Representation and Hesitancy in Population Health Research: Evidence from a COVID-19 Antibody Study

Deniz Dutz, Michael Greenstone, Ali Hortaçsu, Santiago Lacouture, Magne Mogstad, Azeem M. Shaikh, Alexander Torgovitsky & Winnie van Dijk*

February 10, 2023

Abstract

We examine data from a serological study that randomized participation incentives ($0, $100, $500). Minority and poor households are underrepresented at lower incentives. We develop a framework that uses randomized incentives to disentangle non-contact and hesitancy and find that underrepresentation occurs because minority and poor households are more hesitant to participate, not because they are harder to contact. In particular, reservation payments for contacted households in minority and poor neighborhoods are substantially more likely to exceed $100. The $500 incentive closes hesitancy gaps and restores representativeness on observable dimensions including hospitalization and insurance rates, and a COVID-19 risk index.

JEL classification: C83; I14

*Dutz: Department of Economics, University of Chicago (ddutz@uchicago.edu). Greenstone: Department of Economics, University of Chicago (mgreenst@uchicago.edu). Hortaçsu: Department of Economics, University of Chicago (hortacsu@uchicago.edu). Lacouture: Department of Economics, University of Chicago (slacouture@uchicago.edu). Mogstad: Department of Economics, University of Chicago (mogstad@uchicago.edu). Shaikh: Department of Economics, University of Chicago (amshaikh@uchicago.edu). Torgovitsky: Department of Economics, University of Chicago (torgovitsky@uchicago.edu). Van Dijk: Department of Economics, Harvard University and NBER (winnie_vandijk@fas.harvard.edu). Maya Roy provided excellent research assistance. This study is approved by the IRB at the University of Chicago (IRB20-0721). This RCT was registered as AEARCTR-0010630 (Greenstone, 2023). We gratefully acknowledge financial support from the University of Chicago Women's Board, the Sloan Foundation, the Smith Richardson Foundation, the Becker Friedman Institute, the Booz Allen Foundation, and the Flu Lab.

1 Introduction

It is widely documented that scientific studies with human subjects suffer from unequal participation rates across socioeconomic and demographic groups. Nonrepresentative studies undermine the proper allocation of both public and private resources, e.g. by distorting clinical trial results used to inform R&D decisions (NASEM, 2022), and by biasing population statistics that guide policy makers (Wines and Cramer, 2022). Yet, little is known about the key barriers to research participation and how they can be addressed. A recent report for the National Academies of Science, Engineering, and Medicine (NASEM, 2022, p. 107) summarizes the state of the literature:

> *"There is substantial quantitative data demonstrating the size and scope of the problem of underrepresentation and exclusion of populations in research; however, there is a dearth of critical qualitative data about facilitators of successful inclusion."*

The report argues that, as a consequence,

> *"...large swaths of the U.S. population, and those that often face the greatest challenges, are less able to benefit from [new] discoveries because they are not adequately represented in scientific studies."*

There are two broad reasons for non-participation, and thus there are two broad drivers of non-representativeness: either researchers are unable to make contact with a sampled household (non-contact), or a contacted household does not participate because the perceived costs of doing so exceed the perceived benefits (hesitancy). These two reasons for non-participation suggest different approaches to solving the problem of underrepresentation and exclusion in research. Hence, it is important to understand their respective roles. Moreover, in the case of hesitancy, it is important to understand whether financial compensation—a commonly used strategy to increase response rates—is effective in overcoming this barrier.

This paper provides a rare opportunity to learn about causes of and solutions to unequal participation in the context of a COVID-19 serological study. In most practical settings, it is difficult to quantify the relative importance of non-contact and hesitancy for non-participation. For example, in studies with mail, SMS, or online outreach, non-contact is not directly observed. Even when non-contact can be directly measured (as in the less common case of studies with in-person outreach), it is impossible to learn about the distribution of reservation payments for participation.

Unlike these typical settings, the Representative Community Survey Project's (RECOVER) COVID-19 serological study experimentally varied financial incentives for participation. The study was conducted with households in Chicago. Sampled households were sent a package that contained a self-administered blood sample collection kit, and were asked to return the sample by mail to our partner research laboratory in order to be tested for the presence of

COVID-19 antibodies ("seropositivity").[1] Households in the sample were randomly assigned one of three levels of financial compensation for participating in the study: $0, $100, or $500.

The random assignment of incentives creates ex-ante identical groups whose participation rates only differ if the financial compensation causes the benefits to exceed the costs. We develop a framework that uses this experimentally-induced variation together with a simple model of participation behavior to separately identify and estimate the relative importance of non-contact and hesitancy for non-participation. When applied to the RECOVER study, our estimates show that both non-contact and hesitancy are important drivers of non-participation, but that inequality in participation across demographic groups is primarily driven by differences in hesitancy.

Variation in assigned incentives further allows us to learn how contacted households differ in the incentives needed to induce participation. We find that reservation payments for contacted households in minority and poor neighborhoods are substantially more likely to exceed $100. The $500 incentive appears to overcome differences in hesitancy and restore representativeness along observable dimensions, including important plausible confounds in population health studies such as the uninsured rate, hospitalization rates, and an aggregate COVID-19 risk index.

We can rule out differences in perceived benefits of learning one's seropositivity status as an explanation for these differences in hesitancy, since participants were not told their test result. A substantial qualitative literature instead points to differences in trust in the healthcare system and differences in concern about privacy as potential factors limiting study participation among racial minorities.[2]

In Section 2, we describe the RECOVER serological study in greater detail. RECOVER was designed and implemented using best practices in collaboration with NORC, a leading national statistical agency, and the Wilson Antibody Biology Laboratory at the University of Chicago. The collected data consists of the randomly-assigned compensation offer, participation status, and addresses for each sampled household. These addresses are used to link households to a rich set of neighborhood-level (i.e., zipcode-level) characteristics, such as poverty, racial composition, and health, independently of whether the households participated in the study. This allows us to assess to what extent participants are representative of the target population, at least along these observable dimensions.

We estimate participation rates across the randomly assigned incentive levels in Section 3. In the control group, the participation rate is just 6%, which is comparable to the participation rate in other serological studies (Bobrovitz et al., 2021, and our Appendix C). We find that financial compensation has a powerful effect on participation: the $100 incentive

[1] As shown in Appendix C, most COVID-19 serological studies that randomly sample participants use mail or online outreach, as these approaches offer a pragmatic way to obtain contact-free seroprevalence data.

[2] A recent review can be found in Chapter 4 in NASEM (2022). In line with these studies, Alsan et al. (2022) report that Black Americans are less likely to have confidence in research institutions, believe that science is beneficial for them, or enroll in clinical trials. Alsan et al. (2022) also find that Black patients are more likely to cite trust, privacy, and racism as reasons not to enroll in clinical trials, whereas White patients cite logistical barriers and co-morbidities.

increases participation to 17%, and the $500 incentive increases it to 29%. We also find striking differences in participation rates by neighborhood characteristics. For example, in the unincentivized arm, only 2% of households in high poverty neighborhoods participate, compared to 10% in low poverty areas. The $100 incentive substantially increases participation among all groups, but widens differences in participation rates. The $500 incentive increases participation further and, more importantly, it completely closes the participation gap.

We develop a framework for quantifying non-contact and hesitancy as drivers of non-participation in Section 4. In the RECOVER study—as in many other population health studies and social surveys—a sampled household must first be successfully contacted. Because the randomly assigned incentive is only revealed after the household is contacted, non-contact rates do not depend directly on the incentive level. Conditional on being contacted, households decide whether to participate by comparing perceived costs and benefits, which are shifted by randomly assigned incentives. The key assumption is a bound on the proportion of contacted households who would decline to participate at $500. We show how to use the model to quantify non-contact and hesitancy as sources of non-participation, and to learn about differences in the distribution of reservation payments across demographic and income groups. The analysis highlights the critical importance of randomly varying financial incentives, which is rarely done in studies with voluntary random testing.[3]

We apply these methods to the RECOVER study in Section 5. We find that both non-contact and hesitancy are important determinants of low participation. However, these determinants have different implications for who participates. Non-contact rates differ little by household demographics. But households from higher poverty and minority neighborhoods have much higher hesitancy rates, implying that they have higher perceived costs of participation. Decomposing these causes of unequal participation, we find that hesitancy explains 89% of the participation gap at $0, and 93% at $100. Examining the distribution of reservation payments across groups, we find that reservation payments for *contacted* households in poor and minority neighborhoods are substantially more likely to exceed $100.

For example, we estimate that 61% of contacted households in majority minority neighborhoods would not participate for $100, compared to only 14% in majority White neighborhoods. We show that our findings remain qualitatively unchanged even if we assume that a substantial share of contacted households decline to participate at $500. Removing this assumption entirely actually *strengthens* our conclusions about hesitancy, although it weakens our conclusions about the importance of non-contact.

We estimate the extent to which financial incentives overcome the nonrepresentativeness caused by differential hesitancy in Section 6. We stratify average characteristics of participating households by incentive level, and compare these characteristics to those of the entire

[3]In Appendix C, we show that in all publications on COVID-19 serology studies with randomly sampled U.S. subjects, incentives are either not specified, are offered uniformly, or are varied in a non-random way. Dutz et al. (2021) reach a similar conclusion in their systematic review of social surveys used in empirical economics research.

invited sample. We show that without financial incentives, the participating households are highly nonrepresentative of the invited sample along a range of socioeconomic, racial, and health dimensions, including the risk of COVID-19 infection. As one example, only 13% of the participants in the unincentivized arm are from higher poverty neighborhoods, compared to 46% in the target population. Although this gap largely persists at the $100 incentive, the $500 incentive closes it entirely. We find that the $500 arm is representative across a battery of socioeconomic, risk, and health measures.

Our paper contributes to ongoing discussions about the quality of COVID-19 serological studies, and in particular the challenges that stem from a lack of representation. Serological studies were widely used to estimate epidemiological parameters that served as inputs to highly consequential health policy decisions, such as the Infection Fatality Rate (IFR, the likelihood of death conditional on infection) and the share of the population already infected. Although serological studies were implemented in part to address bias due to the existence of asymptomatic and untested infections (Aspelund et al., 2020; Manski and Molinari, 2021), systematic reviews and meta analyses have emphasized that they often relied on nonrepresentative ("convenience") samples, exposing them to a different potential source of bias (see, e.g., Bobrovitz et al., 2021; Chen et al., 2021).[4] Due to the exponential nature of transmission models, even small biases can translate into large forecast errors (see, e.g., Ioannidis et al., 2022).

The problem of nonrepresentativity is not specific to studies that rely on convenience samples. It also occurs in studies with random sampling, because some types of households may be relatively difficult to reach or unwilling to participate. Our paper contributes to this discussion by developing tools to measure the causes of nonrepresentativeness. While we find non-contact rates that are high overall, we are also able to isolate the cause of nonrepresentativeness along racial and poverty lines to differences in hesitancy, rather than differences in non-contact rates. This finding implies that representativeness along these characteristics can be improved by providing higher incentives to participate.

Our paper is also related to a broader literature on unequal representation in scientific studies. A comprehensive review of this literature can be found in a recent report published by the National Academies of Science Engineering and Medicine (NASEM, 2022). As noted above, this report concludes that, to date, there is limited credible evidence on the *causes of* and *solutions to* low and unequal participation across demographic groups. Our study directly addresses these questions.

Our paper is complementary to recent work in economics that studies the *consequences* of unequal representation in clinical trials. Alsan et al. (2022) propose a model that predicts

[4]Concerns about nonrepresentativeness of study samples have also been raised in relation to other types of studies that aimed to inform pandemic-era health policy. For example, Bradley et al. (2021) argued that nonrepresentative surveys substantially overestimated US vaccine uptake due to overrepresentation of highly educated and white participants. Beyond pandemic health policy, concerns have recently been raised about under-counting of Hispanic, Black and Native American residents in the 2020 U.S. Census, which may lead to under-allocation of government resources to these groups (Wines and Cramer, 2022).

that unequal representation in clinical trials reduces the extent to which innovation benefits the underrepresented groups. Key parameters in the model are the perceived net benefits for White and Black patients, which drive differences in recruitment costs across these groups. However, Alsan et al. (2022) note that there is no publicly available data on trial recruitment cost, let alone on how these costs vary across demographic groups.[5] While our setting is not a clinical trial, participation decisions in our setting are likely informed by similar considerations. Our evidence on the causes of non-participation and how they vary by race may be useful for informing clinical trial recruitment.

