HHS Public Access
Author manuscript
J Racial Ethn Health Disparities. Author manuscript; available in PMC 2022 February 01.

Published in final edited form as:
J Racial Ethn Health Disparities. 2021 February ; 8(1): 186–198. doi:10.1007/s40615-020-00770-2.

ChicagO Multiethnic Prevention And Surveillance Study (COMPASS): increased response rates among African American residents in low socioeconomic status neighborhoods

David J. Press[1,2,3,*], **Briseis Aschebrook-Kilfoy**[1,4], **Diane Lauderdale**[1], **Elizabeth Stepniak**[1], **Scarlett Lin Gomez**[5], **Emily Peterson Johnson**[6], **Rajan Gopalakrishnan**[7], **Fabrice Smieliauskas**[8,9], **Donald Hedeker**[1], **Luís Bettencourt**[10,11,12,13], **Luc Anselin**[14,15,16], **Habibul Ahsan**[1,4,7,17,18]

[1]Department of Public Health Sciences, University of Chicago, Chicago, Illinois

[2]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago Illinois

[3]The Center for Health Information Partnerships (CHiP), Institute of Public Health and Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois

[4]Institute for Population and Precision Health, University of Chicago, Chicago, Illinois

[5]Department of Epidemiology and Biostatistics, University of California, San Francisco, School of Medicine, San Francisco, California

[6]Texas Department of State Health Services, Austin, Texas

[7]Comprehensive Cancer Center, University of Chicago, Chicago, Illinois

[8]Department of Economics, Wayne State University, Detroit, Michigan

[*]**Corresponding author:** David J. Press, PhD, MPhil, Department of Public Health Sciences, The University of Chicago Biological Sciences, 5841 S. Maryland Ave., MC2000, Chicago, IL 60637, djpress@uchicago.edu.

Declarations
Not applicable

Conflicts of interest/Competing interests
None declared

Code availability
The code that support the findings of this study may be available on request from the corresponding author, DJP.

Availability of data and material
The data that support the findings of this study may be available on request from the senior author, HA. The data are not publicly available due to their containing information that could compromise the privacy of research participants.

Ethics approval
The study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments. This study was approved by the Institutional Review Board (IRB) of the University of Chicago Biological Sciences Division (http://bsdirb.bsd.uchicago.edu).

Consent for publication
Not applicable

[9]Department of Pharmacy Practice, Wayne State University, Detroit, Michigan

[10]Mansueto Institute for Urban Innovation, University of Chicago, Chicago, Illinois

[11]Department of Ecology and Evolution, University of Chicago, Chicago, Illinois

[12]Department of Sociology, University of Chicago, Chicago, Illinois

[13]Santa Fe Institute, Santa Fe, New Mexico

[14]Department of Sociology, University of Chicago, Chicago, Illinois

[15]The College, University of Chicago, Chicago, Illinois

[16]Center for Spatial Data Science, University of Chicago, Chicago, Illinois

[17]Department of Human Genetics, University of Chicago, Chicago, Illinois

[18]Department of Medicine, University of Chicago, Chicago, Illinois

Abstract

African American (AA) populations experience persistent health disparities in the US. Low representation in bio-specimen research precludes stratified analyses and creates challenges in studying health outcomes among AA populations. Previous studies examining determinants of bio-specimen research participation among minority participants have focused on individual-level barriers and facilitators. Neighborhood-level contextual factors may inform bio-specimen research participation, possibly through social norms and the influence of social views and behaviors on neighbor's perspectives. We conducted an epidemiological study of residents in 5,108 Chicago addresses to examine determinants of bio-specimen research participation among predominantly AA participants solicited for participation in the first six years of ChicagO Multiethnic Prevention And Surveillance Study (COMPASS). We used a door-to-door recruitment strategy by interviewers of predominantly minority race/ethnicity. Participants were compensated with a $50 gift card. We achieved response rates of 30.4% for non-AA addresses and 58.0% for AA addresses, with as high as 80.3% response among AA addresses in low socioeconomic status (SES) neighborhoods. After multivariable adjustment, we found approximately 3 times the odds of study participation among predominantly AA addresses in low vs. average SES neighborhoods (odds ratio (OR)=3.06; 95% confidence interval (CI)=2.20–4.24). Conversely, for non-AA addresses we observed no difference in the odds of study participation in low vs. average SES neighborhoods (OR=0.94; 95% CI=0.71–1.25) after multivariable adjustment. Our findings suggest that AA participants in low SES neighborhoods may be recruited for bio-specimen research through door-to-door approaches with compensation. Future studies may further elucidate best practices to improve bio-specimen research participation among minority populations.

Keywords

Epidemiology; Bio-specimen; Research participation; Health disparities; Study design

Introduction

The African American (AA) population comprises 14% of the US population [1] and is characterized by greater levels of genetic diversity than non-African populations [2]. The AA population experiences health disparities across the lifespan relative to non-Hispanic (NH) White populations, including higher mortality from cancer, cerebrovascular, and heart disease [3–7]. Historically and presently, AA persons have been exposed to disproportionately high levels of social and environmental stressors, including low socioeconomic status (SES), social isolation, interpersonal and institutional discrimination, and residence in under-resourced communities. These factors intersect at individual- and neighborhood- levels and are associated with poor lifestyle factors, healthcare utilization, and health outcomes [8–10].

Although the inclusion of minority populations in clinical and epidemiological research is a growing national priority in the United States [11], AA participants and bio-specimens are under-represented in clinical and basic science research involving genetics, including approximately 6.0% of cancer clinical trials enrollees in ClinicalTrials.gov [12] and <10% of bio-samples in The Cancer Genome Atlas [13–15]. Observational research involving bio-specimens is also hampered by low response rates among AA participants [16–18], which precludes stratified analyses and is a potential source of non-response bias. Recent studies that have examined willingness to participate in biobanking or genetics studies among AA participants have suffered from low response rates ranging from 10% to 32% [16, 19].

Previous studies examining determinants of bio-specimen research participation among minority participants have focused on individual-level barriers and facilitators such as degree of trust towards researchers, knowledge about bio-specimen research, cash remuneration, discomfort with needles, altruism and personal salience of medical research [13, 16–25]. Evidence regarding race/ethnicity as a factor associated with bio-specimen research participation has been mixed [13, 25]. Mistrust of medical professionals among AA persons due to historical racism and institutional discrimination are perceived barriers for AA representation in bio-specimen research [15, 17, 23]. However, a study conducted in Washington, DC found that the association between AA race/ethnicity and research participation was explained by bio-specimen knowledge, which was the only statistically significant determinant after multivariable adjustment. This finding suggests that race/ethnicity may cease to be associated with bio-specimen research participation after accounting for bio-specimen knowledge [17]. Moreover, bio-specimen research participation rates may be improved by tailoring recruitment strategies to specific racial/ethnic groups, including concordance of interviewer race/ethnicity and providing bio-specimen education to community residents [22, 26].

