondents being marked as completing the survey but not as burning their authentication code, which may be why ACES and AWTS have instances of the burn rate being lower than check-in rate.

**Recommendation:** Consider identifying groups of respondents who burn their codes but never complete the survey and sending communication methods focused on those who have burned their codes but have not completed their surveys after they have burned their code for a certain period of time without survey progress.

### Burn Rate by Due Date and Burn Rate by Closeout Over the Years

The burn rate displayed in the previous finding demonstrated the burn rate by the survey closeout. However, the burn rate increases over the course of data collection and not always at a consistent pace. In this section, burn rate at different points in the collection period to give insight on how respondents time burning their codes and whether a burn rate from an earlier subsection of the collection period can indicate the burn rate when the survey closes.



Figure 5: Burn rate by due date and closeout date over the years.



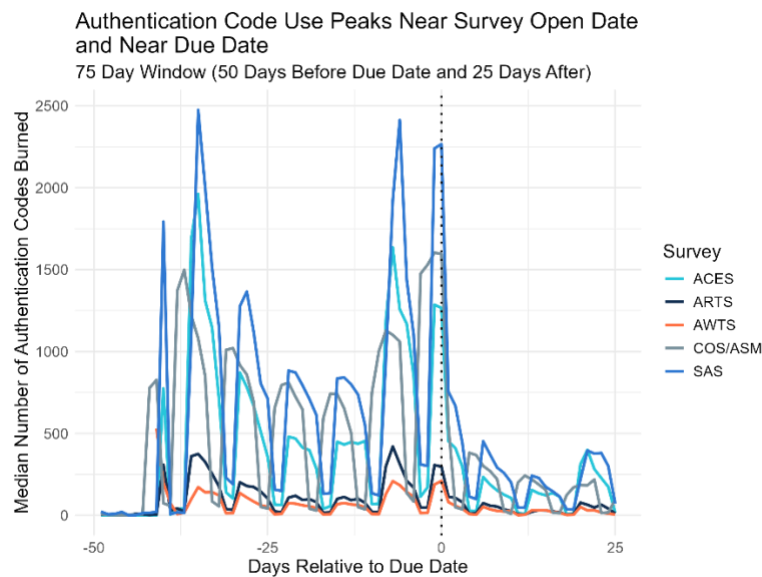
**Finding:** The burn rate by due date and the burn rate by closeout appear to be positively correlated with surveys ACES and ARTS. For the other surveys, there are years where the burn rates by due date and closeout increase or decrease with each other but will then diverge. For example, the burn rates for AWTS in survey years 2017 to 2019 move in the same direction but then move in opposite directions in 2020. This demonstrates that more respondents burning their authentication codes earlier does not always mean more respondents will burn them later in the collection period.

**Recommendation:** Examine possible changes in respondent groups or early communication to provide more insight on why more respondents are either disproportionately burning their codes earlier or later. Focus on early communication for surveys which exhibit a correlation between burn rate by due date and burn rate by close out.

### Authentication Code Use in Relation to Survey Due Date

Looking more closely at daily authentication code usage, we can see how it varies significantly over time. The two figures below, which portray the authentication code use count and the rate, demonstrate that the surveys have similar patterns in engagement that peaks in the few days prior to the survey due date.

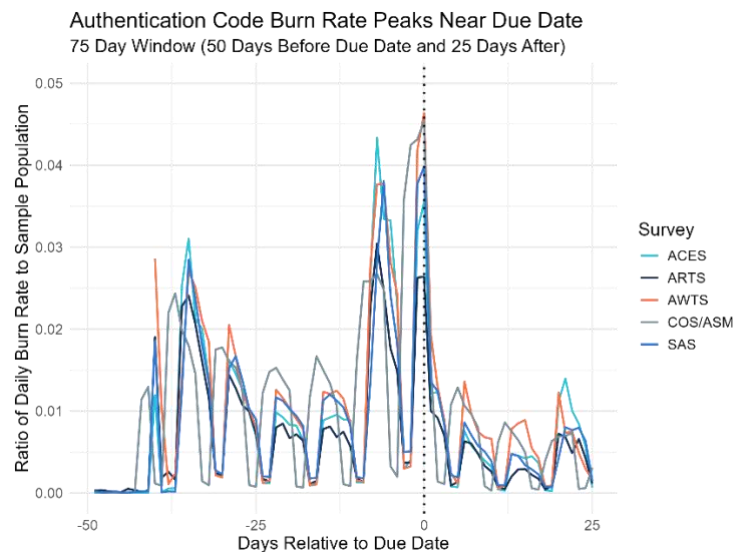
Figure 6: Median Authentication Code Use Across Surveys over a 75-Day Window



