GEORGETOWN UNIVERSITY | McCourt School of Public Policy

Administrative Checkpoints, Burdens, and Human-Centered Design: Increasing Interview Access to Raise SNAP Participation

Better Government Working Paper #2

Jae Yeon Kim, Code for America

Pamela Herd, McCourt School of Public Policy, Georgetown University

Sebastian Jilke, McCourt School of Public Policy, Georgetown University

Donald Moynihan, McCourt School of Public Policy, Georgetown University

Kerry Rodden, Code for America

Abstract

Federal policymakers have urged the use of human-centered design to reduce administrative burdens in policy implementation. In this study we describe the potential of human-centered design principles to identify burdens, reducing the effects of what we label as administrative checkpoints. Administrative checkpoints - mandatory requirements that must be satisfied in order to progress in an administrative process - have disproportionate negative effects in excluding the public from receiving public services. Mandatory interviews are one such checkpoint. Based on consultation with safety net clients and caseworkers, we designed a field experiment (N=1,554) to minimize the exclusionary effects of mandatory interviews for SNAP applicants. Compared to a control group that received a traditional mailer reminder, SNAP applicants who also received texts reminding them of the interview and communicating flexible “interview anytime” scheduling options had a higher interview completion rate by 10 percentage points, a higher approval rate by 6 to 7 percentage points, and also completed interviews sooner. Post experiment surveys show that the text reduced learning costs about the interview requirement.

Introduction

Multiple streams of research and practice have emerged in the implementation of public policies in recent years. First, the nascent administrative burden literature has centered the study of public services on the experiences of the public (Herd et al., 2013; Heinrich, 2018; Heinrich et al., 2022; Pierce & Moulton, 2023). Second, behavioral science studies have sought to find practical solutions to some of the frictions that the public encounters (Linos & Riesch, 2020; Linos, Quan, & Kirkman, 2020). Third, "civic tech" has become a growing influence in government, offering technological solutions to reducing burdens (Pahlka, 2023). Fourth, the practice of "human-centered design" is becoming more visible in government (McGuinness & Schank, 2021).

We employ a case study that illustrates how these streams can merge to improve practice and solve challenging administrative problems during policy implementation. A civic technology nonprofit, Code for America (CfA), worked with Boulder County, Colorado, to identify and reduce administrative burdens in the county's Supplemental Nutrition Assistance Program (SNAP) program on both workers and clients. CfA applied principles of human-centered design – including consultation with clients and workers, identifying user journeys, iterative adoption, and ongoing experimentation – to develop a solution: low-cost texting that communicated flexible interview options. The team also applied a randomized controlled trial to offer a rigorous basis to evaluate the intervention.

This study offers a number of contributions. First, we seek to more directly connect human-centered design with the practice of policy implementation. Human-centered design is growing in influence in government, even as it remains a rarity within both the research and teaching aspects of schools of public policy and administration. In 2021, President Joe Biden signed an Executive Order "*Transforming Federal Customer Experience and Service Delivery to Rebuild Trust in Government*" that identified human-centered design as a means of reducing administrative burdens. The order defined human-centered design as "*an interdisciplinary methodology of putting people, including those who will use or be impacted by what one creates, at the center of any process to solve challenging problems*" and made it federal government policy to use such tools: "*the Federal Government's management of its customer experience and service delivery should be driven fundamentally by the voice of the customer through human-centered design methodologies; empirical customer research; an understanding of behavioral science and user testing, especially for digital services; and other mechanisms of engagement*" (Exec. Order No. 14058, 2021). This study not only summarizes theories of human-centered design, but illustrates its practice in a particular setting, while considering the opportunities and constraints in its application.

Second, we contribute to the "second-generation" of administrative burden theorizing, one which makes legible problems and potential solutions to the experience of burdens (Moynihan et al., 2022). We do this first by developing the concept of administrative checkpoints - mandatory administrative requirements without which an individual cannot proceed in an administrative process. Administrative checkpoints are common (e.g., submission deadlines, voter registration deadlines). Our empirical work illustrates how human-centered design helps to identify and minimize the negative effects of administrative checkpoints. Directing both scholarly and practical attention to administrative checkpoints can generate outsized benefits when it comes to making public programs more accessible.

Third, we offer empirical evidence on the effects of mandatory interviews as an administrative checkpoint, and the value of communicating flexibilities around those interviews. Missed interviews are a key reason why individuals fail to access benefits, or churn on and off programs (Homonoff & Somerville, 2021). We test if making interviews easier to access increases interview completion and program take-up. We find that communicating a recently adopted flexible interview process both saved client time, and increased approval rates for SNAP benefits. Compared to a control group that received a traditional mailer reminder, SNAP applicants that also received texts reminding them of the interview and communicating flexible "interview anytime" scheduling options had a higher interview completion rate by 10 percentage points, a higher SNAP approval rate by 6 to 7 percentage points, and also completed interviews sooner. The results closely mirror a similar experiment in Los Angeles (Giannella et al., 2024).

Follow-up surveys found that those who received the treatment reported lower learning costs: they were more likely to know that the interview was a requirement for receiving services. They also expressed greater confidence that they knew what they were expected to do at the interviews. While the effects of behavioral nudges are increasingly contested (DellaVigna & Linos, 2022), the results suggest that informational nudges remain an effective tool in contexts where interested clients are simply unaware of requirements, or how to fulfill them. In other words, informational nudges become more effective when they convey "news you can use" to their audience.

In the sections that follow we first present the concept of administrative checkpoints, using mandatory SNAP interviews as an example, with a theoretical expectation that clearly communicated flexibilities can increase take up. We then explain the precepts of human-centered design as a tool to solve this problem. Next, we present our case, before presenting our findings from a randomized controlled trial and follow-up surveys, followed by a discussion and conclusion.

Administrative Checkpoints

The emerging study of administrative burdens is attentive to the fact that not all frictions are the same: some are not just onerous, but also exert a disproportionate impact that leaves clients stuck, unable to access benefits for which they are eligible. For example, Heinrich (2016) describes how specific administrative failures lead to "stops." Similarly, Peeters and Widlak (2018) refer to "digital cages" that trap people in an administrative dead-end. Sackett and Lareau (2023) point to "institutional knots" that individuals find themselves in when organizational errors or inconsistencies across organizations compound. It therefore makes sense to better conceptualize and study the administrative conditions that give rise to these outcomes.

We define administrative checkpoints as mandatory administrative requirements that a user must complete in order to proceed in an administrative process. Here we retain Herd and Moynihan's (2018) definition of burdens as experiences, while noting that this does not exclude better conceptualizing state actions that give rise to those experiences. Indeed, a public administration research agenda that seeks to identify and reduce burdens can leverage a richer vocabulary of state actions that give rise to burdens (e.g., Fox, Stayzk, & Feng, 2020). Administrative checkpoints are not administrative burdens, but a type of state action that creates the experience of burdens. Most obviously, administrative checkpoints generate what administrative burden research categorizes as a compliance cost, where the individual provides information including documentation to a state actor (Herd & Moynihan, 2018). To the degree that checkpoints are experienced as unusual, they also increase learning costs for the individual, since they have to be both aware of the checkpoint, and learn how to satisfy its particular demands. As checkpoints are understood to be consequential, they are likely to generate psychological costs, such as fear, stress, frustration and anger.

Administrative checkpoints can take a variety of forms: the requirement to demonstrate identity, to submit paperwork verifying assets, income, or work status, or to appear for an interview. What makes checkpoints deserving of unusual attention is their power. As they are mandatory, they can result in administrative dead-end or refusal. This distinguishes them from other types of frictions that may be onerous, but do not result in a hard stop in the process (e.g., wait times). With an administrative checkpoint, an individual has failed to complete a specific formal step or requirement, leaving them either suspended in or rejected from the administrative process. This implies that among the array of costs that people encounter, administrative checkpoints are deserving of special attention.

Having defined checkpoints, some observations follow. Regardless of the source of the checkpoints, what matters, from the user perspective, is the de facto effect of the checkpoint on their ability to progress. In the safety net domain which we study here, checkpoints can arise during both application and renewal processes. Cumulatively, checkpoints operate akin to veto points in political processes: each additional one increases the chance of failure to complete a process. Reducing or eliminating the number of checkpoints will make it more likely that an individual will succeed in completing the process. Even a seemingly minor checkpoint can exert a disproportionate effect because of the combination of its mandatory nature, and behavioral difficulties that people have when it comes to completing administrative processes, which can arise from conditions of scarcity, inattention, or inability to manage short-term tasks with long-term benefits (Sunstein, 2021).

Since checkpoints are, by definition, mandatory, is there any point in paying attention to them? To address this point, we should think about the value and necessity of checkpoints. In some cases, administrative checkpoints may be mandated by law or formal administrative directive, but of little value to legitimate public goals. In such a case, research can usefully assess the value and necessity of this requirement. This would offer policymakers better information about the degree to which they should impose or maintain administrative checkpoints. For example, many states have eliminated asset tests in SNAP because of evidence that such requirements had limited value, identifying few people with assets that made them ineligible, but imposing large compliance costs on eligible claimants to document the absence of assets (Lin, 2023). Another type of research, one pursued in this study, seeks to soften their impact. In the criminal justice setting, automated reminders about mandatory court appearances reduces missed appointments, warrants, and pretrial incarceration (Chohlas-Wood et al., 2023). The mandatory requirement remains in place in both examples, but the state has taken action to reduce the risk that more people will fail to register to vote, or appear in court.

Mandatory Interviews as an Administrative Checkpoint: The Case of SNAP

In this study we examine one administrative checkpoint, which is the mandatory interview. This is a common requirement for safety net programs. The client for public services must appear at a certain place at a certain time to demonstrate their identity and satisfy the inquiries of a state actor. Interviews might offer a chance for street-level bureaucrats to help clients negotiate a process, but in the domain of SNAP that we study, there is evidence they exert significant negative effects on program take-up. SNAP is a means-tested safety net program used by approximately 41 million Americans, funded by the federal government and administered by states. Federal rules require that to receive benefits, SNAP applicants must complete an interview within 30 days of application. To recertify their status, SNAP clients must complete interviews at least annually. Interviews can be done by phone but are typically scheduled by administrative actors without input from the clients. Missed interviews that are not rescheduled before application or recertification deadlines result in failure to receive benefits.

One way to assess the impact of SNAP interviews is to assess variation in outcomes associated with variation in recertification processes. States can increase the recertification rate to less than a year. More frequent recertification processes are associated with lower take-up (Kabbani & Wilde, 2003; Ribar, Edelhoch, & Liu, 2008). However, such evidence focuses on recertification processes as a whole, rather than just interviews specifically. More recent work has focused on the effects of interviews. Homonoff and Somerville (2021) exploit the random assignment of interview timing to estimate their effect on SNAP take-up. They find that clients assigned mandatory interviews near the end of a recertification deadline, which give the client less time to reschedule an interview, are 22 percent less likely to recertify their status and continue to receive benefits. Some clients eventually return to the program, but the churn off and on creates costs both for the clients and administrators, since clients have to now go through the full application process. Homonoff and Somerville also offer evidence that clients lack of awareness about the initial interview and that difficulty in rescheduling appointments explains a substantial proportion of clients who fail to recertify, implying that both informational nudges and greater flexibility in interview assignment can reduce the negative effect of mandatory interviews on take-up. Lopoo, Heflin and Boskovski (2020) offer evidence that informational reminders via text about SNAP recertification requirements, including interviews, do indeed increase completion of these requirements. Giannella, Homonoff, Rino and Somerville (2024) examine the effects of allowing more flexible interview scheduling among SNAP recipients in Los Angeles. They find that offering more flexible interview scheduling options increases applicant approvals by six percentage points by the 30 day interview deadline, and increases long-term participation in the program by over two percentage points, and expedited benefit receipts through earlier approvals. States can offer the ability to reschedule interviews only with a waiver from the federal government.