Neighborhood-level contextual factors may also inform bio-specimen research participation. As one possible mechanism for how neighborhood-level social determinants of health may underlie many health behaviors and health disparities [27], psychological research posits that behavioral norms may be developed and transferred through observation and verbal communication including positive or negative reinforcement [28]. These social interactions and behavioral norms occur within neighborhoods where individuals reside [29, 30]. This mechanism for neighborhood-level social determinants of health may have relevance to bio-specimen research participation. One previous study of predominantly white participants (<4% AA) found that residence in high level educational attainment zip codes is associated with increased willingness to participate in a bio-specimen study [31]. However, it is unclear whether the neighborhood SES context is associated with bio-specimen research participation among AA participants.

The purpose of this study was to examine the potential association of neighborhood-level factors and bio-specimen research participation. Additionally, this study aimed to explore strategies associated with successful recruitment of racial and ethnic minorities who are traditionally under-represented in bio-specimen research using a cross-sectional study of a larger ongoing cohort study in which participants are recruited for participation in bio-specimen research.

Materials and Methods

This research was conducted in Chicago, Illinois as part of the ongoing ChicagO Multiethnic Prevention And Surveillance Study (COMPASS) (compass.uchicago.edu). Chicago is the third largest city in the US, with approximately 2.7 million residents in 2019 [32]. COMPASS is a population-based longitudinal cohort study designed to identify etiologic associations and opportunities for disease prevention including exposure assessment, early detection, screening, interventions, and survivorship evaluations [33].

Field staff training

The 19 interviewers were of mean age 44.1 (standard deviation=15.5), with 11 females (57.9%), and a diverse racial/ ethnic composition of predominantly minority status (11 AA (57.9%), 6 Hispanic (31.6%), 1 Asian (5.2%), and 1 NH White (5.2%)). All field interviewers had previous experience in field work of various disciplines, although additional training was provided regarding the methodology and procedures of COMPASS. In-person trainings involved PowerPoint presentations with information about the importance of population health research for chronic disease prevention and control. The training emphasized health disparities experienced by AA populations – increased risks of and death from breast cancer, prostate cancer, and overall relative to other populations, in order to underscore the need for increasing participation of minority populations in bio-specimen research.

Additionally, field interviewers were trained in procedures of the consent process, how to use databases such as Research Electronic Data Capture (REDCap) and Dartmouth, and institutional trainings such as Health Insurance Portability and Accountability Act (HIPAA) for complying with medical information privacy laws, and Collaborative Institutional Training Initiative (CITI) for research with human subjects. Field interviewers were also trained in collecting clinical data and specimens, such as blood pressure, urine, and saliva. All field interviewers received phlebotomy training and certification if they did not already have it. Upon completion of the in-person training, field instructors shadowed seasoned interviewers from COMPASS and were supervised by the field manager until it was determined that they were prepared to conduct field interviews independently.

Sampling methodology

In order to match the racial and ethnic distribution desired for participant recruitment (including over-sampling of AA participants), we selected specific census tracts from the original 120 census tracts (~15% of Chicago census tracts) used in COMPASS. A two-stage cluster sampling approach was used: sampling at the census tract level in the first stage to sample tracts that would allow interviewers to efficiently contact households by minimizing travel time and cost; and randomly sampling households within tracts in the second stage to maximize study efficiency while minimizing costs. We used the Stata module *gsample* (Jann, B. 2006) to randomly select the census tracts and addresses within blocks. COMPASS field interviewers were then selected based on having identities that reflected the communities in which research was conducted.

Interviewers were provided with a list of potentially eligible participants compiled from commercially available address lists from the United States Postal Service's Address Management System and demographic data for addresses from commercial sources: NFocus Consulting Inc© [34] and Valassis® [35]. The lists of potentially eligible participants provided partial coverage within the selected census tracts. The recruitment approach was not nested and potential participants at identified addresses were recruited even if they were not on the original vendor list. Within buildings, multiple potential households and units were recruited. Within households, multiple potential participants were approached. Within and across households, multiple interviewers conducted recruitment.

In addition to recruitment from the compiled sampled list of addresses, a snowball sample methodology was utilized in a supplemental recruitment in which field interviewers invited participants to refer friends or family to the study. Snowball recruitment strategy may increase representation of populations that may otherwise be excluded from research due to mistrust of researchers, stigmatization, or social isolation and is a cost-effective strategy to increase participation. Furthermore, during this supplemental recruitment, field interviewers also visited local farmer's markets and YMCAs to recruit participants not on the original vendor list. The above describes the sampling methodology for the early phase used in the present study. Recruitment strategies have subsequently been expanded from targeted community based recruitment to non-targeted recruitment in communities, hospitals and clinics. This strategy will allow for enhanced opportunity for linking with electronic health records in a subsequent phase of recruitment.

Recruitment strategy

Prior to recruiting in a given neighborhood, COMPASS staff spoke with the local Alderman, or elected Chicago city council member, in order to increase participant buy-in through the support and endorsement of trusted community leaders. Outreach and engagement by field staff often included extensive conversations around questions commonly raised by participants including mistrust of research (i.e., Tuskegee trial), individual and community benefits of participation including bi-directional benefit, return of results, and comparisons of community level health. Interviewers were permitted to approach neighborhoods regardless of the interviewer race/ ethnicity or the neighborhood racial/ ethnic composition.

Interviewers went door-to-door during the daytime, Mondays through Fridays and select Saturdays, and recruited all eligible persons within a given household or set of households from a given address. Interviewers recorded the approach date, approach outcome (no contact made, refused participation, revisit planned, all scheduled interviews completed, and no one eligible), notes on refusal and/ or eligibility, and approach date. Individuals were considered eligible for inclusion in COMPASS if they were a resident of the Chicago metropolitan area, age 35 years or older, male or female, English or Spanish speaking, competent to give consent, and permanent resident or citizen of the US. For the purposes of this early phase of the study, we only considered addresses with completed contact attempts, which we defined as addresses where at least 3 approaches were made and/ or an interview that included bio-specimen contribution was successfully completed.