Our paper is also connected to a survey methodology literature that examines how incentives affect both survey participation rates and the demographic composition of survey participants (see Groves et al. (2009); Singer and Ye (2013), and references therein). However, this work has not used randomized incentives to disentangle non-contact and hesitancy as causes of non-participation, nor has it examined how these drivers vary across demographic groups.[6] Existing work that randomizes incentives is typically conducted in lower-stakes settings and the incentive levels over which randomization occurs are orders of magnitude smaller than the incentives we consider. For example, Petrolia and Bhattacharjee (2009) consider a mail survey on consumer preferences for ethanol-based fuels in which invited individuals were randomly offered either no incentives, $1 along with the survey, or a promise of $5 upon completion of the survey. They find that higher incentives affect the participant composition by bringing in a larger percentage of less educated respondents. We are not aware of studies that randomize financial compensation for sharing personal health information for research or policy purposes.

Finally, there is a literature in public economics on how to increase the take-up of social programs. The key barriers are lack of information about eligibility and transaction costs (including stigma) associated with enrollment (see e.g., Moffitt, 1983; Currie, 2006; Dahl et al., 2014; Deshpande and Li, 2018; Finkelstein and Notowidigdo, 2019). In contrast, all sampled households in the RECOVER study are eligible to participate. Instead, the barriers to participation are the researchers' inability to contact the sampled household and their hesitancy to participate conditional on being contacted. There is no private benefit to participation in our context, other than the randomized financial compensation, because participants would not be informed about the result of the test. Thus, participation decisions reflect a tradeoff between perceived private costs and social incentives, rather than eligibility or transaction costs.

[5] On pg. 9 they write: "The cost—in terms of both money and time—of enrolling a new patient in a trial also varies across demographic groups. To the best of our knowledge, there are no publicly available estimates of trial recruitment costs."

[6] There exists, however, some direct measures of the importance of non-contact. For example, some U.S. household surveys run by national statistical agencies are conducted in-person and record reasons for non-participation. Using this type of metadata, Brick and Williams (2013) show that both hesitancy (refusals) and non-contact are drivers of non-participation in surveys outside the domain of public health.

2 Study design and implementation

2.1 Background

The RECOVER serological study was carried out in Chicago between December 2020 and March 2021. The study was designed and conducted in collaboration with two partners from the University of Chicago: NORC, a leading survey and research organization, and the Wilson Antibody Biology Laboratory. The RECOVER study was a pilot study intended to measure participation rates at different levels of compensation across neighborhoods. The results of the pilot were meant to inform the sampling design of a larger study on seropositivity in Chicago. The larger study was never implemented, partly because of low overall participation rates in RECOVER (see Section 3), and partly because the advent of vaccines made seropositivity a lower public health priority.

2.2 Design and implementation

NORC randomly sampled 882 Chicago addresses from United States Postal Service data. Hence, the sampled households were representative of the population of households with a mailing address in the city. Sampled households were sent a package that contained a self-administered blood collection kit, and were asked to return a blood sample to the Wilson Lab to be tested for seropositivity. The package additionally contained a consent form with a short questionnaire, instructions on self-administering and returning a blood sample, a letter explaining the purpose of the study and providing information on financial compensation for participating (i.e., returning a blood sample), and a pre-paid return package.

Households in the sample were randomly assigned one of three levels of compensation: \$0, \$100, or \$500. The latter is quite a high level of compensation for participation in a serological study. The reason for offering such high compensation was to try to maximize participation in this arm of the experiment. In Section 4, we show how to use the \$500 incentive arm to disentangle non-contact and hesitancy rates for the lower incentive arms.

Households were asked to select the adult with the earliest upcoming birthday for participation in the study. The letter informed the household that the returned blood sample would be tested for seropositivity, but that they would not be told the result of the test. Hence, desire to learn about seropositivity status could not contribute to the household's motivation for participating in the study. Appendix A contains copies of the written materials and additional details on sampling, randomization and follow-up procedures.

2.3 Data

Our data consists of the randomly assigned incentive level, participation status, and address for each sampled household. Although we observe additional data via the short questionnaire for participating households, we do not have such information for non-participating households. However, we are able to merge our data with address-based information from external sources. As a result, we can observe a rich set of address-based neighborhood characteristics

for each household, independently of whether they participated in the study. This allows us to analyze non-participation conditional on these characteristics and to compare how the composition of participants differs from that of the invited sample.

We focus our attention on two neighborhood characteristics that feature prominently in discussions of representativeness (NASEM, 2022): poverty status and racial composition. We classify a household as being from a higher poverty neighborhood if the percentage of households below 1.5 times the poverty line is above 30% (and classify it as lower poverty otherwise). We classify a household as being from a majority minority neighborhood if the share of adults identifying as non-Hispanic White is below 50% (and classify it as majority White otherwise). This data is measured at the zipcode level, and is obtained from the American Community Survey.[7] In additional analyses, we also consider other dimensions of neighborhood characteristics, including labor market and health conditions. We obtain this information from the Chicago Health Atlas and the City Health Dashboard (see Appendix B for details).

3 Participation rates and responsiveness to incentives

Figure 1a reports the proportion of households who participated in RECOVER by incentive level. Only 6 percent of unincentivized households participated. This rate is comparable to other serological surveys that invited a random sample of households to be tested for COVID-19 antibodies.[8]

Participation rates increase substantially with the level of the incentive. Offering $100 for participation increases participation rates to 17%, almost triple the rate without incentives. Offering $500 for participation increases participation rates further to 29%.

Figures 1b and 1c report participation rates by neighborhood poverty status and racial composition. While 10% of households in lower poverty neighborhoods participate without financial incentives, only 2% in higher poverty neighborhoods do. While 9% of households in majority White neighborhoods participate without financial incentives, only 4% in majority minority neighborhoods do. The $100 incentive increases participation substantially in lower poverty and majority White neighborhoods, but only modestly in higher poverty and majority minority neighborhoods. However, the $500 incentive increases participation rates to almost parity, with no statistically significant differences by either poverty status or racial composition.

4 A model of study participation

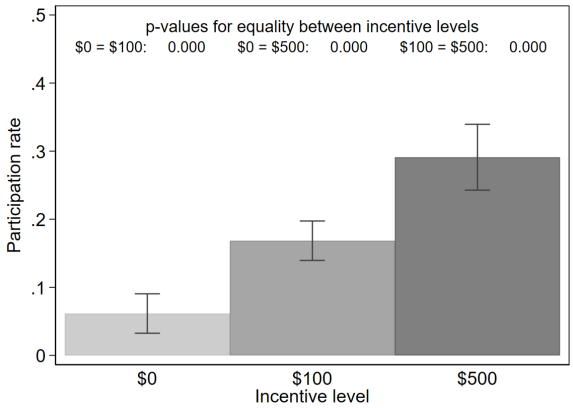
There are two potential explanations for our finding of unequal participation rates. To participate, a household must first be successfully contacted and made aware of the study.

[7] Our conclusions do not materially change if we instead use tract-level data.

[8] In Appendix C, we show that in all serological studies that randomly sampled subjects from a region within the United States, the average (median) participation rate for studies that used mail outreach is 9.0% (8.3%).

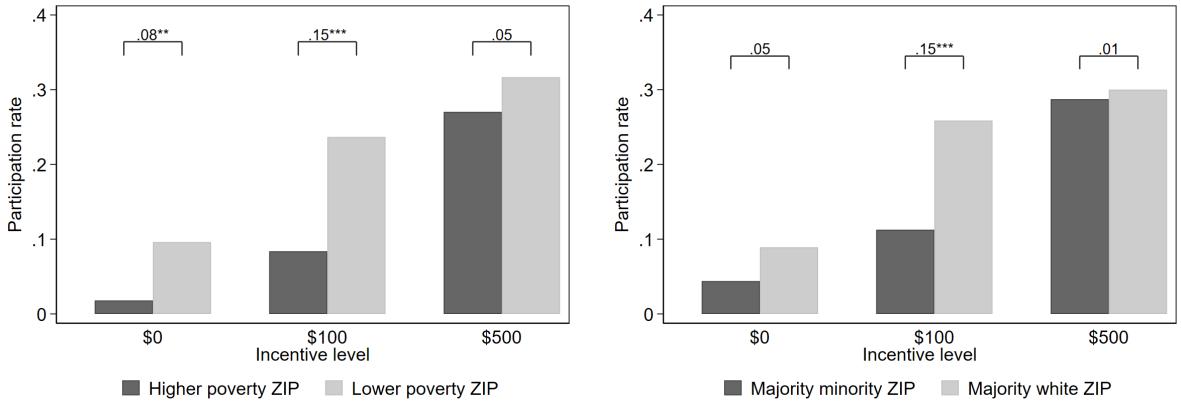
Figure 1: *Participation rates*

(a) *Overall*



(b) *By poverty status*

(c) *By racial composition*



Notes: This figure reports participation rates and 90% confidence intervals by incentive group for the overall sample (a), by neighborhood poverty status (b), and by neighborhood racial composition (c). In panel (a), p–values for testing the pairwise equality in participation rates across incentives are shown in the top. In panels (b) and (c), differences in participation levels across subgroups are depicted above the bars. Stars denote p-values smaller than .1 (*), .05 (**), and .01 (***).

Upon being contacted, the household must then decide to participate. Unequal participation rates are due to some combination of systematic differences in the difficulty of contact and hesitancy to participate. In this section, we model the participation process to quantify non-contact and hesitancy as sources of non-participation. We also measure how contacted households differ in their hesitancy to participate.

4.1 Model

Let $R_i(z) \in \{0,1\}$ denote whether household i would participate if assigned incentive z. Participation is a two-step process in which the household is first *contacted*, and then *decides* to participate. Let $C_i(z) \in \{0,1\}$ denote whether household i would be contacted under incentive level z, and let $D_i(z)$ denote whether they would decide to participate if contacted. Then household i's participation decision is $R_i(z) = C_i(z)D_i(z)$. We will estimate the model separately by demographic groups without any cross-group restrictions, so

we suppress demographic conditioning in the notation.

We impose three baseline assumptions on this model. First, since the assigned incentive is only revealed after the household is contacted and opens the package, we assume that contact does not depend on z, so that $C_i(z) \equiv C_i$. Second, we assume that $D_i(z)$ is non-decreasing in z for all i, so that households are more likely to participate under higher incentives. This is the Imbens and Angrist (1994) monotonicity assumption, which Vytlacil (2002) showed is equivalent to assuming that $D_i(z) = \mathbb{1}[H_i \leq z]$ for some latent variable H_i. Together, these two assumptions imply that

$$R_i(z) = C_i \mathbb{1}[H_i \leq z]. \tag{1}$$

We interpret $z - H_i$ as household i's net benefit from participating, and call H_i their *hesitancy* to participate. Here, a household's hesitancy is the reservation payment that they are willing to accept for participation in the study. If contacted, the household participates if the offered financial incentive exceeds their hesitancy. Third, we assume that the assigned incentive, Z_i, is independent of (C_i, H_i), which is justified by random assignment of incentives.

4.2 Contact and hesitancy rates

We define the *contact rate* as $\gamma \equiv \mathbb{P}[C_i = 1]$ and the *non-contact rate* as $1 - \gamma$. We define the *hesitancy rate* as $\eta(z) \equiv \mathbb{P}[H_i > z | C_i = 1]$, which is the probability that a household would not participate under incentive z if they were contacted. We measure the hesitancy rate conditional on being contacted in order to hold fixed the implementation protocol of the scientific study. Variation in $\eta(z)$ allows us to learn about the distribution of hesitancy (reservation payments) for contacted households.

4.3 Identification and estimation

The researcher does not observe (C_i, H_i), but only the incentive level, Z_i, and the participation decision $R_i \equiv R_i(Z_i)$ under this incentive level. From these observables, they can estimate the *participation rate*

$$\rho(z) \equiv \mathbb{P}[R_i = 1 | Z_i = z] = \mathbb{P}[C_i = 1, H_i \leq z], \tag{2}$$

where the equality follows from the model (1) and random assignment of the incentive, Z_i. Measuring the contact and hesitancy rates requires determining the relative contribution of the unobservables C_i and H_i to ρ, while allowing these unobservables to be dependent.