Prior work therefore suggests both that mandatory SNAP interviews satisfy our conceptualization of a consequential administrative checkpoint, and that clearly communicated flexibilities can reduce the negative impact of interviews on program participation. We state these hypotheses more explicitly after the discussion of the application of human-centered design practices to the case, reflecting the fact that they are informed not just by a limited body of prior research, but also inductively-developed insights from fieldwork. We turn next to a discussion of human-centered design tools.

Human-centered Design

A cousin to our concept of administrative checkpoint are "pain points" which are "real or perceived problems perceived by customers within a system" (General Services Administration and The Lab at OPM, n.d.). Pain points and administrative checkpoints are similar in that they direct attention to specific stages in administrative processes as problematic. The concept of pain point is broader, reflecting any negative experience a client may have, while an administrative checkpoint centers on the mandatory and blocking nature of a requirement. Thus, we might think of administrative checkpoints as a subcategory of pain points, one especially relevant for public settings where identity and eligibility must be determined in multi-stage administrative processes.

Attention to pain points is a central concern of the field of human-centered design (sometimes referred to as user centered design). "*Human-centered design is a technique to understand administrative processes from the user perspective, using those insights to adjust those processes to better match human capacities. Human centered design will often employ distinct stages of discovery, design, delivery, and measurement as part of an iterative and ongoing process*" (Herd, Moynihan, & Widman, 2023). Discovery involves directly engaging with, through observation and conversation, the people involved in the process. Design involves building prototypes of solutions that draw on client insights, testing those solutions, and soft launches to gain further feedback. The delivery stage shares the outcome, and appropriate goals and expectations. Measurement tracks how well the change is doing, relying on a mix of qualitative and quantitative indicators (General Services Administration and The Lab at OPM, n.d.). Within each stage there are additional tools and practices when it comes to consulting with clients, presenting new designs and testing them (Rappin et al., 2020). These include how to effectively interview clients and frontline staff, connect with stakeholders, having administrators experience administrative processes themselves, creating journey maps of paths that clients pursue through a process, sprint meetings that pull together key stakeholders to solve problems, and building prototypes of new processes to gain feedback on (wireframing) (Sullivan & Soka, 2022).

The field of human-centered design is amorphous, and rigorous scholarship around its use in public settings is rare. Fundamental principles include using ethnographic techniques to understand the underlying reasons why people encounter problems with the systems they engage in (which can include public servants and not just clients), and using rapid iterations of process changes with equivalent feedback loops, reviewing and checking hypotheses, acting transparently, and testing prototypes with users (Mergel, 2022; Norman, 2024). The emphasis on iteration and continual improvement echoes long-standing management practices such as Total Quality Management. But as the name suggests, human-centered design emphasizes designing around the human, and not the process or technology (Norman, 2024). This makes it suitable for understanding and fixing administrative processes that shape human experiences, especially those that have been historically inattentive to how to do so well. The emphasis on iteration and experimentation also mirrors applied behavioral science work. A key difference with human centered design is the emphasis on ethnographic modes of data collection, and inductively-deduced insights in addition to broader theorizing.

Proponents of human-centered design see it as a set of practices that puts the public at the center of citizen-state interactions. Relying on user insights protects against blind spots, unanticipated consequences, and slow feedback loops that they see as characteristic of standard modes of public management that reflect a separation of policy design from policy implementation (McGuinness & Schank, 2021). In this respect, human-centered design shares overlapping assumptions with the administrative burden framework (Herd & Moynihan, 2018). One is that the delivery of public services can only be best understood by closely studying how it is experienced by clients. Another is that such experiences are often characterized by frictions that public officials do not recognize. Human-centered design also represents a form of participation in government, but one different from public comments, or town hall meetings, centered on the idea of co-design in building experiences. It offers a potential remedy against the downsides of automation, where algorithms that are disconnected from human experience become decision factories that generate burdens rather than solving them (Peeters & Widlak, 2023)

The growing attention to customer experience and administrative burden reduction within the Biden administration has advanced the use of human-centered design principles. But it was already an established practice in parts of the federal government that focused on customer experience or improving digital interfaces (Sinai, Leftwich, & McGuire, 2020). Federal units, such as the US General Service Administration, and US Digital Service have developed human-centered design guides for federal employees. The Office of Management and Budget uses journey mapping to understand key “life experiences” that members of the public encounter. Some of this growth in influence reflects the growing presence of civic technologists in government. In a “Civic Technologists Practice Guide” Cyd Harrell (2020, 114) writes that “*There’s no question that design is an essential discipline. Pretty much the only way that software, however well built, is actually useful for its intended purpose is to design it for (and ideally with) the people who will use it*”.

Case Background: Applying Human-Centered Design Principles

This section describes the intervention, as is standard with any description of a field experiment. But to convey the way in which human-centered design played a role, we also describe how the intervention was developed, which involved a partnership between a local government and a civic tech organization, CfA, drawing on administrative burden and behavioral science insights. Human centered design practices in the case included a) field studies that interviewed clients and caseworkers to identify pain points, which surfaced interviews and text-based communication, b) developing and prototyping an alternative solution, c) journey mapping, d) working with users to explain and build buy-in for the intervention, e) a randomized controlled trial to assess the effect of the intervention, f) short feedback loops to develop and test the intervention.

Colorado is one of the state partners for CfA’s Safety Net Innovation Lab project, which seeks to use digital expertise to increase access to safety net benefits. States apply to work with CfA based on shared interests, and then identify specific innovation opportunities. CfA staff made field trips to the state in July and September of 2022 to gather insights about improvement opportunities in PEAK (Program Eligibility and Application Kit), Colorado's integrated safety net program website that includes SNAP and other services. In the first research trip, CfA investigated how SNAP clients used digital tools to apply and maintain their benefits and documented the experience of workers and administrators as they processed benefit applications. Specifically, the team shadowed caseworkers, visited county offices, and interviewed community members. By doing so, CfA discovered that applicants faced significant administrative burdens due to the difficulties in understanding the SNAP processes and requirements. Applicants often felt lost and frustrated, leading to a decline in their trust in the administrative process. Although the PEAK website was widely used (accounting for almost

50% of benefit applications in the state), its design did not make applicants confident about how they were progressing through the process.

During the second research trip, CfA staff examined the communication between the county and applicants regarding policy requirements, such as the SNAP interview requirement and protocol. The research validated the findings from the previous round of in-person research and extended them by conducting co-design activities with stakeholders. For instance, CfA researchers mapped the application journey with community members to understand their mental model of their experience, identifying pain points. In the next step, community members were invited to co-design solutions for improving communication between the county and applicants. CfA facilitated discussions by providing several communication options. Community members chose their preferred options and responded to broad questions about how to improve communication channels (e.g., if you could redesign this part, what would it look like?). In the process, CfA discovered that, as in many other places, mail notices, the standard communication method to SNAP applicants, are unreliable. They often arrive late, causing confusion and anxiety, and result in SNAP applicants missing their interviews. Furthermore, CfA observed that online applications are convenient but challenging. Applicants cannot seek assistance as easily as they can when filling out necessary forms in person at an office. Mandatory interviews emerged as an important pain point where applicants failed. Descriptive data analyses undertaken by CfA staff based on early 2022 administrative data found that missed interviews were a primary reason for SNAP procedural denials, accounting for half of denials in some counties, including Boulder. Missed interviews were more likely among younger and online applicants, and for clients experiencing homelessness. Community members also expressed their desire to have more agency in the process by being able to select a flexible interview date and time that works with their schedule.

CfA chose Boulder County as a site for the pilot test, with a focus on applicants who applied for SNAP through PEAK. Colorado is a state with county-level administration of SNAP benefits. Boulder County was selected as a county that was large enough to assess meaningful impacts for a test, and was willing to try new approaches. Boulder County had previously attempted to use texting as a communication method. However, they were unable to gain insights from this experience because they did not employ a causal research design. They adopted flexible interview processes in December of 2022, about three months prior to the field experiment. This allowed clients to call to reschedule appointments but also to call the county to conduct an unscheduled interview. One concern was that individuals were unaware of flexible interview opportunities. The intervention therefore involved an informational nudge – about the interview requirement and opportunity for flexible interviews – that was built on a change in administrative processes. The discovery process pointed to confusion about how interviews were to be scheduled, and whether the client or the county official would initiate the appointment. Thus, not only did communicating the information matter, so too did the clarity and usefulness of the information.

Research Design and Data

Our field experiment estimates whether texting information about interview requirements and deadlines, combined with an opportunity to schedule a flexible interview, increases SNAP interview completion rate and leads to earlier interview completions, resulting in a higher rate of receiving benefits compared to their counterparts. This research focuses on applicants who applied to the program through the state's integrated benefit website. These applicants had a higher rate of missing interviews compared to those who applied to the same program through other means, such as calling, mailing, or walking in. Our hypotheses are consistent both with the discovery about the specific processes, as well as the limited prior research, in terms of the effects of mandatory interviews (Homonoff & Somerville, 2021), the potential for text reminders (Lopoo et al., 2020) and more flexible interview scheduling opportunities (Giannella et al., 2024) to reduce learning costs in a way that increase take up. The first three hypotheses focus on client outcomes (interview completion, timing of interview completion, and take-up of benefits), while the latter two seek to convey insights on the mechanisms that explain differences in those outcomes (knowledge about interview requirement, confidence about interview process).

1. H1 (compliance): SNAP applicants who received text reminders and flexible scheduling opportunities will comply with the interview requirement at a higher rate than applicants who did not receive text reminders.
2. H2 (scheduling): SNAP applicants who received text reminders and flexible scheduling opportunities will schedule the interview earlier than applicants who did not receive text reminders.
3. H3 (take-up): SNAP applicants who received text reminders and flexible scheduling opportunities will receive benefits at a higher rate than applicants who did not receive text reminders.
4. H4 (awareness): SNAP applicants who received text reminders will be more aware of the interview requirement compared to applicants who did not receive text reminders.

5. H5 (ease): SNAP applicants who received text reminders will be more likely to believe that it is easier to prepare for the interview than applicants who did not receive text reminders.

Field Experiment

The field experiment was conducted between March 22 and June 22, 2023. Residents were divided into treatment and control groups based on a randomly assigned variable: their case number. Applicants with odd case numbers were assigned to the treatment group. Applicants with even case numbers were assigned to the control group. Balance tests based on 2022 administrative data show no difference between treatment and control groups on observable variables, such as age, gender, language preference, homelessness status, the number of benefit programs clients applied for, the number of people on their cases, whether they missed SNAP interviews for previous year applications, and whether they were denied SNAP benefits (see Appendix A in the Online Appendix).

All applicants received traditional mailers that contained the scheduled date and time for their interviews. The letter did convey that *"if you miss your appointment or need to reschedule, you must contact us…to reschedule your interview."* The mailer did not mention the recently created flexible "interview anytime" interview option created by Boulder County. The treatment groups received the official mailer, as well as texts. We should consider the treatment to include not just information that reduced learning costs about the interview process, but also that provided easy access to flexible and unscheduled interview opportunities. Control group applicants could call to reschedule interviews, but this was presented as arising from a need to do so. Given that the flexible "call anytime" interview option had been adopted just months previously, few would have been aware of this option.