Measures

We consolidated our dataset to the addresses for all building types (single family, multi-unit, etc.) using a long-to-wide transform. This transformation generated summarized data for interviewer-, household-, and design factors at the address level. Interviewer-level factors were based on the interviewer with the most approaches to the address (age, sex, race/ ethnicity). Household-level factors were the potential participant characteristics from both the vendor list and the participant data from recruited subjects, which were summarized as the average household data at the address unit level (mean age [continuous], predominant race/ethnicity [categorical], majority interviewer-household concordant on race/ethnicity [yes, no, no majority for household and/or interviewer]). Design factors were the household units within an address (1, 2–3, and 4+), and whether the address was in the original target sample from the vendor contact or part of the snowball sample.

Neighborhood SES was defined as a time-invariant measure of neighborhood SES developed and made publicly available by the authors of Miles et al. (2016) [36]. For this measure, neighborhood SES was defined at the census tract level on a scale from 0 to 100 with 50 being the national average. An unconstrained single factor model was used, according to the equation (1 × [ln{Median Household Income}]) + (−1.129 × [ln{% Female-Headed Households}]) + (−1.104 × [ln{% Workers 16 years or older who are unemployed}]) + (−1.974 × [ln{% of households in poverty}]) + 0.451 × (1 ×[%high school grads but not bachelors holders] + 2 × [%bachelors holders]) [36].

We categorized these census tract-level neighborhood SES data into quintiles of neighborhood SES at the Chicago area level. We separately defined neighborhood SES categories based on cluster cores of neighborhood SES as average, low, and high using the local Getis-Ord G* statistic [37]. The local G* statistic represents the degree to which the neighborhood characteristic of that tract is more similar (high or low) to the characteristics of the neighboring tracts than would be the case under spatial randomness [41]. Tracts were considered neighboring if they shared an edge with the focal tract. We defined neighborhood SES categories with different levels of statistical significance: a pseudo P-Value<0.05 with randomization of 999 permutations as the analytical measure; and a separate measure with 99,999 permutations and a pseudo P-value<0.0001 for a robustness check. Local G* statistics for each tract were calculated and neighborhoods were categorized using GeoDa 1.12 [38].

Outcome

Our outcome was defined as whether one or more participants enrolled at the address (yes/ no). Participants completed a ~1-hour interview and contributed bio-specimens, including a blood sample, collection of cheek cells using the Oragene DNA self-collection kit, and an optional urine sample.

Statistical Analyses

We used chi-square statistics to generate descriptive characteristics by household, interviewer, and design factors, for whether 1+ participant was enrolled at the address or not. We also used multivariable logistic regression analyses to examine the association between neighborhood-level factors (separately and combined) and bio-specimen research participation, adjusting for design, interviewer, and household factors. To examine the robustness of our main neighborhood-level findings to different classifications of neighborhood level measures, we performed sensitivity analyses of our final models using the three neighborhood SES classifications described above. To assess whether measured characteristics modified the neighborhood level effects of interest, we performed sensitivity analyses by separately stratifying our full model according to predominant interviewer age (<40, 40+ years), predominant interviewer race/ethnicity (AA, NH White, and Hispanic), predominant household age (<55, 55+ years), predominant household race/ethnicity (AA, NH White, and Hispanic), and whether the address was in the original sample (yes/ no). In our multivariable logistic regression models, we calculated odds ratios (ORs) and 95% confidence intervals (CIs). Statistical significance was defined at a nominal $p<0.05$. All analyses were conducted in SAS 9.4 (Cary, NC).

Institutional Review Board (IRB)

All study participants signed a consent form prior to enrollment. All study procedures and materials were reviewed and approved by the University of Chicago Biological Sciences Division IRB (http://bsdirb.bsd.uchicago.edu). The $50 gift card compensation, also approved by the IRB and considered non-coercive, was consistent with studies that have required similar and substantial time input and bio-specimen collection from participants. An example of the Recruitment form used by the interviewers during door-to-door recruitment is provided in **Online Resource**.

Results

A flow diagram of the study sample is presented in Figure 1. Of 14,413 addresses with potentially eligible participants from two commercial vendors and participant referral, we excluded: 7,183 (49.8%) not contacted by interviewers in three screening rounds; 1,417 (9.8%) with contact made and re-visit planned, but interview not yet completed; 423 (6.2%) with household members who did not meet eligibility criteria; 137 (<1%) with missing information on interviewer; and 22 (<1%) with typographical errors in address. As such, our study included 5,108 addresses with potentially eligible persons. Of those addresses, 2,748

(53.8%) refused actively or passively and 2,360 (46.2%) had 1 or more participants enrolled. All 5,108 addresses were included in our analysis, with 4,521 participants total within those participating addresses. Differences in addresses approached in all three screening rounds vs. less than three screening rounds were observed by neighborhood SES category, with the lowest inclusion among addresses in high SES neighborhoods (24.6%) and the highest inclusion among addresses in low SES neighborhoods (47.8%) (P<0.001).

Descriptive characteristics of the study sample are presented in Table 1. Of the 5,108 addresses approached, 2,923 (57.2%) were comprised of predominantly AA households. Descriptive household, interviewer, design and neighborhood characteristics of addresses by whether at least one participant was enrolled within addresses were statistically significantly different for each factor included in the analysis (all p-values <0.001). Of note, response rates varied across predominant household race/ethnicity, with 58.0% of AA addresses and 30.4% of non-AA addresses enrolling 1+ participant. Response rates also varied by neighborhood SES, with 57.3% of low neighborhood SES addresses, 41.4% of average neighborhood SES addresses and 32.2% of high neighborhood SES addresses enrolling 1+ participant. Approximately 80.3% of AA addresses in low SES neighborhoods participated.

Table 2 presents results from the multivariable logistic regression analyses for neighborhood characteristics with outcome of bio-specimen research participation within addresses (yes/no), among the 5,108 addresses. In fully adjusted models, AA addresses had 2.19 times the odds of participation as NH White addresses (OR= 2.19; 95% CI=1.75–2.74). Relative to addresses located in average SES neighborhoods, addresses in low SES neighborhoods had 2.23 times the odds of participation (95% CI=1.91–2.60), and addresses in high SES neighborhoods had 0.30 times the odds of participation (95% CI=0.21–0.42), after full adjustment for household, interviewer, and design factors.

Table 3 presents results from the fully adjusted multivariable logistic regression analyses jointly stratified by whether the address was part of the original or snowball sample, and whether the address was or was not a predominantly AA address. Statistically significantly high odds of participation in low SES neighborhoods relative to average SES neighborhoods were driven by predominantly AA addresses (OR=4.13; 95% CI=3.30–5.16), with statistically significantly high odds in both the original and snowball samples (P-values <0.01). However, no statistically significant difference in odds of participation were observed among predominantly non-AA addresses in low SES neighborhoods relative to those in average SES neighborhoods after full adjustment (OR=1.02; 95% CI=0.81–1.29). In fully adjusted models including only AA addresses in the original target sample, we observed 3.06 times the odds of participation in low SES neighborhoods relative to AA addresses in average SES neighborhoods (OR=3.06; 95% CI=2.20–4.24). In fully adjusted models including only non-AA addresses in the original target sample, we observed no difference in the fully adjusted odds of participation in low vs. average SES neighborhoods (OR=0.94; 95% CI=0.71–1.25). Conversely, statistically significantly low odds of participation for addresses in high SES neighborhoods were driven by addresses that were not predominantly AA, with 0.20 times the odds of participation relative to predominantly non-AA addresses in average SES neighborhoods (95% CI=0.14–0.31).