To make progress, we consider what can be said about the contact and hesitancy rates under assumptions on the magnitude of the hesitancy rate at the highest incentive, $\bar{z}$. In the RECOVER survey, $\bar{z} = \$500$ is large, suggesting that $\eta(\bar{z})$ is small, and that non-participation in the \$500 treatment arm is primarily or solely due to non-contact. Since contact is not affected by the incentive level, the participation model then allows us to infer the hesitancy

rates at lower incentives as well.

To see how this works, suppose that we know $\eta(\bar{z})$ exactly and decompose it as

$$\eta(\bar{z}) = \frac{\mathbb{P}[C_i = 1, H_i > \bar{z}]}{\mathbb{P}[C_i = 1]} = \frac{\mathbb{P}[C_i = 1] - \overbrace{\mathbb{P}[C_i = 1, H_i \leq \bar{z}]}^{= \rho(\bar{z}) \text{ by (2)}}}{\underbrace{\mathbb{P}[C_i = 1]}_{\equiv \gamma}} = 1 - \frac{\rho(\bar{z})}{\gamma}.$$

Rearranging shows that the contact rate γ (and non-contact rate $1 - \gamma$) is identified:

$$\gamma = \frac{\rho(\bar{z})}{1 - \eta(\bar{z})}. \tag{3}$$

Hesitancy rates at other incentive levels can then be identified by the following argument:

$$\begin{aligned} \eta(z) &= \mathbb{P}[z < H_i \leq \bar{z} | C_i = 1] + \mathbb{P}[H_i > \bar{z} | C_i = 1] \\ &= \frac{\rho(\bar{z}) - \rho(z)}{\gamma} + \eta(\bar{z}) = \left(\frac{\rho(\bar{z}) - \rho(z)}{\rho(\bar{z})} \right) (1 - \eta(\bar{z})) + \eta(\bar{z}), \end{aligned} \tag{4}$$

where the second equality used (2), and the third equality substituted in the identified contact rate from (3). We estimate (3) and (4) through their sample analogs by substituting the estimated participation rates $\rho(z)$ and $\rho(\bar{z})$.

Our baseline estimates set $\eta(\bar{z}) = 0$, which corresponds to the assumption that any household would have participated at \$500 incentive had they been aware of it (had they been contacted). Given the generosity of the incentive, we view this as a reasonable assumption. However, we also report estimates that allow $\eta(\bar{z})$ to vary in the set $[0, \alpha]$, where α is a number smaller than $1 - \rho(\bar{z})$, the largest value that keeps γ a proper probability via (3). Under this assumption, bounds on γ and $\eta(z)$ are given by

$$\rho(\bar{z}) \leq \gamma \leq \frac{\rho(\bar{z})}{1 - \alpha} \quad \text{and} \quad \frac{\rho(\bar{z}) - \rho(z)}{\rho(\bar{z})} \leq \eta(z) \leq \frac{\rho(\bar{z}) - \rho(z)(1 - \alpha)}{\rho(\bar{z})}. \tag{5}$$

The widest "worst-case" bounds are obtained at $\alpha = 1 - \rho(\bar{z})$. In Appendix D, we prove that these bounds are sharp (best possible, given the assumptions) for any choice of α, as long as observed participation rates $\rho(z)$ are increasing in z.

5 The causes of low and unequal participation rates in RECOVER

We now use the method in the previous section to separately estimate non-contact and hesitancy in the RECOVER study. We then use the variation in assigned incentives to learn about the distribution of hesitancy (reservation payments).

Table 1: *Estimated non-contact and hesitancy rates*

		Hesitancy rate	
	Non-contact rate	At $0	At $100
Overall	0.71	0.79	0.42
	(0.66, 0.76)	(0.68, 0.89)	(0.28, 0.56)
Higher poverty	0.73	0.93	0.69
	(0.68, 0.78)	(0.80, 1.00)	(0.55, 0.83)
Lower poverty	0.68	0.70	0.25
	(0.60, 0.76)	(0.54, 0.85)	(0.02, 0.49)
Difference	0.05	0.24	0.44
	(-0.05, 0.14)	(0.03, 0.45)	(0.17, 0.71)
Majority minority	0.71	0.85	0.61
	(0.66, 0.76)	(0.73, 0.97)	(0.47, 0.74)
Majority White	0.70	0.70	0.14
	(0.60, 0.80)	(0.50, 0.90)	(0.00, 0.47)
Difference	0.01	0.14	0.47
	(-0.09, 0.12)	(-0.08, 0.36)	(0.14, 0.80)

Notes: This reports estimates of non-contact and hesitancy rates under the baseline assumption that all contacted households would choose to participate if offered $500. 90% CIs are shown in parentheses.

5.1 Baseline estimates

Table 1 reports our baseline estimates, which are constructed under the assumption that all households would choose to participate at $500 if they were aware of the study ($\eta(\bar{z}) = 0$).

The first column of Table 1 shows estimates of the non-contact rate. Under our baseline assumption, all households who did not participate at $500 did so because they were not contacted, and so our estimates of the non-contact rate are the complement of the participation rates at $500 shown in Figure 1. Participation rates of 29% under the $500 incentive arm correspond to non-contact rates of 71%. Consistent with Figure 1, we find no large or statistically significant differences in non-contact rates across households by neighborhood poverty or racial composition.

The second column of Table 1 shows estimates of the hesitancy rate when no financial incentive is offered. With no financial incentive, 79% of contacted households would not participate. This figure increases to 93% for households in high poverty neighborhoods and to 85% for households in majority minority neighborhoods. These findings suggest that that the perceived costs of participation are empirically relevant barriers to participation, especially for poor and minority households.

Variation in assigned incentives allows us to learn about the distribution of hesitancy and how it varies across groups. The third column of Table 1 shows that a $100 incentive

sharply decreases the overall hesitancy rate from 79% to 42%. However, the decrease is largely driven by households in lower poverty and majority White neighborhoods: in majority White neighborhoods, only 14% of contacted households decline to participate when offered the $100 incentive. Hesitancy rates remain substantial among households in higher poverty and majority minority neighborhood. Whereas reservation payments for contacted households in poor and minority neighborhoods are somewhat more likely to exceed $0, they are 2.5-4 times more likely to exceed $100. These findings suggest that the perceived costs of participation are high in general, and much higher for poor and minority households.

In interpreting what may explain these differences in hesitancy, we can rule out differences in perceived benefits of learning one's seropositivity status as an explanation for these differences in hesitancy, since participants were not told their test result. A substantial qualitative literature instead points to differences in trust in the healthcare system and differences in concern about privacy as potential factors limiting study participation among racial minorities (see, e.g., Chapter 4 of NASEM, 2022; Alsan et al., 2022).

5.2 Decomposing the causes of unequal participation

A decomposition exercise helps clarify the relative importance of non-contact and hesitancy for explaining unequal participation by poverty status and racial composition. Suppose that majority minority households had the same hesitancy at $0 as majority White households. Then, instead of a participation rate of .043 at $0, majority minority households would have a $(1 - .71) \times (1 - .70) = .085$ participation rate, only slightly lower than the .090 participation rate for majority White households, and eliminating 89% of the participation gap. The same calculation for the $100 incentive brings participation for majority minority households from 11.3% to 24.8%, relative to 25.8% for majority White households, eliminating 93% of the participation gap. Similarly, if higher poverty households had the same hesitancy as lower poverty households, their participation would rise from 1.8% to 8.2% at $0 and 8.3% to 20.2% at $100, compared to 9.6% at $0 and 24.0% at $100 for lower poverty households. In all cases, setting hesitancy rates equal across households largely closes participation gaps across poverty status and racial composition. These results suggest that unequal participation rates are primarily driven by differences in hesitancy.

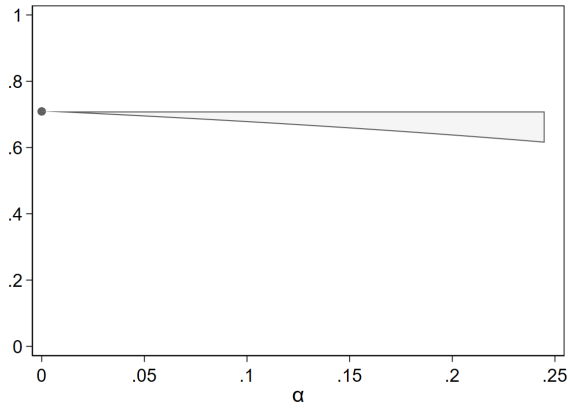
5.3 Sensitivity analysis

The estimates in Table 1 use the assumption that all contacted households would choose to participate if offered the $500 incentive. That is, the hesitancy rate at $500 is zero or, in the notation of Section 4, $\eta(\bar{z}) = 0$. In this section, we conduct a sensitivity analysis that estimates bounds on the same parameters under the weaker assumption that $\eta(\bar{z}) \leq \alpha$.

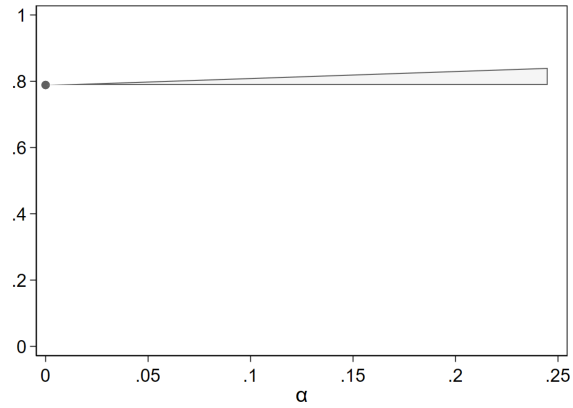
Figure 2 plots the estimated bounds on the overall non-contact and hesitancy rates for α up to .25. For example, allowing $\alpha = .20$ means assuming that up to 20% of contacted households decline to participate at $500 because they find the incentive not high enough to overcome their perceived costs. Even under this conservative assumption, Figure 2a shows

Figure 2: *Bounds on non-contact and hesitancy rates*

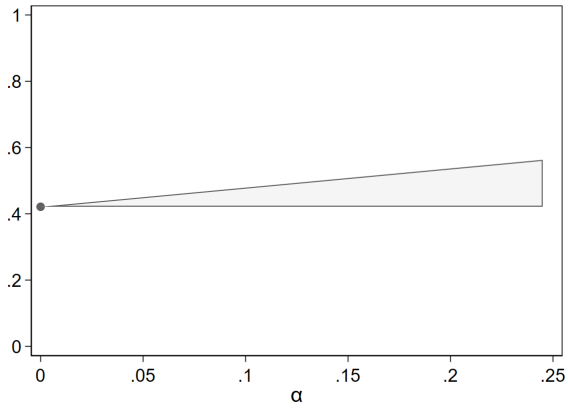
(a) *Non-contact rate*



(b) *Hesitancy rate at $0*



(c) *Hesitancy rate at $100*



Notes: These figures report estimates of the bounds in (5) for different levels of α.

that non-contact rates remain high at 64%. Higher hesitancy rates at $500 also rationalize higher hesitancy rates at lower incentive values (see (5)), reinforcing the conclusion that hesitancy is also an important source of non-participation. At $\alpha = .20$, between 79% and 83% of contacted households would not participate without an incentive.

Figure 3 plots estimated bounds by demographic group. Figures 3a and 3b show that the bounds on non-contact rates by demographic group largely overlap for all $\alpha \leq .25$, reinforcing the conclusion that non-contact rates do not vary systematically by demographics. Figures 3c–3f show that the opposite is true for hesitancy rates: even at $\alpha = .25$, hesitancy rates at both $0 and $100 differ markedly by both poverty status and racial composition. These results are consistent with the conclusions from the baseline case.

