Online Appendix:

Non-Representativeness 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‘

A Additional exhibits

Table A.1: *Balance test*

	Incentive level			p-value
	$0	$100	$500	of equality
Share Non-White (%)	62.3	63.1	67.6	0.16
	[27.6]	[27.6]	[27.6]	
Share poor (%)	35.3	35.5	38.1	0.20
	[15.9]	[15.9]	[15.9]	
Share working age (%)	59.4	59.0	58.7	0.65
	[8.0]	[8.0]	[8.0]	
Share Female (%)	51.5	51.7	51.8	0.45
	[2.6]	[2.6]	[2.6]	
Share unemployed (%)	8.3	8.5	9.1	0.36
	[5.3]	[5.3]	[5.3]	
Share uninsured (%)	8.6	8.6	9.0	0.59
	[4.2]	[4.2]	[4.2]	
Drug-related hospitalization rate (per 10k)	29.1	30.4	33.2	0.38
	[29.0]	[28.8]	[28.9]	
Preventable hospitalization rate (per 10k)	188.7	191.6	203.1	0.21
	[80.4]	[80.5]	[80.9]	
COVID local risk index	5.2	5.3	5.7	0.17
	[3.0]	[3.0]	[3.0]	
Joint test				0.80
N	374	374	134	

Notes: This table presents the average neighborhood characteristics for the invited sample by incentive group. Standard deviations are presented in square brackets below the estimated means. The last column presents the p-value for the null hypothesis of equality of means across incentive groups. The number of invited households per incentive group is presented at the bottom of the table.

Table A.2: *Sample sizes*

		Incentive level		
	Pooled	$0	$100	$500
Invited	882	374	374	134
Participants	125	23	63	39
and Majority white	62	13	37	12
and Majority minority	63	10	26	27
and Lower poverty	84	19	47	18
and Higher poverty	41	4	16	21

Notes: This table presents cell counts (pooled and by incentive level) for the invited sample, for participants, and for participants in the subgroups examined in Panels B and C of Table 2.

Table A.3: *Correlation between neighborhood characteristics*

	Share Non-White (%)	Share poor (%)	Share working age (%)	Share Female (%)	Share unemployed (%)	Share uninsured (%)	Drug-related hosp. rate (per 10k)	Preventable hosp. rate (per 10k)	COVID local risk index
Share Non-White (%)	1.000	0.907	-0.720	0.447	0.859	0.629	0.586	0.749	0.935
Share poor (%)	0.907	1.000	-0.669	0.322	0.821	0.692	0.662	0.765	0.890
Share working age (%)	-0.720	-0.669	1.000	-0.513	-0.690	-0.468	-0.373	-0.431	-0.827
Share Female (%)	0.447	0.322	-0.513	1.000	0.567	-0.118	0.479	0.437	0.535
Share unemployed (%)	0.859	0.821	-0.690	0.567	1.000	0.355	0.736	0.825	0.883
Share uninsured (%)	0.629	0.692	-0.468	-0.118	0.355	1.000	0.112	0.327	0.550
Drug-related hosp. rate (per 10k)	0.586	0.662	-0.373	0.479	0.736	0.112	1.000	0.872	0.639
Preventable hosp. rate (per 10k)	0.749	0.765	-0.431	0.437	0.825	0.327	0.872	1.000	0.763
COVID local risk index	0.935	0.890	-0.827	0.535	0.883	0.550	0.639	0.763	1.000

Notes: This table presents, for the invited sample, the degree of correlation between the nine neighborhood characteristics we consider.