Upon further investigation, recruitment within high SES neighborhoods was found to be mostly comprised of addresses that were part of the snowball sample (73.0%). Robustness checks based on different measures of neighborhood-level factors produced broadly similar results (Table 4). Removing leverage points and large residuals did not substantively affect findings.

Discussion

There is longstanding concern about low research study participation rates in epidemiological studies in general [20, 26] and for bio-specimen research among AA participants in particular [22]. In a Chicago study of predominantly AA participants that utilized door-to-door recruitment by predominantly minority status interviewers and $50 gift card compensation for research participation including bio-specimen collection, we examined associations between census tract-level SES on bio-specimen research participation at the address level, after adjustment for summarized address-level data on households, interviewers, and design characteristics. We achieved response rates of 30.4% for non-AA addresses and 58.0% for AA addresses, with as high as 80.3% response among AA addresses in low SES neighborhoods. In fully adjusted models of AA addresses in the original target sample, we observed approximately 3 times the odds of participation for AA addresses in low SES neighborhoods relative to AA addresses in average SES neighborhoods. Conversely, for non-AA addresses we observed no statistically significant difference in the fully adjusted odds of participation in low vs. average SES neighborhoods. Stratifying by original sample (yes/ no) allowed us to minimize the threat that our findings were driven by the snowball sampling approach. Our findings suggest that door-to-door recruitment and compensation may be effective strategies to recruit traditionally under-represented racial/ethnic minorities for bio-specimen research in low SES neighborhoods in Chicago.

To our knowledge, this is the first study to examine the association between small-area level (census tract or block group) features of the neighborhood context and bio-specimen research participation, as well as the first study of its kind among predominantly AA addresses. Our results differs from a recent study in which respondents within ZIP codes of higher educational attainment had higher self-reported willingness to participate in a bio-specimen research study. Several design features differed between this San Diego Blood Bank study and ours: the sociodemographic characteristics of the population base (predominantly White population vs. predominantly AA population); the method of recruitment (email vs. door-to-door in-person recruitment); compensation ($0 vs. $50 gift card); demands of time (5 minutes vs. ~1 hour); demands of bio-specimen collection (no vs. yes); demands of sensitive information (limited sensitive information vs. detailed sensitive information) [31].

Our findings may help to clarify the previously mixed evidence for the individual-level effect of AA race/ethnicity on bio-specimen research participation [13, 17, 18, 22, 25]. According to previous qualitative research, some AA persons hold the view that bio-specimen research “[benefits] white populations, while minority groups have been unjustly used as “lab rats””[17]. Such views may reflecting mistrust among AA populations following mainstream media coverage about the development of the HeLa cell line [39]. Previous evidence does however suggest that bio-specimen knowledge moderates racial/ ethnic disparities in bio-specimen research participation [22]. A recent qualitative examination of interviewer-reported factors of research participation in the early phase of COMPASS study found that interviewer race and skills were interviewer-perceived facilitators and that fear of the blood draw and mistrust of medical researchers were interviewer-perceived barriers [40]. These perceptions from our field staff are consistent with findings from the present study, which together suggest that culturally informed door-to-door recruitment in combination with compensation and education to community members and potential participants may be an effective strategy for bio-specimen research participation among AA participants. Additional research in other regions and with other design features may help to confirm and clarify the results presented here.

One possible mechanism for our observed results could be that low neighborhood SES served as a surrogate for low individual SES and hence, potential participants residing in low SES neighborhoods were more motivated by the $50 gift card compensation than potential participants residing in neighborhoods of average neighborhood SES. This is a sensitive issue that is ethically complex. Importantly, our recruitment strategy incorporated bio-specimen research education within both the community engagement and individual participant levels as integral aspects of protecting the welfare, rights, and privacy of human subjects. The $50 gift card compensation in our study was approved by the IRB as consistent with studies that have required similar time input and bio-specimen collection from participants.

A related but possibly distinct mechanism for our observed findings may be the increased availability of potential participants who were unemployed and physically located at the address during the day/ time of interviewer approaches. Low SES neighborhoods may also be a surrogate for population bases with low employment. Among early phase COMPASS participants, approximately 81% self-reported that they were not currently working full-time. Moreover, increased potential participant availability to interviewers in low SES neighborhoods presumably increased the opportunities for recruitment contact relative to potential participants who were full-time employees in average and high SES neighborhoods. Additionally, it is possible that social interactions and behavioral norms that occurred within neighborhoods where individuals resided had an impact on the willingness for bio-specimen research participation.

Our study is strengthened by the geographically and socio-culturally diverse Chicago setting with address information that was used to link participants and non-participants with geo-spatial data at a small-area census tract level. Furthermore, our recruitment strategy emphasized cultural sensitivity training and implementation of interviewers in concert with active community engagement from our institution, including conversations with community block clubs, Aldermen, and community-based police in targeted tracts. While our findings are provocative, we suggest that they be interpreted as specific to our recruitment strategy, compensation, time/ bio-specimen demands for participants, and population base. Chicago has been characterized by unique historical, cultural, and social conditions, as well as specific geospatial differences, particularly for segregated AA neighborhoods [41, 42].

A common limitation of epidemiological studies is the lack of individual-level data from non-responders. Our study lacked individual-level data from non-responders, as well as household-level data on potentially important characteristics, such as SES. Since bio-specimens are provided by individuals after informed consent, the use of address (aggregating individuals, depending on the address) as the unit of analysis is a major weakness of this study. Moreover, our reported response rates were expressed at an address level and do not reflect response rates at an individual level. Our measures may, however, be a surrogate for household-level unmeasured income and employment status. Nevertheless, geospatial methods to elucidate features of the small-area level social environment are useful in research to help quantify potential self-selection bias such as when a population is non-responsive or difficult to contact [27]. An additional limitation of our study is that we did not collect information on point-of-contact for referrals ("who referred who"), which may have been used to inform a network analysis of social connections within communities.