**Finding:** Across the legacy surveys, there is increased daily authentication code usage prior to the due date compared to the period after. There are two particularly notable peaks in usage during the 75-day window included in the survey, which correspond to when the initial postal mailouts and emails and the due date reminder mailouts and emails are sent.

**Recommendation:** None.

Figure 7: Median Authentication Code Burn Rate Across Surveys over a 75-Day Window



**Finding:** Across the legacy surveys, authentication code usage (measured by the daily burn count divided by the sample population that has not yet used their code) peaks near and on the due date. Compared to the previous chart, this ratio normalizes the count that can vary significantly based on the size of the survey and demonstrates that the legacy surveys follow

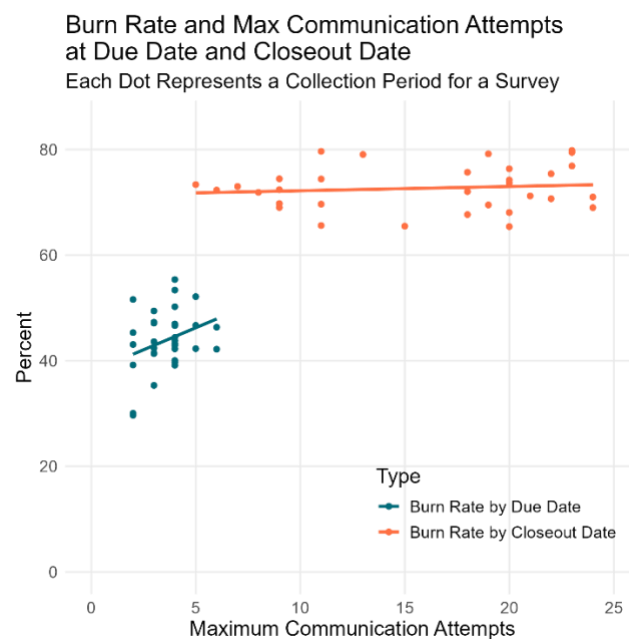
similar trends during the 75-day window. This chart suggests that the messaging sent around the due date may be most effective at increasing burn rate compared to other periods.

**Recommendation:** None.

## Communication and Burn Rate

The relationship between burn rate and the maximum number of communication attempts is observed. The graph below observes the entire collection timeline for each survey year as well as the shorter timeline from the open to the due date of the collection period. There is no pattern observed until the data is segmented into the burn rate and maximum number of communication attempts during the period from only the start of the collection to the survey due date.

*Figure 8: Burn rate vs. maximum number of communication attempts for different collection periods in a survey year.*



**Finding:** For the entire collection period of the surveys, there does not appear to be a correlation between the number of maximum communication attempts and the burn rate. However, when sub setting the data to the period of time between the open and due date of a survey year, there appears to be a positive correlation between the maximum number of communication attempts and burn rate by due date.

It is also observed that the maximum number of communication attempts for the entire collection period varies much more than the period by the due date, demonstrating the number of communications after the due date greatly increases for several survey years.