As discussed in Section 4, the largest value that we can set α to while still rationalizing the model is $1 - \rho(\bar{z})$, which we estimate to be 71% among the overall population. This value of α represents the "worst-case" assumption that everyone in the $500 incentive arm was contacted, but 71% declined to participate because $500 was not a sufficient incentive. If this were true, then non-contact rates would be zero, and hesitancy rates at lower incentives

would be even larger; for example between 79% and 94% at $0. Thus, even without taking a stand on α, we can conclude that hesitancy is an important barrier to participation. However, our view is that allowing for the possibility that 71% of contacted households would not trade $500 for a quick at-home blood sample is unreasonable. Under smaller—but still large—values of α, we find non-contact to also be an important cause of non-participation.

6 How financial incentives affect representativeness

The findings in the previous section show that both non-contact and hesitancy are important causes of low participation, but that only hesitancy differs by poverty status and racial composition. This result implies that different perceived costs of participation lead to non-representativeness in unincentivized studies, but that incentivized studies can improve representativeness. In this section, we quantify the extent to which this is the case.

Figure 4 reports average neighborhood-level characteristics of participating households. The estimates are stratified by incentive level; intuitively, one can think of each incentive level as representing different studies conducted among ex-ante identical populations. The dotted horizontal lines show the mean neighborhood-level characteristics of the target population. Estimates closer to the dotted line are more representative.

The results show that an unincentivized study would be highly nonrepresentative. A study with a $100 incentive would also be highly nonrepresentative on all six characteristics.[9] However, a study with a $500 incentive is representative on all six characteristics, with average participant characteristics that are statistically indistinguishable from those in the target population. For example, with no incentive only 13% of the participants are from higher poverty neighborhoods, compared to a target population average of 46%. This figure would increase to 22% with a $100 incentive, and to 51% with a $500 incentive.

Given the RECOVER study's goal, a particularly important dimension of nonrepresentativeness is the COVID-19 local risk index (Figure 4f). This index directly relates to the goals of our partners at the Wilson Antibody Biology Laboratory.[10] A study with no incentive or a $100 incentive would understate the average COVID-19 risk index in the target population by more than 1.5 points on a 10-point scale, but a study with a $500 incentive would be almost exactly representative.

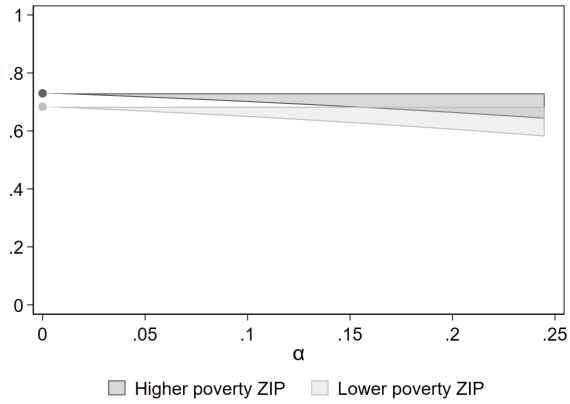
Appendix Figure E.1 shows similar patterns across a variety of additional risk and health measures. A study with a $500 incentive would be representative for the share of smoking adults, hospitalization rates related to alcohol, behavioral health, and mood or depressive orders, as well as for a credit insecurity index. In contrast, both an unincentivized study and one with a $100 incentive would not be representative for any of these measures, except the

[9] The six characteristics are positively correlated but typically not highly so; Appendix Table E.1 shows that the correlations lie within 0.11 and 0.87.

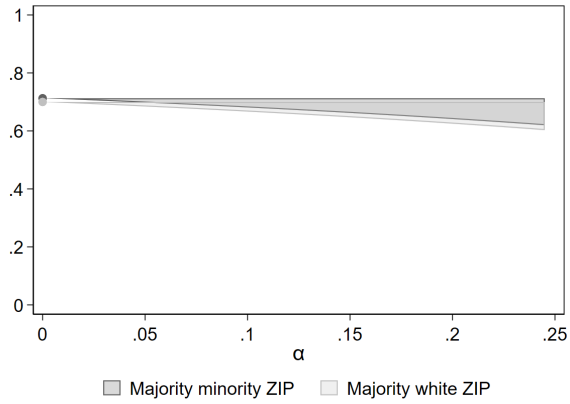
[10] Although we observe antibody test results for participants, we do not observe them for non-participants. Our findings suggest one should be extremely cautious in using test results from only the self-selected group of participants.

Figure 3: *Bounds on non-contact and hesitancy rates by demographics*

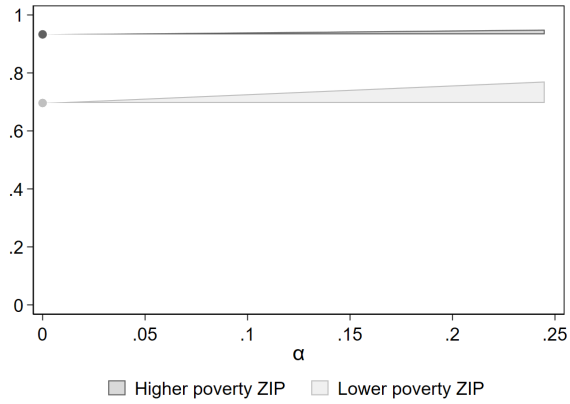
(a) *Non-contact rate by poverty status*



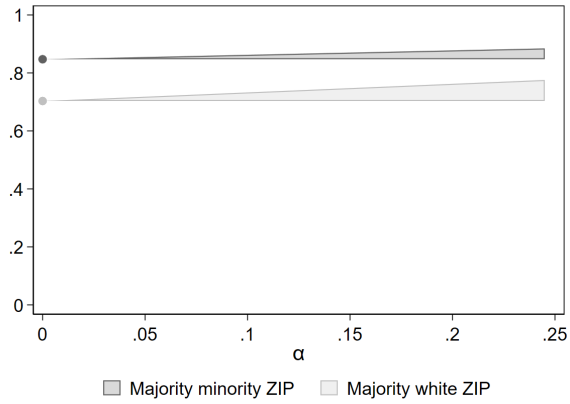
(b) *Non-contact rate by racial composition*



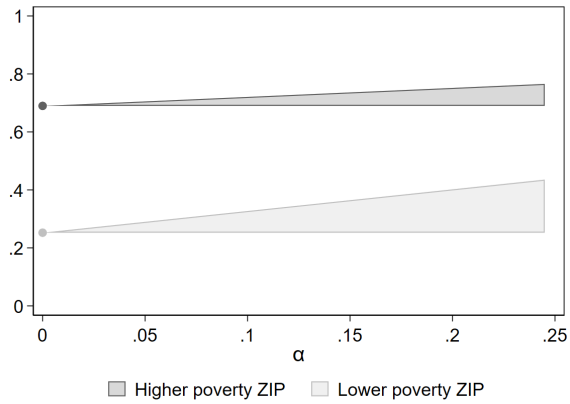
(c) *Hesitancy rate at $0 by poverty status*



(d) *Hesitancy rate at $0 by racial composition*



(e) *Hesitancy rate at $100 by poverty status*



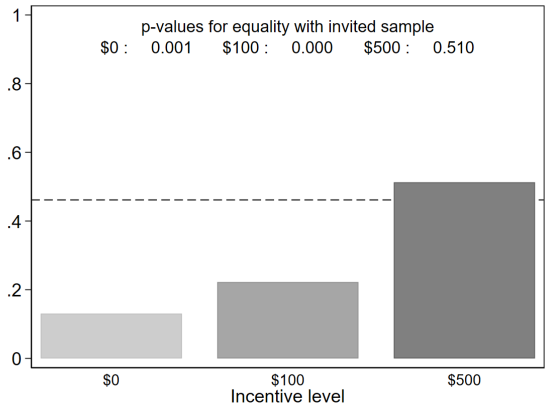
(f) *Hesitancy rate at $100 by racial composition*

Notes: These figures report estimates of the bounds in (5) for different levels of α broken down by demographic group.

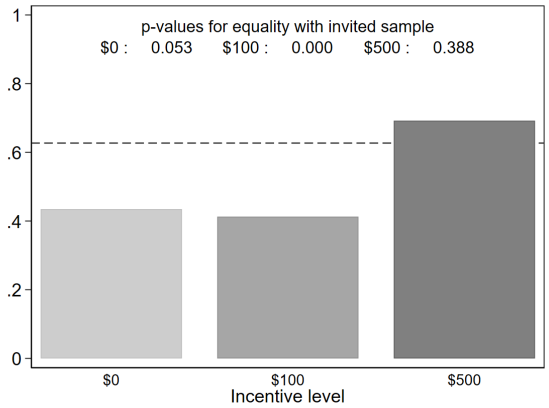
hospitalization rate related to alcohol use. Overall, these findings suggest that similar concerns will arise in studies or surveys designed to measure outcomes that are highly correlated with these measures.

Figure 4: *Effect of hesitancy on representativeness: main characteristics*

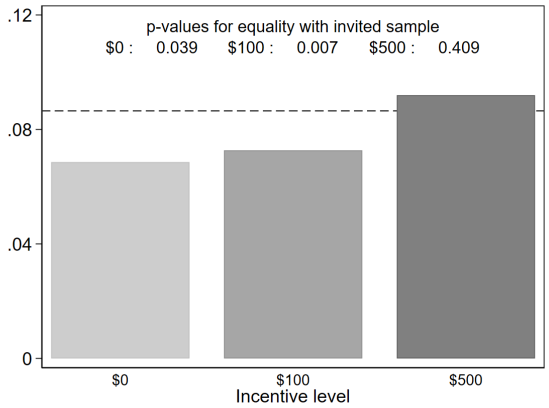
(a) *Indicator for higher poverty neighborhood*



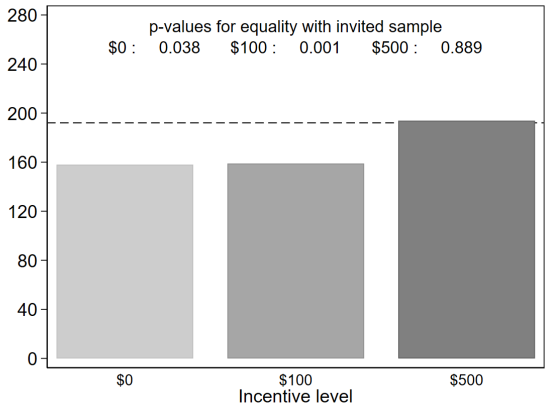
(b) *Indicator for minority neighborhood*



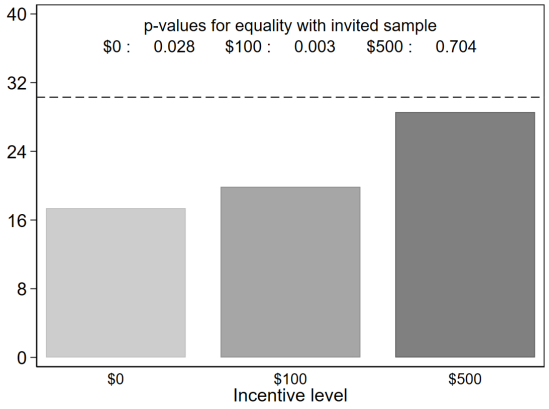
(c) *Uninsured rate*



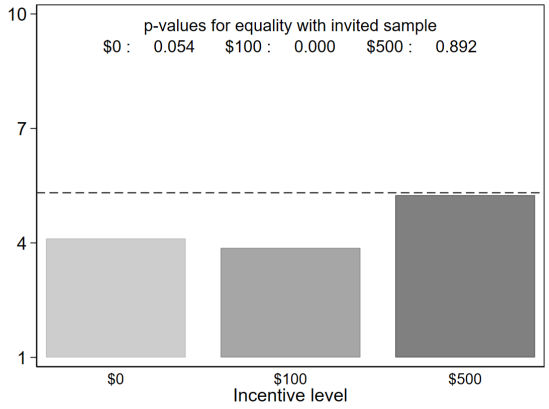
(d) *Preventable hospitalization rate (per 10,000)*



(e) *Drug-related hospitalization rate (per 10,000)*



(f) *COVID-19 local risk index (from 1-10)*



Notes: These figures show participant average neighborhood-level measures stratified by incentive group. The horizontal dashed line depicts the average among the invited sample. Variable definitions are given in Appendix B.