Another limitation of our study is our inability to completely control for interviewer-level features of the research recruitment strategy successes and challenges that are missing or not captured in the COMPASS household database. In particular, we were unable to measure the interviewer's recruitment perceptions in different communities or tracts in order to examine how such perceptions may have impacted interviewer recruitment strategies. These implicit factors may have impacted the interviewer's approach and communications, as well as the communities where interviewers focused their efforts (our interviewer team was predominantly of minority racial/ ethnic background and the interviewers were permitted to approach addresses in low, average, and high SES neighborhoods). By anecdote there was an incident where a COMPASS interviewer of minority race/ethnicity was approached by the police in a more affluent community because of a call by a concerned resident who was suspicious about the interviewer's door-to-door activity. This suggests that the mechanism by which contextual factors may have improved research participation in minority communities may have been complex and due to the interplay of individual-, household-, interviewer-, and contextual factors (i.e, minority status interviewers experienced successes in low SES neighborhoods and challenges in high SES neighborhoods). Furthermore, our household database relied on interviewers for quality data collection. Our method of consolidating interviewer-level factors to the address and tract level partially reduced the threat of potential differential bias by interviewer data collection.

Additionally, our study is limited by the choice of census tracts as administrative units to define neighborhoods in our analysis. For example, sociodemographic features, such as SES, impact outcomes at various levels of spatial aggregation from household, street, block group, tract, and above, and may vary for different communities, risk factors, and outcomes [43]. As such, it is possible that the administrative geographic units in our study may not have effectively captured neighborhood boundaries. We partially minimized the threat this posed to our estimates by smoothing the analytical contextual measures based on the characteristics of the neighborhood features in a census tract's neighbors. Nevertheless, further work may elucidate whether contextual features impact bio-specimen research participation at differing levels of geographic detail.

Conclusions

Our findings suggest that characteristics of the neighborhood SES context, possibly as a surrogate for individual SES, influence bio-specimen research participation in Chicago. Efforts to include a sufficient representation from minority populations in clinical research [11, 43] would benefit from careful attention to study design, sampling frame, and potential differential sociodemographic and contextual-level factors [43]. Enriching research studies with information on the small-area level neighborhood context may elucidate bio-specimen research participation and help to tailor effective recruitment strategies to specific households and neighborhoods. Our findings suggest that AA participants, who are traditionally under-represented in bio-specimen research, may be responsive to door-to-door study recruitment approaches with compensation for bio-specimen research participation. This study has implications for future bio-specimen research with traditionally under-represented populations.

Acknowledgements

We thank Tomide Owolabi and Siam Rezwan for literature review support and Michele Thompson for administrative support. This work was supported by the National Cancer Institute [grant numbers 5P30CA014599-44 and T32CA193193], the National Institute of Environmental Health Sciences [grant number P30ES027792-03], the National Institute on Aging [grant number T32AG000243], the Susan G. Komen Foundation [grant number GTDR16376189], and the University of Chicago Department of Public Health Sciences.

References

1. Rastogl S, Johnson T, Hoeffel E, Drewery M. The Black Population: 2010 In: 2010 Census Briefs. US Department of Commerce: Economics and Statistics Administration; 2011.
2. Campbell MC, Tishkoff SA. African genetic diversity: Implications for human demographic history, modern human origins, and complex disease mapping. Annu. Rev. Genom. Hum. Genet 2008, 9:403–433.
3. O'Keefe EB, Meltzer JP, Bethea TN. Health disparities and cancer: Racial disparities in cancer mortality in the United States, 2000–2010. Frontiers Public Health 2015, 3(51).
4. DeSantis CE, Siegel RL, Sauer AG, Miller KD, Fedewa SA, Alcaraz KI, et al. Cancer statistics for African Americans, 2016: Progress and opportunities in reducing racial disparities. CA Cancer J. Clin 2016.
5. Daly B, Olopade OI. A perfect storm: How tumor biology, genomics, and health care delivery patterns collide to create a racial survival disparity in breast cancer and proposed interventions for change. CA Cancer J. Clin 2015, 65(3):221–238. [PubMed: 25960198]
6. Hirschman J, Whitman S, Ansell D. The black:white disparity in breast cancer mortality: the example of Chicago. Cancer Cause Control 2007, 18(3):323–333.
7. Cunningham TJ, Croft JB, Liu Y, Lu H, Eke PI, Giles WH. Vital Signs: Racial disparities in age-specific mortality among blacks or African Americans - United States, 1999–2015. MMWR Morb. Mortal. Wkly 2017, 66(17):444–456.
8. Morris AM, Rhoads KF, Stain SC, Birkmeyer JD. Understanding racial disparities in cancer treatment and outcomes. J. Am. Coll. Surg 2010, 211(1):105–113. [PubMed: 20610256]
9. Firebaugh G, Acciai F. For blacks in America, the gap in neighborhood poverty has declined faster than segregation. Proc. Nat. Academy Sci 2016, 113(47):13372–13377.
10. Jackson JS, Knight KM, Rafferty JA. Race and unhealthy behaviors: Chronic stress, the HPA axis, and physical and mental health disparities over the life course. Am. J. Public Health 2010, 100(5):933–939. [PubMed: 19846689]
11. Amendment: NIH policy and guidelines on the inclusion of women and minorities as subjects in clinical research. Notice Number: NOT-OD-18–014. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-18-014.html