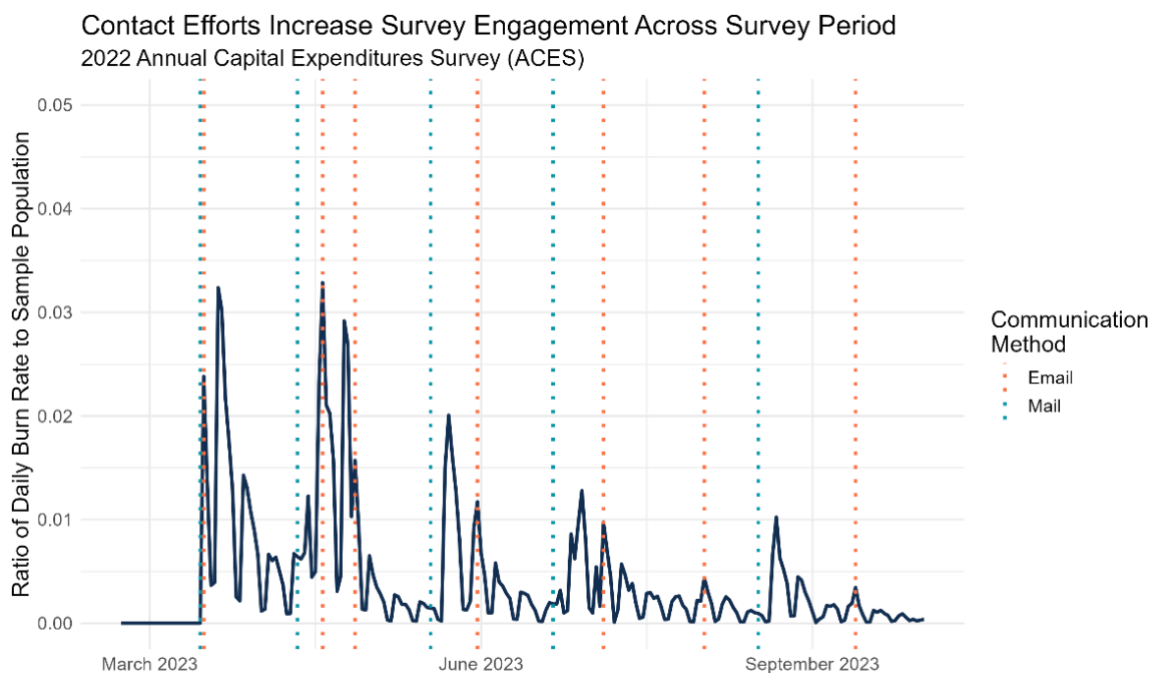
**Limitation:** Given that we know that max communication attempts have increased over time (see figure 1), we also know that the observations with 20-25 max communication attempts are

also more recent survey years. Therefore, while there appears to be no correlation between overall max communication attempts and burn rate at closeout date, there could be external factors effecting the relationship, such as the covid-19 pandemic and changes in industry since survey year 2015. It could be that because of the external factors, the increase in number of communication attempts was needed to maintain a burn rate around 70 percent in more recent years.

**Recommendation:** None.

While looking at communication plans and burn rate at closeout does not clearly show a relationship, when we review communication over the course of a collection period, we can see that contact efforts result in increased survey engagement. The figure below, which represents the 2022 ACES collection period, depicts the ratio of daily authentication code use divided by the number of authentication codes remaining in the sample that have yet to be burned. This ratio provides a better metric of new code usage, which is the goal of each communication attempt. The vertical lines in the figure represent dates that communications were sent out. Across the collection period, the peaks in survey engagement occur on or shortly after the communications are sent out.

*Figure 9: Survey Engagement Across the Statistical Period*



**Finding:** Across the survey year, communication efforts increase respondent engagement. There are peaks in authentication code usage on or near the vertical lines, which represent dates of communication efforts.