7 Conclusion

Scientific studies with human subjects often report lower participation rates among Black, Hispanic, and low-income households. Lack of representation can cause public resources to be

misdirected, and as such poses a risk to public health and to the proper allocation of resources in other policy domains. Yet, little is known about the causes of under-representation, or about solutions to this problem.

In this paper, we showed how randomized financial incentives can be used to measure the relative importance of non-contact and hesitancy for non-participation, and to quantify how they contribute to underrepresentation of minority low-income groups. In the context of the RECOVER serological study, we found that non-contact is a cause for low participation, but not of underrepresentation. Hesitancy among contacted households is a cause of both low participation and underrepresentation. Examining how hesitancy varies with incentives, we find that reservation payments for contacted households in poor and minority neighborhoods are substantially more likely to exceed $100. High incentives ($500) for participation appear to restore representativeness on dimensions that are important and plausible confounds in population health studies, including the uninsured rate, hospitalization rates, and an aggregate COVID-19 risk index.

References

Alsan, M., M. Durvasula, H. Gupta, J. Schwartzstein, and H. L. Williams (2022). Representation and Extrapolation:. *NBER working paper*. 3, 5, 6, 13

Aspelund, K., D. M.C., J. Stock, and C. Walker (2020). Identification and estimation of undetected covid-19 cases using testing data from iceland. *working paper*. 5

Bobrovitz, N., R. K. Arora, C. Cao, E. Boucher, M. Liu, C. Donnici, M. Yanes-Lane, M. Whelan, S. Perlman-Arrow, J. Chen, H. Rahim, N. Ilincic, M. Segal, N. Duarte, J. V. Wyk, T. Yan, A. Atmaja, S. Rocco, A. Joseph, L. Penny, D. A. Clifton, T. Williamson, C. P. Yansouni, T. G. Evans, J. Chevrier, J. Papenburg, and M. P. Cheng (2021, June). Global seroprevalence of SARS-CoV-2 antibodies: A systematic review and meta-analysis. *PLOS ONE 16*(6), e0252617. 3, 5, 35

Bradley, V. C., S. Kuriwaki, M. Isakov, D. Sejdinovic, X.-L. Meng, and S. Flaxman (2021, December). Unrepresentative big surveys significantly overestimated US vaccine uptake. *Nature 600*(7890), 695–700. 5

Brick, J. M. and D. Williams (2013). Explaining rising nonresponse rates in cross-sectional surveys. *The ANNALS of the American academy of political and social science 645*(1), 36–59. 6

Chen, X., Z. Chen, A. S. Azman, X. Deng, R. Sun, Z. Zhao, N. Zheng, X. Chen, W. Lu, T. Zhuang, J. Yang, C. Viboud, M. Ajelli, D. T. Leung, and H. Yu (2021, May). Serological evidence of human infection with SARS-CoV-2: A systematic review and meta-analysis. *The Lancet. Global Health 9*(5), e598–e609. 5

Currie, J. (2006). The take up of social benefits. In A. Auerbach, D. Card, and J. Quigley (Eds.), *Public Policy and the Income Distribution*, Chapter 3. New York: Russell Sage Foundation. 6

Dahl, G. B., K. V. Loken, and M. Mogstad (2014). Peer effects in program participation. *The American Economic Review 104*(7), 2049–2074. 6

Deshpande, M. and Y. Li (2018). Who is screened out? application costs and the targeting of disability programs. *AEJ: Policy 11*(4), 213–248. 6

Dutz, D., I. Huitfeldt, S. Lacouture, M. Mogstad, A. Torgovitsky, and W. van Dijk (2021). Selection in surveys. Technical report, National Bureau of Economic Research. 4

Finkelstein, A. and M. J. Notowidigdo (2019). Take-up and targeting: Experimental evidence from snap. *The Quarterly Journal of Economics 134*(3), 1505–1556. 6

Groves, R. M., F. J. Fowler Jr., M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau (2009). *Survey Methodology*. Wiley New York. 6

Imbens, G. W. and J. D. Angrist (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica 62*(2), pp. 467–475. 10

Ioannidis, J. P., S. Cripps, and M. A. Tanner (2022). Forecasting for COVID-19 has failed. *International Journal of Forecasting 38*(2), 423–438. 5

Manski, C. and F. Molinari (2021). Estimating the covid-19 infection rate: Anatomy of an inference problem. *Journal of Econometrics 220*(S2), 181–192. 5

Moffitt, R. (1983). An economic model of welfare stigma. *The American Economic Review 73*(5), 1023–1035. 6

NASEM (2022). *Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups*. Washington, DC: The National Academies Press. 2, 3, 5, 8, 13

Petrolia, D. R. and S. Bhattacharjee (2009). Revisiting incentive effects: evidence from a random-sample mail survey on consumer preferences for fuel ethanol. *Public Opinion Quarterly 73*(3), 537–550. 6

Singer, E. and C. Ye (2013). The use and effects of incentives in surveys. *The ANNALS of the American Academy of Political and Social Science 645*(1), 112–141. 6

Torgovitsky, A. (2019, January). Partial identification by extending subdistributions. *Quantitative Economics 10*(1), 105–144. 37

Vytlacil, E. (2002). Independence, Monotonicity, and Latent Index Models: An Equivalence Result. *Econometrica 70*(1), 331–341. 10

Wines, M. and M. Cramer (2022, March). 2020 Census Undercounted Hispanic, Black and Native American Residents. *The New York Times*. 2, 5

Appendix (for online publication)

A Study implementation

This Appendix describes the design and implementation of the RECOVER serological study. The study was designed and implemented in collaboration with NORC at the University of Chicago, and the University of Chicago Wilson Antibody Biology Laboratory. Appendix A.1 discusses the construction of the sampling frame and the sampling and randomization procedures. Appendix A.2 describes outreach and follow-up procedures, and additionally discusses the materials sent to invited households. These materials are reproduced in Appendix A.3. This study, its design, and its implementation were approved by the IRB at the University of Chicago (IRB20-0721).

A.1 Sampling and randomization procedures

NORC constructed a sampling frame of approximately 1.2 million household addresses in the city of Chicago based on address data from the United States Postal Service Computerized Delivery Sequence File (CDSF).[11] The CDSF contains a record for every mail delivery point in the U.S. and these records are updated monthly.

NORC then randomly sampled 882 household addresses from the sampling frame for the RECOVER study. All addresses had an equal probability of being randomly sampled. These 882 household addresses were randomly (and with equal probability) assigned to one of three compensation arms: 374 addresses were assigned to the $0 arm, 374 addresses were assigned to the $100 arm, and 134 addresses were assigned to the $500 arm.

A.2 Outreach and follow up procedures

Each household in the RECOVER study sample was sent a package that contained a self-administered blood collection kit, an invitation, and a consent form with a short questionnaire. All households received material that was identical in all aspects except for minor modifications relating to compensation for participating (i.e. returning a blood sample) depending on the assigned incentive arm. In particular, households in the $0 arm were not told about financial compensation for participating, and households in the $100 and $500 compensation arms were notified that they would receive $100 and $500 for participating, respectively.

The blood collection kit included instructions on self-administering and returning a blood sample.[12] The written material explained the purpose of the study, provided information on financial compensation for participating (if applicable), and explained which member of the household should participate and how to participate, and provided contact information.

[11] The software program used to create the sampling frame is retained by NORC.

[12] On the instruction card, a link to video instructions for taking the sample was provided (https://vimeo.com/286513641), and invitees were reminded that they could call the toll-free line to have a phone interviewer from NORC walk them through the sample-taking process.

Invitees were additionally provided a toll-free phone number to call with any questions about the study, procedures, their participation, or rights as a research participant. Appendix Exhibits A.1, A.2, and A.3 respectively depict the invitations sent to households in the $0, $100 compensation, and $500 compensation arms.

The consent form noted that the purpose of the study was to learn how many people had already been exposed to the virus, that the study had received IRB approval, that participants' data would be securely stored, that they would not receive the result of the test, and that compensation (if offered) would be received when the Wilson laboratory received the blood sample. The consent form concluded with a request for the participant's signature and a short questionnaire. The consent form differed slightly depending on the assigned incentive arm. The first two pages of the consent form differed slightly depending on the assigned incentive arm. Appendix Exhibits A.4, A.5, and A.6 respectively depict the first page of the consent forms sent to households in the $0, $100 compensation, and $500 compensation arms. Appendix Exhibits A.7, A.8, and A.9 respectively depict the second page of the consent forms sent to households in the $0, $100 compensation, and $500 compensation arms. The third and fourth pages of the consent form were uniform across compensation arms and are respecitvely depicted in Appendix Exhibits A.10 and A.11.

After all packages were sent, NORC additionally sent up to three reminder postcards to all sampled households who had not yet returned a kit. Appendix Exhibits A.12, A.13, and A.14 respectively depict the postcards sent to households in the no compensation, $100 compensation, and $500 compensation arms. NORC also conducted up to three weekly phone calls to these households.[13] If a non-usable sample was received by the laboratory, NORC contacted households to inform them that their sample was not usable. Households were offered the option of receiving a replacement kit to attempt to take their sample again.[14] If households refused to receive a replacement kit, the interviewer would explain that they would not receive payment. If households agreed to receive a replacement kit, a new kit was mailed to the participant. There was no additional payment or penalty for having to retake one's sample. As soon as the second sample was received by the laboratory, the participant was sent their payment, even if this second sample was also unusable.

[13]Since the sampling frame itself does not contain any telephone number information, phone numbers were appended to the sample once it was selected using data from commercial providers.

[14]Reasons for a non-usable sample include an empty or not-attempted kit or a kit with insufficient blood sample. The research team was informed by the kit manufacturer that their product returns a total of 1.8% unusable samples. When empty or not-attempted kits were sent back, NORC also attempted to assist the participant in understanding the required conditions of participation and compensation via phone call.

A.3 Materials

Exhibit A.1: *Invitation sent to households, incentive level $0*

THE UNIVERSITY OF CHICAGO

HELP US LEARN ABOUT COVID-19 IN CHICAGO!

Dear Fellow Chicagoan,

We launched the RECOVER study to learn about how many Chicagoans have had COVID-19. You can help your community by helping us understand how the virus has affected different parts of the city.

It is fast, easy, and safe to take part in this study.

You are among a small number of people scientifically chosen to represent others in your area.

Your participation is important!

HERE'S HOW IT WORKS:

1 **Collect a small blood sample at home in just a few minutes using the kit you received.**

2 **Mail the kit to our research team using the pre-paid packaging.**

We know your time is valuable. We thank you for your participation!

- **Who can participate?** To make sure the study represents all Chicagoans, one adult per household – the one with the next upcoming birthday – can take part.
- **Is there a deadline?** Please return your kit within 10 business days of receiving this invitation. Samples must be mailed back the same day that they are taken.
- **Where can I learn more?** For more information, visit the study website at recover.uchicago.edu or call us toll-free at 800-483-2565 (Se Habla Español).
  Your unique access code is:

Your participation is a way to help our community, neighbors, and friends overcome this pandemic together. We hope you will join us and participate in this important study!

Michael Greenstone, PhD, Lead Principal Investigator
On behalf of the RECOVER Study team, a collaboration between the University of Chicago Social Sciences Division (SSD), Biological Sciences Division (BSD), and Department of Medicine

Exhibit A.2: *Invitation sent to households, incentive level $100*

THE UNIVERSITY OF CHICAGO

HELP US LEARN ABOUT COVID-19 IN CHICAGO!

Dear Fellow Chicagoan,

We launched the RECOVER study to learn about how many Chicagoans have had COVID-19. You can help your community by helping us understand how the virus has affected different parts of the city.

It is fast, easy, and safe to take part in this study.

You are among a small number of people scientifically chosen to represent others in your area.

Your participation is important!

We know your time is valuable.

As thanks for your participation, we will send you a check for

$100

after we receive your kit.

HERE'S HOW IT WORKS:

1 **Collect a small blood sample at home in just a few minutes using the kit you received.**

2 **Mail the kit to our research team using the pre-paid packaging.**

- **Who can participate?** To make sure the study represents all Chicagoans, one adult per household – the one with the next upcoming birthday – can take part.
- **Is there a deadline?** Please return your kit within 10 business days of receiving this invitation. Samples must be mailed back the same day that they are taken.
- **Where can I learn more?** For more information, visit the study website at recover.uchicago.edu or call us toll-free at 800-483-2565 (Se Habla Español).
  Your unique access code is:

Your participation is a way to help our community, neighbors, and friends overcome this pandemic together. We hope you will join us and participate in this important study!