Table A.4: *Representativeness of participants across incentive groups - reweighting*

		$0		$100		$500	
	Invited	UNW	RW	UNW	RW	UNW	RW
Share Non-White (%)	63.5	48.6	60.3	49.1	59.9	62.3	59.6
	(0.9)	(4.9)	(5.4)	(3.0)	(3.1)	(3.8)	(4.1)
Share poor (%)	35.8	25.8	31.8	27.7	35.0	36.9	35.2
	(0.5)	(2.9)	(2.4)	(1.7)	(1.9)	(2.2)	(2.4)
Share uninsured (%)	8.6	6.9	7.3	7.3	8.9	9.2	8.7
	(0.1)	(0.9)	(0.6)	(0.5)	(0.6)	(0.7)	(0.7)
Drug-related hospitalization rate (per 10k)	30.3	17.4	23.4	19.9	25.5	28.6	27.4
	(1.0)	(4.4)	(3.4)	(2.7)	(3.0)	(3.4)	(4.4)
Preventable hospitalization rate (per 10k)	192.1	158.0	173.1	158.8	180.0	193.8	189.5
	(2.7)	(15.0)	(13.2)	(9.1)	(9.1)	(11.5)	(13.9)
COVID local risk index	5.3	4.1	5.1	3.9	5.0	5.3	5.0
	(0.1)	(0.5)	(0.6)	(0.3)	(0.3)	(0.4)	(0.5)

Notes: This table presents the average neighborhood characteristics of participants across incentive groups with and without reweighting. The first column presents the average in the invited sample. The next three pairs of columns present unweighted ('UNW') and reweighted ('RW') averages for unincentivized participants and participants in the $100 and the $500 incentive groups. We compute the probability of participation by racial composition and poverty status of the neighborhood and reweight participants by the inverse of this probability. Standard errors are shown in parentheses; we compute these via bootstrapping for reweighted estimates.

B Robustness results

B.1 Robustness to measures of race and poverty status

Appendix Table B.1 presents the same results as in Table 3 for different measures of race (Panel A) and poverty (Panel B). We see that the findings and conclusions discussed in Section 2.5 are not sensitive to how we defined these measures: relative to households that participate without incentives, households that participate with $500 are more likely to reside in neighborhoods with higher shares of racial minorities and poverty, and these patterns are often monotonic across all three incentive levels.

Table B.1: *Representativeness of participants across incentive groups: Robustness to alternative definitions*

	Incentive level				p-value of	p-value of non-rep		
	$0	$100	$500	Invited	selection	$0	$100	$500
Panel A: *Racial composition*								
Share Non-White (%)	48.6	49.1	62.3	63.5	0.02	0.01	0.00	0.80
	(4.9)	(3.0)	(3.8)	(0.9)				
Share Black (%)	20.6	16.1	27.3	30.0	0.13	0.17	0.00	0.61
	(5.6)	(3.4)	(4.3)	(1.1)				
Share Hispanic (%)	18.4	21.2	24.5	24.4	0.50	0.20	0.25	1.00
	(4.2)	(2.5)	(3.2)	(0.8)				
Panel B: *Poverty status*								
Share poor [below 2x PL] (%)	25.8	27.7	36.9	35.8	0.00	0.00	0.00	0.66
	(2.9)	(1.7)	(2.2)	(0.5)				
Share below PL (%)	12.3	13.9	18.8	17.9	0.00	0.00	0.00	0.51
	(1.5)	(0.9)	(1.2)	(0.3)				
Share below 1.5xPL (%)	19.1	21.0	28.2	27.3	0.00	0.00	0.00	0.64
	(2.2)	(1.3)	(1.7)	(0.4)				
Share below 1.85xPL (%)	23.9	25.7	34.5	33.4	0.00	0.00	0.00	0.65
	(2.7)	(1.6)	(2.1)	(0.5)				
Share below 3xPL (%)	38.2	39.5	50.8	49.6	0.00	0.01	0.00	0.70
	(3.7)	(2.3)	(2.9)	(0.7)				

Notes: This table presents the average neighborhood characteristic of participants across incentive groups (first three columns), the average characteristic of the invited sample (fourth column), the p-value for equality of participant averages across incentive groups (fifth column), and the p-value for equality of the invited and the participant averages for each incentive group (last three columns). Standard errors are presented below in parentheses. Panel A examines alternative measures on the racial composition of neighborhoods. Panel B examines alternative measures on the poverty status of neighborhoods.