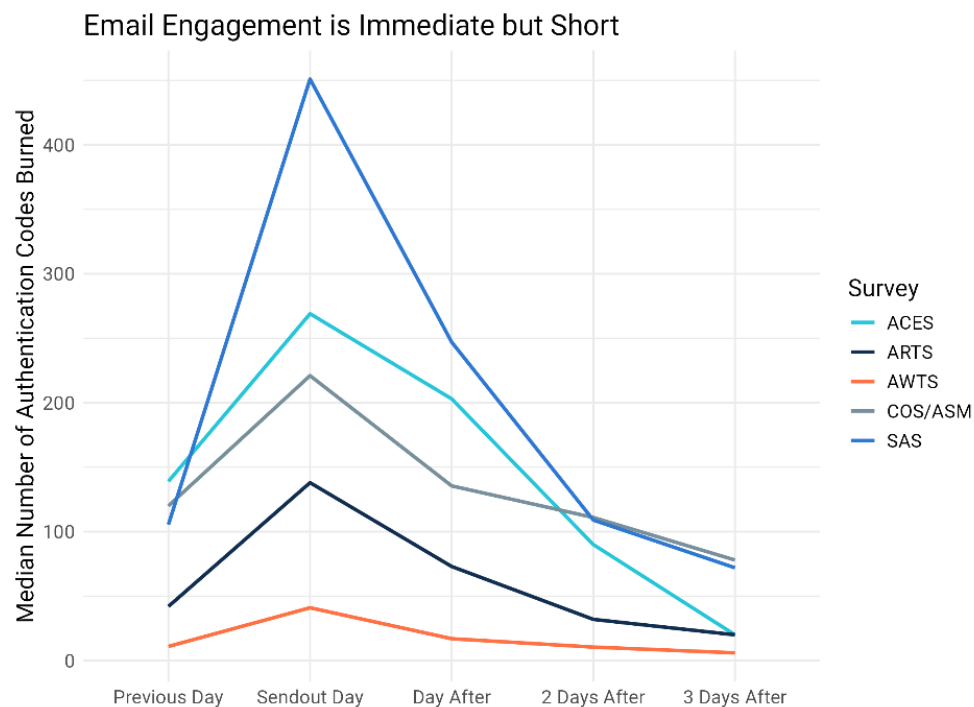
**Recommendation:** None.

## Communication and Engagement Window

Lastly, we reviewed burn rate in relation to specific types of communication method to better understand each method's window for survey engagement. For each observation in the email and mail tables, new columns were created to track the number of authentication codes being burn during the days previous, during, and after send out.

### Email.

Figure 10: Engagement Window for Emails

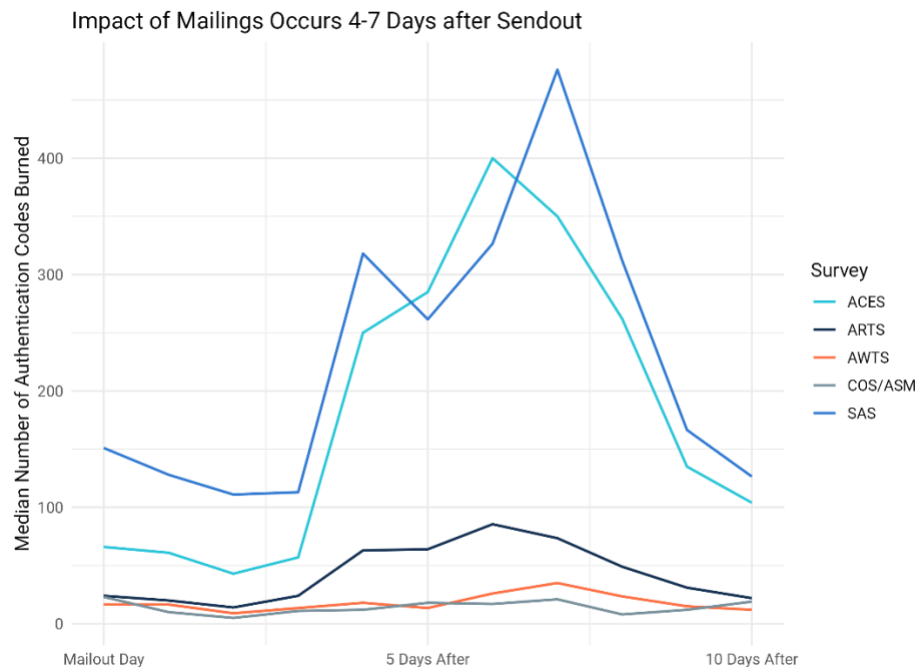


**Finding:** When surveys use email communication, the median number of authentication codes burned after an email is sent reaches its peak on the day of send out and returns to levels before the day the email is sent 2 days after. It is also observed that across the surveys there is a slight decrease in the median number of authentication codes burned 3 days after send out in comparison to the median number of authentication codes burned the day previous to send out.

**Recommendation:** Space emails at least 2 days apart to isolate the impact of emails. However, there are not findings on overlapping emails.