Michael Greenstone, PhD, Lead Principal Investigator
On behalf of the RECOVER Study team, a collaboration between the University of Chicago Social Sciences Division (SSD), Biological Sciences Division (BSD), and Department of Medicine

Exhibit A.3: *Invitation sent to households, incentive level $500*

THE UNIVERSITY OF CHICAGO

HELP US LEARN ABOUT COVID-19 IN CHICAGO!

Dear Fellow Chicagoan,

We launched the RECOVER study to learn about how many Chicagoans have had COVID-19. You can help your community by helping us understand how the virus has affected different parts of the city.

It is fast, easy, and safe to take part in this study.

You are among a small number of people scientifically chosen to represent others in your area.

Your participation is important!

We know your time is valuable.

As thanks for your participation, we will send you a check for

$500

after we receive your kit.

HERE'S HOW IT WORKS:

1 **Collect a small blood sample at home in just a few minutes using the kit you received.**

2 **Mail the kit to our research team using the pre-paid packaging.**

- **Who can participate?** To make sure the study represents all Chicagoans, one adult per household – the one with the next upcoming birthday – can take part.
- **Is there a deadline?** Please return your kit within 10 business days of receiving this invitation. Samples must be mailed back the same day that they are taken.
- **Where can I learn more?** For more information, visit the study website at recover.uchicago.edu or call us toll-free at 800-483-2565 (Se Habla Español).
  Your unique access code is:

Your participation is a way to help our community, neighbors, and friends overcome this pandemic together. We hope you will join us and participate in this important study!

Michael Greenstone, PhD, Lead Principal Investigator
On behalf of the RECOVER Study team, a collaboration between the University of Chicago Social Sciences Division (SSD), Biological Sciences Division (BSD), and Department of Medicine

Exhibit A.4: *Consent form: key information, incentive level $0*

THE UNIVERSITY OF CHICAGO
RECOVER Study

Investigators: Michael Greenstone, Magne Mogstad, Azeem Shaikh,
Alex Torgovitsky, Ali Hortaçsu, Sarah Cobey, Patrick Wilson
University of Chicago, 1126 E. 59th Street, Chicago, IL 60637
Phone Number: (773) 702-0759 | Protocol Number: IRB20-0721

Consent Form for Participation in a Research Study

KEY INFORMATION

You are being invited to participate in a research study about COVID-19 in Chicago. The purpose of this section is to give you key information to help you decide whether to participate.

WHAT IS THE STUDY ABOUT?
The purpose of this study is to learn how many people have already been exposed to the virus that causes COVID-19, known as SARS-CoV-2, including those that had mild or no symptoms. We will conduct antibody tests in blood samples collected from a random sample of Chicagoans.

WHAT DOES PARTICIPATION INVOLVE?
Participation involves taking a small blood sample, using the kit we sent you. This is done from your home and completed in just a few minutes. The kit is safe and easy to use. You will mail the sample back to our laboratory, following the instructions provided in the kit box.

WHO CAN PARTICIPATE?
One adult (18+) from your household can participate. It is important that our sample is random. We ask that the adult with the next upcoming birthday participates. If that person is not available or does not want to participate, we ask that no other person in the household participates.

WHAT ARE KEY REASONS I MIGHT CHOOSE TO PARTICIPATE?
We hope that you will participate in this study to help us understand how the disease has spread in Chicago. We know your time is valuable, and we appreciate your participation.

WHAT ARE KEY REASONS I MIGHT CHOOSE NOT TO PARTICIPATE?
This study involves collecting a small blood sample by self-administering a finger prick. You may feel momentary, mild discomfort which will fade within a minute or two. You may decide not to participate if you are not comfortable self-administering the finger prick.

ARE THERE ANY COSTS FOR PARTICIPATING?
No. You will not incur any costs for participation.

WHAT ABOUT CONFIDENTIALITY?
Study data will be handled confidentially. While there are always risks of data breach in any research study, our data is stored on secure servers that minimize this risk.

DO I HAVE TO TAKE PART IN THE STUDY?
Participation is completely voluntary. You will not lose any services, benefits, or rights that you normally have if you choose not to participate, or if you choose to leave the study at any time.

QUESTIONS?
Call to speak to a member of our team: 800-483-2565. For questions about your rights as a research subject, call the Biological Sciences Division (BSD) Institutional Review Board (IRB): 773-702-6505.

If you would like to participate, please return the final page of this consent form with your sample. We will not be able to process your sample without a signed consent form.

Exhibit A.5: *Consent form: key information, incentive level $100*

THE UNIVERSITY OF CHICAGO
RECOVER Study

Investigators: Michael Greenstone, Magne Mogstad, Azeem Shaikh, Alex Torgovitsky, Ali Hortaçsu, Sarah Cobey, Patrick Wilson
University of Chicago, 1126 E. 59th Street, Chicago, IL 60637
Phone Number: (773) 702-0759 | Protocol Number: IRB20-0721

Consent Form for Participation in a Research Study

KEY INFORMATION

You are being invited to participate in a research study about COVID-19 in Chicago. The purpose of this section is to give you key information to help you decide whether to participate.

WHAT IS THE STUDY ABOUT?
The purpose of this study is to learn how many people have already been exposed to the virus that causes COVID-19, known as SARS-CoV-2, including those that had mild or no symptoms. We will conduct antibody tests in blood samples collected from a random sample of Chicagoans.

WHAT DOES PARTICIPATION INVOLVE?
Participation involves taking a small blood sample, using the kit we sent you. This is done from your home and completed in just a few minutes. The kit is safe and easy to use. You will mail the sample back to our laboratory, following the instructions provided in the kit box.

WHO CAN PARTICIPATE?
One adult (18+) from your household can participate. It is important that our sample is random. We ask that the adult with the next upcoming birthday participates. If that person is not available or does not want to participate, we ask that no other person in the household participates.

WHAT ARE KEY REASONS I MIGHT CHOOSE TO PARTICIPATE?
We hope that you will participate in this study to help us understand how the disease has spread in Chicago. We know your time is valuable. To thank you for your participation, we will send you a check for **$100** when our laboratory receives your blood sample.

WHAT ARE KEY REASONS I MIGHT CHOOSE NOT TO PARTICIPATE?
This study involves collecting a small blood sample by self-administering a finger prick. You may feel momentary, mild discomfort which will fade within a minute or two. You may decide not to participate if you are not comfortable self-administering the finger prick.

ARE THERE ANY COSTS FOR PARTICIPATING?
No. You will not incur any costs for participation.

WHAT ABOUT CONFIDENTIALITY?
Study data will be handled confidentially. While there are always risks of data breach in any research study, our data is stored on secure servers that minimize this risk.

DO I HAVE TO TAKE PART IN THE STUDY?
Participation is completely voluntary. You will not lose any services, benefits, or rights that you normally have if you choose not to participate, or if you choose to leave the study at any time.

QUESTIONS?
Call to speak to a member of our team: 800-483-2565. For questions about your rights as a research subject, call the Biological Sciences Division (BSD) Institutional Review Board (IRB): 773-702-6505.

If you would like to participate, please return the final page of this consent form with your sample. We will not be able to process your sample or compensate you without a signed consent form.

Exhibit A.6: *Consent form: key information, incentive level $500*

THE UNIVERSITY OF CHICAGO
RECOVER Study

Investigators: Michael Greenstone, Magne Mogstad, Azeem Shaikh, Alex Torgovitsky, Ali Hortaçsu, Sarah Cobey, Patrick Wilson
University of Chicago, 1126 E. 59th Street, Chicago, IL 60637
Phone Number: (773) 702-0759 | Protocol Number: IRB20-0721

Consent Form for Participation in a Research Study

KEY INFORMATION

You are being invited to participate in a research study about COVID-19 in Chicago. The purpose of this section is to give you key information to help you decide whether to participate.

WHAT IS THE STUDY ABOUT?
The purpose of this study is to learn how many people have already been exposed to the virus that causes COVID-19, known as SARS-CoV-2, including those that had mild or no symptoms. We will conduct antibody tests in blood samples collected from a random sample of Chicagoans.

WHAT DOES PARTICIPATION INVOLVE?
Participation involves taking a small blood sample, using the kit we sent you. This is done from your home and completed in just a few minutes. The kit is safe and easy to use. You will mail the sample back to our laboratory, following the instructions provided in the kit box.

WHO CAN PARTICIPATE?
One adult (18+) from your household can participate. It is important that our sample is random. We ask that the adult with the next upcoming birthday participates. If that person is not available or does not want to participate, we ask that no other person in the household participates.

WHAT ARE KEY REASONS I MIGHT CHOOSE TO PARTICIPATE?
We hope that you will participate in this study to help us understand how the disease has spread in Chicago. We know your time is valuable. To thank you for your participation, we will send you a check for **$500** when our laboratory receives your blood sample.

WHAT ARE KEY REASONS I MIGHT CHOOSE NOT TO PARTICIPATE?
This study involves collecting a small blood sample by self-administering a finger prick. You may feel momentary, mild discomfort which will fade within a minute or two. You may decide not to participate if you are not comfortable self-administering the finger prick.

ARE THERE ANY COSTS FOR PARTICIPATING?
No. You will not incur any costs for participation.

WHAT ABOUT CONFIDENTIALITY?
Study data will be handled confidentially. While there are always risks of data breach in any research study, our data is stored on secure servers that minimize this risk.

DO I HAVE TO TAKE PART IN THE STUDY?
Participation is completely voluntary. You will not lose any services, benefits, or rights that you normally have if you choose not to participate, or if you choose to leave the study at any time.

QUESTIONS?
Call to speak to a member of our team: 800-483-2565. For questions about your rights as a research subject, call the Biological Sciences Division (BSD) Institutional Review Board (IRB): 773-702-6505.

If you would like to participate, please return the final page of this consent form with your sample. We will not be able to process your sample or compensate you without a signed consent form.

Exhibit A.7: *Consent form: detailed information, page 1, incentive level $0*

DETAILED INFORMATION

INTRODUCTION
We are a team of researchers from the University of Chicago, conducting a study to learn how many people in Chicago have been exposed to the virus that causes COVID-19, known as SARS-CoV-2, including those that had mild or no symptoms. We will conduct antibody tests in blood samples collected from a random sample of Chicagoans. You are among approximately 3000 households invited to participate.

ELIGIBILITY
If you received an invitation box in the mail, it means that your household has been randomly selected. One adult (age 18+) from your household can participate. It is important that our sample is random. For this reason, we ask that the adult with the next upcoming birthday participates. If that person is not available or does not want to participate, we ask that no other person in the household participates. Past history of COVID-19 testing, diagnosis, or exposure does not matter. As long as the person is 18+, resides in Chicago, and has the next birthday, they are eligible.

WHAT IS INVOLVED IN THE STUDY?
Participants will self-administer a finger prick to take a small blood sample, following the instructions provided in the box. The kit is safe and easy to use and takes about 5-10 minutes to complete. Participants will mail their sample to our laboratory where it will be tested for antibodies against the virus that causes COVID-19, which may indicate prior exposure to the virus. This test does not indicate active infection. This study does not involve genetic testing. You will not be notified of your test results. We may contact you about your participation in the study, or to invite you to participate in later phases of the research.

WHAT ARE THE BENEFITS?
We hope that you will participate in this study to help us understand how the disease has spread in Chicago. We know your time is valuable, and we thank you for your participation.

WHAT ARE THE RISKS?
The discomfort of self-administering the finger prick is minimal and will fade within a minute or two. The kit is safe and commercially available, used routinely to collect at-home blood samples. As with any finger prick, there is a small risk of bruising, tenderness, and a rare risk of infection. While there are always risks of data breach in any research study, we take all possible precautions to minimize this risk including storing data on secure servers only accessible to authorized personnel.

ARE THERE ANY COSTS FOR PARTICIPATING?
No. You will not incur any costs for participation. You will use pre-paid packaging in the box to return your sample to our laboratory. There will be no costs to you or your insurance company for the costs of tests or services that are being performed solely for the purposes of this study. You or your insurance company remain responsible for costs related to your usual medical care.