Treatment applicants received two versions of the same message in English and Spanish, sent three times: 24 hours and three days after they submitted their application, and 24 hours before the scheduled interview. The exact wording of text messages are provided in Appendix C. The wording drew from a human-centered design knowledge base that CfA had developed in texting safety net clients in other settings (Code for America, 2022). In the first text, applicants are asked to call to schedule their interview. In the second text, applicants are told "*you have the option to complete your interview at your own convenience*." In the final text, which comes the day before a scheduled interview, applicants are told "*Can't make it? Just call anytime Monday through Friday between 8:30am-3pm to conduct an unscheduled interview.*" The treatment therefore emphasized not just the possibility of rescheduling, but framed it as a positive choice rather than just a function of a need to reschedule. It also communicated the new “call anytime” alternative, allowing clients to pursue an unscheduled interview, simply by clicking the number in the text.

We track three outcomes: the SNAP interview attendance rate, the timing of SNAP interviews, and the approval rate of SNAP applications. The impact of texting was estimated by comparing the difference in means of the outcome variables between the individuals who received the treatment and those who did not. The estimation strategy is ITT (intention-to-treat) because we only know who received the texts, not who read them. To estimate heterogeneous treatment effects, we also estimated the ITT conditional on applicants’ language preference, gender, homeless status, and household size.

Sample size calculations are described in Appendix D in the Online Appendix. Our simulation-based design diagnostics informed us that a sample size around 1,500 would be sufficient to achieve 80% statistical power. The final sample size is 1,554. A total of 772 applicants were assigned to the treatment and 782 applicants were assigned to the control group. The group numbers are not exactly identical because some applicants' texts bounced, and we did not include them.

Supplementary Survey Analysis

Our primary interest is in the effect of the treatment on client outcomes. But we are also interested in gaining insights about the mechanisms that underpin those outcomes, or why the texts were effective, reflected in hypothesis 4 and 5. We therefore conducted a follow-up survey with the study participants to assess the learning cost associated with the interview portion of the SNAP application process (for the full questionnaire, see Appendix E in the Online Appendix). We sent SMS interactive (two-way) surveys to 1,537 applicants. We dropped 17 applicants because their phone numbers were found to be unreliable during the earlier messaging process.

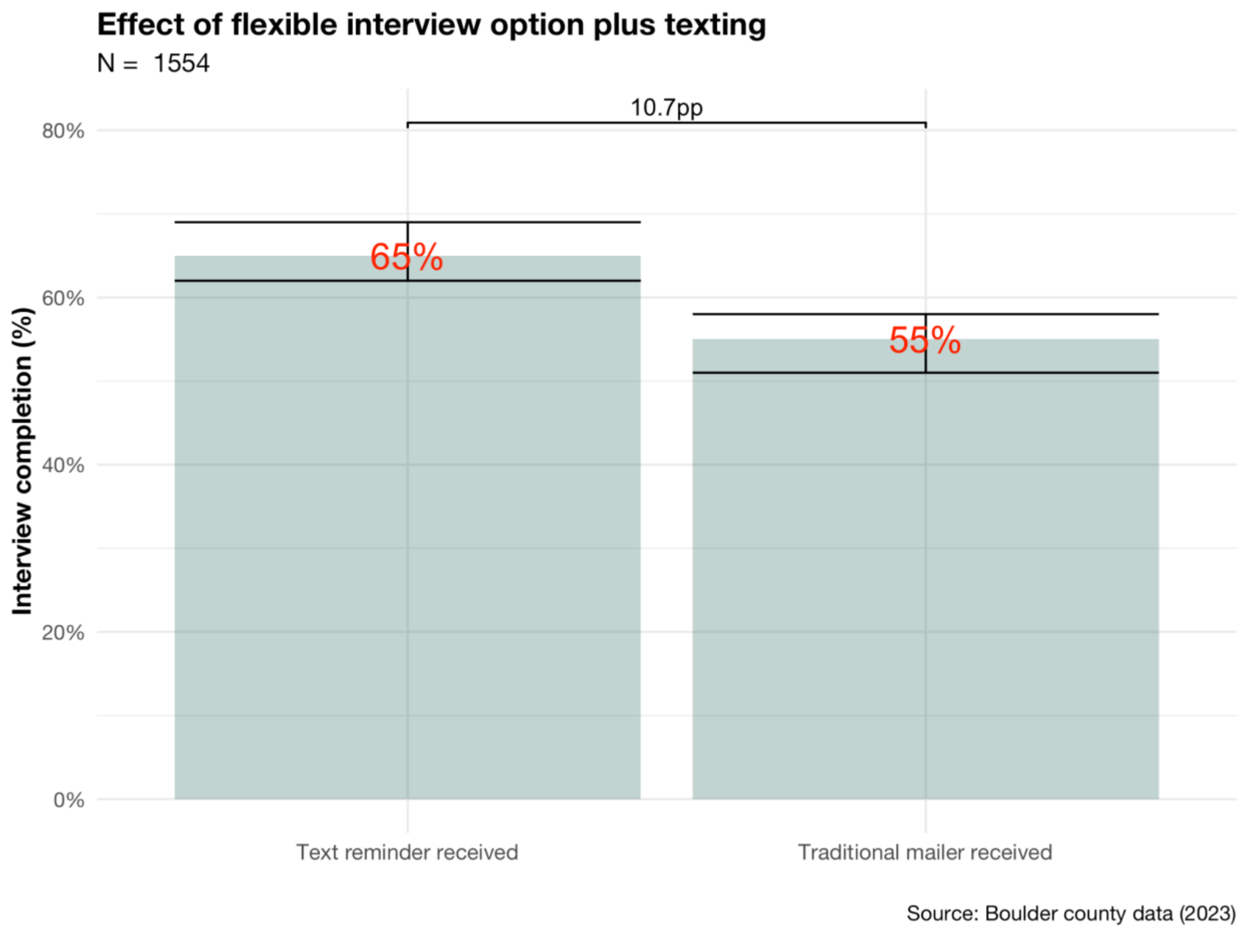
- Question 1: When you applied, did you know you had to do an interview? (Yes/No)
- Question 2: How easy or difficult was it to know what to do for your interview? (1-4, scale running from "very easy" to "very difficult" based on Jilke et al.’s (2024) administrative burden scale)

We estimated effects by comparing the difference in survey response means between individuals who received the treatment and those who did not. The survey was launched on June 22, 2023, after the field experiment was completed. It remained active until July 6, 2023. A total of 470 respondents answered the first question (response rate: 30.6%), and 225 respondents answered the second question (response rate: 14.6%). Power analyses suggest the sample is sufficient to detect effects (see online Appendix F). Treatment status is not statistically associated with the response rate for the questions (see online Appendix G).

Results

We first examine the effects of the treatment on interview completion rate and timing (H1 and H2). The difference in means test indicates that SNAP applicants who received text reminders were 10.7 percentage points (low 95% CI: 5.8pp, high 95% CI: 15.5pp, p-value < 0.0001) more likely to complete the interview than those who did not receive the reminders. Figure 1 is based on the county administrative data which tracked completion rate for the initially scheduled interview. However, for a robustness check, we also estimated the difference in means between the treated and control groups using the state's administrative database, which also includes rescheduled interviews that were later completed. Using the state administrative data produces a slightly smaller, but still significant estimated treatment effect (the treatment group's interview rate: 70.4%, the control group's interview rate: 62%). The treated group was 8.4 percentage points more likely to complete an interview even when including rescheduled interviews (low 95% CI: 2.8pp, high 95% CI: 14pp, p-value: 0.003). As the application information was not fully retrievable from the state's database, the number of observations decreased from 1,554 to 1,115.

Figure 1



We test for heterogeneous effects on both interview compliance and approval rates. We do not find that the effect on interview completion rate varies by gender, language preference, or homeless status of the applicant (see Appendix H). We do find that the effect on interview completion rate varies by the applicant's household size (-12pp, low 95% CI: -21pp, high 95% CI: -3, p.value = 0.01), but above standard significance levels for approval rates. Applicants with a larger household who received the treatment show up for the interview at a lower rate than their counterparts. Since larger households have greater need and can potentially claim more benefits, the finding is inconsistent with neoclassical assumptions about people rationally investing time to overcome burdens. Rather, and consistent with Finkelstein and Notowidigdo (2019), the result is consistent with a behavioral framework where applicants from larger households may be struggling with more temporal scarcity that distracts them from completing processes, even with additional help.

Hypothesis 2 proposed that the treatment would generate quicker interview completion. We cannot directly observe which clients called the county, but any interviews completed before the scheduled date could only have occurred via a phone call, and so early interviews serve as a proxy for use of the flexible interview option. If a client missed an interview, they receive a letter notice from the county informing them of the missed interview, and that they are responsible for rescheduling the interview. Clients may call the county to do so.

We found that SNAP applicants who received text reminders tended to show up for their initially scheduled interview, on average, 3-4 days earlier (low CI: 3 days, high CI: 5 days, p-value < 0.0001) compared to those who did not receive reminders. Figure 2 plots the count of SNAP applicants who attended the interview based on the difference between their interview attendance and the scheduled interview dates. In the x-axis, "-1" indicates that the applicant completed the interview one day earlier than the scheduled date. The distribution pattern shows that the majority of early interviewed SNAP applicants were those who received text reminders. Among the SNAP applicants who completed their interviews earlier (n = 178), 88% of them belonged to the treatment group, and 57% did their interviews between the first and second texts. Among those SNAP applicants who completed their interview later than initially scheduled (n = 417), 60% of them belonged to the control group. The composition between the two groups is almost identical for those SNAP applicants who completed their interview on the scheduled date.