12. Duma N, Aguilera JV, Paludo J, et al. Representation of minorities and women in oncology clinical trials: Review of the past 14 years. J. Oncol. Pract 2018, 14(1):e1–e10. [PubMed: 29099678]
13. Halbert CH, McDonald J, Vadaparampil S, Rice L, Jefferson M. Conducting precision medicine research with African Americans. PLoS One 2016, 11(7):e0154850. [PubMed: 27441706]
14. Haga SB. Impact of limited population diversity of genome-wide association studies. Genet. Med 2009, 12:81.
15. Erwin DO, Moysich K, Kiviniemi MT, et al. Community-based partnership to identify keys to biospecimen research participation. J. Cancer Educ 2013, 28(1):43–51. [PubMed: 23055133]
16. McDonald JA, Barg FK, Weathers B, et al. Understanding participation by African Americans in cancer genetics research. J. Nat. Medical Assoc 2012, 104(0):324–330.
17. Luque JS, Quinn GP, Montel-Ishino FA, et al. Formative research on perceptions of biobanking: What community members think. J. Cancer Educ 2012, 27(1):91–99. [PubMed: 21927867]
18. Dang JHT, Rodriguez EM, Luque JS, Erwin DO, Meade CD, Chen MS. Engaging diverse populations about biospecimen donation for cancer research. J. Community Genet 2014, 5(4):313–327. [PubMed: 24664489]
19. McDonald JA, Vadaparampil S, Bowen D, et al. Intentions to donate to a biobank in a national sample of African Americans. Public Health Genomics 2014, 17(3):173–182. [PubMed: 24942180]
20. Galea S, Tracy M. Participation rates in epidemiologic studies. Ann. Epidemiol 2007, 17(9):643–653. [PubMed: 17553702]
21. Corbie-Smith G, Thomas S, Williams M, Moody-Ayers S. Attitudes and beliefs of African Americans toward participation in medical research. J. Gen. Intern. Med 1999, 14:537–546. [PubMed: 10491242]
22. Dash C, Wallington SF, Muthra S, Dodson E, Mandelblatt J, Adams-Campbell LL. Disparities in knowledge and willingness to donate research biospecimens: a mixed-methods study in an underserved urban community. J. Community Genetics 2014, 5(4):329–336. [PubMed: 24771039]
23. Goldman RE, Kingdon C, Wasser J, et al. Rhode Islanders' attitudes towards the development of a statewide genetic biobank. Personalized Med. 2008, 5(4):339–359.
24. Lemke AA, Halverson C, Ross LF. Biobank participation and returning research results: Perspectives from a deliberative engagement in South Side Chicago. Am. J. Medical Genetics 2012, 158A(5):1029–1037.
25. Asare M, Heckler CE, Culakova E, et al. Racial/ethnic differences in comprehension of biospecimen collection: A nationwide University of Rochester Cancer Center NCI community oncology research program study. J. Cancer Educ 2019 doi: 10.1007/s13187-018-1664-z.
26. Moorman P, Newman B, Millikan R, Tse C-K, Sandler D. Participation rates in a case-control study: The impact of age, race, and race of interviewer. Ann. Epidemiol 1999, 9(3):188–195. [PubMed: 10192651]
27. National Academies of Sciences and Medicine. Improving health research on small populations: Proceedings of a workshop. Washington, DC: The National Academies Press; 2018.
28. Oostveen T, Knibbe R, De Vries H. Social influences on young adults' alcohol consumption: Norms, modeling, pressure, socializing, and conformity. Addict. Behav 1996, 21(2):187–197. [PubMed: 8730520]
29. Dulin A, Risica PM, Mello J, et al. Examining neighborhood and interpersonal norms and social support on fruit and vegetable intake in low-income communities. BMC Public Health 2018, 18(1):455. [PubMed: 29621989]
30. Davey-Rothwell MA, Siconolfi DE, Tobin KE, et al. The role of neighborhoods in shaping perceived norms: An exploration of neighborhood disorder and norms among injection drug users in Baltimore, MD. Health Place 2015, 33:181–186. [PubMed: 25840353]
31. Bloss C, Stoler J, Schaierer C, et al. Characteristics of likely precision medicine initiative participants drawn from a large blood donor population. Health Affairs 2018, 37(5):786–792. [PubMed: 29733733]
32. Reyes C and O'Connell PM. "Chicago slips in population but is still third-largest city in U.S. – for now". Chicago Tribune, 5 23, 2019.

33. Aschebrook-Kilfoy B, Kibriya MG, Jasmine F, et al. Cohort Profile: The ChicagO Multiethnic Prevention and Surveillance Study (COMPASS). Under review 2020.
34. U.S. Department of Health and Human Services - National Cancer Institute. Cancer Incidence – Surveillance, Epidemiology, and End Results (SEER) Registries Research Data. Bethesda, MD: National Cancer Institute [cited November 2015]. Available from: http://seer.cancer.gov/data/metadata.html.
35. Valassis. Valassis Communications, Inc.; http://valassis.com/about-us/. Accessed January 2019.
36. Miles JN, Weden MM, Lavery D, Escarce JJ, Cagney KA, Shih RA. Constructing a time-invariant measure of the socio-economic status of U.S. census tracts. J. Urban Health 2016, 93(1):213–232. [PubMed: 26678071]
37. Getis A, Ord JK. The analysis of spatial association by use of distance statistics. Geog. Anal 1992, 24(3):189–206.
38. Anselin L “GeoDa Software 1.12,” 2018 [Online]. Available: https://geodacenter.github.io/. Accessed January 2019.
39. Skloot R The immortal life of Henrietta Lacks. New York: Crown Publishers; 2010.
40. Owolabi T, Aschebrook-Kilfoy B. Determinants of African-American participation in a multiethnic biobanking cohort in chicago. In: Short-term research experience for underrepresented persons (STEP-UP) Symposium: 7 2017.
41. Schnell C, Braga AA, Piza EL. The influence of community areas, neighborhood clusters, and street segments on the spatial variability of violent crime in Chicago. J. Quantitative Criminol 2017, 33(3):469–496.
42. Qato D, Daviglus M, Wilder J, Lee T, Qato D, Lambert B. ‘Pharmacy deserts’ are prevalent in Chicago’s predominantly minority communities, raising medication access concerns. Health Affairs 2014, 33(11):1958–1965. [PubMed: 25367990]
43. Polite BN, Adams-Campbell LL, Brawley OW, et al. Charting the future of cancer health disparities research: A position statement from the American Association for Cancer Research, the American Cancer Society, the American Society of Clinical Oncology, and the National Cancer Institute. J. Clin. Oncol 2017, 35(26):3075–3082. [PubMed: 28737975]

Author Manuscript

Author Manuscript

Author Manuscript

Consent to participate

Informed consent was obtained from all individual participants included in the study.