## Mail.

Figure 11: Engagement Window for Mailings



**Finding:** Postal mailings have a delayed effect on respondent engagement. We suspect that this is due to the logistical factors such as the time it takes to mail to be delivered to an establishment. Across the surveys, the increase in survey engagement occurs between 4-7 days after mailout. The increased authentication code usage period lasts for a few days whereas with emails the authentication code use count goes back to average levels 2 days after send out.

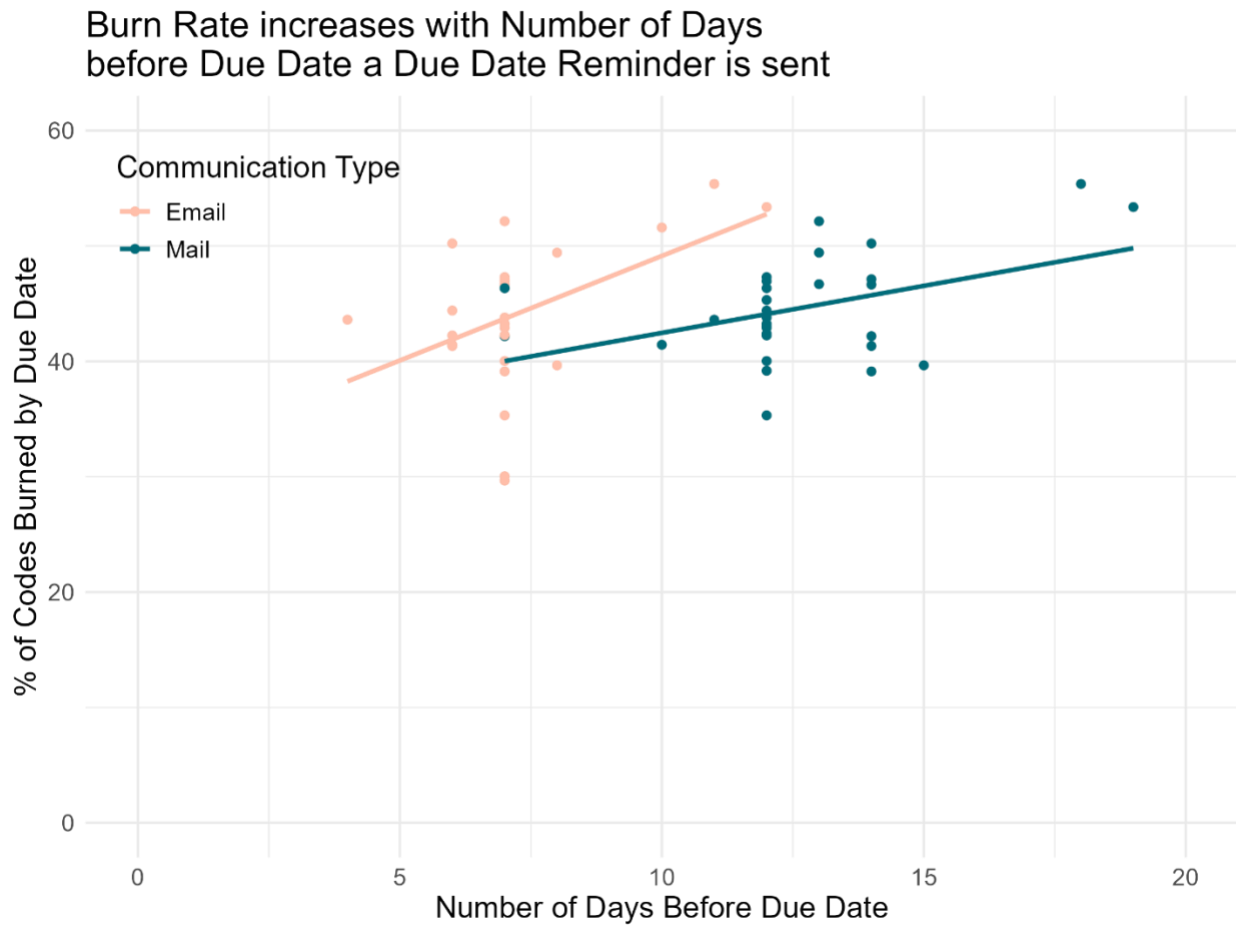
**Limitation:** There is a lot of overlap between email and mail which makes it hard to isolate the impact of mail since the engagement shelf life is much longer. Approximately half of mailouts had an email sent within a week of mailout and a quarter of mailouts had an email sent 4-7 days after mailout, which we know to be peak mailout time. Since the impact of email is immediate, it is harder to observe the engagement shelf life of mail alone.