B.2 Robustness to binarizing racial composition and poverty status

Appendix Table B.2 presents the same results as in Panel B of Table 2 for different binarizations of our considered measure of race. The first set of results is as in the main paper, and the second and third set of results respectively change the cutoff to 45% and 55%. Appendix Table B.3 presents the same results as in Panel C of Table 2 for different binarizations of our considered measure of poverty. The first set of results is as in the main paper, and the following three results vary how we define a household as poor (150% or 200% of the

poverty line), and whether the share of households in the neighborhood is greater than the median share (roughly 34%) or greater than 30%. We consistently find the same results and conclusions.

Table B.2: *Participation rates (in %) across incentive levels and neighborhood racial composition: robustness*

	Incentive level			Incentive difference	
	\$0	\$100	\$500	\$100 − \$0	\$500 − \$100
Majority non-white (above 50%)					
Majority white	8.9	25.9	30.0	17.0	4.1
	(2.8)	(2.8)	(5.3)	(4.0)	(6.0)
Majority minority	4.4	11.3	28.7	6.9	17.5
	(2.2)	(2.2)	(3.5)	(3.1)	(4.1)
Majority non-white (above 45%)					
Majority white	7.2	25.4	30.3	18.2	4.9
	(3.0)	(3.0)	(5.9)	(4.3)	(6.6)
Majority minority	5.6	12.5	28.7	6.9	16.2
	(2.1)	(2.1)	(3.4)	(3.0)	(4.0)
Majority non-white (above 55%)					
Majority white	8.7	24.6	33.3	15.9	8.8
	(2.5)	(2.5)	(4.7)	(3.6)	(5.4)
Majority minority	3.7	10.1	26.5	6.4	16.5
	(2.4)	(2.4)	(3.7)	(3.4)	(4.4)

Notes: This table presents participation rates by incentive group and alternative racial composition definitions. Standard errors are presented in parentheses below the estimated rates.

Table B.3: *Participation rates (in %) across incentive levels and neighborhood poverty status: robustness*

	Incentive level			Incentive difference	
	\$0	\$100	\$500	\$100 − \$0	\$500 − \$100
Share below 200% PL is above median					
Lower poverty	9.7	23.7	31.6	14.0	7.8
	(2.4)	(2.4)	(4.5)	(3.4)	(5.1)
Higher poverty	2.2	9.1	27.3	6.8	18.2
	(2.5)	(2.5)	(3.8)	(3.6)	(4.6)
Share below 200% PL is above 30%					
Lower poverty	10.4	23.4	26.1	13.0	2.7
	(2.7)	(2.7)	(5.0)	(3.8)	(5.7)
Higher poverty	3.2	12.0	30.7	8.9	18.6
	(2.3)	(2.3)	(3.6)	(3.2)	(4.3)
Share below 150% PL is above median					
Lower poverty	9.7	23.7	32.1	14.0	8.4
	(2.4)	(2.4)	(4.5)	(3.4)	(5.1)
Higher poverty	2.2	9.4	26.9	7.2	17.5
	(2.5)	(2.5)	(3.8)	(3.6)	(4.6)
Share below 150% PL is above 30%					
Lower poverty	9.6	23.7	31.7	14.1	8.0
	(2.3)	(2.3)	(4.3)	(3.3)	(4.9)
Higher poverty	1.8	8.4	27.0	6.6	18.6
	(2.6)	(2.6)	(3.9)	(3.7)	(4.7)

Notes: This table presents participation rates by incentive group and alternative poverty status definitions. Standard errors are presented in parentheses below the estimated rates.

C 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 C.1 discusses the construction of the sampling frame and the sampling and randomization procedures. Appendix C.2 describes outreach and follow-up procedures, and additionally discusses the materials sent to invited households. These materials are reproduced in Appendix C.3. This study, its design, and its implementation were approved by the IRB at the University of Chicago (IRB20-0721).

C.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).[7] 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.

C.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.[8] 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. 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

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

[8] 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.