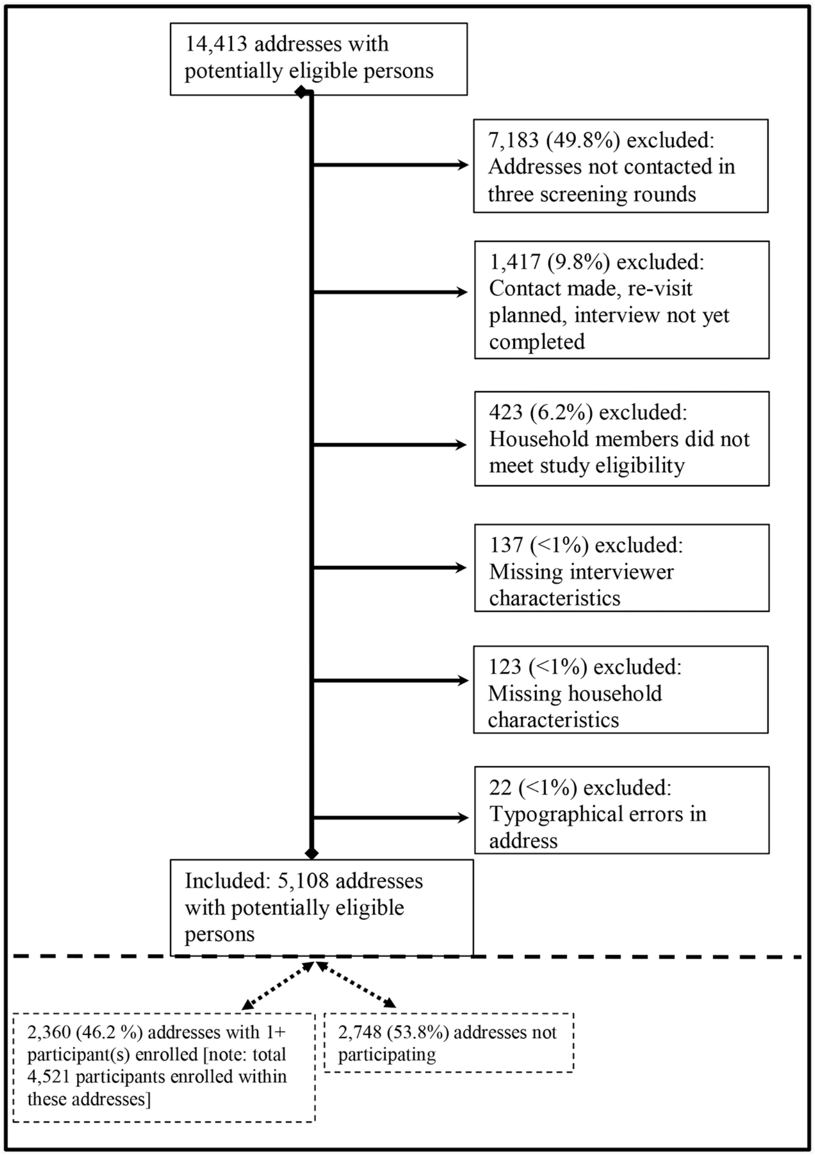

Figure 1.
Flow diagram of Chicago study sample

Table 1

Descriptive household, interviewer and design characteristics of 5,108 addresses in Chicago study, by whether one or more participant was enrolled in the study

Address level characteristics	1+ participant enrolled at address				Total
	No		Yes		
	n	Row%	n	Row%	
Design characteristics					
Address units (apartments) approached					
1	1,385	54.6	1,153	45.4	2,538
2–3	1,183	57.8	863	42.2	1,513
4+	180	34.4	344	65.7	1,057
χ2 P-value		<.001			
Original target sample					
Yes	555	25.3	1,639	74.7	2,194
No	2,193	75.3	721	24.7	2,914
χ2 P-value		<.001			
Interviewer characteristics					
Predominant interviewer age (majority; years)					
<30	664	73.5	239	26.5	903
30 to <40	899	60.4	589	39.6	1,488
40 to <50	247	35.5	448	64.5	695
50 to <60	401	41.4	567	58.6	968
60+	537	51.0	517	49.1	1,054
χ2 P-value		<.001			
Predominant interviewer sex (majority)					
Female	1,598	51.7	1,492	48.3	3,090
Male	1,003	55.3	810	44.7	1,813
No majority	147	71.7	58	28.3	205
χ2 P-value		<.001			
Predominant interviewer R/E (majority)					
African American	1,609	51.4	1,524	48.6	3,133
Non-Hispanic White	165	37.3	277	62.7	442
Hispanic	835	64.6	457	35.4	1,292
Asian	10	19.6	41	80.4	51
No majority	129	67.9	61	32.1	190
χ2 P-value		<.001			
Household characteristics					
Household age (mean; years)					
<30	<10	~	<10	~	17
30 to <40	150	31.5	326	68.5	476
40 to <50	719	52.6	647	47.4	1,366
50 to <60	640	45.5	768	54.6	1,408

Address level characteristics	1+ participant enrolled at address				
	No		Yes		Total
	n	Row%	n	Row%	
60+	1,231	66.9	610	33.1	1,841
χ2 P-value		<.001			
Household race/ ethnicity (majority)					
African American	1,227	42.0	1,696	58.0	2,923
Non-Hispanic White	796	71.4	319	28.6	1,115
Hispanic	472	66.9	234	33.1	706
Asian	63	91.3	<10	8.7	69
Other/unknown	19	39.6	29	60.4	48
No majority	171	69.2	76	30.8	247
χ2 P-value		<.001			
Interviewer and household concordant on race/ethnicity (majority)					
No	1,093	55.9	861	44.1	1,954
Yes	1,366	50.0	1,367	50.0	2,733
No majority	289	68.7	132	31.4	421
χ2 P-value		<.001			
Neighborhood characteristics					
Neighborhood socioeconomic status (SES)°					
Average	1,884	58.6	1,333	41.4	3,217
Low	710	42.7	954	57.3	1,664
High	154	67.8	73	32.2	227
χ2 P-value		<.001			
Total	2,748	53.8	2,360	46.2	5,108

°Categories defined as cores or clusters of similar census tracts in geographic and feature space by Local G* statistic at the Chicago area level using 999 permutations test with pseudo P-value <0.05 and queen contiguity, using 2010 Census tract boundaries

~Statistic suppressed due to cell frequency <10

Table 2:

logistic regression analyses with outcome of bio-specimen research participation within addresses (yes/no), among addresses approached in Chicago study sample, adjusting for design, interviewer, household, and other neighborhood-level factors

Characteristics of addresses	1+ participant enrolled at address		OR	(95% CI)
	No	Yes		
Design characteristics				
Address units (apartments) approached				
1	1,385	1,153	1.00	(Reference)
2–3	1,183	863	1.25	(1.06 – 1.46)
4+	180	344	3.16	(2.46 – 4.05
Sample				
Original	2,193	721	1.00	(Reference)
Snowball	555	1,639	10.3	(8.78 – 12.0)
Interviewer characteristics				
Predominant interviewer age (majority; years)	[continuous]		1.02	(1.02 – 1.03)
Predominant interviewer sex (majority)				
Male	1,003	810	1.00	(Reference)
Female	1,598	1,492	0.85	(0.71 – 1.02)
No majority	147	58	0.66	(0.41 – 1.04)
Predominant interviewer R/E (majority)				
Non-Hispanic White	165	277	1.00	(Reference)
African American	1,609	1,524	0.79	(0.56 – 1.11)
Hispanic	835	457	1.11	(0.77 – 1.61)
Asian	10	41	9.89	(4.10 – 23.9)
No majority	129	61	2.69	(0.72 – 9.97)
Household characteristics				
Predominant household age (mean; years)	[continuous]		0.99	(0.99 – 1.00)
Household race/ ethnicity (majority)				
Non-Hispanic White	796	319	1.00	(Reference)
African American	1,227	1,696	2.19	(1.75 – 2.74)
Hispanic	63	<10	0.84	(0.65 – 1.09)
Asian	472	234	0.23	(0.09 – 0.58)
Other/unknown	19	29	2.43	(1.18 – 5.00)
No majority	171	76	2.86	(0.82 – 9.94)
Interviewer and household concordant on R/E (majority)				
No	155	438	1.00	(Reference)
Yes	1,039	1,212	0.79	(0.65 – 0.97)
No majority for household and/or interviewer	33	46	0.24	(0.07 – 0.87)
Neighborhood characteristics				
Neighborhood SES°				
Average	1,884	1,333	1.00	(Reference)