Figure 2

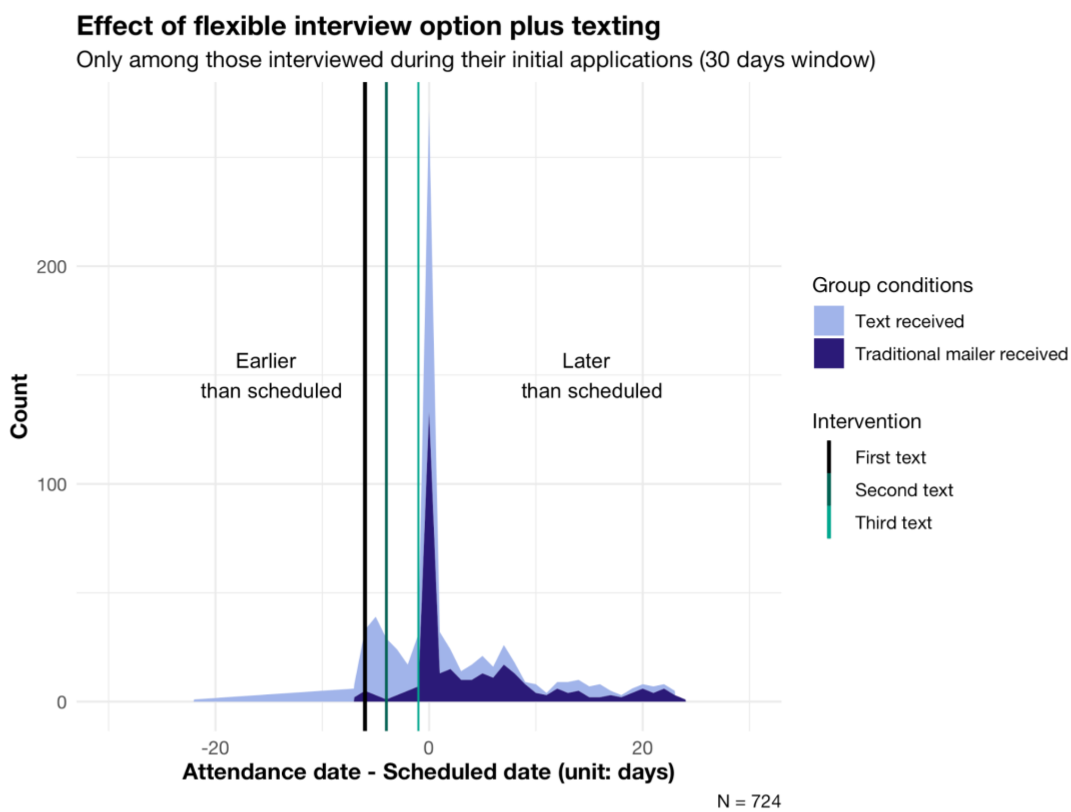
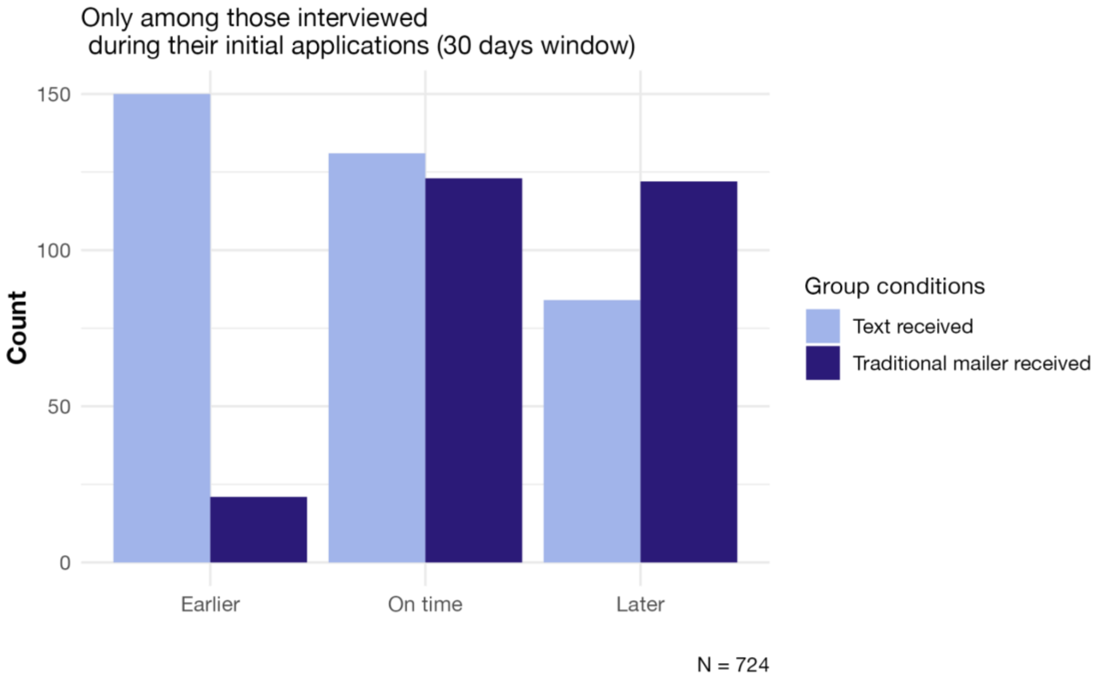


Figure 3



Hypothesis 3 predicted that the treatment would increase the SNAP take-up of benefits. Table 1 presents the results. The estimates in square brackets represent 95% confidence intervals, while the estimates in parentheses indicate p-values. We used robust standard errors (HC2[1]) in our regression analysis. Model 1 focuses on the difference between the treatment and control groups, but excludes 37 cases (20 cases from the treatment group, and 17 from the treatment group) where we couldn't observe their outcomes due to missing information in the state's administrative data. Model 2 includes 16 imputed cases[2] where we at least knew their interview

[1] HC2 stands for heteroskedasticity-consistent errors that are less sensitive to high-leverage points.
[2] We utilized conditional logic to impute the missing values in the approval status variable. If an applicant's approval status was missing and we were aware that they had missed the interview according to the Boulder county data, we assumed that their case had been denied. The most ideal variable for determining an applicant's interview status is the state's administrative data, as it is used directly for eligibility purposes. However, for applicants whose benefit status is unknown, their interview status is also missing. Therefore, this option was not feasible. The county and state's interview variables are highly correlated (Pearson's product-moment correlation coefficient = 0.83, low-95% CI: 0.81, high-95% CI: 0.85, p-value < 0.0001).

status. If an applicant missed their interview, their case was denied (known as a procedural denials). Model 3 further modifies the data used in Model 2 by only considering approved cases during applicants' first application cycle. Cases approved during resubmissions are not counted. Since text reminders were only used for initial applications, we believe this outcome is the most accurate representation of the effect of the intervention. The results indicate that the treatment effect on the approval rate is nearly identical, increasing takeup by just between six to seven percentage points (low 95% CI: 1pp, high 95% CI: 11pp) across all three models, despite slight variations in the underlying data (p-values ranging from 0.003 to 0.01).

Table 1: Impact of Texting and Flexible Interviewing on Take-Up

	Approved	Approved (procedural denials imputed)	Approved during initial applications (procedural denials imputed)
(Intercept)	0.3399*** [0.3057, 0.3741] p = 0.0000	0.3355*** [0.3016, 0.3694] p = 0.0000	0.2121*** [0.1819, 0.2423] p = 0.0000
Treatment status	0.0631* [0.0145, 0.1116] p = 0.0110	0.0643** [0.0160, 0.1125] p = 0.0091	0.0691** [0.0262, 0.1121] p = 0.0016
Num.Obs.	1517	1533	1539
R2	0.004	0.004	0.006
R2 Adj.	0.004	0.004	0.006
AIC	2097.3	2112.4	1771.8
BIC	2113.3	2128.4	1787.8
Log.Lik.	-1045.647	-1053.200	-882.891
RMSE	0.48	0.48	0.43
Std.Errors	HC2	HC2	HC2

*Note: +=.1, *=.05, **=.01, ***=0.001*

We also estimated the impact of interviewing on SNAP take-up among compliers using a Local Average Treatment Effect estimator (Imbens & Angrist, 1994). Among those who complied (i.e., when texted showed up for the interview, and when only received traditional mailers did not show up for the interview), it is possible to estimate the impact of interviewing on SNAP take-up using two-stage least squares instrumental variable estimation. In this case, we used the texting assignment variable as an instrumental variable since it is associated with the interview status but not with other confounders. The results are summarized in the following table. Interviewing increased the approval rate by 61-67 percent (p-value < 0.01). Note that we used the county-provided interview status variable as an instrument. We did not use the state-provided interview variable because it includes rescheduled interviews, which made it difficult to estimate how texting caused variation in the interview completion status.

Table 2. Impact of Interviewing on Take-up Among Compliers

	Approved	Approved (procedural denials imputed)	Approved during initial applications (procedural denials imputed)
(Intercept)	0.0016 [-0.2542, 0.2574] p = 0.9899	0.0016 [-0.2443, 0.2474] p = 0.9899	-0.1533 [-0.4143, 0.1077] p = 0.2495
Interview status	0.6146** [0.1892, 1.0400] p = 0.0047	0.6147** [0.2014, 1.0279] p = 0.0036	0.6693** [0.2322, 1.1065] p = 0.0027
Num.Obs.	1517	1533	1536
R2	0.197	0.202	-0.066
R2 Adj.	0.197	0.201	-0.067
Std.Errors	HC2	HC2	HC2

*Note: +=.1, *=.05, **=.01, ***=0.001. The interview status variable comes from Boulder County.*

Hypotheses 4 and 5 related to how the treatment affected learning costs of SNAP applicants, and rely on survey responses collected after the completion of the interview stage. Table 3 presents the results of the models used for analysis. Model 1 shows that treated SNAP applicants were, on average, 9 percentage points (low 95% CI: 0pp, high 95% CI: 18pp, p-value: 0.04) more likely to know that an interview is required (question 1) compared to those who were not treated. Model 2 indicates that the treated SNAP applicants were, on average, 11 percentage points (low 95% CI: 2pp, high 95% CI: 20pp, p-value: 0.01) more likely to believe they knew what they needed to do during the interview (question 2). In Model 3, we transformed the 1-4 scale interview question into a binary one (easy or difficult) to observe a clearer pattern. Model 4 utilized an index variable that averages the two questions as an outcome variable to enhance reliability. The results consistently show that texting reduced the perceived learning costs for individuals.

Table 3: Impact of Texting on Learning Costs

	Know interview	Interview easy (4-point)	Interview easy (binary)	Learning cost index
(Intercept)	0.5171*** [0.4528, 0.5814] p = 0.0000	0.4116*** [0.3540, 0.4692] p = 0.0000	0.4087*** [0.3180, 0.4994] p = 0.0000	0.4362*** [0.3699, 0.5026] p = 0.0000
Treated	0.0931* [0.0034, 0.1828] p = 0.0420	0.1096* [0.0220, 0.1972] p = 0.0144	0.1277+ [-0.0031, 0.2584] p = 0.0556	0.1335** [0.0370, 0.2299] p = 0.0069
Num.Obs.	470	225	225	225
R2	0.009	0.027	0.016	0.032
R2 Adj.	0.007	0.022	0.012	0.028
AIC	676.4	146.8	328.1	191.2
BIC	688.8	157.1	338.4	201.4
RMSE	0.49	0.33	0.50	0.37
Std.Errors	HC2	HC2	HC2	HC2

*Note: +=.1, *=.05, **=.01, ***=0.001. The response rate for each question in the two-way text survey varies as respondents can opt out at any time. The initial question in the survey was about the respondents' knowledge of the requirement for SNAP interview, so it had a higher response rate than the question about the ease of the interview.*

Discussion: Nudging with News You Can Use

Prior work has pointed to mandatory interviews being an administrative checkpoint that causes people to exit out of processes and associated benefits that they otherwise desire. While ordeal mechanism models (Zeckhauser, 2021) suggest that this might reflect different elasticities in preferences for benefits - the people unwilling to deal with the interviews simply don't want the associated benefits as much as those that do – empirical work suggests otherwise, with a failure to overcome interviews driven partly by time constraints (Homonoff & Somerville, 2021). In our study, the interviews remain mandatory, but relaxing time constraints by allowing clients to reschedule or call anytime dampened the negative effects of interviews on completion of interview and benefit approval. The empirical findings suggest that even when checkpoints cannot be removed, there are ways to make them more forgiving for clients.

We consider why the intervention was relatively successful. At one level, our experiment can be considered as testing the effect of adding texts to traditional mailers. One interpretation is that the results reflect the sheer volume of information rather than mode or content. For some clients, the texts may simply have arrived earlier than the mailers, thereby giving clients more time to manage the administrative checkpoint. The earlier interview completion times among the treatment group is consistent with such an interpretation. However, we believe that the different content of the information within the texts also help to explain the results. This content made applicants not just aware that they could reschedule their appointments, but offered a new "call anytime" option and framed the choice in positive terms ("at your convenience"). Even the ability to complete the interview earlier depended upon being able to leverage non-standard appointment processes. Such encouragement and conveying of options should increase confidence about managing the administrative checkpoint.

