Journal of Behavioral
Public Administration
Vol 6, pp. 1-22
DOI: 10.30636/jbpa.61.349

Research Article

Decreasing Compliance Costs Using Behavioral Interventions: An Evaluation of Lead Testing in Chicago

Sarah Souders[*], Leonard M. Lopoo[*], Sheena Mirpuri[†], Patrick Schwer[‡], and Emily Cardon[†]

Abstract: Exposure to lead is a well-documented public health concern. One source of exposure is water transported via lead pipes. As the city with the most lead service lines, Chicago offers at-home lead testing. However, at-home testing is a multi-step process with considerable compliance costs for residents who choose to engage with the city's service. The onerous process is reflected in low return rates of requested lead test kits. This study has two components designed to reduce compliance costs and increase return rates: a randomized control trial (RCT) testing the effectiveness of a text message reminder on lead test kit returns and a pre/post evaluation of the redesign of instructions for completing the lead testing kit. Results show the text reminder increased test kit returns by around 67 percent and the redesigned kits increased use by 20 percent. Subsequent analyses show that the text message intervention was effective across the income distribution. These large, universal effect sizes indicate the usefulness of behavioral interventions to reduce resident compliance costs in multi-step processes at the local level.

Keywords: Compliance costs, RCT, Behavioral intervention, Simplification, SMS reminders

---

Introduction

Exposure to lead is a well-documented public health concern.[i] There is no known safe level of lead, and exposure to lead risks lower academic achievement, behavioral problems, slowed growth, and anemia among children (Hollingsworth et al., 2020; Sorensen et al., 2019; O'Connor et al., 2018; Hauptman et al., 2017). While high levels of exposure have been limited by banning the use of lead in paint and gasoline, lower doses of lead can still be found in food, air, toys, contaminated soils, and water (Dietrich et al., 2021; EPA, 2015). Those living in places with leaded pipe and solder (permitted before 1986) are at a high risk as lead ions leach into the water they transport (Clark et al., 2015).

The EPA estimates there are still 6 to 10 million lead services lines across the country, with the City of Chicago having the most (EPA, 2023). In 2022, city crews replaced less than one-half of 1% of the approximately 400,000 lines responsible for contaminating Chicagoans' tap water, making small progress toward mandated replacement of all lead service lines by 2077 (Chase, 2023).

In response, Chicago currently operates an at-home lead test kit program that alerts residents to exposure and provides the City the evidence needed to request replacement service lines. At-home testing is a key part of Chicago's strategy to reduce lead exposure as use of lead pipes varies property-to-property, rather than at the street or neighborhood level. While lead test kits are free and mailed

* Maxwell School, Syracuse University
† The Behavioral Insights Team
‡ City of Chicago Department of Water Management
Address correspondence to Sarah Souders at sasouder@syr.edu.

directly to residents that request them, correctly testing for lead and arranging pick-up of the completed lead tests, requires the execution of several multi-step processes, which can be conceptualized as compliance costs in an administrative burden framework (Burden et al., 2012; Heinrich 2018; Herd and Moynihan 2018; Moynihan et al., 2015).

This paper contributes to the behavioral public administration literature by reporting results from a randomized control trial and, separately, a simplification intervention to assess whether compliance costs among participants can be reduced with behavioral interventions. While there is much evidence that interventions to limit administrative burden increase program awareness, enrollment, and take-up (Herd et al., 2013; Remler et al., 2001), less is known about whether behavioral interventions effectively mitigate compliance costs for those that opt-into but now must move through multi-step government processes. In this study, we assess the effectiveness of two staggered interventions introduced to improve the return of requested lead testing kits in Chicago in 2022.[ii] First, residents that previously requested but did not return lead test kits were exposed to reminders sent as short message service (SMS) messages or what are commonly referred to as "texts". Second, initial test kits created by the city included complex instructions that made completing lead testing difficult. In the second intervention, the kits were redesigned to simplify instructions and included tools to increase compliance (Mirpuri et al., 2022). We evaluate whether lead test kit returns increased following SMS message and kit instruction redesign treatments as well as whether the interventions prompted more interaction with Chicago's Department of Water Management (DWM). Additionally, we consider whether compliance costs were especially eased for residents most at-risk of lead exposure.

Literature Review

Administrative burden, or the costs of government interaction, is typically divided into three separate categories: learning costs, compliance costs, and psychological costs (Herd & Moynihan 2018; Moynihan et al. 2015). Learning costs occur as individuals seek information necessary to participate in government services. Compliance costs accumulate as participants attempt to follow the requirements for participation, and psychological costs arise as participants experience stigma, feelings of dependence, as well as stress and frustration that may arise when navigating bureaucratic processes.[iii]

Empirical papers documenting administrative burden have focused on learning costs and participant take-up of social programs for which they are eligible (Heinrich, 2016; Nisar, 2018; Currie, 2006; Herd and Moynihan, 2019; Moynihan et al., 2015). Attempts to limit compliance costs have been effective. Simplified reporting and recertification led to an increase in successful claims (Kabbani & Wilde, 2003; Ratcliffe et al., 2007), simplified forms and application processes (Kopczuk & Pop-Eleches, 2007; Leininger et al., 2011; Schwabish, 2012), and in-person assistance and auto-enrollment significantly increased take-up (Aizer, 2003; Bettinger et al., 2012; Bhargava & Manoli, 2015; Dorn et al., 2009; Herd et al. 2013; Schanzenbach 2009).

Our paper contributes to a growing literature that considers the potential of behavioral interventions to increase compliance among participants in existing burdensome systems. Behavioral nudges through SMS messages or behaviorally informed communication notices are effective ways to increase recertification rates among SNAP enrollees and increase compliance with complicated housing regulations (Lopoo et al., 2020; Linos et al., 2020). Beyond prodding participants through government interactions, behavioral science interventions may ease compliance by improving participant's understanding. Simplifying instructions and information gathering, for instance, have increased completion of multi-step processes. Written notices using personalization, indirect reciprocity, and simplified language have been especially successful (Linos et al., 2020; Lopoo et al, 2020; Behavioral Insights Team 2015; Haynes et al. 2013; Sanders & Kirkman, 2019; Yoeli et al., 2013). Furthermore, informing participants about the benefits of compliance and including planning prompts that emphasize social norms have also improved completion of multi-step processes (Lasky-Fink et al. 2021; Linos et al., 2020; Milkman et al. 2011). These tactics have been applied to increase completion of college financial aid applications, K-12 school choice applications, and applications for the Earned Income Tax Credit (Bettinger et al. 2012; Hastings & Weinstein 2008; Bhargava & Manoli 2015).

Behavioral interventions could also be a powerful tool to remedy inequality in experiences of administrative burden. Often, programs aimed to assist vulnerable groups impose costs that exclude many, decrease take-up, and lead to eventual drop-out (Heinrich & Brill, 2015; Brodkin & Majmundar, 2010). Many programs create negative citizen-state interactions (Barnes & Henly, 2018) and inequalities persist in positive interactions (Holt & Vinopal, 2019). If behavioral interventions are especially effective in moving vulnerable groups through a burdensome process, perhaps by decreasing citizens' decision-making bias (Chudnovsky & Peeters, 2021; Christensen et al., 2020), they could help achieve a program's goal and reduce inequality in experiences with government.

Examining the City of Chicago's Water Lead Test Kit program allows us to assess whether decreasing compliance costs using behavioral interventions improves test kit return, which types of interventions (SMS nudges and simplification) are effective, and whether interventions are especially successful for members of the program's target population. As we explain below, we define residents in low-income census tracts and in census tracts with old housing stock as the target population because research has found that children whose families are poor and who live in census tracts with old housing stock are at the highest lead poisoning risk (Jones et al. 2009; Sorensen et al. 2021). If the aim of lead test kits is to reduce exposure, we would expect the program to target low-income residents and residents in areas with old housing stock.

Lead Testing in Chicago

Since 1966, the Chicago Department of Public Health has offered lead testing, inspection, and abatement programs (Chicago Department of Public Health, 2016). The current program, where citizens request and execute at-home lead test kits, began in April 2016. Test kits are delivered to home residences by the United States Postal Service and are picked up by Chicago DWM staff once scheduled by the resident. The DWM dispatches about 15,000 test kits annually and most requested test kits are not returned.[iv] Before SMS reminders and simplified kits were introduced, test kits requested in January of 2022 had a 29.7 percent cumulative return rate after 111 days. Low return rates are costly—each test kit costs $12.50—and indicate compliance barriers for residents concerned about elevated lead levels in their water.

The test kit includes three bottles for sampling water at specific time intervals, a testing form providing information to the Chicago DWM staff about the samples collected and water service lines in the home, and instructions to avoid water use (including showers, toilets, and faucets) prior to testing for at least six hours.

Residents that request the kit may not return it for several reasons. They may not receive the kit, may lose it, or may never open it. Losing or forgetting about the requested test kit are especially likely as time since the request increases. Others may find the materials included in the test kit confusing. Some may face challenges carving out time to stagnate water for six hours, and so avoid testing. Those that test may never schedule a test kit for pickup, or the kit may be damaged or lost prior to being picked up by DWM staff. Some may complete testing but not complete the water testing form, so their kit is rejected. Residents that successfully complete lead testing are sent results by mail in six to eight weeks.

In preparation for this intervention, the Behavioral Insights team interviewed four Chicago DWM staff members as well as water quality staff in three peer cities (Cincinnati, DC, and NYC), and conducted two focus groups with a total of eight Chicago residents (five residents who had not returned their kits and three residents who had) to understand the most salient challenges to test kit return. The interviews and focus groups revealed that residents dropped out of the process at different steps, residents who returned their kits did so within a week or two, and that peer programs used reminders to encourage returns. These findings led to the development of the two interventions: redesigning the test kit and sending SMS reminders.