Exhibits C.1, C.2, and C.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 first two pages of the consent form differed slightly depending on the assigned incentive arm. Appendix Exhibits C.4, C.5, and C.6 respectively depict the first page of the consent forms sent to households in the $0, $100 compensation, and $500 compensation arms. Appendix Exhibits C.7, C.8, and C.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 respectively depicted in Appendix Exhibits C.10 and C.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 C.12, C.13, and C.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.[9] 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.[10] 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.

[9]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.

[10]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.

C.3 Materials

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

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

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

D Data sources and variable definitions

Our empirical analysis uses data collected from the serology study described in Section 2. We also link the study data to a set of neighborhood (five digit zipcodes) characteristics we collect from three sources: the American Community Survey, the Chicago Health Atlas, and the City Health Dashboard. Below, we provide additional information on how we collect and use this data to define the individual and neighborhood characteristics we consider.

D.1 Individual characteristics

As described in Section 2, the RECOVER study included a short questionnaire that households were asked to complete (Greenstone et al., 2023). The questionnaire elicited the participant's age, their gender, their race, whether they are Hispanic, and their 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, and $100,000 or more). Although completion of the questions was required to receive compensation, all questions excluding age included a 'Prefer not to answer' option. See Exhibit C.11 of Appendix C for a copy of the questionnaire.

Of the 125 participants, 121 (97%) provided at least one response that was not 'Prefer not to answer,' and 109 (87%) provided responses to all questions that were not 'Prefer not to answer' for all questions. Specifically regarding race and income, 119 (95%) of participants provided a response to race that was not 'Prefer not to answer,' and 109 (87%) provided a response to income that was not 'Prefer not to answer.' In our analyses of individual characteristics, we drop responses that are either missing or 'Prefer not to answer.'

For participants for which we observe valid responses, we measure whether they are non-White (race is not White or they are Hispanic), whether they are poor (their household yearly income is below $50,000), whether they are of working age (ages 25-60), and whether they are female.

D.2 Neighborhood characteristics

American Community Survey. We obtain neighborhood demographics from the 2019 American Community Survey 5-year estimates (2015-2019) (U.S. Census Bureau, 2020). We collect, for each neighborhood: (1) the share of individuals not identifying as non-Hispanic white, (2) the share of households below 200% the poverty line, (3) the share of individuals between 25 and 60 years old, and (4) the share of individuals who identify as female.

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 (Chicago Department of Public Health, 2022). More specifically, we obtain the uninsurance rate and two diagnosis-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 diagnosis per 10,000 people in 2017, excluding discharges to Veterans Administration hospitals. We obtain these hospitalization rates for the following diagnoses: (1) drug-related (which include amphetamines, cannabis, cocaine, drug-induced mental disorders, hallucinogens, opioids, sedatives, hypnotics, anxiolytics, tranquilizers, barbiturates, and other drugs); and (2) 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 Department of Population Health, NYU Langone Health (2021). 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 (U.S. Census Bureau, 2021b). 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; and (2) 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.

D.3 Aligning individual and neighborhood characteristics

We constructed individual and neighborhood characteristics to minimize differences in definitions. Measures of race and gender naturally align. For our measure of poverty, we choose the cutoff at 200% of the poverty line–rather than 150% or 100%–to more closely align with the income bins elicited in the study's questionnaire. In particular, for a household of three, 200% of the poverty line was $41,122 in 2020 (U.S. Census Bureau, 2021a), which, relative to other cutoffs, is closer to the $50,000 cutoff from the study's questionnaire. Finally, for our measure of age, we consider an indicator for working age because the ACS does not provide a natural neighborhood-level measure of average age but does provide the share of working age individuals. As we show in Appendix B, our results are not sensitive to how we define neighborhood characteristics.

E 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-19 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.

E.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.[11] 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.

E.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

[11] 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.

E.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.