Our findings come amidst something of a recalibration of the value of informational nudges. We interpret the relatively large (cf. DellaVigna & Linos, 2022) effects of our findings as reflecting the point that the efficacy of nudges depends upon the context and content. For example, informational nudges are more likely to be useful when learning costs matter a great deal, as is the case here (Herd et al., 2023). Our treatment group was not just reminded about the interviews but also given information about how to reschedule. Giving clients information and tools to manage a necessary requirement is “news they can use”: information that helps them to solve a large and real problem of inflexible and time-constrained interviews.

This interpretation is consistent with prior evidence on the impact of mandatory interviews (Homonoff & Somerville, 2021), and other evidence on the impact of making interviews more flexible (Giannella et al., 2024). Indeed, the effect size we observe (6-7 percentage points) on approval rates is very similar to those seen by Giannella et al.’s study (6 percentage points). In both cases, new SNAP applicants received both standard messaging about an appointment, while the treated group was offered information about flexible scheduling options. The treatment even increased client beliefs that it was easy to know what to do at an interview. Since the texts provided minimal information about the content of the interview, this result might be interpreted as clients feeling more security and confidence about their interactions with the state as a result of the communication, rather than increased knowledge of how to actually engage during the interview.

Administrative burden research acknowledges federalism as adding potential veto points and variation in burden outcomes, generally focusing on tensions between states and the federal government (Herd et al., 2023). Our study adds additional insight about a state-local tension in the delivery of a federally funded program. The state government designed an integrated benefit website and determines how it functions. However, the services are delivered at a lower level. As a result, state level actors cannot establish a full understanding of the connection between the application process and its consequences, or observe the impact of experimentation. Conversely, the county is heavily involved in case management. They are aware of how interviews are conducted and know who completed interviews and earned benefits. As noted above, the county could not change mailers to inform clients about a county innovation on flexibility that offered a “call anytime” option. This reflects a reasonable concern from the state about consistency in messaging, but constrains the capacity of the county to communicate and experiment. No single actor possesses all the knowledge pertaining to the benefit application journey and the power to reshape that journey. Such dispersions of responsibility and knowledge across different levels of government makes it more challenging to coordinate change efforts.

Conclusion

Our analysis finds that a text message that provided information about a mandatory interview process and offered flexible interview opportunities substantially increased interview completion and take-up of benefits. Some caveats apply about the nature of the sample and inferences about the causal mechanisms that lead to the result. Our study took place in one county in a single state, with a relatively homogenous population. However, the results are consistent with Giannella's et al.'s (2024) findings in the more diverse Los Angeles County. We interpret the effects as reflecting a bundled intervention, combining more timely and better communication and interview flexibility. Such an approach is common with much behavioral research, especially in real-world settings where there is little prior evidence. Nevertheless, future research could do more to establish the degree to which the volume and timing of information versus awareness of the flexibility of interviews drive the results.

Another generalizability concern centers on administrative capacity and the deployment of different tools in our case (Pierce & Moulton, 2023). The costs of texting are minimal, typically ranging between 1 and 5 cents per text depending on service used (Code for America, 2022). But other costs may be more forbidding. The analysis involved human-centered design tools of discovery and journey mapping, along with social science tools of evaluation. In practice, these combined resources are not easily available to many governments, especially at the local level. The county was not charged for technological experience and human-centered design skills offered by CfA, or for the expertise of data scientists who managed the evaluation component. Boulder is a reasonably large and well-resourced county. Missed interviews were a significant problem in tying up staff time, making a flexible interviewing approach more attractive. But they needed external help to assess if it made a difference for clients.

The importance of the desire to experiment at the county level cannot be overstated. Indeed most behavioral interventions, even successful ones, are not adopted by local governments and adoptions are largely driven by evidence supporting existing practices rather than new interventions (DellaVigna, Kim, & Linos 2022). Without county leadership and knowledge, the discovery phase would have floundered, administrative data would not have been available, and the experiment would not have worked. Pursuing more ambitious solutions to reduce the effects of state actions on the experience of burdens will often require partnerships with governments willing to change internal administrative processes. For other governments to incorporate such practices requires either building internal capacity, or developing external partnerships of the kind described in this case, or a combination of the two.

For the field of public administration and policy, incorporating tools of human-centered design offers both an educational skill to incorporate into education of students and practitioners, as well as a research tool. Future research could seek to validate both, i.e., evaluate the impact of human-centered designing training and tools at an aggregate level, as well as study them in the context of specific changes efforts. It could also consider how to build capacity for this skill within government, as well as identify barriers to doing so. For example, traditional public sector procurement processes specify the delivery of a specific product at the outset, a model of managing that is not conducive to processes of discovery and iterative experimentation (Pahlka, 2023).

References

Blevins, C., & Mullen, L. (2015). Jane, John... Leslie? A historical method for algorithmic gender prediction. *DHQ: Digital Humanities Quarterly*, 9.

Burden, B., Canon, D. Mayer, K., & Moynihan, D. (2014). Election laws, mobilization, and turnout: The unanticipated consequences of election reform. *American Journal of Political Science*, 58, 95-109.

Chohlas-Wood, A., Coots M., Nudell J., Nyarko J., Brunskill E., Rogers T., & Goel, S. (2023). Automated reminders reduce incarceration for missed court dates: Evidence from a text message experiment. *arXiv preprint arXiv:2306.12389.*

Code for America. (2022). Texting playbook: Basics of texting safety net clients. Retrieved March 23, 2024, from *https://codeforamerica.org/resources/texting-playbook.*

DellaVigna, S., & Linos, E. (2022). RCTs to scale: Comprehensive evidence from two nudge units. *Econometrica*, 90, 81-116.

DellaVigna, S., Kim W., & Linos, E. (2022). Bottlenecks for evidence adoption. NBER Working Paper No. 30144. National Bureau of Economic Research.

Finkelstein, A., & Notowidigdo, M. (2022). Take-up and targeting: Experimental evidence from SNAP. *The Quarterly Journal of Economics,* 134, 1505-1556.

Fox, A., Stazyk, E., & Feng, W. (2020). Administrative easing: Rule reduction and medicaid enrollment. *Public Administration Review,* 80, 104-117.

Giannella, E., Homonoff, T., Rino, G., & Somerville, J. (2024). Administrative burden and procedural denials: Experimental evidence from SNAP. American Economic Journal: Economic Policy. Retrieved March 23, 2024, from https://www.aeaweb.org/articles?id=10.1257/pol.20220701

Harrell, C. (2020). A civic technologist's practice guide. San Francisco, California: Five Seven Five Books.

Heinrich, C., Camacho, S., Henderson S., Hernández M., & Joshi, E. (2022). Consequences of administrative burden for social safety nets that support the healthy development of children. *Journal of Policy Analysis and Management,* 41, 11-44.

Heinrich, C. (2016). The bite of administrative burden: A theoretical and empirical investigation. *Journal of Public Administration Research and Theory,* 26, 403-420.

Heinrich, C. (2018). "A thousand petty fortresses": Administrative burden in US immigration policies and its consequences. *Journal of Policy Analysis and Management*, *37*, 211-239.

Herd, P., & Moynihan, D. (2019). Administrative burden: Policymaking by other means. New York, New York: Russell Sage Foundation.

Herd, P., DeLeire T., Harvey H., & Moynihan, D. (2013). Shifting administrative burden to the state: The case of medicaid take-up. *Public Administration Review,* 73, S69-S81.

Herd, P., Hoynes, H., Michener, J., & Moynihan, D. (2023). Introduction: Administrative Burden as a Mechanism of Inequality in Policy Implementation. *RSF: The Russell Sage Foundation Journal of the Social Sciences.* 9, 1-31.

Homonoff, T., & Somerville, J. (2021). Program recertification costs: Evidence from SNAP. *American Economic Journal: Economic Policy,* 13, 271-298.

General Services Administration and The Lab at OPM. (n.d.). Human centered design (HCD) discovery stage field guide V.1. Retrieved January 24, 2024, from https://www.gsa.gov/system/files/HCD-Discovery-Guide-Interagency-v12-1.pdf

Imbens, G., & Angrist, J. (1994). Identification and estimation of local average treatment effects. *Econometrica,* 62, 467-475.

Jilke, S., Bækgaard, M., Herd, P., & Moynihan, D. (2023). Short and sweet: Measuring experiences of administrative burden. *Journal of Behavioral Public Administration,* 7.

Kabbani, N., & Wilde, P. (2003). Short recertification periods in the US Food Stamp Program. *Journal of Human Resources,* 1112-1138.

Lin, J. (2023). Designing effective welfare programs: Evidence from SNAP's BBCE expansion. Retrieved March 23, 2024, from https://jouchunlin.github.io/jobmarket/jou-jmp-snap-bbce-paper.pdf

Linos, E., Quan, L., & Kirkman, E. (2020). Nudging early reduces administrative burden: Three field experiments to improve code enforcement. *Journal of Policy Analysis and Management*, 39, 243-265.

Linos, E., & Riesch, N. (2020). Thick red tape and the thin blue line: A field study on reducing administrative burden in police recruitment. *Public Administration Review,* 80, 92-103.

Linos, E., Lasky-Fink, J., Larkin, C., Moore, L., & Kirkman, E. (2023). The Formality Effect. *Nature Human Behavior,* 1-11.

Lopoo, L., Heflin, C., & Boskovski, J. Testing behavioral interventions designed to improve on-time SNAP recertification. *Journal of Behavioral Public Administration,* 3.

McGuinness, T., & Schank, H. (2021). Power to the public: The promise of public interest technology. Princeton, New Jersey: Princeton University Press.

Mergel, I. (2022). Human centricity in digital delivery: Enhancing agile governance. IBM Center for the Business of Government. Retrieved March 24, 2024, from https://www.businessofgovernment.org/sites/default/files/Human-Centricity%20in%20Digital%20Delivery-Enhancing%20Agile%20Governance.pdf.

Moynihan, D., Giannella, E., Herd, P., and Sutherland, J. (2022). Matching to categories: Learning and compliance costs in administrative processes. *Journal of Public Administration Research and Theory,* 32, 750-764.

Moynihan, D. (2023). The new progressives? The emergence of civic tech in the United States and its implications for governing. In the author's possession.

Mullen, L. (2021). Gender: Predict gender from names using historical data. R. Retrieved on March 24, 2024, from https://github.com/lmullen/gender.

Norman, D. (2024). Design for a better world: Meaningful, sustainable, humanity centered. Cambridge, Massachusetts: MIT Press.

Nowok, B., Raab, G. & Dibben, C. (2016). Synthpop: Bespoke creation of synthetic data in R. *Journal of Statistical Software*, 74, 1-26.

Pahlka, Jennifer. (2023). Recoding America: why government is failing in the digital age and how we can do better. New York, New York: Metropolitan Books.

Pierce, S., & Moulton, S. (2023). The effects of administrative burden on program equity and performance: Evidence from a natural experiment in a foreclosure prevention program. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *9*, 146-178.

Peeters, R., & Widlak, A. (2023). Administrative exclusion in the infrastructure-level bureaucracy: The case of the Dutch daycare benefit scandal. *Public Administration Review,* 83, 863-877.

Peeters, R., & Widlak, A. (2018). The digital cage: Administrative exclusion through information architecture–The case of the Dutch civil registry's master data management system. *Government Information Quarterly,* 35, 175-183.

Rappin, N., Bernius, M., Hirschtritt, D., Joshi A., kaur, t., Paredes, C., & Sutherland, J. (2020). The Code for America qualitative research practice guide. Code for America. Retrieved March 24, 2024, from https://f.hubspotusercontent30.net/hubfs/5622333/CFA_QualitativeResearchGuide_v1.pdf.