Lead Test Kit Interventions

To increase return rates, two behavioral interventions were introduced. The interventions were staggered, so residents were only ever exposed to one treatment at a time. First, among residents that received a requested lead test kit after September 1, 2021, but before March 30, 2022, a randomly selected group received a SMS message reminder in May 2022 to return their requested test kit (Mirpuri et al., 2022). Please see Figure 1 for the content of the SMS message. Second, residents that received a requested a lead test kit after March 30, 2022, were sent new kits. The kit's exterior, directions, and paperwork were redesigned using behavioral science principles to increase test kit completion and return.

> "DWM: If you still want to check the water in your home, complete the lead test kit we sent you and schedule a pickup in the next two weeks.
>
> Reply "s" to stop messages. Visit ChicagoWaterQuality.org or call (312) 742-2406 to get tips, request a new kit, or close your request.[1]"

Figure 1. Text Message

Intervention 1. SMS Messages

Using Chicago DWM administrative data, we identified Chicago residents who had requested a test kit between September 1, 2021, and March 29, 2022, but who had not returned them as of March 29, 2022. In this period, 5,427 tests were shipped. In total, 3,601 residents were included the trial; 1,800 received the SMS message and 1,801 did not.[v] Randomization occurred within three blocks or "sets" defined by the test kit shipment date. Those who had a test kit shipped in September 2021 and had not returned it by March 29, 2022, are defined as Set 3. Within Set 3, we randomly selected half of the test requestors to receive a reminder text and half, the control, received nothing. Those in the treatment group received the text on May 11, 2022. Similarly, those who had a test kit shipped between October and December 2021 constitute Set 2. We randomly selected half of the requestors in Set 2 to receive a text (sent on either May 17 or 18, 2022). The other half served as the control. Set 1 is composed of those who were shipped a test kit between January and March 2022 and had not returned it by March 30, 2022. We also randomly selected half of this set to receive a text, which was sent on May 19, 2022.

Empirical Approach

We run linear probability models (LPM) to estimate the average treatment effect of the SMS message and report results for all households that requested a lead test kit in our sample.[vi] Model (1) estimates the overall treatment effect of the SMS intervention.

$$(1)\quad Y_i = \beta_0 + \beta_1 Treatment_i + \alpha_{Set\,S} + \varepsilon_i$$

The dependent variable, Y_i, is one of three outcomes for the SMS RCT: lead test kit return, whether a pickup was scheduled[vii], and engagement with DWM after requesting a test kit. Kits are classified as returned if the test kit was received by DWM and no errors in testing were determined. Pickup scheduled is an outcome equal to one if a pickup with Chicago DWM was scheduled regardless of whether the kit was received or errors in testing were determined. Engagement is a broader variable set equal to one if any of the following occurred after the test kit was requested: the kit was returned, a pickup was scheduled, the service request was cancelled, a new kit was ordered, or a home visit for water testing was scheduled. We also include a vector of indicator variables to account for the randomization blocks (Glennerster & Takavarasha 2013). One might also expect heterogeneous treatment effects by set. The longer it has been since the test was shipped, the less effective the treatment may be due to lost tests or a realization among residents that the test requires several inconveniences. Therefore, we also report models by set (block) as well.

We specify Model (2) to assess whether households in our target population (low-income residents) are more likely to return test kits following the SMS treatment. We did not have individual level information on the respondents other than an address, so we merged census tract level data to each respondent using 2021 American Community Survey data obtained via IPUMS National Historic Geographical Information (NHGIS). Specifically, we use census tract data on median household income and use the entire city to generate mutually exclusive groups that capture responses to treatment by those residing in census tracts that are within the 0-25th percentile of the city's census tract median income, those residing in tracts that are within the 25th-50th percentile of the city's census tract median income, and those residing in tracts that are within the 50-75th percentile of the city's census tract median income.[viii] The omitted category is those are those in tracts above the 75th percentile of the city's census tract median income. We interact the treatment indicator variable with the different census tract income indicators to determine if there are heterogeneous treatment effects by income. We also include a vector of indicator variables to account for the randomization blocks, $\alpha_{Set\ S}$.

(2) $$Y_{ict} = \beta_0 + \beta_1 Treatment_i + \beta_2 < 25thPercentile_{ct} + \beta_3 25{-}< 50thPercentile_{ct} + \beta_4 50{-}< 75thPercentile_{ct} + \beta_5 Treatment_i * < 25thPercentile_{ct} + \beta_6 Treatment_i * 25{-}< 50thPercentile_{ct} + \beta_7 Treatment_i * 50{-}< 75thPercentile_{ct} + \alpha_{Set\ S} + \varepsilon_i$$

We use a similar model, Model (3), to estimate average treatment effects for those in census tracts where lead exposure is more likely, i.e., those where the median age of housing stock, AgeStock, is 1989 or older. [ix, x]

(3) $$Y_{ict} = \beta_0 + \beta_1 Treatment_i + \beta_2 AgeStock < 1989_{ct} + \beta_3 Treatment_i * AgeStock < 1989_{ct} + \alpha_{Set\ S} + \varepsilon_i$$

Because of amendments to the Safe Drinking Water Act and the U.S. EPA's Lead and Copper Rule banned use of leaded pipe to transport water in 1986, we examine differential treatment effects for tracts with a median build year of 1989 or less. We use 1989 instead of 1986 because the ACS data provide the count of housing stock in a census track built within a bin of five years and 1989 is the bin that allows us to observe all housing stock built up to 1986.

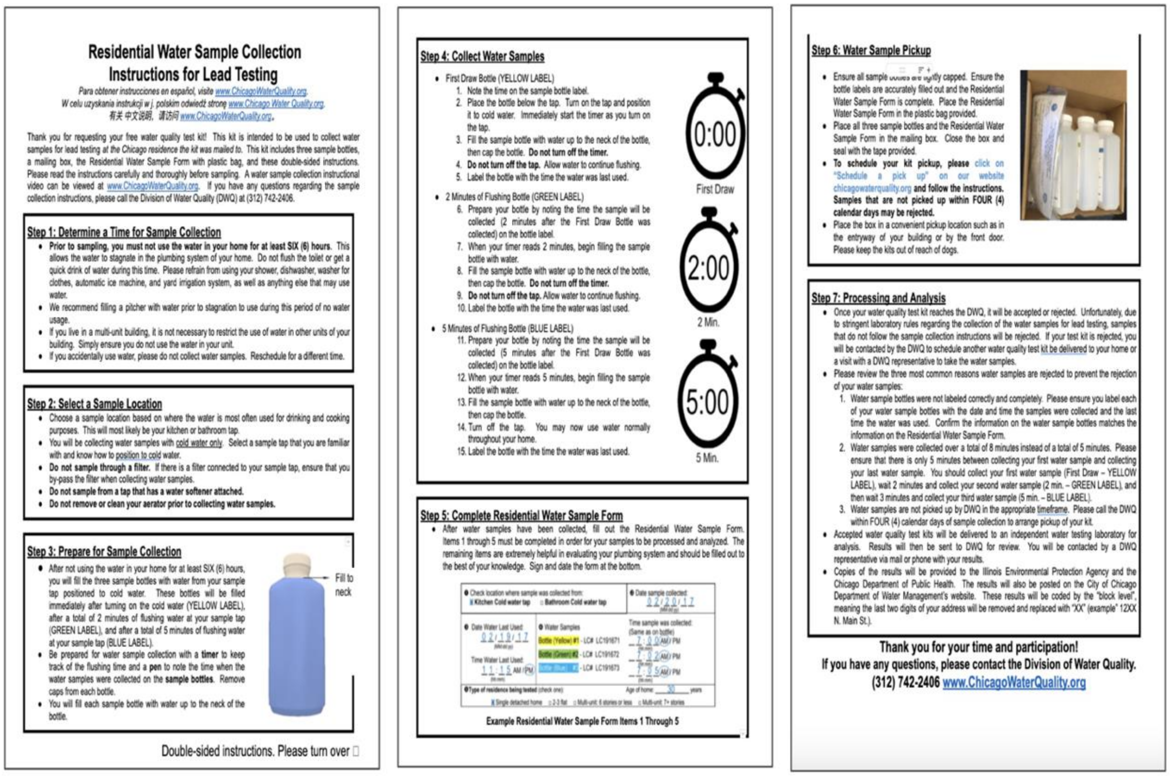
Intervention 2. Test Kit Redesign

To assess the effectiveness of the redesign intervention, we compare return rates for old kits to return rates for redesigned test kits.[xi] Old test kits included a brochure and water testing form. New kits included a three-panel color brochure which replaced three-page written instructions, an updated water testing form, two stickers to place on water faucets or toilets to enforce water stagnation, and two new stickers were added to the outside of the box emphasizing the three-step lead testing process and how the kit could be returned.

In focus groups conducted prior to the implementation, the old test kit brochure was described as intimidating, confusing, and overwhelming, and while the tasks were clear, residents found them challenging. In response, and to reduce the cognitive load of the reader, language on brochures was simplified. Larger text was used to make it easy to read and fill in, and the directions had lists to guide residents through the process. To increase the likelihood that instructions were followed, and errors reduced, the redesigned brochure included a "tips and tricks" section, added graphics to depict instructions visually, and provided a large QR code that links to an instructional video. See Figure 2 below for illustrations of the brochure before and after the redesign.

The water testing form residents we required to submit with the physical tests was also redesigned. The original form was challenging for two reasons. First, the form asked questions residents may not know how to answer, such as the year interior plumbing was installed in the home. Second, the form included many optional questions, confusing residents as to what information was required to complete the form. The redesigned water testing form included multiple-choice options to guide resident answers to questions they may not know much about and excluded optional questions. To encourage completion of the form the bottom included a callout—**"You've done the hard part!"—**to make progress salient and encourage kit return. Figure 3 below compares the old and new water testing form.