F Decomposing non-contact and hesitancy

Nonrepresentativeness relative to the invited sample is caused by differential non-participation. Non-participation occurs for one of two reasons: either a sampled household is unable to be contacted (non-contact), or a contacted household does not participate because the perceived costs of doing so exceed the perceived benefits (hesitancy). This Appendix develops and applies a method for separating the roles of non-contact and hesitancy in determining non-participation (and nonrepresentativeness).

F.1 A model of study participation

F.1.1 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 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.

F.1.2 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.

F.1.3 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.

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 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). Although we have suppressed demographic conditioning, we emphasize that when we estimate the model separately by demographic group, $\eta(\bar{z})$ can take any value lower than the upper bound α for each group, and can vary across groups. 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 $\alpha = 1-\rho(\bar{z})$. 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.

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.[12] 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

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

of (C_i, H_i) that is independent of Z_i and given by

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

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

F.2 The causes of low and unequal participation rates in RECOVER

We now use the method above 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).

F.2.1 Baseline estimates

Table F.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 F.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 Table 2 in the main body. Participation rates of 29% under the $500 incentive arm correspond to non-contact rates of 71% and we find no large or statistically significant differences in non-contact rates across households by neighborhood racial composition and poverty status.

The second column of Table F.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 85% for households in majority non-White neighborhoods and 92% for households in higher poverty neighborhoods. These findings suggest that the perceived costs of participation are empirically relevant barriers to participation, especially for minority and lower-income 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 F.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 majority White and lower poverty 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 majority non-White and higher poverty neighborhoods. Whereas reservation payments for contacted households in minority and lower-income 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 minoity and lower-income 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

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

	Non-contact rate	Hesitancy rate At $0	Hesitancy rate At $100
All	0.71	0.79	0.42
	(0.66, 0.76)	(0.68, 0.89)	(0.28, 0.56)
Majority non-White	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)	(-0.20, 0.47)
Difference	0.01	0.14	0.47
	(-0.09, 0.12)	(-0.08, 0.36)	(0.14, 0.80)
Higher poverty	0.73	0.92	0.67
	(0.67, 0.78)	(0.79, 1.05)	(0.52, 0.81)
Lower poverty	0.68	0.69	0.25
	(0.60, 0.77)	(0.53, 0.86)	(0.00, 0.49)
Difference	0.04	0.22	0.42
	(-0.05, 0.14)	(0.01, 0.44)	(0.14, 0.69)

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.

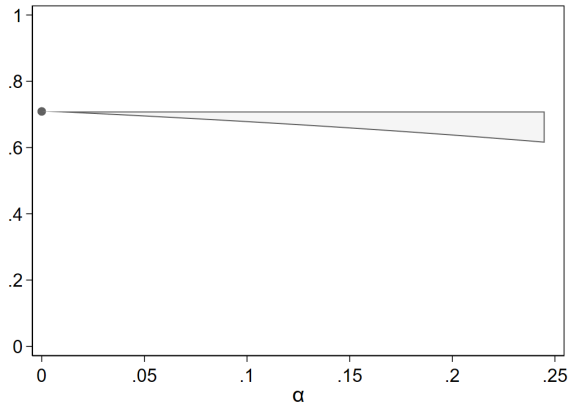
differences in hesitancy, since participants were informed that they would not be told their test result. A substantial qualitative literature instead points to differences in trust in the healthcare system and differences in concern about privacy as potential factors limiting study participation among racial minorities (see, e.g., Chapter 4 of NASEM, 2022; Alsan et al., 2022).