WILL I BE PAID FOR MY PARTICIPATION?
No, you will not be compensated for your participation.

HOW WILL YOU MAINTAIN CONFIDENTIALITY?
Data you provide will be handled confidentially. While there are always risks of data breach in any research study, our data is stored on secure servers that minimize this risk and we will strip the data of identifiable information, such as your name. We will retain a link (or "key") to the identifiable information in password-protected, encrypted computer files accessible only to

Exhibit A.8: *Consent form: detailed information, page 1, incentive level $100*

DETAILED INFORMATION

INTRODUCTION
We are a team of researchers from the University of Chicago, conducting a study to learn how many people in Chicago have been exposed to the virus that causes COVID-19, known as SARS-CoV-2, including those that had mild or no symptoms. We will conduct antibody tests in blood samples collected from a random sample of Chicagoans. You are among approximately 3000 households invited to participate.

ELIGIBILITY
If you received an invitation box in the mail, it means that your household has been randomly selected. One adult (age 18+) from your household can participate. It is important that our sample is random. For this reason, we ask that the adult with the next upcoming birthday participates. If that person is not available or does not want to participate, we ask that no other person in the household participates. Past history of COVID-19 testing, diagnosis, or exposure does not matter. As long as the person is 18+, resides in Chicago, and has the next birthday, they are eligible.

WHAT IS INVOLVED IN THE STUDY?
Participants will self-administer a finger prick to take a small blood sample, following the instructions provided in the box. The kit is safe and easy to use and takes about 5-10 minutes to complete. Participants will mail their sample to our laboratory where it will be tested for antibodies against the virus that causes COVID-19, which may indicate prior exposure to the virus. This test does not indicate active infection. This study does not involve genetic testing. You will not be notified of your test results. We may contact you about your participation in the study, or to invite you to participate in later phases of the research.

WHAT ARE THE BENEFITS?
We hope that you will participate in this study to help us understand how the disease has spread in Chicago. We know your time is valuable. To thank you for your participation, we will send you a check for $100 when our laboratory receives your blood sample.

WHAT ARE THE RISKS?
The discomfort of self-administering the finger prick is minimal and will fade within a minute or two. The kit is safe and commercially available, used routinely to collect at-home blood samples. As with any finger prick, there is a small risk of bruising, tenderness, and a rare risk of infection. While there are always risks of data breach in any research study, we take all possible precautions to minimize this risk including storing data on secure servers only accessible to authorized personnel.

ARE THERE ANY COSTS FOR PARTICIPATING?
No. You will not incur any costs for participation. You will use pre-paid packaging in the box to return your sample to our laboratory. There will be no costs to you or your insurance company for the costs of tests or services that are being performed solely for the purposes of this study. You or your insurance company remain responsible for costs related to your usual medical care.

WILL I BE PAID FOR MY PARTICIPATION?
Yes. We know your time is valuable and to thank you for your participation, we will send you a check for $100 when we receive your blood sample at our laboratory. For each participant, the compensation amount was determined by a lottery.

Exhibit A.9: *Consent form: detailed information, page 1, incentive level $500*

DETAILED INFORMATION

INTRODUCTION
We are a team of researchers from the University of Chicago, conducting a study to learn how many people in Chicago have been exposed to the virus that causes COVID-19, known as SARS-CoV-2, including those that had mild or no symptoms. We will conduct antibody tests in blood samples collected from a random sample of Chicagoans. You are among approximately 3000 households invited to participate.

ELIGIBILITY
If you received an invitation box in the mail, it means that your household has been randomly selected. One adult (age 18+) from your household can participate. It is important that our sample is random. For this reason, we ask that the adult with the next upcoming birthday participates. If that person is not available or does not want to participate, we ask that no other person in the household participates. Past history of COVID-19 testing, diagnosis, or exposure does not matter. As long as the person is 18+, resides in Chicago, and has the next birthday, they are eligible.

WHAT IS INVOLVED IN THE STUDY?
Participants will self-administer a finger prick to take a small blood sample, following the instructions provided in the box. The kit is safe and easy to use and takes about 5-10 minutes to complete. Participants will mail their sample to our laboratory where it will be tested for antibodies against the virus that causes COVID-19, which may indicate prior exposure to the virus. This test does not indicate active infection. This study does not involve genetic testing. You will not be notified of your test results. We may contact you about your participation in the study, or to invite you to participate in later phases of the research.

WHAT ARE THE BENEFITS?
We hope that you will participate in this study to help us understand how the disease has spread in Chicago. We know your time is valuable. To thank you for your participation, we will send you a check for $500 when our laboratory receives your blood sample.

WHAT ARE THE RISKS?
The discomfort of self-administering the finger prick is minimal and will fade within a minute or two. The kit is safe and commercially available, used routinely to collect at-home blood samples. As with any finger prick, there is a small risk of bruising, tenderness, and a rare risk of infection. While there are always risks of data breach in any research study, we take all possible precautions to minimize this risk including storing data on secure servers only accessible to authorized personnel.

ARE THERE ANY COSTS FOR PARTICIPATING?
No. You will not incur any costs for participation. You will use pre-paid packaging in the box to return your sample to our laboratory. There will be no costs to you or your insurance company for the costs of tests or services that are being performed solely for the purposes of this study. You or your insurance company remain responsible for costs related to your usual medical care.

WILL I BE PAID FOR MY PARTICIPATION?
Yes. We know your time is valuable and to thank you for your participation, we will send you a check for $500 when we receive your blood sample at our laboratory. For each participant, the compensation amount was determined by a lottery.

Exhibit A.10: *Consent form: detailed information, page 2*

HOW WILL YOU MAINTAIN CONFIDENTIALITY?
Data you provide will be handled confidentially. While there are always risks of data breach in any research study, our data is stored on secure servers that minimize this risk and we will strip the data of identifiable information, such as your name. We will retain a link (or "key") to the identifiable information in password-protected, encrypted computer files accessible only to authorized members of the research team, all of whom receive training in maintaining strict standards of confidentiality.

The study data and key will be retained for at least 10 years after study funding ends. Your name and any identifying information will not be used in any publications or presentations. The laboratory may retain a portion of your blood sample for future testing. Data and specimens that are stripped of identifying information may be used for future research by our team or other investigators without additional informed consent.

Authorized representatives from the University of Chicago and NORC at the University of Chicago may review your research data for purposes such as monitoring or managing the conduct of this study. Your records may be reviewed by federal agencies whose responsibility is to protect human subjects in research including the Office of Human Research Protections (OHRP). Your records may be viewed by representatives of the University including the Institutional Review Board (a committee that oversees the research) and the Office of Clinical Research.

WHAT ARE MY RIGHTS AS A PARTICIPANT?
Participation is completely voluntary. You will not lose any services, benefits, or rights you normally have if you choose not to participate, or if you choose to leave the study at any time. If you decide not to sign this consent form, you will not be able to participate in the study. You will be given a copy of this document after you have signed. Your consent does not have an expiration date.

If the need arises, we will tell you about significant new information that may affect your willingness to stay in this study. If you would like to speak to a member of our study team, you may contact: **800-483-2565**. For questions about your rights as a research subject, you may contact the University of Chicago Biological Sciences Division (BSD) Institutional Review Board (IRB) at 773-702-6505.

PARA ESPAÑOL: Si desea ver este formulario de consentimiento en español, puede visitar recover.uchicago.edu e iniciar sesión con el código de acceso único de la carta de invitación, o llamar a este número para hablar con un miembro de nuestro equipo: 800-483-2565

Exhibit A.11: *Consent form and questionnaire*

RETURN THIS PAGE WITH THE COMPLETED BLOOD SAMPLE KIT.

CONSENT: I have read and understood the information in the consent form, including the objectives of the study and procedures. I understand that my participation is voluntary, and that I do not have to sign this form if I do not want to be part of this research study.

By signing below, I confirm my eligibility and I agree to participate in this study.

* REQUIRED FIELD

First name: * _____

Middle name: _____

Last name: * _____

Signature: * _____

Date: * _____ **Time:** * _____

Email: * _____ **Phone #:** * _____

We will only use your email & phone number to contact you about your participation in this study.

Please provide the following information that will help us better analyze COVID-19 infection rates in Chicago:

1. **What is your age?** * _____ years
2. **What is your gender?** *
   - ☐ Male
   - ☐ Female
   - ☐ Other
   - ☐ Prefer not to answer
3. **Which one or more of the following best describes your race? Please check all that apply.** *
   - ☐ White
   - ☐ Black or African American
   - ☐ American Indian or Alaska Native
   - ☐ Asian
   - ☐ Native Hawaiian or Pacific Islander
   - ☐ Other
   - ☐ Prefer not to answer
4. **Are you of Hispanic, Latino, or Spanish origin?** *
   - ☐ Yes
   - ☐ No
   - ☐ Prefer not to answer
5. **What was your household's approximate total income from all sources in 2019?** *
   - ☐ Less than $20,000
   - ☐ $20,000 to less than $50,000
   - ☐ $50,000 to less than $100,000
   - ☐ $100,000 or more
   - ☐ Prefer not to answer

Would you like to be contacted in the future about other COVID-19 research opportunities?

If yes, you may be contacted by other researchers and provided with more information and a separate consent form. Your answer to this question will not affect your participation in this study.

- ☐ Yes, I would like to be contacted about other COVID-19 research studies
- ☐ No, please do not contact me about other COVID-19 research studies

LAB USE ONLY **KIT CODE:** ___ ___ ___ - ___ ___ ___ - ___ ___ ___ **TEST CODE:** ___ ___ ___

Exhibit A.12: *Reminder postcard, incentive level $0*

Exhibit A.13: *Reminder postcard, incentive level $100*

Exhibit A.14: *Reminder postcard, incentive level $500*

B Data sources and variable definitions

Our empirical analysis uses data collected from the serology study described in Section 2. We link the study data to a set of (five digit) zipcode-level characteristics we collect from three sources: the American Community Survey, the Chicago Health Atlas, and the City Health Dashboard. We provide information on each source and the obtained variables below.

American Community Survey. We obtain neighborhood demographics from the 2019 American Community Survey 1-year estimates. We classify a zipcode as: (1) higher poverty if the percentage of households below 1.5 times the poverty line is above 30% (and lower poverty otherwise); and (2) majority minority if the share of adults identifying as non-Hispanic white is below 50% (and majority White otherwise).

Chicago Health Atlas. We obtain zipcode-level health measures from the Chicago Health Atlas, a portal developed by the Chicago Department of Public Health and Population Health Analytics Metrics Evaluation Center at University of Illinois Chicago. More specifically, we obtain the uninsurance rate and five diagnose-specific hospitalization rates. The uninsurance rate is defined as the average percentage of residents without health insurance between 2016 and 2020. Diagnose-specific hospitalization rates are defined as the age-adjusted number of hospitalizations discharges for a given diagnose per 10,000 people in 2017, excluding discharges to Veterans Administration hospitals. We obtain these hospitalization rates for the following diagnoses: (1) alcohol-related; (2) drug-related (which include amphetamines, cannabis, cocaine, drug induced mental disorders, hallucinogens, opioids, sedatives, hypnotics, anxiolytics, tranquilizers, barbiturates, and other drugs); (3) mood and depressive disorders (which include bipolar and manic depressive disorders); (4) behavioral health (which include substance use disorder and mental disorders); and (5) preventable (defined as conditions that could be managed in a clinic setting).

City Health Dashboard. We obtain additional health and labor market measures from the City Health Dashboard, a portal developed by NYU Langone Health. The dashboard provides data at the census tract level, which we aggregate to the ZIP code level via population-weighted averages using Census relationship files. The following measures are obtained from this source: (1) annual unemployment rate, defined as the percentage of individuals at least 16 years that were unemployed and seeking work at any point in 2020; (2) credit insecurity index, defined as the proportion of local residents who have limited access to credit, either because they have no credit history or have negative credit outcomes; (3) COVID-19 local risk index, which measures, on a scale between 1 and 10, the potential for COVID-19 infection and risk for more severe COVID-19 outcomes and risks at the zipcode-level; and (4) Share of smoking adults, defined as the percentage of adults aged 18 years or above reporting to be current smokers in 2019.