Characteristics of addresses	1+ participant enrolled at address		OR	(95% CI)
	No	Yes		
Low	710	954	2.23	(1.91 – 2.60)
High	154	73	0.30	(0.21 – 0.42)
Total	2,748	2,360		

[a] Categories defined as cores or clusters of similar census tracts in geographic and feature space by Local G* statistic at the Chicago area level using 999 permutations test with pseudo P-value <0.05 and queen contiguity, using 2010 Census tract boundaries

Table 3:

logistic regression analyses for neighborhood characteristics with outcome of bio-specimen research participation within addresses (yes/no), among addresses approached in Chicago study sample, adjusting for design, interviewer, household, and other neighborhood-level factors - jointly stratified by sample type (original and/ or snowball) and address type (predominantly AA households and/ or not predominantly AA households) [•]

Characteristics of addresses	Original and snowball sample				Original sample				Snowball sample			
	1+ participant within address		OR	(95% CI)	1+ participant within address		OR	(95% CI)	1+ participant within address		OR	(95% CI)
	No	Yes			No	Yes			No	Yes		
All addresses												
Neighborhood SES [°]												
Average	1,884	1,333	1.00	(Reference)	1,513	412	1.00	(Reference)	371	921	1.00	(Reference)
High	154	73	0.30	(0.21 – 0.42)	39	15	0.75	(0.39 – 1.47)	115	58	0.47	(0.30 – 0.73)
Low	710	954	2.23	(1.91 – 2.60)	641	294	1.61	(1.30 – 1.99)	69	660	4.32	(3.20 – 5.82)
Total	2,748	2360	5,108		2,193	721	2,194		555	1,639	2,914	
Predominantly AA addresses												
Neighborhood SES [°]												
Average	1,038	903	1.00	(Reference)	774	171	1.00	Reference	264	732	1.00	(Reference)
High	<10	26	7.28	(0.96 – 55.1)	-	-		~	<10	26	7.93	(1.03 – 61.1)
Low	188	767	4.13	(3.30 – 5.16)	153	190	3.06	(2.20 – 4.24)	35	577	6.78	(4.59 – 10.0)
Total	1,227	1,696	2,923		927	361	1,288		300	1,335	1,635	
Excluding predominantly AA addresses												
Neighborhood SES [°]												
Average	846	430	1.00	(Reference)	739	241	1.00	(Reference)	107	189	1.00	(Reference)
High	153	47	0.20	(0.14 – 0.31)	39	15	0.79	(0.58 – 1.07)	114	32	0.20	(0.11 – 0.36)
Low	522	187	1.02	(0.81 – 1.29)	488	104	0.80	(0.41 – 1.55)	34	83	1.64	(0.94 – 2.85)
Total	1,521	664	2,185		1,266	360	1,626		255	304	559	

[•] Interviewer-level factors: interviewer with the majority approaches at the address unit level (age (continuous), sex, race/ ethnicity). Household-level factors: average potential participant characteristics at the address unit level (mean age (continuous), predominant race/ ethnicity (categorical) in unstratified models, majority interviewer-household concordant on R/E (yes, no, no majority for household and/or interviewer) in unstratified models). Design factors: potential participants within addresses (ordinal) and whether the address was in the original target sample from the vendor contact list (yes/no), in unstratified models.

[°] Categories defined as cores or clusters of similar census tracts in geographic and feature space by Local G* statistic at the Chicago area level using 999 permutations test with pseudo P-value <0.05 and queen contiguity

°Categories defined as cores or clusters of similar census tracts in geographic and feature space by Local G* statistic at the Chicago area level using 999 permutations test with pseudo P-value <0.05 and queen contiguity, using 2010 Census tract boundaries

~Statistic suppressed due to cell frequency <10

Table 4

Logistic regression analyses for neighborhood haracteristics with outcome of bio-specimen research participation within addresses (yes/no), among 5,108 addresses approached in Chicago study sample. Models are adjusted for design, interviewer, household, and other neighborhood-level factors [•]

Characteristic	1+ participant within the address		OR	(95% CI)
	No	Yes		
Local G at Chicago level using 99,999permutations; pseudo P-value <0.0001*				
Neighborhood socioeconomic status (SES)[1]				
Average	2,484	1,944	1.00	(Reference)
Low	264	407	1.79	(1.46–2.20)
High	-	<20		~
Local G at Chicago level using 999permutations; pseudo P-value <0.05*				
Neighborhood SES [2]				
Average	1,884	1,333	1.00	(Reference)
Low	710	954	2.23	(1.91–2.60)
High	154	73	0.30	(0.21–0.42)
Quintiles at the Chicago level				
Neighborhood SES [3]				
Quintile 1 (Low)	264	424	1.00	(Reference)
Q2	578	771	1.23	(0.96–1.57)
Q3	1,617	933	0.38	(0.30–0.47)
Q4	289	217	0.51	(0.37–0.69)
Q5 (High)	-	<20		~
Total	2,748	2,360		

[•] Interviewer-level factors: interviewer with the majority approaches at the address unit level (age (continuous), sex, race/ ethnicity). Household-level factors: average potential participant characteristics at the address unit level (mean age (continuous), predominant race/ ethnicity (categorical), majority interviewer-household concordant on R/E (yes, no, no majority for household and/or interviewer)). Design factors: potential participants within addresses (continuous) and whether the address was in the original target sample from the vendor contact list (yes/no), using 2010 Census tract boundaries

[1.] Categories defined as cores or clusters of similar census tracts in geographic and feature space by Local G* statistic at the Chicago area level using 999 permutations test with pseudo P-value <0.05 and queen contiguity, using 2010 Census tract boundaries

[2.] Categories defined as cores or clusters of similar census tracts in geographic and feature space by Local G* statistic at the Chicago area level using 999 permutations test with pseudo P-value <0.05 and queen contiguity, using 2010 Census tract boundaries

[3.] Categories defined as quintiles of nSES at Chicago area level, using 2010 Census tract boundaries

[~] Statistic suppressed due to cell frequency <20