The redesigned kit also included three types of stickers to encourage kit completion. The first type emphasized the benefits of lead testing and an arbitrary two-week deadline. The second type of sticker was a set of large, red circular "Don't Use" stickers to put on water faucets or toilets. The third type were large, blue circular "Do Not Move" stickers to place on the outside of the box so that once put out for collection, the test kit would not be tampered with by other members of the household or residence. Appendix A includes illustrations of the stickers.



Residential Water Sample Collection
Instructions for Lead Testing

Para obtener instrucciones en español, visite www.ChicagoWaterQuality.org.
W celu uzyskania instrukcji w j. polskim odwiedź stronę www.Chicago Water Quality.org.
有关中文说明，请访问 www.ChicagoWaterQuality.org。

Thank you for requesting your free water quality test kit! This kit is intended to be used to collect water samples for lead testing *at the Chicago residence the kit was mailed to.* This kit includes three sample bottles, a mailing box, the Residential Water Sample Form with plastic bag, and these double-sided instructions. Please read the instructions carefully and thoroughly before sampling. A water sample collection instructional video can be viewed at www.ChicagoWaterQuality.org. If you have any questions regarding the sample collection instructions, please call the Division of Water Quality (DWQ) at (312) 742-2406.

Step 1: Determine a Time for Sample Collection

- **Prior to sampling, you must not use the water in your home for at least SIX (6) hours.** This allows the water to stagnate in the plumbing system of your home. Do not flush the toilet or get a quick drink of water during this time. Please refrain from using your shower, dishwasher, washer for clothes, automatic ice machine, and yard irrigation system, as well as anything else that may use water.
- We recommend filling a pitcher with water prior to stagnation to use during this period of no water usage.
- If you live in a multi-unit building, it is not necessary to restrict the use of water in other units of your building. Simply ensure you do not use the water in your unit.
- If you accidentally use water, please do not collect water samples. Reschedule for a different time.

Step 2: Select a Sample Location

- Choose a sample location based on where the water is most often used for drinking and cooking purposes. This will most likely be your kitchen or bathroom tap.
- You will be collecting water samples with cold water only. Select a sample tap that you are familiar with and know how to position to cold water.
- **Do not sample through a filter.** If there is a filter connected to your sample tap, ensure that you by-pass the filter when collecting water samples.
- **Do not sample from a tap that has a water softener attached.**
- **Do not remove or clean your aerator prior to collecting water samples.**

Step 3: Prepare for Sample Collection

- After not using the water in your home for at least SIX (6) hours, you will fill the three sample bottles with water from your sample tap positioned to cold water. These bottles will be filled immediately after turning on the cold water (YELLOW LABEL), after a total of 2 minutes of flushing water at your sample tap (GREEN LABEL), and after a total of 5 minutes of flushing water at your sample tap (BLUE LABEL).
- Be prepared for water sample collection with a **timer** to keep track of the flushing time and a **pen** to note the time when the water samples were collected on the **sample bottles**. Remove caps from each bottle.
- You will fill each sample bottle with water up to the neck of the bottle.

Fill to neck

Double-sided instructions. Please turn over

Step 4: Collect Water Samples

- First Draw Bottle (YELLOW LABEL)
  1. Note the time on the sample bottle label.
  2. Place the bottle below the tap. Turn on the tap and position it to cold water. Immediately start the timer as you turn on the tap.
  3. Fill the sample bottle with water up to the neck of the bottle, then cap the bottle. **Do not turn off the timer.**
  4. **Do not turn off the tap.** Allow water to continue flushing.
  5. Label the bottle with the time the water was last used.
- 2 Minutes of Flushing Bottle (GREEN LABEL)
  6. Prepare your bottle by noting the time the sample will be collected (2 minutes after the First Draw Bottle was collected) on the bottle label.
  7. When your timer reads 2 minutes, begin filling the sample bottle with water.
  8. Fill the sample bottle with water up to the neck of the bottle, then cap the bottle. **Do not turn off the timer.**
  9. **Do not turn off the tap.** Allow water to continue flushing.
  10. Label the bottle with the time the water was last used.
- 5 Minutes of Flushing Bottle (BLUE LABEL)
  11. Prepare your bottle by noting the time the sample will be collected (5 minutes after the First Draw Bottle was collected) on the bottle label.
  12. When your timer reads 5 minutes, begin filling the sample bottle with water.
  13. Fill the sample bottle with water up to the neck of the bottle, then cap the bottle.
  14. Turn off the tap. You may now use water normally throughout your home.
  15. Label the bottle with the time the water was last used.

0:00 First Draw
2:00 2 Min.
5:00 5 Min.

Step 5: Complete Residential Water Sample Form

- After water samples have been collected, fill out the Residential Water Sample Form. Items 1 through 5 must be completed in order for your samples to be processed and analyzed. The remaining items are extremely helpful in evaluating your plumbing system and should be filled out to the best of your knowledge. Sign and date the form at the bottom.

Check location where sample was collected from: X Kitchen Cold water tap / Bathroom Cold water tap		Date sample collected: 02/20/17
Date Water Last Used: 02/19/17; Time Water Last Used: 11:15 AM/PM	Water Samples: Bottle (Yellow) #1 - LC# LC191671; Bottle (Green) #2 - LC# LC191672; Bottle (Blue) #3 - LC# LC191673	Time sample was collected (Same as on bottle): 7:00 AM/PM; 7:02 AM/PM; 7:05 AM/PM
Type of residence being tested (check one): X Single detached home / 2-3 flat / Multi-unit: 6 stories or less / Multi-unit: 7+ stories		Age of home: 30 years

Example Residential Water Sample Form Items 1 Through 5

Step 6: Water Sample Pickup

- Ensure all sample bottles are tightly capped. Ensure the bottle labels are accurately filled out and the Residential Water Sample Form is complete. Place the Residential Water Sample Form in the plastic bag provided.
- Place all three sample bottles and the Residential Water Sample Form in the mailing box. Close the box and seal with the tape provided.
- **To schedule your kit pickup, please** click on "Schedule a pick up" on our website chicagowaterquality.org **and follow the instructions. Samples that are not picked up within FOUR (4) calendar days may be rejected.**
- Place the box in a convenient pickup location such as in the entryway of your building or by the front door. Please keep the kits out of reach of dogs.

Step 7: Processing and Analysis

- Once your water quality test kit reaches the DWQ, it will be accepted or rejected. Unfortunately, due to stringent laboratory rules regarding the collection of the water samples for lead testing, samples that do not follow the sample collection instructions will be rejected. If your test kit is rejected, you will be contacted by the DWQ to schedule another water quality test kit be delivered to your home or a visit with a DWQ representative to take the water samples.
- Please review the three most common reasons water samples are rejected to prevent the rejection of your water samples:
  1. Water sample bottles were not labeled correctly and completely. Please ensure you label each of your water sample bottles with the date and time the samples were collected and the last time the water was used. Confirm the information on the water sample bottles matches the information on the Residential Water Sample Form.
  2. Water samples were collected over a total of 8 minutes instead of a total of 5 minutes. Please ensure that there is only 5 minutes between collecting your first water sample and collecting your last water sample. You should collect your first water sample (First Draw – YELLOW LABEL), wait 2 minutes and collect your second water sample (2 min. – GREEN LABEL), and then wait 3 minutes and collect your third water sample (5 min. – BLUE LABEL).
  3. Water samples are not picked up by DWQ in the appropriate timeframe. Please call the DWQ within FOUR (4) calendar days of sample collection to arrange pickup of your kit.
- Accepted water quality test kits will be delivered to an independent water testing laboratory for analysis. Results will then be sent to DWQ for review. You will be contacted by a DWQ representative via mail or phone with your results.
- Copies of the results will be provided to the Illinois Environmental Protection Agency and the Chicago Department of Public Health. The results will also be posted on the City of Chicago Department of Water Management's website. These results will be coded by the "block level", meaning the last two digits of your address will be removed and replaced with "XX" (example" 12XX N. Main St.).

Thank you for your time and participation!
If you have any questions, please contact the Division of Water Quality.
(312) 742-2406 www.ChicagoWaterQuality.org