Ribar, D., Edelhoch, M., & Liu, Q. (2008). Watching the clocks: The role of food stamp recertification and TANF time limits in caseload dynamics." *Journal of Human Resources,* 43, 208-238.

Rinker, T. (2021). Sentimentr: Calculate text polarity sentiment. R. Retrieved March 24, 2024, from https://github.com/trinker/sentimentr.

Sackett, B., & Lareau, A. (2023). Institutional entanglements: How institutional knots and reverberating consequences burden refugee families." *RSF: The Russell Sage Foundation Journal of the Social Sciences,* 9, 114-132.

Sinai, N., Leftwich, D., & McGuire, B. (2020). Human-centered policymaking: What government policymaking can learn from human-centered design and agile software development. Harvard Kennedy School. Retrieved March 24, 2024, from https://www.belfercenter.org/sites/default/files/2020-04/HumanPolicyMaking.pdf

Sullivan, K. & Soka, S. (2022). Starting small with human-centered redesign." Beeck Center for Social Impact + Innovation and Civilla. Retrieved March 24, 2024, from https://beeckcenter.georgetown.edu/wp-content/uploads/2022/11/Small-ScaleHumanCenteredRedesign.pdf.

Sunstein, C. (2021). Sludge: What stops us from getting things done and what to do about it. Cambridge, Massachusetts: MIT Press.

Sweeney, L. K-anonymity: A model for protecting privacy. *International journal of uncertainty, fuzziness and knowledge-based systems,* 10, 557-570.

Templ, M., Kowarik, A., & Meindl, B. (2015). Statistical disclosure control for micro-data using the R package sdcMicro. *Journal of Statistical Software,* 67, 1-36.

Zeckhauser, R. (2021). Strategic sorting: The role of ordeals in health care. *Economics & Philosophy,* 37, 64-81.

Appendix

Appendix A. Covariate balance checks

Table A.1. Covariate balance check based on the actual data

		Control Mean	Control Std. Dev.	Treatment Mean	Treatment Std. Dev.	Diff. in Means	Std. Error
Age		35.93	13.36	35.17	13.20	-0.76	0.61
Female		0.58	0.49	0.55	0.50	-0.02	0.02
Homeless		0.11	0.32	0.12	0.32	0.00	0.01
Spanish application		0.03	0.17	0.03	0.16	0.00	0.01
Expedited application		0.02	0.14	0.01	0.12	-0.01	0.01
Number of programs applied		1.19	0.42	1.24	0.49	0.05	0.02
Number of people on applications		1.10	0.48	1.13	0.58	0.03	0.02
Missed interviews		0.30	0.46	0.29	0.45	-0.01	0.02
Missed documentation		0.11	0.31	0.10	0.30	-0.01	0.01
SNAP denied		0.55	0.50	0.56	0.50	0.02	0.02
		N	Pct.	N	Pct.		
	Boulder	968	100.0	919	100.0		
	PEAK	968	100.0	919	100.0		

Source: Colorado Benefits Management System (January-June, 2022)

To protect applicants’ privacy and comply with the data-sharing agreement between Code for America and the government partner, we created the synthetic data based on the actual data used for covariate balance checks. We decided that the observational data we analyzed to this end gives away too much information about SNAP applicants. The synthetic data were created by using the “synthpop” package in R programming language (Nowok, B., G.M. Raab, and C. Dibben 2016): https://cran.r-project.org/web/packages/synthpop/index.html

As Figures A.1-4 demonstrate, the synthetic data closely follow the actual data for their distributions. Tables A.1-2 confirm that the covariate balanced outputs from these two datasets are almost identical.

We will share this synthetic data for replication purposes. We also logged that we executed the R markdown file containing the covariate balance checks. The log file provides evidence that we created balance tables based on the actual and synthetic data.

Figure A.1. Distribution comparison between actual (light blue) and synthetic data (dark blue) for applicants' age, expedited applications, and the number of programs applicants applied for.

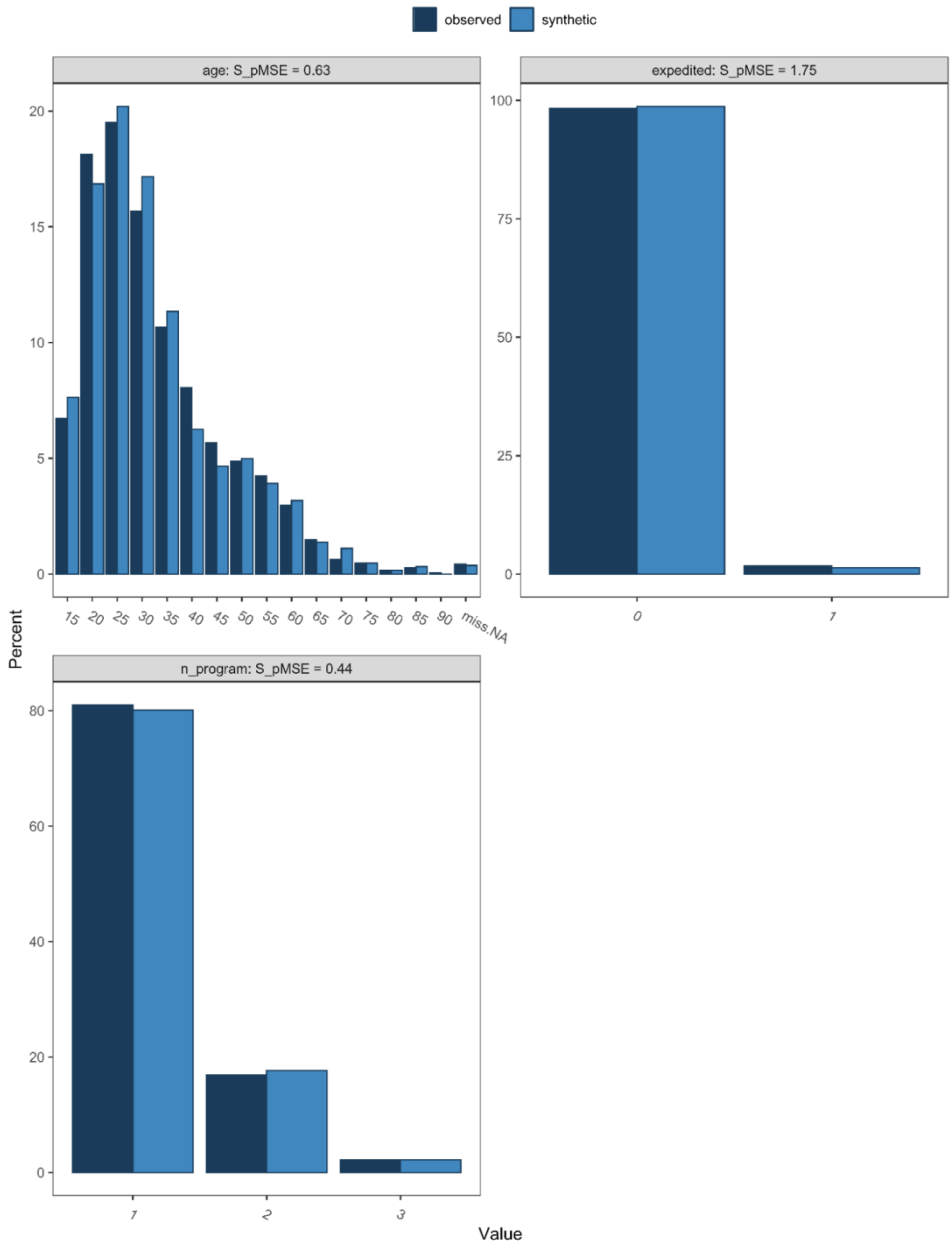


Figure A.2. Distribution comparison between actual (light blue) and synthetic data (dark blue) for the number of people on applications, missing interview rates, and missing verification rates.

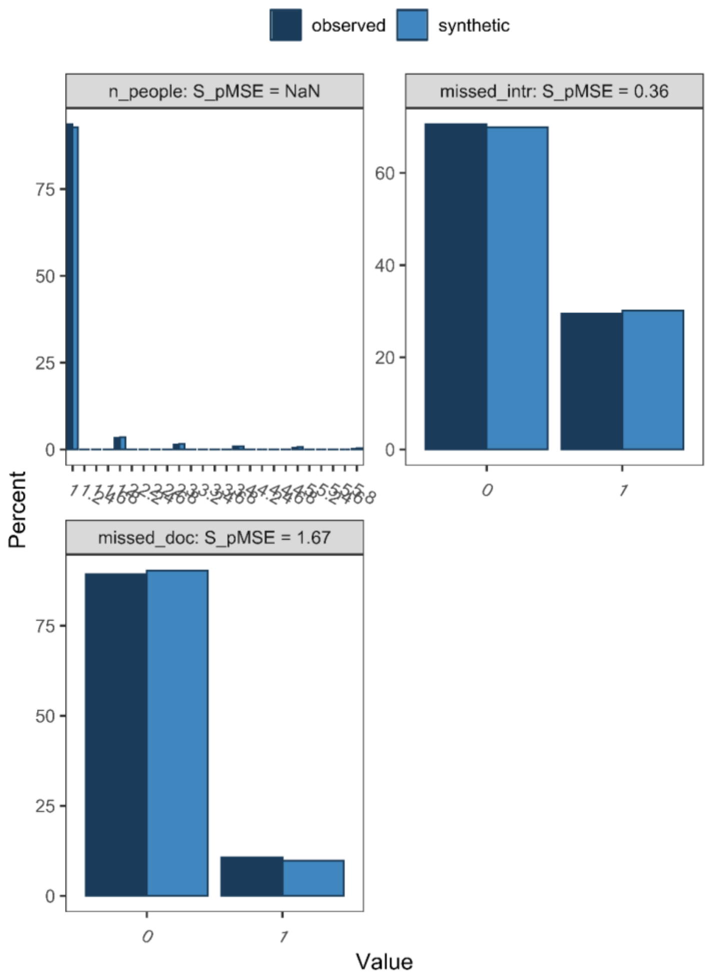


Figure A.3. Distribution comparison between actual (light blue) and synthetic data (dark blue) for applications applied for multiple programs, applications with multiple people, denial rates, and assignment statuses.

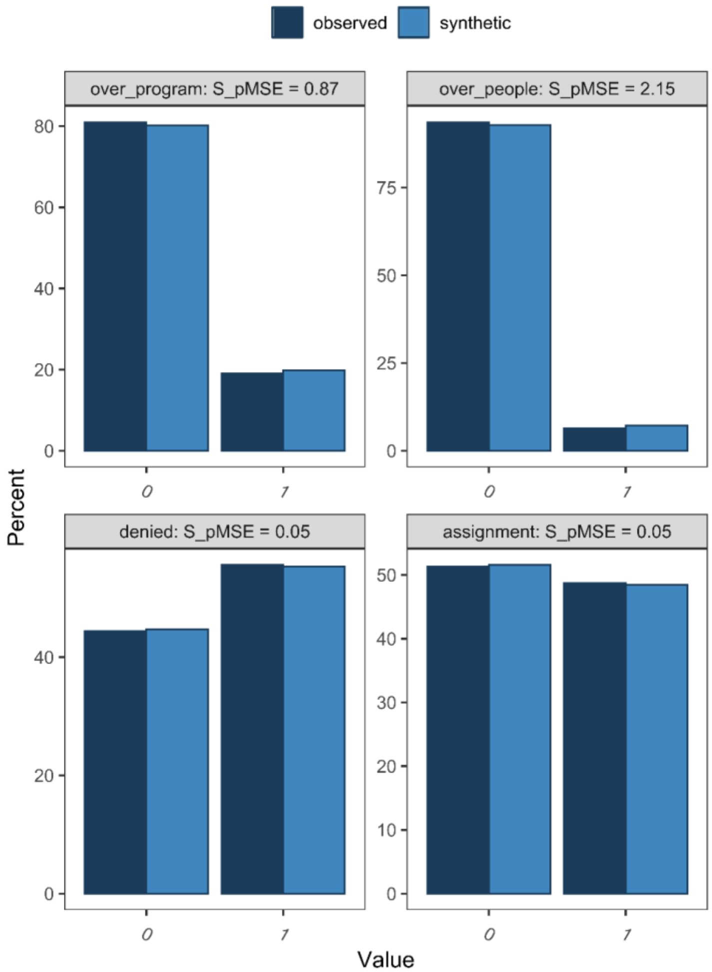


Figure A.4. Distribution comparison between actual (light blue) and synthetic data (dark blue) for female, homeless status, and Spanish preference.