C Comparable COVID-19 serological surveys

Bobrovitz et al. (2021) perform a systematic review of serological studies with the goal of identifying and subsequently synthesizing studies that tested for COVID-10 antibodies. We use their metadata to identify studies that, like ours, invited a random sample of subjects from a pre-specified geographic region in the United States to be tested for COVID-19 antibodies. Our goal in doing so is to understand common practices of such serological surveys and to contextualize our serological survey. In what follows, we first describe our process of identifying such studies using metadata from Bobrovitz et al. (2021)'s systematic review. We then discuss the data we collected for each study we identify. We conclude by presenting our findings.

C.1 Identification of comparable serological studies

Bobrovitz et al. (2021) identify 968 serosurveys conducted between January 1, 2020 and December 31, 2020 that, among other requirements, tested participants for COVID-19 antibodies and reported a sample size, study date, location, and seroprevalence estimate (see Figure 1 of Bobrovitz et al. (2021) for additional details). The metadata for these studies is publicly-available.

We seek to identify studies which invited a random sample of subjects from a geographic region in the United States to be tested for COVID-19 antibodies. We accomplish this goal in two steps. First, we use variables constructed by Bobrovitz et al. (2021) to restrict to studies that were (1) conducted in the United States, (2) used an appropriate sample frame, and (3) used a probability sample.[15] Nineteen studies satisfy these restrictions.

Second, we restrict to the subset of these studies that (1) were published in a scientific journal, (2) defined the target population to be subjects in a geographic region (up to age restrictions, such as excluding children), and (3) invited either the entire target population or a random subsample of the target population. Thus, of the nineteen studies, we excluded three studies that were not from scientific journals, two studies whose target population were respectively prisoners and hospital and/or clinic patients, two studies that constructed their invited samples using market research firms that maintain proprietary samples, and three studies that constructed their invited samples using participants from other surveys. The remaining nine studies satisfy our requirements, and constitute our analysis sample of studies.

C.2 Measuring survey implementation and participation rates

For each study in our analysis sample, we use the metadata of Bobrovitz et al. (2021) to collect (when possible) the outreach method, the number of invited subjects, the number of

[15] Bobrovitz et al. (2021) code a study as using an appropriate sample frame if the sample frame 'described and it approximated the target population' (see item 1 of the metadata) and code a study as using a probability sample if the study used a probability sampling method or the entire sample (see item 2 of the metadata). See the supplementary materials of Bobrovitz et al. (2021) for additional details.

participant subjects, and the offered incentive for participation. Outreach methods could be mail, in-person, online, phone, or any combination of these. We take the number of invited subjects to be the number of subjects who were initially invited to participate in the study, and take the number of participant subjects to be the number of subjects who submitted to be tested for COVID-19 following the study's implementation. The unit for subjects is defined based on the unit targeted by the initial serosurvey invitation. For example, if invites were sent to households but the invitation allowed multiple individuals within a household to participate, subjects correspond to households. When the study includes mail-only as an outreach method and reports invited and participant numbers for mail-only, we use the mail-only results. Two members of the research team independently performed these data collection steps, and there were no conflicts.

C.3 Results of our systematic review

We obtained outreach methods and number of invited subjects and participants for all nine studies. The average participation rate over the nine studies is 12.5% (median: 11.3%, min: 0.4%, max: 23.6%). Four studies either exclusively used mail or reported mail-only results, and the average participation rate for these is 9.0% (median: 8.3%, min: 3.1%, max: 16.5%). These participation rates are comparable to the participation rates we obtained in our serosurvey without financial incentives (6.2%) and with $100 in financial incentives (16.8%). The participation rate we obtain when offering $500 in financial incentives (29.1%) is greater than the maximum participation rate of these studies.

Only three studies explicitly reported financial incentives (or lack thereof) for participation. The offered incentive (participation rate) for each of these three studies was: $10 (16.5%), $50-$100 (7.8%), and $60-$100 (11.3%). For the latter two studies, variations in the amounts were non-random and were used to increase participation rates for certain groups.

Taken together, our results yield three conclusions. First, participation rates in serological surveys that invite a random sub-sample of subjects from a geographic region in the United States are typically low and consistent with the participation rates we obtained in our study. Second, mail is a common form of outreach in serological surveys, with 44% of studies employing this method. Third, financial incentives for participation are rarely explicitly mentioned. In the few studies that do explicitly mention financial incentives, the amounts range from $10-100 and are either assigned uniformly or varied non-randomly.

D Proof of sharpness

Equations (3) and (4) show that γ and $\eta(z)$ are point identified for any value of $\eta(\bar{z})$ such that these expressions remain in the $[0,1]$ interval for each z. From (3), we see that $\gamma \in [0,1]$ if and only if $\eta(\bar{z}) \in [0, 1-\rho(\bar{z})]$. When $\eta(\bar{z}) = 0$, (4) reduces to $(\rho(\bar{z}) - \rho(z))/\rho(\bar{z})$, which is between 0 and 1 as long as $\rho(z)$ is an increasing function of z. On the other hand, when $\eta(\bar{z}) = 1 - \rho(\bar{z})$, (4) reduces to $\eta(z) = 1 - \rho(z)$, which is also between 0 and 1. We conclude that if $\rho(z)$ is increasing in z, then setting $\rho(\bar{z}) = \alpha$ for any $\alpha \in [0, 1-\rho(\bar{z})]$ implies that γ and $\eta(z)$ are point identified via (3) and (4). Taking the union of these points across all $\alpha \in [0, 1-\rho(\bar{z})]$ produces the bounds given in (5).

It remains to be shown that the model can rationalize the data when $\eta(\bar{z})$ is set to any $\alpha \in [0, 1-\rho(\bar{z})]$, and $\rho(z)$ is given, and weakly increasing. To show this, we take α as given and construct a distribution of (C_i, H_i) that is independent of Z_i and (i) reproduces the given $\rho(z)$ for each z, when responses are determined via (1), while (ii) satisfying $\eta(\bar{z}) = \alpha$. The construction proceeds by reversing the logic of the identification argument. First, set the marginal contact rate to be

$$\gamma \equiv \mathbb{P}[C_i = 1] = \frac{\rho(\bar{z})}{1-\alpha}.$$

Next, set the hesitancy rate at each z to be

$$\eta(z) = \left(\frac{\rho(\bar{z}) - \rho(z)}{\rho(\bar{z})} \right) (1-\alpha) + \alpha.$$

Any increasing function defined on a subset of the real line and contained between 0 and 1 can be extended (perhaps non-uniquely) to a proper distribution function.[16] As noted above, both γ and $\eta(z)$ are within 0 and 1, and $\eta(z)$ is decreasing in z, because $\rho(z)$ is increasing in z. Extend $1-\eta(z)$ to a proper distribution function Φ. We use Φ to define a joint distribution of (C_i, H_i) that is independent of Z_i and given by

$$\begin{aligned} & \mathbb{P}[C_i = 1, H_i \leq h] = \gamma \Phi(h) \\ \text{and} \qquad & \mathbb{P}[C_i = 0, H_i \leq h] = (1-\gamma)\Phi(h). \end{aligned}$$

This joint distribution satisfies (i) and (ii) by construction. *Q.E.D.*

[16] The proof is trivial in the scalar case; see Lemma 2 of Torgovitsky (2019) for a generalization to the vector case.

E Additional results

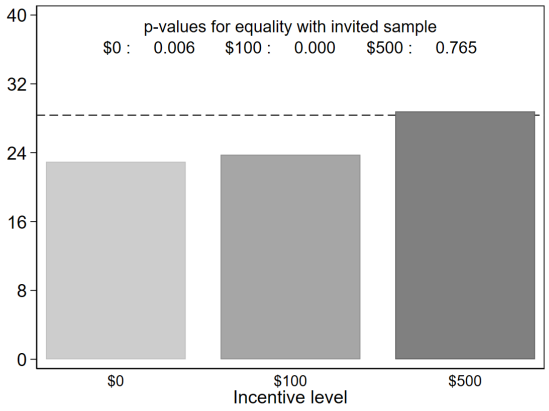
Table E.1: *Correlation between neighborhood characteristics*

	Majority Non-White ZIP	Higher poverty ZIP	Uninsured rate	Preventable hospitalization rate	Drug-related hospitalization rate	COVID local risk index
Majority Non-White ZIP	1.000	0.714	0.688	0.565	0.369	0.761
Higher poverty ZIP	0.714	1.000	0.518	0.649	0.550	0.826
Uninsured rate	0.688	0.518	1.000	0.327	0.112	0.550
Preventable hospitalization rate	0.565	0.649	0.327	1.000	0.872	0.763
Drug-related hospitalization rate	0.369	0.550	0.112	0.872	1.000	0.639
COVID local risk index	0.761	0.826	0.550	0.763	0.639	1.000

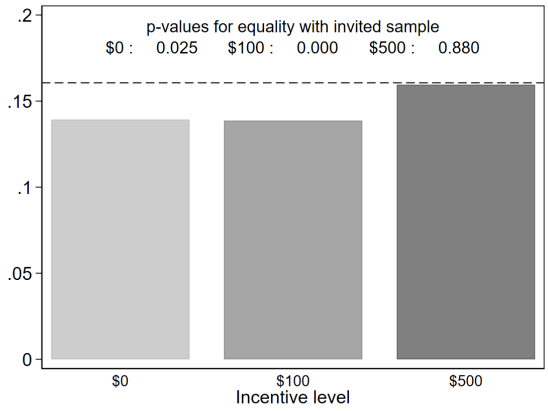
Notes: This table presents, for the invited sample, the degree of correlation between the household characteristics considered in Section 6.

Figure E.1: *Effect of hesitancy on representativeness: additional characteristics*

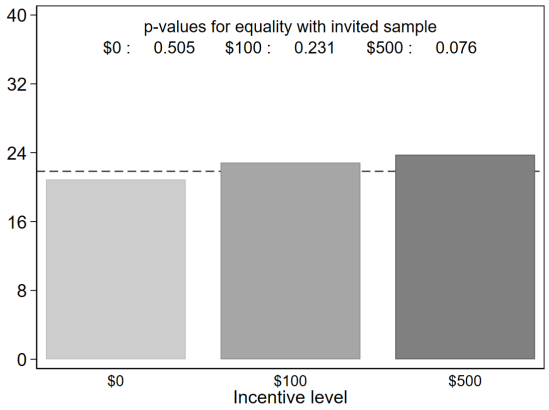
(a) *Credit insecurity index*



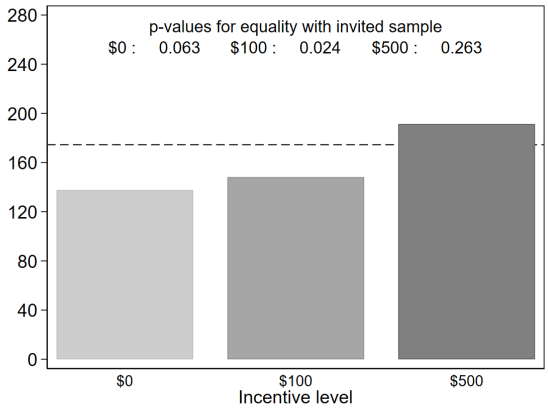
(b) *Share of smoking adults*



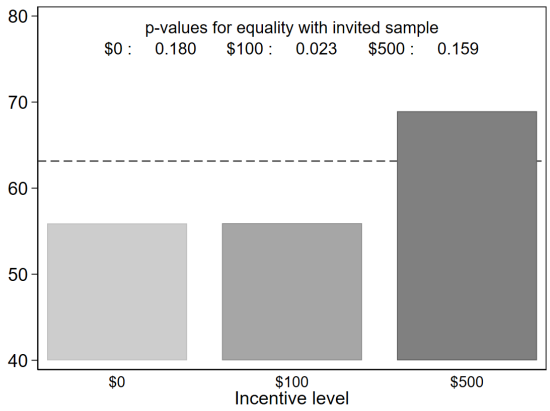
(c) *Alcohol related hospitalization rate*



(d) *Behavioral health hospitalization rate*



(e) *Mood and depressive disorder hospitalization rate*



Notes: See notes for Figure 4