Figure 2A. Lead Test Kit Brochure: Before Redesign

Figure 2B. Lead Test Kit Brochure: After Redesign

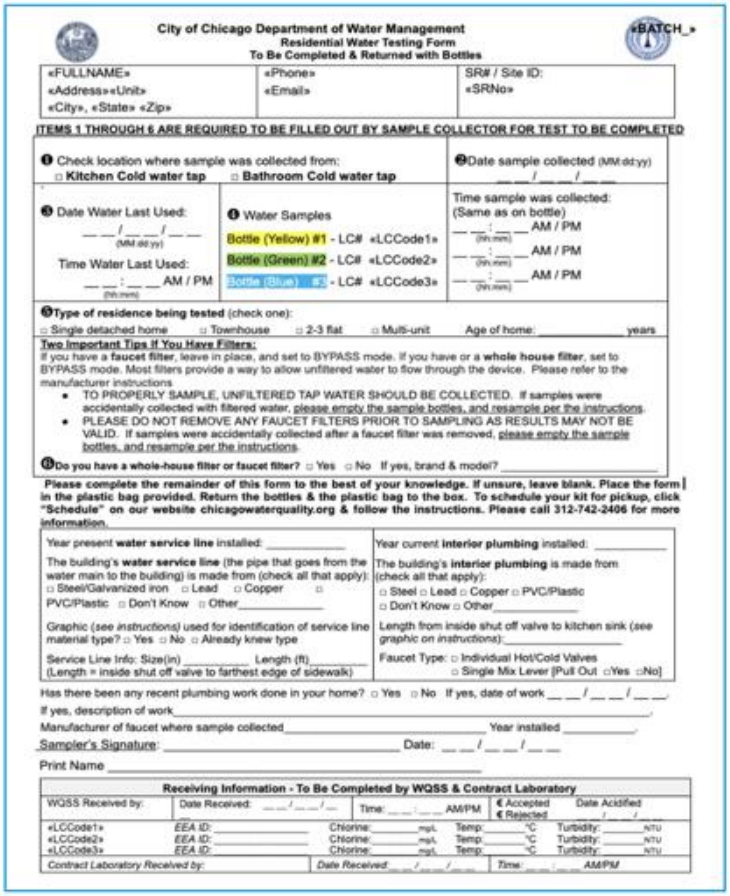

City of Chicago Department of Water Management
Residential Water Testing Form
To Be Completed & Returned with Bottles
«BATCH_»
«FULLNAME» «Phone» SR# / Site ID:
«Address»«Unit» «Email» «SRNo»
«City», «State» «Zip»
ITEMS 1 THROUGH 6 ARE REQUIRED TO BE FILLED OUT BY SAMPLE COLLECTOR FOR TEST TO BE COMPLETED
❶ Check location where sample was collected from: □ Kitchen Cold water tap □ Bathroom Cold water tap
❷ Date sample collected (MM:dd:yy) __ __ / __ __ / __ __
❸ Date Water Last Used: __ __ / __ __ / __ __ (MM:dd:yy)
Time Water Last Used: __ __ : __ __ AM / PM (hh:mm)
❹ Water Samples
Bottle (Yellow) #1 - LC# «LCCode1»
Bottle (Green) #2 - LC# «LCCode2»
Bottle (Blue) #3 - LC# «LCCode3»
Time sample was collected: (Same as on bottle)
__ __ : __ __ AM / PM
__ __ : __ __ AM / PM
__ __ : __ __ AM / PM
❺ Type of residence being tested (check one):
□ Single detached home □ Townhouse □ 2-3 flat □ Multi-unit Age of home: ____ years
Two Important Tips If You Have Filters:
If you have a faucet filter, leave in place, and set to BYPASS mode. If you have or a whole house filter, set to BYPASS mode. Most filters provide a way to allow unfiltered water to flow through the device. Please refer to the manufacturer instructions
• TO PROPERLY SAMPLE, UNFILTERED TAP WATER SHOULD BE COLLECTED. If samples were accidentally collected with filtered water, please empty the sample bottles, and resample per the instructions.
• PLEASE DO NOT REMOVE ANY FAUCET FILTERS PRIOR TO SAMPLING AS RESULTS MAY NOT BE VALID. If samples were accidentally collected after a faucet filter was removed, please empty the sample bottles, and resample per the instructions.
❻ Do you have a whole-house filter or faucet filter? □ Yes □ No If yes, brand & model?
Please complete the remainder of this form to the best of your knowledge. If unsure, leave blank. Place the form in the plastic bag provided. Return the bottles & the plastic bag to the box. To schedule your kit for pickup, click "Schedule" on our website chicagowaterquality.org & follow the instructions. Please call 312-742-2406 for more information.
Year present water service line installed: ____
The building's water service line (the pipe that goes from the water main to the building) is made from (check all that apply): □ Steel/Galvanized iron □ Lead □ Copper □ PVC/Plastic □ Don't Know □ Other ____
Graphic (see instructions) used for identification of service line material type? □ Yes □ No □ Already knew type
Service Line Info: Size(in) ____ Length (ft) ____ (Length = inside shut off valve to farthest edge of sidewalk)
Year current interior plumbing installed: ____
The building's interior plumbing is made from (check all that apply): □ Steel □ Lead □ Copper □ PVC/Plastic □ Don't Know □ Other ____
Length from inside shut off valve to kitchen sink (see graphic on instructions): ____
Faucet Type: □ Individual Hot/Cold Valves □ Single Mix Lever [Pull Out □Yes □No]
Has there been any recent plumbing work done in your home? □ Yes □ No If yes, date of work __ __ / __ __ / __ __
If yes, description of work ____
Manufacturer of faucet where sample collected ____ Year installed ____
Sampler's Signature: ____ Date: __ __ / __ __ / __ __
Print Name ____
Receiving Information - To Be Completed by WQSS & Contract Laboratory
WQSS Received by: Date Received: __ __ / __ __ / __ __ Time: __ __ : __ __ AM/PM □ Accepted □ Rejected Date Acidified __ / __ / __
«LCCode1» EEA ID: Chlorine: ____ mg/L Temp: ____ °C Turbidity: ____ NTU
«LCCode2» EEA ID: Chlorine: ____ mg/L Temp: ____ °C Turbidity: ____ NTU
«LCCode3» EEA ID: Chlorine: ____ mg/L Temp: ____ °C Turbidity: ____ NTU
Contract Laboratory Received by: Date Received: __ / __ / __ Time: __ __ : __ __ AM/PM

Figure 3A. Water Testing Form: Before Redesign

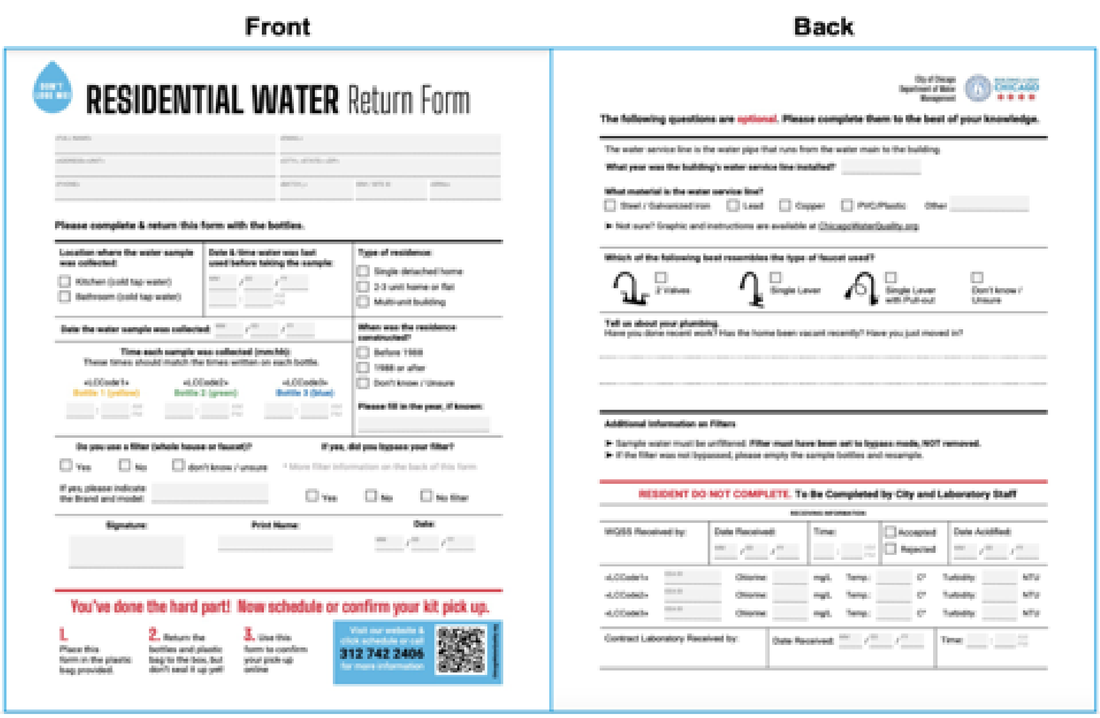


Figure 3B. Water Testing Form: After Redesign

Empirical Approach

Receipt of redesigned test kits was not randomly assigned. Rather, all test kits shipped out on or after March 30, 2022, contained the redesigned materials. To estimate the efficacy of the new materials, we compare the cumulative probability of test kit return at various lengths of time since the kit was shipped for the old and new packets. The kits were not shipped every day, but rather in batches. The three most recent batch shipments before the redesign were January 20, February 8, and February 28, 2022. We compare return rates for those to batch shipments of the redesign sent on March 30, April 19, and May 10, 2022. Model (4) compares cumulative return rates of old kits, those shipped January 20, 2022-Feburary 28, 2022, to return rates of redesigned kits, those shipped March 30, 2022 – May 10, 2022. [xii] We omit the batches shipped on February 28, 2022, which serves as the comparison for the return rates for the redesigned kits. In addition, to reduce heterogeneity among those who received the old and new designs, we also compare those who received the last shipment of old test kits to those that received the first shipment of redesigned kits in our data—those shipped on February 28th and March 30th – to assess differences in return rates under the most similar conditions.

$$(4) \quad Y_i = \beta_0 + \beta_1 KitShipJan20_i + \beta_2 KitShipFeb8_i + \beta_3 KitShipMarch30_i + \beta_4 KitShipApril19_i + \beta_5 KitShipMay10_i + \varepsilon_i$$

Y_i is the cumulative probability of kit return within 7, 21, and 35 days of kit shipment date. These lengths of time were chosen somewhat arbitrarily, but as shown in Figure 4 below, these differences persist across the time horizon of our data. β_0 reflects kit return for kits shipped on February 28th. Because these shipments before and after March 30, 2022 were not assigned randomly, one should not necessarily interpret the changes observed as causal. Our analysis does not account for unobserved differences between those who requested test kits after March 30th from those who requested test kits prior to March 30th.