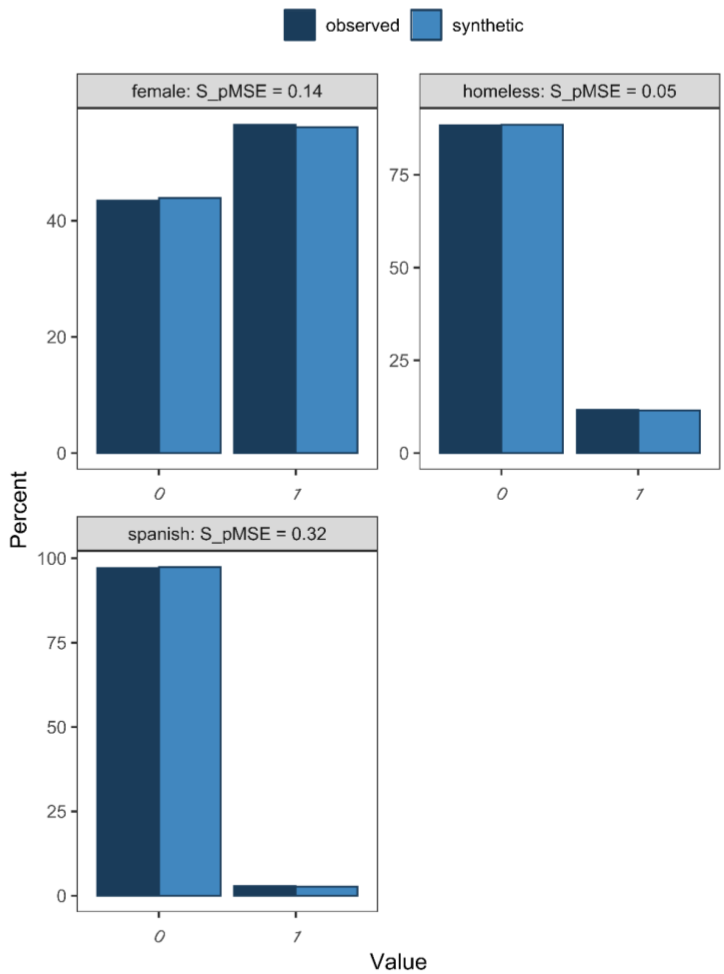


Table A.2. Covariate balance check based on the synthetic data

		Control Mean	Control Std. Dev.	Treatment Mean	Treatment Std. Dev.	Diff. in Means	Std. Error
Age		35.39	12.99	35.26	13.92	-0.13	0.62
Female		0.56	0.50	0.56	0.50	0.00	0.02
Homeless		0.11	0.31	0.12	0.33	0.01	0.01
Spanish application		0.03	0.16	0.03	0.16	0.00	0.01
Expedited application		0.01	0.11	0.01	0.11	0.00	0.01
Number of programs applied		1.21	0.46	1.23	0.47	0.01	0.02
Number of people on applications		1.13	0.58	1.16	0.65	0.03	0.03
Missed interviews		0.31	0.46	0.29	0.45	-0.03	0.02
Missed documentation		0.10	0.30	0.09	0.29	-0.01	0.01
SNAP denied		0.57	0.50	0.54	0.50	-0.03	0.02
		N	Pct.	N	Pct.		
	Boulder	973	100.0	914	100.0		
	PEAK	973	100.0	914	100.0		

Source: Synthetic data

Appendix B. County mailer content

Dear _____,

We have received your application or recertification for assistance. An interview is required for ______(program)_____.

You have been scheduled for a phone interview at (time) on (date). Please call (303) 441-1660 at this time. Please contact our office at (303) 441-1000 if you like a face-to-face interview instead of a phone interview.

If you miss your appointment or need to reschedule, you must contact us at (303) 441-1660 to reschedule your interview. If you do not complete a required interview for your application or recertification, your benefits will be denied.

Appendix C. Text reminders

In the text below, Code for America and Boulder couldn't mention SNAP specifically, but instead used PEAK, a broader term, to comply with the FCC guideline on telephone consumer protection (TCP), particularly in the context of human and health services.

Initial text (sent 24 hours after initial application)

Boulder County:Hi {First Name} we've processed your PEAK application!

Please call 303-441-1660 in the next week to complete a phone interview. The interview should take about 30 minutes and you'll learn if you qualify for benefits.

Our hours: 8:30am-3pm, Mon-Fri

Learn more about the process here: http://www.boco.org/SNAP-FAQ

First reminder (sent 3 days after initial application)

Boulder County: Hi {First Name} reminder that you have the option to complete your interview at your own convenience by calling 303-441-1660 any time between 8:30am-3pm, Mon-Fri.

Learn more about the process here: www.boco.org/SNAP-FAQ

Second reminder (24 hours before scheduled interview)

Hi {First Name}, your PEAK phone interview is tomorrow. This should take about 30 minutes. Call us at 303-441-1660 during your scheduled time to complete your interview.

Can’t make it? Just call anytime Monday through Friday between 8:30am-3pm to conduct an unscheduled interview.

Learn more about the process here: www.boco.org/SNAP-FAQ

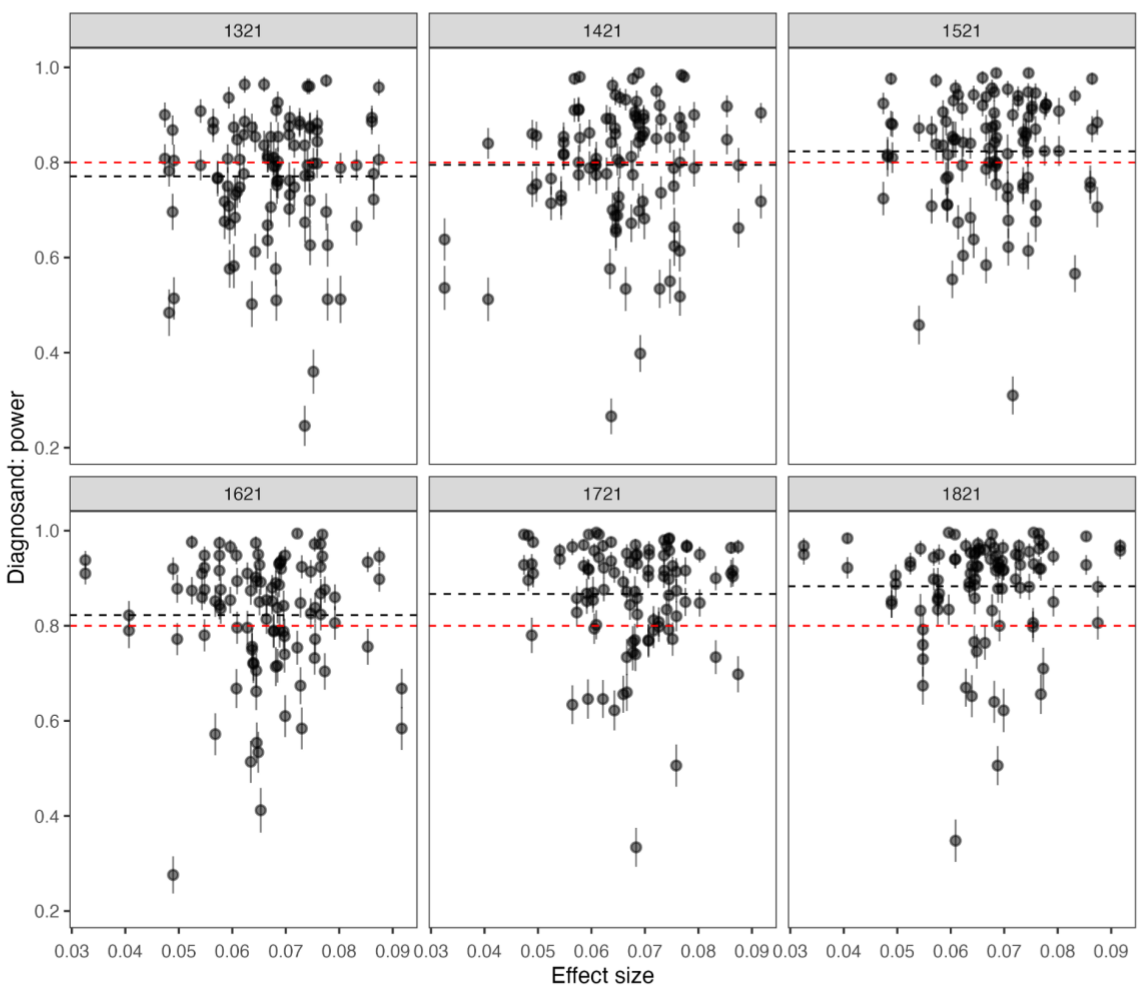
Appendix D. Power analysis for field experimentation

Figure D.1. Power analysis simulation for field experimentation

We calculated the sample size using CfA's prior research experience in California (Sacramento) and the similar experience from Hunger Free Colorado, a community-based organization, in Colorado (Arapahoe). Based on prior research we assumed the effect size would be between 0.05 and 0.08, following a normal distribution with a mean equal to the average of these two point estimates, and a standard deviation of 0.01. Based on these assumptions, we used bootstrap simulations to examine the relationship between effect size and statistical power. Below are the simulation results. Each data point is based on 500 bootstrap simulations.

- Red line: represents the power for the minimum detectable power (the power that yields a p-value equals or smaller than 0.05).
- Black line: represents the weighted average power from a specific set of simulations.

As the sample size increased, the two lines started to converge, and eventually, the black line surpassed the red line.



Appendix E. Questionnaire

Respondents needed to opt-in before taking the survey.

First administrative burden question

Q1 Hi {First Name}, this is Boulder County. We see you applied for PEAK. **When you applied, did you know you had to do an interview?** Your answer will help us improve and won't change your status. Messages and date rates may apply. Text STOP to no longer get surveys. Text STOP ALL to unsubscribe from all messages from us.

- ◯ YES
- ◯ NO

Q2 Thanks! Would you be ok to answer a few more questions? This should only take 5 minutes.

- ◯ YES
- ◯ NO

Second administrative burden question

Q3 **How easy or difficult was it to know what to do for your interview?** Text the NUMBER that matches your answer

- ◯ It was very easy
- ◯ It was easy
- ◯ It was difficult
- ◯ It was very difficult

Q4 is only for those completed their interviews

Q4 Almost done. If you're comfortable, **would you mind telling us how you felt during your interview?** You can reply to this question by texting us in your own words OR just text SKIP to go to the last question

Q5 is only for those missed their interviews.