**Recommendation:** Sending emails at least 2 days before or at least a week after mailouts would better isolate the impact of mailouts on engagement.

## Due Date Reminder Timing

Looking at engagement by period rather than communication message, the graph below observes overall respondent engagement in the period from opening to due date. Due date reminders are sent in the weeks previous to a survey's due date to remind respondents about the upcoming survey due date. A new column was created to track the number of days before a due date a due date reminder is sent.

Figure 12: Scatter plot of early burn rate and the number of days before the due date a due date reminder is sent.



**Finding:** The burn rate by the due date is positively correlated with the number of days before due date a due date reminder is sent. Emails are typically sent in the week before the due date while mailings are sent 2 weeks before.

**Recommendation:** Send email DDRs closer to a week or a little over a week before the due date and mail DDRs 2-3 weeks before the due date.

## Limitations and Recommendations

Throughout the project, we faced two key limitations: data accessibility and overlap of communications.

### Limitation Theme 1: Data Accessibility

**Limitation 1A.** We did not have access to records that linked businesses to specific communications or communication methods. For example, businesses must opt in for email correspondence. It is unclear whether these records were not maintained as part of the legacy surveys' operations or if the data exists and their location is unknown to the AIES team. As a result, the communication methods respondents received could only be estimated and we were unable to model the effect of communications and communication methods at the business level.

**Limitation 1B.** The only modeling that we were able to conduct took place at the survey level. We had records of 42 survey collection periods, but because we did not have authentication code information for survey years 2015 and 2016 and had minimal information on M3UFO's communication efforts, the sample size for modeling communication and respondent engagement was 28.

**Limitation 1C.** We were unable to include outbound call data in the analysis. While we had access to outbound call data, which was recorded at the business level, it was only for a few survey collection periods and not available for all surveys. Therefore, we had to exclude call data from the analysis.

**Recommendation:** It is recommended that there are individual records of communication strategies which record which respondents receive what communication to analyze the direct relationship between communication and whether a respondent burns their authentication code or not. Additionally, it is recommended that for further research of respondent rates, that all communication information would be stored in the same parent folder for convenience and efficiency.

There was no documentation on how people use or view the links provided through mail communication. Additionally, there was no way to observe which or how many respondents were clicking on links provided in email communication.

### Limitation Theme 2: Overlap of Communications

**Limitation 2A.** When email communication is sent in the days following a mail communication, it makes it hard to isolate the causal effect of different methods.

**Limitation 2B.** We were unable to explore the combined effect of overlapping communications compared to a single method.

**Recommendation:** Avoid sending emails during the peak activity days 4-7 days after mailouts to get a better sense of the impact both methods have by themselves. To observe the effect of dual communication with email and mail, implement experimental groups that receive varying levels of overlapping where mail is timed with email to be received at in the same time frame.



## Conclusion

Contact efforts are a key component of survey production but can also be highly burdensome. Between survey years 2015-2022, the communication strategies have significantly increased the number and methods of potential outreach attempts to survey respondents, but we had to look closely at each communication attempt rather than the overall strategy to see how communication affects survey engagement.

Based on our analysis, we find that contact efforts have a positive effect on survey engagement that varies based on the contact method. However, while there are several trends that can be identified, the explanations for them are unknown due to how much ambiguity there is in the data (such as external factors like the covid-19 pandemic or overlapping communications). Because of overlapping communication and lack of individual records, this impact can only be estimated on a higher-level. A next step in this project would be to move forward to examine the impact of outgoing calls as a communication method on engagement. Detailed records on outgoing calls to remind respondents about the survey could help better isolate the impact of other methods as well as compare the impact of outgoing calls to emails and mails, which is especially important considering they are a more expensive communication tactic.