Results

Intervention 1. SMS Text Message Results

In Table 1, we report results from the RCT. We find positive estimates for recipients of the SMS message. Overall results (column 4) show that those who received the SMS text message were 3.7 percentage points more likely to return the test kit compared to the group that requested a lead test kit but did not receive a SMS message. Compared to the return rate for the control group 1.4 percent, this is a 177 percent (((0.037-0.014)/0.014)x100) increase. Those who received the SMS text message were also 4.1 percentage points more likely to schedule a pick-up and 4.4 percentage points more likely to engage with Chicago's DWM after requesting the lead test kit than test kit requesters that did not receive the SMS text. Pick-ups scheduled also increased by 173 percent compared to the control group and the engagement rate increased 159 percent. Point-estimates are statistically significant at conventional levels.[xiii]

In the first three columns, we compare treatment effect estimates by set to determine if the duration between the ship date and the treatment date produces heterogeneous treatment effects. While we see some decline in the size of the estimate, F-tests of equivalence show no statistically significant difference between the estimates (please see the bottom line in each panel). However, the control group means are vastly different by sets. For instance, while the treatment effect estimate in Set 3 is less (0.032 compared to 0.042) than the estimate for Set 1 or Set 2, the mean return rate for the control group is much smaller in Set 3 than either Set 1 or Set 2. Therefore, the likelihood of returning a test kit is ten times larger for the treated group in Set 3 than the control group. The comparable treatment sizes are twice as large for those in Set 2 and 1.35 times as large in Set 1. Thus, while treatment was effective for all sets, it may be most effective for those with the longest time since the kits were shipped.

Table 1. Estimated effect of SMS text on outcomes, by group.

	Set 1	Set 2	Set 3	Overall
Requested Kits				
Effect on lead test kit return	0.042***	0.042***	0.032***	0.037***
	(0.015)	(0.013)	(0.006)	(0.006)
Test equivalent effects	Set 1 = Set 2	Set 2 = Set 3	Set 1 = Set 3	
Kit return	0.978	0.508	0.539	
Effect on pickup scheduled	0.052***	0.044***	0.035***	0.041***
	(0.016)	(0.013)	(0.007)	(0.006)
Test equivalent effects	Set 1 = Set 2	Set 2 = Set 3	Set 1 = Set 3	
Pickup scheduled	0.704	0.520	0.320	
Effect on engagement rate	0.057***	0.048***	0.037***	0.044***
	(0.017)	(0.014)	(0.007)	(0.006)
Test equivalent effects	Set 1 = Set 2	Set 2 = Set 3	Set 1 = Set 3	
Engagement rate	0.696	0.457	0.273	
Mean return	0.031	0.021	0.003	0.014
Mean pickup	0.033	0.021	0.003	0.015
Mean engagement	0.035	0.023	0.004	0.017
N	851	957	1,793	3,601

Notes: p<0.10, **p<0.05, ***p<0.01. Robust standard errors in parentheses below coefficient estimate. All coefficients in a column from a single OLS regression. All sets are pooled together to estimate the "Overall" column and dummy variables for each set are included in this regression. Set 2 is the omitted set. Mean values reflect those for members of the control group. We also test statistical differences in treatment effects by set when a fully interacted model is estimated. P-values for these tests are reported for each outcome.

Table 2 examines heterogeneity of the SMS treatment by economic disadvantage, and Table 3 assesses the differential effect of the treatment for those in census tracts where the median age of housing stock is built before 1989. Both tables show that the SMS message intervention was equally effective for residents regardless of the median income and age of housing stock in their home census tract. Coefficients on the interaction variables are not statistically significant at conventional levels. We also failed to reject a joint test with the null hypothesis that all are treatment interactions equal zero.

While we cannot rule out the possibility of heterogenous treatment effects by analyzing tract-level differences, this analysis does provide confidence that there are not systematic differences in the treatment across neighborhoods in Chicago.

Table 2. Estimated Effect of SMS Text on Outcomes, Heterogeneity by Income

Median Income	**Test Kit Return**	**Pickup Scheduled**	**Engagement Rate**
Treatment	0.054***	0.059***	0.058***
	(0.012)	(0.013)	(0.013)
Treatment * 25th Percentile	-0.028	-0.027	-0.010
	(0.015)	(0.017)	(0.017)
Treatment *25th-50th Percentile	-0.014	-0.020	-0.018
	(0.017)	(0.018)	(0.018)
Treatment* 50th-75th Percentile	-0.025	-0.029	-0.025
	(0.018)	(0.018)	(0.018)
25th Percentile	-0.010	-0.007	-0.013**
	(0.006)	(0.007)	(0.009)
25th-50th Percentile	-0.000	-0.000	0.001
	(0.008)	(0.008)	(0.009)
50th-75th Percentile	0.007	0.007	0.004
	(0.008)	(0.008)	(0.009)
Constant	0.024	0.023	0.027
Mean Outcome	0.033	0.036	0.039
Joint F-test that all Treatment Interactions = 0	0.303	0.353	0.558
N	3,601	3,601	3,601

Notes: ***p<0.01. **p<0.05. Robust standard errors in parentheses below coefficient estimate. Census tract median income was $48,799 at the 25th percentile of the median income distribution, $68,444 at the 50th percentile, and $95,077 at the 75th percentile. All values in 2021 dollars. All sets are pooled together to estimate each column and dummy variables for each set are included in each regression. Set 2 is the omitted set. Treatment refers to those who received the treatment and resided in the 75th percentile of the census tract median income distribution and above. We also test whether the differential response to treatment among census tracts below the 75th percentile of median income is equal to zero with an F-test of all interaction variables. P-values of this test are reported.

Table 3. Estimated Effect of SMS Text on Outcomes, Housing Stock Heterogeneity

	Test Kit Return	**Pickup Scheduled**	**Engagement Rate**
Treatment	0.037***	0.042***	0.045***
	(0.009)	(0.010)	(0.010)
Built Before 1990	-0.006	-0.005	-0.006
	(0.006)	(0.006)	(0.006)
Treatment*Built Before 1990	-0.000	-0.002	-0.000
	(0.012)	(0.012)	(0.013)
Constant	0.026	0.025	0.028

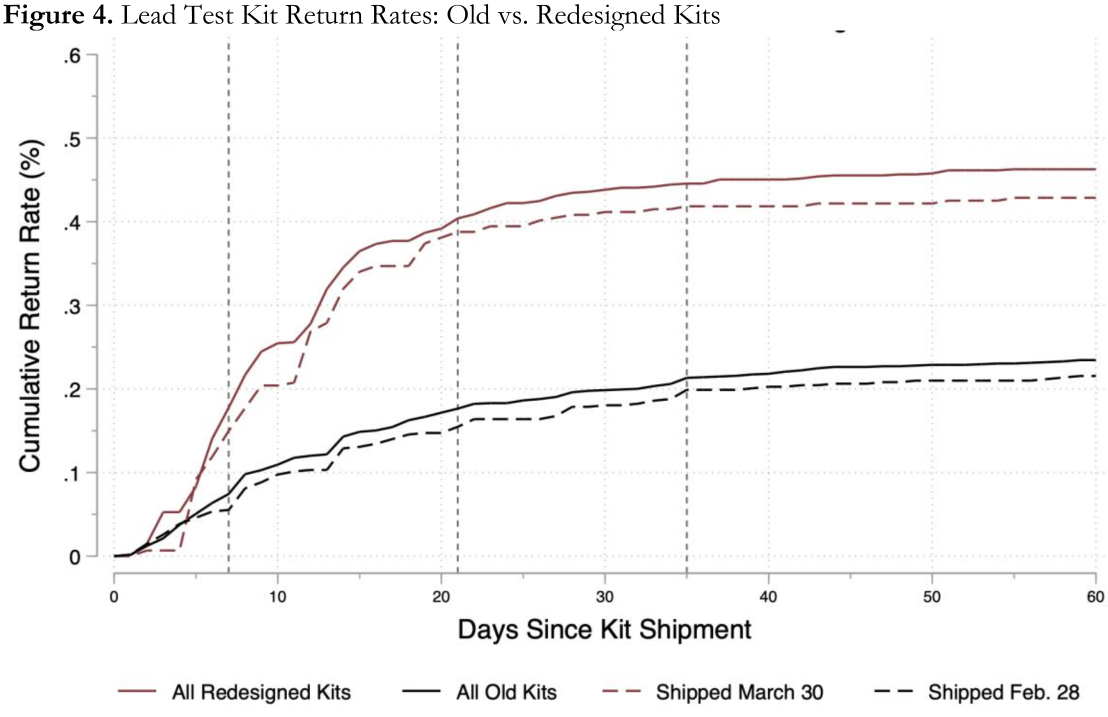
Notes: $p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$. Robust standard errors clustered by resident in parentheses below coefficient estimate. All sets are pooled together to estimate each column and dummy variables for each set are included in each regression. Set 2 is the omitted set. Old Housing Stock is an indicator set to one if a resident resides in a census tract whose median housing stock was built before 1990. Treatment those who received the treatment and lived in census tracts where median housing stock was built before 1990.

Intervention 2. Test Kit Redesign Results

Figure 4 shows differences in the cumulative probability of return for those who received the old and those who received the redesigned kits. Return rates for old test kits are in black and return rates for redesigned test kits in maroon. The dashed lines compare kits shipped on March 30, 2022, to old test kits shipped on February 28, 2022, the two most proximate and, therefore, homogenous comparisons. The figure shows the success of the redesigned kits by the first week, and reaches a maximum by the third week, after which the gap persists throughout the time period over which we have data. Return rates 7, 21, and 35 days after shipment for redesigned kits were 12, 27, and 25 percent higher among those receiving redesigned kits, respectively.