Q5 Almost done. We see you need to do an interview. **Are any of the reasons below why you haven't done one yet?** Text ONE NUMBER that matches your reason.

- ○ I have one scheduled, but it hasn't happened yet
- ○ I've been too busy
- ○ I have issues with my phone or connection
- ○ I thought someone would call me
- ○ No, it's something else

Q6 Last question. Would you be willing to talk more to someone from our team about your experience? You would be paid for your time.

- ○ YES
- ○ NO

Q7 Thanks for sending us a response. When you're done saying everything you need to say, you can text I'M READY to answer the next question

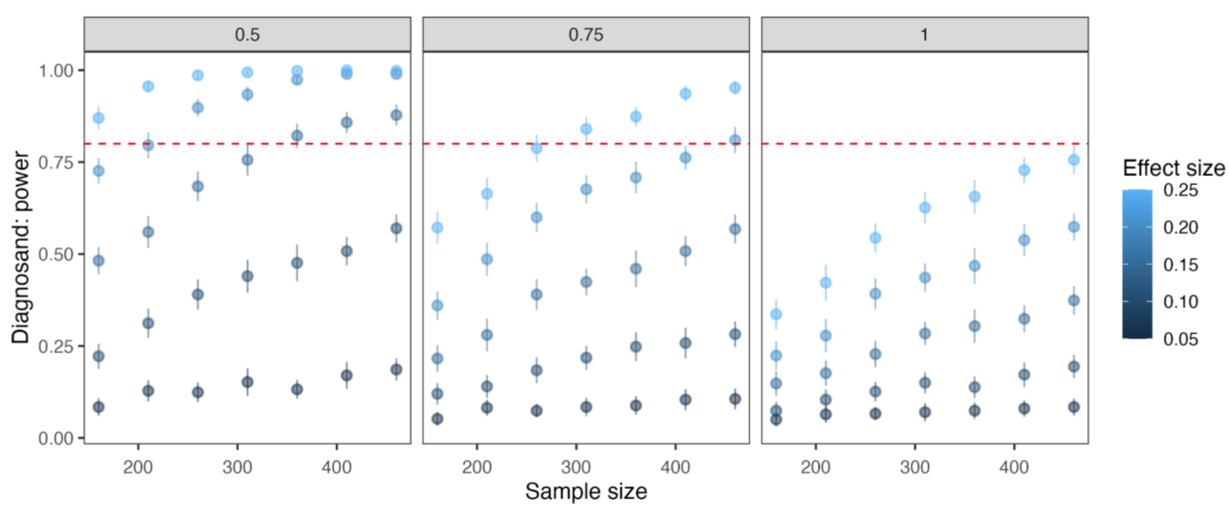
Appendix F. Power analysis for follow-up survey

We simulated the relationship between sample size and statistical power, taking into account the expected effect size (ranging from 0.05 to 0.25) and the standard deviation of the outcome measures (ranging from 0.5 to 1). The results show that a sample size of around 300 would be sufficient to achieve 80% statistical power.

Below are the simulation results. Each data point is based on 500 bootstrap simulations.

- Red line: representing the desired power level of 0.8 (which corresponds to a p-value of 0.05).
- Interpretation: As the effect size (legend) increases, the power also increases (y-axis). Similarly, as the sample size (x-axis) increases, the power also increases (y-axis). Finally, the power increases as the standard deviation of Y (panels) decreases.
-

Figure F.1. Power analysis simulation for follow-up survey



Appendix G. Response rates for the survey questions

- Q1 = “Know the interview requirement”
- Q2 = the 4 point administrative burden scale

Figure G1. Response rates by the treatment status

	Control Mean	Control SE	Treated Mean	Treated SE	Diff. in Means	p-value
Q1 response rate	19%	1%	20%	1%	1%	0.810
Q2 response rate	15%	1%	14%	1%	-1%	0.756

Appendix H. Heterogeneous treatment effect

In this analysis, we estimated the conditional ITT by considering the applicants' language preferences, gender, homeless status, and household size as covariates.

1. Language and gender: The county provided us with the language preference, but not the gender category of each applicant. In addition, we could not retrieve the gender information for 28% of the applicants in the state dataset. To address this, we used the clients' names to infer their presumed gender using the “gender” package in R (Mullen et al. 2021): https://cran.r-project.org/web/packages/gender/index.html The model uses historical databases of the first names and dates of birth to make these inferences. We validated the accuracy of this prediction model using the existing gender information, which yielded a 94% accuracy rate.
2. Homeless status: 28% of the applicants’ homeless status information was missing in the sample. 9% of the applicants in our study were homeless.
3. Household size (exact matching): Since the administrative eligibility outcome table does not provide applicants’ income, we used their household size as a proxy variable. The applicants’ average household size is 1.3 in the sample. 38% of the applications’ household size information was missing in the sample when we combined the study data and administrative data based on the applicants’ case identifiers and their application dates (exact matching).
4. Household size (fuzzy matching): Assuming that the applicants’ household size didn’t change substantially within the three months, we imputed the missing values using their case identifiers even though their application dates did not match (fuzzy matching). This exercise reduces the missing rate from 38% to 2%. The applicants’ average household size in this imputed data is also 1.3.

In the following regression models, the standard errors were calculating using the HC2 standard errors.

Table H.1.Heterogeneous treatment effects on interview completion rates

	Language preference	Gender	Homeless status	Household size (exact matching)	Household size (fuzzy matching)
Treatment status	0.107*** [0.058, 0.156] p = <0.001	0.120*** [0.059, 0.182] p = <0.001	0.095** [0.034, 0.155] p = 0.002	0.179* [0.021, 0.337] p = 0.026	0.251*** [0.123, 0.380] p = <0.001

Spanish	-0.003 [-0.210, 0.204] p = 0.977				
Treatment status × Spanish	-0.064 [-0.377, 0.248] p = 0.687				
Male		-0.060+ [-0.131, 0.011] p = 0.096			
Treatment status × Male		-0.031 [-0.131, 0.070] p = 0.551			
Homeless status			-0.047 [-0.184, 0.090] p = 0.503		
Treatment status × Homeless status			0.097 [-0.098, 0.293] p = 0.329		
Household size				0.426*** [0.346, 0.507] p = <0.001	
Treatment status × Household size				-0.090 [-0.202, 0.023] p = 0.119	
Household size (imputed)					0.421*** [0.353, 0.488] p = <0.001

Treatment status × Household size (imputed)					-0.123** [-0.215, -0.030] p = 0.009
Num.Obs.	1556	1523	1115	967	1518
R2	0.012	0.018	0.012	0.162	0.144
R2 Adj.	0.010	0.016	0.009	0.159	0.143
AIC	2185.4	2128.7	1574.3	1140.5	1912.1
BIC	2212.2	2155.3	1599.4	1164.8	1938.8
Log.Lik.	-1087.716	-1059.327	-782.149	-565.225	-951.065
F	6.201	9.425	4.530	61.893	85.162
RMSE	0.49	0.49	0.49	0.43	0.45

Table H.2.Heterogeneous treatment effects on approval rates

	Language preference	Gender	Homeless status	Household size (exact matching)	Household size (fuzzy matching)
Treatment status	0.066** [0.018, 0.115] p = 0.008	0.081** [0.020, 0.143] p = 0.010	0.047 [-0.012, 0.106] p = 0.118	0.091 [-0.068, 0.250] p = 0.261	0.158* [0.034, 0.282] p = 0.013
Spanish	0.029 [-0.175, 0.233] p = 0.781				
Treatment status × Spanish	-0.097 [-0.417, 0.223] p = 0.553				

Male		0.017 [-0.053, 0.087] p = 0.637			
Treatment status × Male		-0.030 [-0.131, 0.070] p = 0.553			
Homeless status			0.036 [-0.097, 0.170] p = 0.592		
Treatment status × Homeless status			0.178+ [-0.012, 0.369] p = 0.066		
Household size				0.446*** [0.365, 0.527] p = <0.001	
Treatment status × Household size				-0.019 [-0.132, 0.095] p = 0.747	
Household size (imputed)					0.458*** [0.393, 0.523] p = <0.001
Treatment status × Household size (imputed)					-0.084+ [-0.173, 0.005] p = 0.065
Num.Obs.	1533	1501	1115	965	1516
R2	0.005	0.005	0.013	0.197	0.186

R2 Adj.	0.003	0.004	0.011	0.195	0.184
AIC	2116.0	2069.4	1509.5	1149.9	1793.9
BIC	2142.7	2096.0	1534.6	1174.3	1820.5
Log.Lik.	-1053.016	-1029.706	-749.759	-569.948	-891.956
F	2.397				115.080
RMSE	0.48	0.48	0.47	0.44	0.44

Appendix I. Data privacy protection

The replication data does not contain any personally identifiable information (PII), such as the applicants' either case or application identifiers. We took additional steps to process, generalize, and suppress data to strengthen anonymity. For instance, the interview timing field in the replication data only refers to the day gaps between the scheduled and completion dates. The actual dates are unknown.

In addition, we used the "sdcMicro" package in R (Templ et al. 2024) to test the degree of anonymity: https://cran.r-project.org/web/packages/sdcMicro/index.html We did so because combining these anonymous fields could help identify the study participants. We assumed that applicants' language preference, gender, homeless status, and household size could be risk factors.

The test shows that the number of observations that violate 2- or 3-anonymity is 0, and the number of observations that violate 5-anonymity is only 1 (6%). In this context, "2," "3," or "5" refers to the number of people sharing similar characteristics within a group. This concept of privacy is called k-anonymity (Sweeney 2002): https://dl.acm.org/doi/10.1142/S0218488502001648 This *k* number matters for privacy protection because it indicates that it is difficult for an intruder to identify individuals in the dataset unless they know at least k-1 individuals in the same group.

We also considered creating synthetic data but didn't apply it here as the method didn't produce reliable results that show similar distributions between actual and synthetic data. To provide more clarity, in Appendix A, we employed a synthetic data approach to check covariate balances based on the administrative records of the Colorado government's Benefits Management System (CBMS) from January to June 2022. In this appendix (Appendix I), we applied the same approach to our field experiment and survey data collected from March 22 to June 22, 2023. We speculate that it may not have worked because some covariates (such as household size) included imputed values, as described in the main text and Appendix H.

Appendix J. Pre-registration

Code for America uses the Hex platform (https://hex.tech/) for internal data analysis as it provides industry-grade security and privacy protections. Code for America's research conducted a pre-analysis of the field experiment, including covariate balance check, power analysis, and estimation strategies. This platform is not publicly accessible as it is designed to be used for internal purposes. However, we share the screenshot of the Hex notebook (similar to a Jupyter notebook), which contains the pre-analysis and its timestamp (March 23, 2023).

In Progress Feedback

Boulder pilot study

Published **Mar 23, 2023** by **Jae Yeon Kim**

This experiment, which is taking place in Boulder, Colorado, is designed to measure the effectiveness of a text messaging campaign in increasing the rate of completion of interviews for the Supplemental Nutrition Assistance Program (SNAP). The goal of this campaign is to encourage more individuals to participate in SNAP by making it easier for them to complete their interviews. The experiment will involve sending text messages to a group of participants, reminding them of their upcoming appointments and providing them with useful information about the interview process. By doing so, the hope is to increase the overall completion rate of SNAP interviews and ultimately improve the lives of those who rely on this important program.

Jae Yeon Kim (jykim@codeforamerica.org)

[Insert date]

- Program: SNAP
- Area: Feedback
- State: Colorado
- Methods: simulations and experimental data analysis