F.2.2 Decomposing the causes of unequal participation

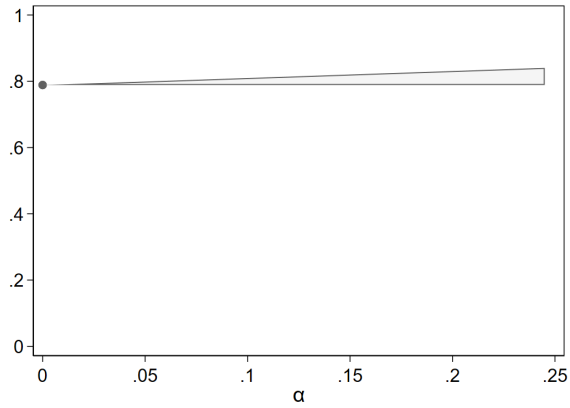
A decomposition exercise helps clarify the relative importance of non-contact and hesitancy in explaining unequal participation by racial composition and income level. Suppose that majority non-White households had the same hesitancy at $0 as majority White households. Then, instead of a participation rate of .043 at $0, majority non-White 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 non-White 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 3.2% to 8.4% at $0 and 12.0% to 20.3% at $100, compared to 10.4% at $0 and 23.4% at $100 for higher income households. In all cases, setting hesitancy rates equal across households largely closes participation gaps across racial composition and income level. These results suggest that unequal participation

Figure F.1: *Bounds on non-contact and hesitancy rates*

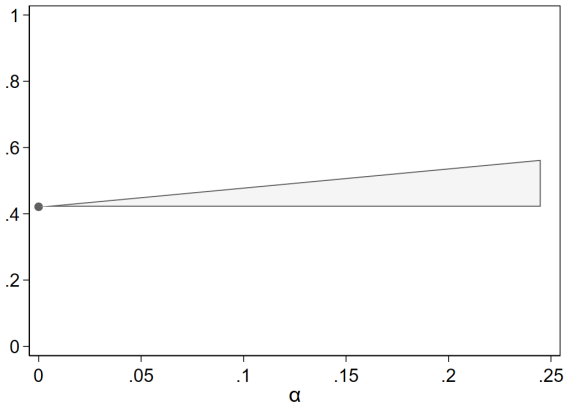
(a) *Non-contact rate*



(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.

rates are primarily driven by differences in hesitancy.

F.2.3 Sensitivity analysis

The estimates in Table F.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 our notation, $\eta(\bar{z}) = 0$. 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 F.1 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 F.1a shows 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.

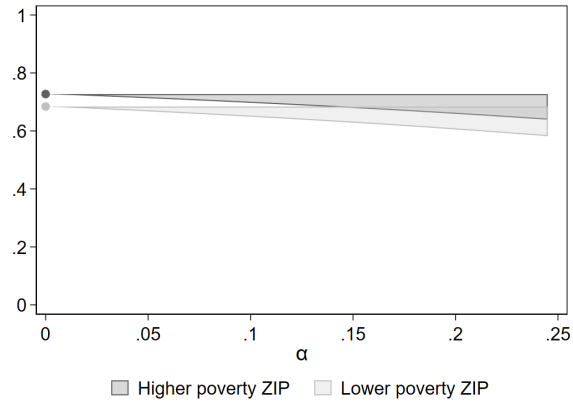
Figure F.2 plots estimated bounds by demographic group. For any α, the share of contacted households who decline to participate at \$500 can take any value lower than α for each group, and this share is allowed to vary freely across demographic groups. For example, we allow for households in higher poverty neighborhoods to decline to participate at \$500 at higher (or lower) rates than households in lower poverty neighborhoods. Figures F.2a and F.2b 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 F.2c–F.2f show that the opposite is true for hesitancy rates: even at $\alpha = .25$, hesitancy rates at both \$0 and \$100 differ markedly by both racial composition and poverty status. These results are consistent with the conclusions from the baseline case.[13]

As discussed in Appendix Section F.1, 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.

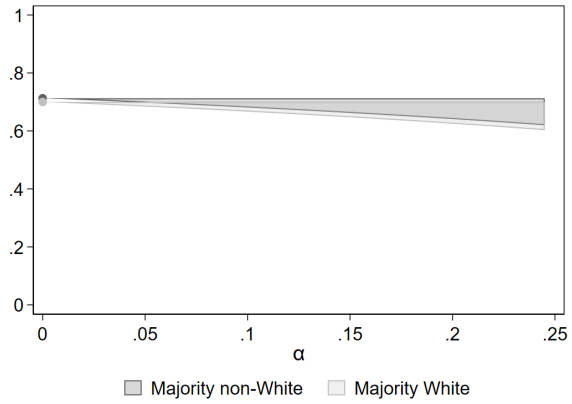
[13]The choice of letting α go as high as .25 was for illustrative purposes. These conclusions continue to hold even if we allow α to be as high as .45. Even at this value, it is still the case that hesitancy rates at \$0 and \$100 do not intersect–and thus differ–by racial composition and poverty status.