We quantify the differences in return rates in Table 5 to see that kit return rates among redesigned kits (March 30, April 19, May 10) are larger and statistically significantly different from old kits shipped on February 28th. Kits shipped in the March 30th group, the first shipment after the redesign, were 9.4 percentage points more likely to be returned within 7 days, 23.3 percentage points more likely to be returned within 21 days, and 21.9 percentage points more likely to be returned within 35 days than kits shipped on February 28th. Kits shipped in the April 19th and May 10th groups were 13.1 and 8.4 percentage points more likely to be returned within 7 days, 28.1 and 23.8 percentage points more likely to be returned within 21 days, and 28.1 and 24.5 percentage points more likely to be returned within 35 days, respectively, compared to kits shipped on February 28th. Differences in return rates remain when redesigned kits are compared to the other groups of old kits. Table 5 includes p-values for statistical tests of the differences that would be observed if earlier comparison groups (Jan. 20 and Feb. 8) were used instead of February 28th.

Figure 4. Lead Test Kit Return Rates: Old vs. Redesigned Kits



Notes: Figure shows differences in cumulative return rates among those receiving old kits (black lines) and those receiving redesigned kits (maroon lines). Solid lines compare average cumulative return rates among kits shipped January 20, February 8, and February 28, 2022, and kits shipped March 30, April 19, and May 10, 2022. Days since kit shipped is a relative measure to account for differences in timing of kit receipt. Dashed lines compare kit return rates of batches shipped closest to March 30 (date redesigned kits were first shipped).

Table 5. Estimated Effect of Kit Redesign on Kit Return, Days After Shipment

	7 Days After	**21 Days After**	**35 Days After**
Jan. 20th Shipment	0.032	0.049	0.038
	(0.020)	(0.027)	(0.028)
Feb. 8th Shipment	0.040	0.018	-0.003
	(0.027)	(0.036)	(0.036)
Feb. 28th Shipment	0.055***	0.155***	0.199***
(Constant)	(0.014)	(0.018)	(0.019)
March 30th Shipment	0.094***	0.233***	0.219***
	(0.023)	(0.031)	(0.032)
April 19th Shipment	0.131***	0.281***	0.281***
	(0.025)	(0.033)	(0.034)
May 10th Shipment	0.084***	0.238***	0.245***
	(0.024)	(0.032)	(0.033)
Test Jan. 20 = March 30	0.007	0.000	0.000
Test Jan. 20 = April 19	0.000	0.000	0.000
Test Jan. 20 = May 10	0.005	0.000	0.000
Test Feb. 8 = March 30	0.058	0.000	0.000
Test Feb. 8 = April 19	0.000	0.000	0.000
Test Feb. 8 = May 10	0.009	0.000	0.000
N	2,041	2,041	2,041

Notes: $p<0.10$, $^{**}p<0.05$, $^{***}p<0.01$. Robust standard errors clustered by resident in parentheses below coefficient estimate. Sample in bottom panel limited to test kits shipped in Feb. 28th and March 30th groups. We also test statistical differences in cumulative return rates 7, 21, and 35 days after kit shipment to assess differences that would be observed if earlier comparison groups (Jan. 20 and Feb. 8) were used to evaluate differences in return rates instead of Feb. 28th. P-values for these tests are reported.

Conclusion

Collectively, our results show that the introduction of behavioral interventions to decrease compliance costs improves lead test kit return, the number of pickups scheduled, and engagement with Chicago DWM. The SMS message intervention increased return rates by 3.7 percentage points overall and was effective across the target population. Residents in census tracts with the lowest median income and median housing stock built before 1990 did not respond differentially to the treatment. The test kit redesign intervention was also effective. Those receiving a redesigned kit were about 25 percentage points more likely to return it within 35 days, on average, indicating that simplification is an effective way to move requesters through the testing and return process. While we are more confident in the SMS intervention results given the RCT, both behavioral approaches, the SMS nudge and simplification, increased kit returns.

However, even after these interventions overall return rates remain modest. In the case of the SMS nudge, return rates do not exceed 10 percent after treatment. Varying when SMS text messages are sent out relative to kit shipment, say after return rates peak and begin to slow down (21 days after shipment), could further increase return rates. Targeting interventions close to the date of kit receipt could be especially effective as our analysis shows return rates decrease over time. Other approaches to increase kit returns, like reducing the number of steps in the at-home test kit process or increasing education and communication campaigns could be costly. The low-cost and large response to the SMS and brochure redesign interventions are attractive, however, there is much more work to be done to ensure that all residents return their requested kits.

The low cost of behavioral interventions means that these interventions often produce cost savings for governments. In the case of Chicago, these interventions were cost neutral. The City used an existing text message platform to send out messages and printed redesigned brochures using the same in-house group used to create the old brochures. Thus, our results show that without incurring additional costs the SMS intervention increased return rates by 3.7 percentage points, or an additional 555 kits returned a year.

Given the size of the effect from both the SMS text message RCT and the lead test kit redesign, public administrators might consider other ways to introduce simplification and SMS nudges to move participants through multi-step compliance processes at the local level. Further, as this case demonstrates, behavioral interventions can be used by governments to ease participation in current burdensome services as they simultaneously tackle larger challenges, like the replacement of all leaded water pipes by 2077.

Acknowledgments

We are grateful to the following staff at The Behavioral Insights Team for their work on this project: Maggie McNease for creating the redesigned kit materials, Jasmine Pineda for data cleaning, early analysis, and assistance with evaluation design, and Anna Keleher for developing the interventions. We would like to thank the Chicago Department of Water Management for their invaluable support and implementation, specifically Andrea Cheng, Sophie Manley, Andrew Ristau, and Teresa Muldoon. We would also like to express our gratitude to Bloomberg Philanthropies for supporting this work and Bloomberg Associates for their guidance throughout the project, specifically Jacob Koch and Adam Freed.

Notes

[i] EPA (2013) provides a thorough review of hundreds of studies investigating the effects of lead from epidemiology, toxicology, public health, neuroscience, and other medical disciplines. Lead exposure has negative effects across different measures of cognition and academic performance, such as IQ tests (Schnaas et al. 2006; Lanphear et al. 2005; Ris et al. 2004; Canfield et al. 2003; Bellinger et al. 1992), and primary school assessments (Aizer et al. 2018; Rau et al. 2015; McLaine et al. 2013; Reyes 2011; Chandramouli et al. 2009; Miranda et al. 2007a; Nilsson 2009; Miranda et al. 2007b). Lead exposure decreases rates of high school graduation (Nilsson 2009; Fergusson at al., 1997; Needleman et al. 1990), and results in lower adult earnings (Nilsson 2009). Early life lead exposure also impacts externalizing behaviors such as attention, impulsivity, and hyperactivity in young children (Froehlich et al. 2009; Chen et al. 2007); increased delinquent and antisocial activity and higher rates of arrest (Aizer and Currie 2019; Reyes 2015; Wright et al. 2008; Fergusson et al 2008; Needleman et al. 2002; Needleman et al. 1996).

[ii] Interventions were designed by the Chicago Department of Water Management in partnership with the Behavioral Insight Team.

[iii] Additional costs, like redemption costs, may arise if individuals face learning costs to use or redeem public benefits (Barnes 2021).

[iv] Between January 2021-January 2022, 11,546 at-home lead test kits were requested. Averages before the sample period indicate about 15,000 test kits were requested each year.
[v] It is important to note that 1,826 test kits requested between September 1, 2021, and March 29, 2022, are excluded from the sample as the test kits were returned before March 29, 2022. One should, therefore, consider these treatment effects estimates to be appropriate for Chicago residents who requested test kits and did not return them, not all test kit requestors.
[vi] We also estimate the average treatment effect of the SMS message using a probit model and the results are substantively similar.
[vii] A person may have scheduled pick-up, but the test kit could not be counted as returned for several reasons. It could be that the person did not put the kit out for collection, the city may have failed to pick up a test kit, or the scheduled pick-up date was outside of the trial window.
[viii] We also define economic disadvantage by a census tract's share of homeownership and single-family homes. We evaluate differential responses to treatment by again comparing characteristics of each requester's census tract to the median within the city, evaluating differences in responses to treatment by quartile. Results (available in Appendix B) indicate smaller differences between the most and least disadvantaged neighborhoods by this definition, but still find higher return rates for those in the most advantaged tracts (75^{th} percentile and above) compared to those just below (50^{th} and 75^{th} percentile).
[ix] There are plenty of old homes where leaded pipes have been removed in wealthy neighborhoods. To assess the relationship between age of home and wealth of neighborhood, we correlate a census tract's median household income and share of housing stock in the census tract built before 1989. The correlation is r=-0.51, indicating that as a tract's median household income increases the share of old housing stock decreases.
[x] While we do not have individual level data to test balance between the treatment and control groups, we did use these neighborhood measures to test balance and found no statistical difference in any of the neighborhood characteristics we had. These measures included median household income, racial and ethnic composition, educational attainment, the use of public assistance, household age, structure, and number of bedrooms. A table of these balance tests is available in Appendix B.
[xi] SMS text messages were sent to old test kits—kits requested between September 2021 and March 2022—between May 11-19, 2022. SMS text messages were sent to new test kits—kits requested after March 2022—June 2, 2022. Return rates are calculated prior to "old" and "new" kit exposure to the SMS intervention.
[xii] While we would expect test kit redesign to affect scheduled pick-ups, engagement, as well as the error rate, a binary measure of return was the only outcome collected.
[xiii] Results are consistent when time elapsed is included as a quadratic as well.
[xiii] This benefit estimate does not account for improved outcomes that follow lead detection, abatement, and remediation, nor the costs to ship kits. If 10 test kits are picked up an hour and DWM employees earn a wage of $21.08 an hour (ZipRecruiter, 2023), picking up additional test kits would cost about $1,166, indicating increased return rates have a more than $5,000 benefit to the city.

References

Aizer, A. (2003). Low take-up in Medicaid: Does outreach matter and for whom?. *American Economic Review*, *93*(2), 238-241.

Aizer, A., & Currie, J. (2019). Lead and juvenile delinquency: new evidence from linked birth, school, and juvenile detention records. *Review of Economics and Statistics*, *101*(4), 575-587.

Barnes, C. Y. (2021). "It takes a while to get used to": The costs of redeeming public benefits. *Journal of Public Administration Research and Theory*, *31*(2), 295-310.

Barnes, C., Halpern-Meekin, S., & Hoiting, J. (2023). "I Used to Get WIC... But Then I Stopped": How WIC Participants Perceive the Value and Burdens

of Maintaining Benefits. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *9*(5), 32-55.
Barnes, C. Y., & Henly, J. R. (2018). "They are underpaid and understaffed": How clients interpret encounters with street-level bureaucrats. *Journal of Public Administration Research and Theory*, *28*(2), 165-181.
Behavioural Insights Team (BIT). (2015). The Behavioral Insights Team Update 2013–2015. London, UK: Cabinet Office.
Bellinger, D. C., Stiles, K. M., & Needleman, H. L. (1992). Low-level lead exposure, intelligence and academic achievement: a long-term follow-up study. *Pediatrics*, *90*(6), 855-861.
Bettinger, E. P., Long, B. T., Oreopoulos, P., & Sanbonmatsu, L. (2012). The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment. *The Quarterly Journal of Economics*, *127*(3), 1205-1242.
Bhargava, S., & Manoli, D. (2015). Psychological frictions and the incomplete take-up of social benefits: Evidence from an IRS field experiment. *American Economic Review*, *105*(11), 3489-3529.
Brodkin, E. Z., & Majmundar, M. (2010). Administrative exclusion: Organizations and the hidden costs of welfare claiming. *Journal of Public Administration Research and Theory*, *20*(4), 827-848.
Burden, B. C., Canon, D. T., Mayer, K. R., & Moynihan, D. P. (2012). The effect of administrative burden on bureaucratic perception of policies: Evidence from election administration. *Public Administration Review*, *72*(5), 741-751.
Canfield, R. L., Henderson Jr, C. R., Cory-Slechta, D. A., Cox, C., Jusko, T. A., & Lanphear, B. P. (2003). Intellectual impairment in children with blood lead concentrations below 10 µg per deciliter. *New England journal of medicine*, *348*(16), 1517-1526.
Chandramouli, K., Steer, C. D., Ellis, M., & Emond, A. M. (2009). Effects of early childhood lead exposure on academic performance and behaviour of school age children. *Archives of disease in childhood*, *94*(11), 844-848.
Chase, Brett. (2022, December 2). *Lightfoot's plan to replace Chicago's lead water pipes has switched out 280 of an estimated 390,000 service lines*. Chicago Sun-Times. https://chicago.sun-times.com/2022/12/2/23488902/lead-water-service-line-replacement-chicago
Chen, A., Cai, B., Dietrich, K. N., Radcliffe, J., & Rogan, W. J. (2007). Lead exposure, IQ, and behavior in urban 5-to 7-year-olds: does lead affect behavior only by lowering IQ?. *Pediatrics*, *119*(3), e650-e658.
Chicago Department of Public Health. (2016). *Fifty Years Fighting Lead in Chicago: The Plan for a Lead Free Generation.* Healthy Chicago 2.0 Partnering to Improve Equity. https://www.chicago.gov/content/dam/city/depts/cdph/food_env/general/Lead_Poison_Prevention_Program/CDPH_LeadBrochure_10172016.pdf
Christensen, J., Aarøe, L., Baekgaard, M., Herd, P., & Moynihan, D. P. (2020). Human capital and administrative burden: The role of cognitive resources in citizen-state interactions. *Public Administration Review*, *80*(1), 127-136.
Chudnovsky, M., & Peeters, R. (2021). The unequal distribution of administrative burden: A framework and an illustrative case study for understanding variation in people's experience of burdens. *Social Policy & Administration*, *55*(4), 527-542.
Clark, B. N., Masters, S. V., & Edwards, M. A. (2015). Lead Release to Drinking Water from Galvanized Steel Pipe Coatings. *Environmental Engineering Science*, *32*(8), 713–721. https://doi.org/10.1089/ees.2015.0073
Dietrich, M., Krekeler, M. P., Kousehlar, M., & Widom, E. (2021). Quantification of Pb pollution sources in complex urban environments through a multi-source isotope mixing model based on Pb isotopes in lichens and road sediment. *Environmental Pollution*, *288*, 117815.
Dorn, S., Hill, I., & Hogan, S. (2009). The Secrets of Massachusetts' Success: Why 97 Percent of State Residents Have Health Coverage State Health Access Reform Evaluation. Robert Wood Johnson Foundation.
US EPA. (2023). *Biden-Harris Administration Announces $332 Million for Drinking Water Infrastructure Upgrades in Illinois* (Illinois) [News Release]. https://www.epa.gov/newsreleases/biden-harris-administration-announces-332-million-drinking-water-infrastructure
US EPA. (2015). *Use of Lead Free Pipes, Fittings, Fixtures, Solder, and Flux for Drinking Water* [Overviews and Factsheets]. https://www.epa.gov/sdwa/use-lead-free-pipes-fittings-fixtures-solder-and-flux-drinking-water
US EPA. (2013). *Learn about Lead* [Overviews and Factsheets]. https://www.epa.gov/lead/learn-about-lead
Fergusson, D. M., Boden, J. M., & Horwood, L. J. (2008). Dentine lead levels in childhood and criminal behaviour in late adolescence and early adulthood. *Journal of Epidemiology & Community Health*, *62*(12), 1045-1050.
Fergusson, D. M., Horwood, L. J., & Lynskey, M. T. (1997). Early dentine lead levels and educational outcomes at 18 years. *Journal of Child Psychology and Psychiatry*, *38*(4), 471-478.
Froehlich, T. E., Lanphear, B. P., Auinger, P., Hornung, R., Epstein, J. N., Braun, J., & Kahn, R. S. (2009). Association of tobacco and lead exposures with attention-deficit/hyperactivity disorder. *Pediatrics*, *124*(6), e1054-e1063.

Glennerster, R. & Takavarasha, K. (2013). *Running randomized evaluations: A practical guide.* Princeton University Press.

Hauptman, M., Bruccoleri, R., & Woolf, A. D. (2017). An Update on Childhood Lead Poisoning. *Clinical Pediatric Emergency Medicine*, *18*(3), 181–192. https://doi.org/10.1016/j.cpem.2017.07.010

Hastings, J. S., & Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly journal of economics*, *123*(4), 1373-1414.

Haynes, L. C., Green, D. P., Gallagher, R., John, P., & Torgerson, D. J. (2013). Collection of delinquent fines: An adaptive randomized trial to assess the effectiveness of alternative text messages. *Journal of Policy Analysis and Management*, 32, 718–730.

Heinrich, C. J. (2016). The bite of administrative burden: A theoretical and empirical investigation. *Journal of Public Administration Research and Theory*, *26*(3), 403-420.

Heinrich, C. J. (2018). Presidential address: "A thousand petty fortresses": Administrative burden in US immigration policies and its consequences. *Journal of Policy Analysis and Management*, *37*(2), 211-239.

Heinrich, C. J., & Brill, R. (2015). Stopped in the name of the law: Administrative burden and its implications for cash transfer program effectiveness. *World Development*, *72*, 277-295.

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

Herd, P., DeLeire, T., Harvey, H., & Moynihan, D. P. (2013). Shifting administrative burden to the state: The case of medicaid take-up. *Public Administration Review*, *73*(s1), S69-S81.

Jones, R. L., Homa, D. M., Meyer, P. A., Brody, D. J., Caldwell, K. L., Pirkle, J. L., & Brown, M. J. (2009). Trends in blood lead levels and blood lead testing among US children aged 1 to 5 years, 1988–2004. *Pediatrics*, *123*(3), e376-e385.

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

Kopczuk, W., & Pop-Eleches, C. (2007). Electronic filing, tax preparers and participation in the Earned Income Tax Credit. *Journal of public Economics*, *91*(7-8), 1351-1367.

Lanphear, B. P., Hornung, R., Khoury, J., Yolton, K., Baghurst, P., Bellinger, D. C., & Roberts, R. (2005). Low-level environmental lead exposure and children's intellectual function: an international pooled analysis. *Environmental health perspectives*, *113*(7), 894-899.

Lasky-Fink, J., Robinson, C. D., Chang, H. N. L., & Rogers, T. (2021). Using behavioral insights to improve school administrative communications: The case of truancy notifications. *Educational researcher*, *50*(7), 442-450.

Leininger, L. J., Friedsam, D., Dague, L., Mok, S., Hynes, E., Bergum, A., & DeLeire, T. (2011). Wisconsin's BadgerCare Plus Reform: Impact on Low-Income Families' Enrollment and Retention in Public Coverage. *Health services research*, *46*(1p2), 336-347.

Linos, E., Quan, L. T., & Kirkman, E. (2020). Nudging early reduces administrative burden: Three field experiments to improve code enforcement. *Journal of Policy Analysis and Management*, *39*(1), 243-265.

Lopoo, L. M., Heflin, C., & Boskovski, J. (2020). Testing behavioral interventions designed to improve on-time SNAP recertification. *Journal of Behavioral Public Administration*, *3*(2).

McLaine, P., Navas-Acien, A., Lee, R., Simon, P., Diener-West, M., & Agnew, J. (2013). Elevated blood lead levels and reading readiness at the start of kindergarten. *Pediatrics*, *131*(6), 1081-1089.

Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2011). Using implementation intentions prompts to enhance influenza vaccination rates. *Proceedings of the National Academy of Sciences*, *108*(26), 10415-10420.

Miranda, M. L., Keating, M., Kim, D., Hull, A. P., Paul, C. J., & Galeano, M. A. O. (2007)a. Blood lead and water treatment: Miranda et al. respond. *Environmental health perspectives*, *115*(10), A488-A489.