Figure F.2: *Bounds on non-contact and hesitancy rates by demographics*

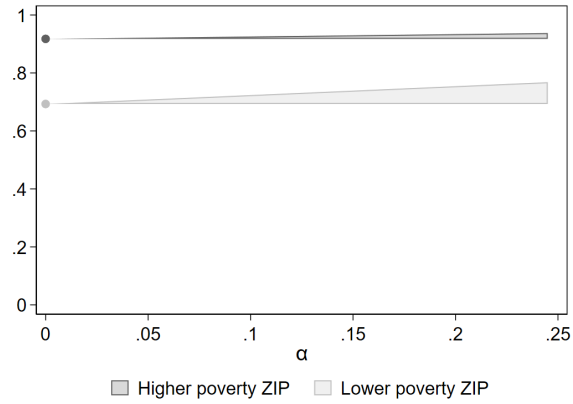
(a) *Non-contact rate by income level*



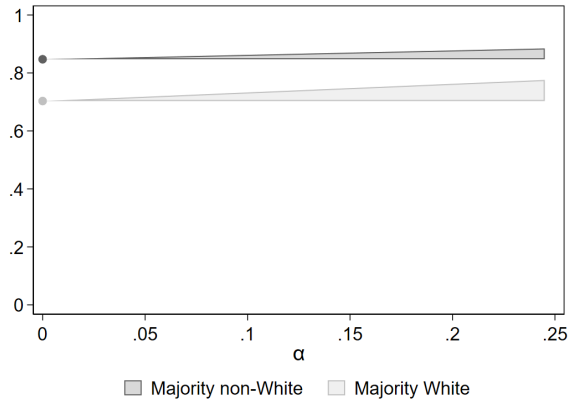
(b) *Non-contact rate by racial composition*



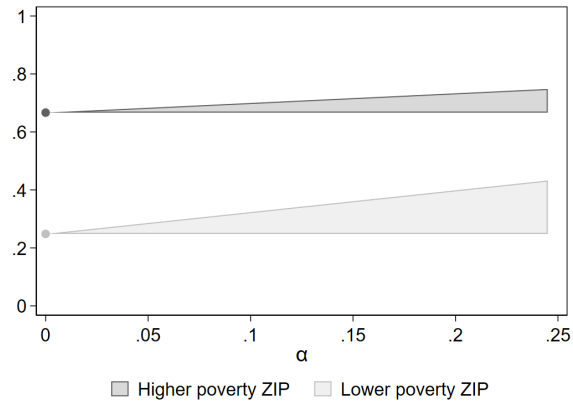
(c) *Hesitancy rate at $0 by income level*



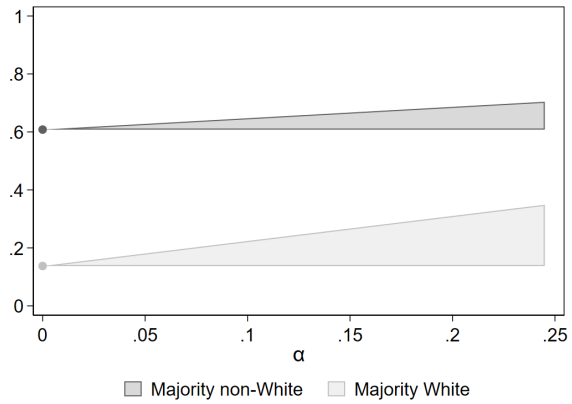
(d) *Hesitancy rate at $0 by racial composition*



(e) *Hesitancy rate at $100 by income level*



(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.