Miranda, M. L., Kim, D., Galeano, M. A. O., Paul, C. J., Hull, A. P., & Morgan, S. P. (2007)b. The relationship between early childhood blood lead levels and performance on end-of-grade tests. *Environmental health perspectives*, *115*(8), 1242-1247.

Mirpuri, S., Pineda, J., McNease, M., Cardon, E., (2022). Increasing lead kit return rates among Chicago residents. The Behavioral Insights Team.

Moynihan, D., Herd, P., & Harvey, H. (2015). Administrative burden: Learning, psychological, and compliance costs in citizen-state interactions. *Journal of Public Administration Research and Theory*, *25*(1), 43-69.

Needleman, H. L., McFarland, C., Ness, R. B., Fienberg, S. E., & Tobin, M. J. (2002). Bone lead levels in adjudicated delinquents: A case control study. *Neurotoxicology and teratology*, *24*(6), 711-717.

Needleman, H. L., Riess, J. A., Tobin, M. J., Biesecker, G. E., & Greenhouse, J. B. (1996). Bone lead levels and delinquent behavior. *Jama*, *275*(5), 363-369.

Needleman, H. L., Schell, A., Bellinger, D., Leviton, A., & Allred, E. N. (1990). The long-term effects of exposure to low doses of lead in childhood: an 11-year follow-up report. *New England journal of medicine*, *322*(2), 83-88.

Nilsson, J. P. (2009). The long-term effects of early childhood lead exposure: Evidence from the phase-out of leaded gasoline. *Institute for Labour Market Policy Evaluation (IFAU) Work. Pap.*

Nisar, M. A. (2018). Children of a lesser god: Administrative burden and social equity in citizen–state interactions. *Journal of Public Administration Research and Theory*, *28*(1), 104-119.

O'Connor, D., Hou, D., Ye, J., Zhang, Y., Ok, Y. S., Song, Y., Coulon, F., Peng, T., & Tian, L. (2018). Lead-based paint remains a major public health concern: A critical review of global production, trade, use, exposure, health risk, and implications. *Environment International*, *121*, 85–101. https://doi.org/10.1016/j.envint.2018.08.052

Rau, T., Urzúa, S., & Reyes, L. (2015). Early exposure to hazardous waste and academic achievement: Evidence from a case of environmental negligence. *Journal of the Association of Environmental and Resource Economists*, *2*(4), 527-563.

Remler, D., Rachlin, J. E., & Glied, S. (2001). What can the take-up of other programs teach us about how to improve take-up of health insurance programs?. *NBER working paper*, (w8185).

Reyes, J. W. (2011). Childhood lead and academic performance in Massachusetts. *Childhood*, *11*(3).

Reyes, J. W. (2015). Lead exposure and behavior: Effects on antisocial and risky behavior among children and adolescents. *Economic Inquiry*, *53*(3), 1580-1605.

Ris, M. D., Dietrich, K. N., Succop, P. A., Berger, O. G., & Bornschein, R. L. (2004). Early exposure to lead and neuropsychological outcome in adolescence. *Journal of the International Neuropsychological Society*, *10*(2), 261-270.

Sanders, M., & Kirkman, E. (2019). I've booked you a place, good luck: A field experimentapplying behavioral science to improve attendance at high impact recruitment events.Journal of Behavioral Public Administration, 2(1).

Schanzenbach, D. (2009). *Experimental estimates of the barriers to Food Stamp enrollment*. Institute for Research on Poverty, University of Wisconsin-Madison.

Schnaas, L., Rothenberg, S. J., Flores, M. F., Martinez, S., Hernandez, C., Osorio, E., & Perroni, E. (2006). Reduced intellectual development in children with prenatal lead exposure. *Environmental health perspectives*, *114*(5), 791-797.

Schwabish, J. A. (2012). Downloading benefits: The impact of online food stamp applications on participation. In *Association for Public Policy Analysis and Management conference, Baltimore, November* (pp. 8-10).

Sorensen, L. C., Fox, A. M., Jung, H., & Martin, E. G. (2019). Lead exposure and academic achievement: Evidence from childhood lead poisoning prevention efforts. *Journal of Population Economics*, *32*, 179-218.

Vinopal, K., & Holt, S. B. (2019). Rookie mistakes: The interplay of teacher experience and racial representation. *Educational Researcher*, *48*(7), 421-437.

Wright, J. P., Dietrich, K. N., Ris, M. D., Hornung, R. W., Wessel, S. D., Lanphear, B. P., ... & Rae, M. N. (2008). Association of prenatal and childhood blood lead concentrations with criminal arrests in early adulthood. *PLoS medicine*, *5*(5), e101.

Yoeli, E., Hoffman, M., Rand, D. G., & Nowak, M. A. (2013). Powering up with indirect reciprocity in a large-scale field experiment. Proceedings of the National Academy of Sciences of the United States of America, 110(Suppl 2), 10424–10429.Journal of Policy Analysis and Management. DOI: 10.1002/pam. Published on behalf of the Association for Public Policy Analysis and Management.

Appendix

Appendix A. Redesigned Kit Stickers

Figure 1A. Stickers for Resident Use

Figure A2. On-the-box stickers

Appendix B. Balance Tests & Additional Analysis

Table B1. Summary Statistics, 2021, Chicago Census Tracts in SMS Text Message Sample

	All People	Treatment	Control	Treat=Control P-Value
Med. Income (2021 dollars)				
Household	$79,461	$78,542	$80,380	0.15
Race/Ethnicity^				
Asian	5.78	6.76	5.94	0.91
Black	26.88	26.85	26.91	0.67
Hispanic	27.38	28.24	26.51	0.71
White	37.24	36.60	37.88	0.64
Other	0.42	0.42	0.42	0.98
Two or more races	2.30	2.27	2.34	0.65
Educational Attainment^				
High School or Less	12.34	11.91	12.76	0.68
High School Graduate	21.70	21.78	21.63	0.68
Some College	23.20	23.33	23.07	0.69
Bachelor's Degree	24.83	24.54	25.12	0.69
Master's Degree	12.57	12.49	12.65	0.70
Professional or Doctorate	5.36	5.27	4.17	0.76
Public Assistance				
Share Participants	17.12	17.07	17.18	0.81
Share Owner-Occupied				
Household	0.90	0.90	0.90	0.31
Share Single Family Home				
Household	0.42	0.41	0.42	0.38
Med. Year Structure Built				
Household	1959	1959	1959	0.84
Med. Value (2021 dollars)				
Household	$323,731	$328,253	$319,205	0.11
Number of Bedrooms^				
No bedroom	4.08	4.22	3.95	0.70
One	14.69	14.94	14.43	0.66
Two	33.18	33.07	33.30	0.72
Three	31.37	31.20	31.54	0.71
Four	11.71	11.65	11.77	0.62
Five or more	4.96	4.92	5.00	0.76
No. of Tracts	3,601	1800	1801	

Notes: Summary statistics are presented for all Chicago census tracts included in the SMS text message intervention as a treatment or control tract. High School or Less includes all who completed any grades between K-11 but did not graduate from high school. High School Graduate includes those with a traditional high school diploma or a high school equivalency certificate (GED). Some College includes those with one or two years of college or an associate degree. Census tracts are determined as "treated" if any test kits in the Census tract received the SMS text message treatment. ^ indicates t-tests for these variables are chi-square tests comparing the distribution of each categorical variable in treatment and control groups.

Table B2. Effects of SMS Intervention, Heterogeneity by Share Ownership & Single-Family Home, 2021

	Test Kit Return	**Pickup Scheduled**	**Engagement Rate**
Share Owner Occupied			
Treatment	0.050***	0.053***	0.050***
	(0.015)	(0.016)	(0.016)
Treatment*25th Percentile	-0.032	-0.030	-0.020
	(0.019)	(0.020)	(0.020)
Treatment*25-50th Percentile	-0.001	0.002	0.005
	(0.019)	(0.020)	(0.020)
Treatment*50-75th Percentile	-0.018	-0.018	-0.007
	(0.018)	(0.018)	(0.019)
25th Percentile	0.002	0.004	0.003
	(0.009)	(0.010)	(0.011)
25th-50th Percentile	-0.002	-0.002	-0.004
	(0.009)	(0.009)	(0.009)
50th-75th Percentile	-0.012**	-0.013**	-0.015
	(0.008)	(0.008)	(0.009)
Constant	0.026	0.025	0.029
Share Single Family	**Test Kit Return**	**Pickup Scheduled**	**Engagement Rate**
Treatment	0.048***	0.055***	0.055***
	(0.013)	(0.014)	(0.014)
Treatment*25th Percentile	0.000	-0.006	-0.003
	(0.018)	(0.019)	(0.019)
Treatment*25-50th Percentile	-0.020	-0.018	-0.011
	(0.018)	(0.019)	(0.020)
Treatment*50-75th Percentile	-0.024	-0.027	-0.027
	(0.016)	(0.017)	(0.018)
25th Percentile	0.004	-0.000	0.004
	(0.007)	(0.007)	(0.008)
25th-50th Percentile	0.019	0.016	0.015
	(0.009)	(0.009)	(0.009)

50th-75th Percentile	-0.001	-0.004	-0.001
	(0.006)	(0.006)	(0.007)
Constant	0.016	0.018	0.019

Notes: Census tract median share of owner-occupied units was 0.87 at the 25th percentile, 0.92 at the 50th percentile, and 0.95 at the 75th percentile of the homeowner distribution across all Chicago census tracts. Census tract median share of single-family homes was 0.18 at the 25th percentile, 0.35 at the 50th percentile, and 0.64 at the 75th percentile of the single-family home share distribution across all Chicago census tracts. ***p<0.01. **p<0.05. Treatment refers to the omitted 75th percentile and above group. In each column all sets are pooled together and dummy variables for each set are included in each regression. Set 2 is the omitted set.