BMJ Open

# Cohort Profile: The ChicagO Multiethnic Prevention and Surveillance Study (COMPASS)



| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2020-038481 |
| Article Type: | Cohort profile |
| Date Submitted by the Author: | 23-Mar-2020 |
| Complete List of Authors: | Aschebrook-Kilfoy, Briseis; University of Chicago, Department of Public Health Science; The University of Chicago, Public Health Sciences<br>Kibriya, Muhammad; The University of Chicago, Public Health Sciences<br>Jasmine, Farzana ; The University of Chicago, Public Health Sciences<br>Stepniak, Liz; The University of Chicago, Public Health Sciences<br>Gopalakrishnan, Rajan; The University of Chicago, Public Health Sciences<br>Craver, Andrew; The University of Chicago, Public Health Sciences<br>Zakin, Paul; The University of Chicago, Public Health Sciences<br>Tasmin, Saira; The University of Chicago, Public Health Sciences<br>Kim, Karen; University of Chicago<br>Goss, Kathleen; The University of Chicago, Public Health Sciences<br>List, Marcy; The University of Chicago, Public Health Sciences<br>LeBeau, Michelle; The University of Chicago, Public Health Sciences<br>Ahsan, Habibul; University of Chicago, Health Studies |
| Keywords: | EPIDEMIOLOGY, PREVENTIVE MEDICINE, PUBLIC HEALTH |
| | |

*I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.*

*The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.*

*Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.*

# Cohort Profile: The ChicagO Multiethnic Prevention and Surveillance Study (COMPASS)

Briseis Aschebrook-Kilfoy[1,2], Muhammad G. Kibriya[1,2], Farzana Jasmine[1,2], Liz Stepniak[1,2], Rajan Gopalakrishnan[3], Andrew Craver[2], Paul Zakin[2], Saira Tasmin[1,2], Karen Kim[3], Kathleen H. Goss[3], Marcy List[3], Michelle LeBeau[3], Habibul Ahsan[1,2]

[1] Department of Public Health Sciences, University of Chicago, Chicago, IL, USA.
[2] Institute for Population and Precision Health, University of Chicago, Chicago, IL, USA.
[3] Comprehensive Cancer Center, University of Chicago, Chicago, IL, USA.
[4] Department of Medicine, University of Chicago, Chicago, IL, USA.

**Address for correspondence and reprints:**
Dr. Briseis Aschebrook-Kilfoy, Institute for Population and Precision Health, The University of Chicago, 5841 S Maryland Ave, Chicago IL 60636; email: brisa@uchicago.edu

**Running Head**: Cohort Profile for COMPASS

## ABSTRACT

**Purpose:** The Chicago Multiethnic Prevention and Surveillance Study or **"**COMPASS" is a population-based cohort study with a goal to examine the risk and determinants of cancer and chronic disease. COMPASS aims to address factors causing and/or exacerbating health disparities using a precision health approach by recruiting diverse participants in Chicago, with an emphasis on those historically underrepresented in biomedical research.

**Participants:** Nearly 8,000 participants have been recruited from 72 of the 77 Chicago Community Areas. Enrollment entails the completion of an hour-long survey, consenting for past and future medical records from all sources, the collection of clinical and physical measurement data, and the on-site collection of biological samples including blood, urine, and saliva. Indoor air monitoring data and stool samples are being collected from a subset of participants. Upon collection, all biological samples are processed and aliquoted within 24 hours before long-term storage and subsequent analysis.

**Findings to Date:** The cohort reported an average age of 53.7 years, while 80.5% identified as African American, 5.7% as Hispanic, and 47.8% as male. Over 50% reported earning less than $15,000 yearly, 35% were obese, and 47.8% were current smokers. Moreover, 38% self-reported having had a diagnosis of hypertension, while 66.4% were measured as hypertensive at enrollment.

**Future Plans:** We plan to expand recruitment up to 100,000 participants from the Chicago metropolitan area in the next decade using a hybrid community and clinic-based recruitment framework that incorporates data collection through mobile medical units. Follow-up data collection from current cohort members will include serial samples, as well as longitudinal health, lifestyle, and behavioral assessment. We will supplement self-reported data with electronic medical records (EMRs), expand the collection of biometrics and biosamples to facilitate increasing digital epidemiologic study designs, and link to state and/or national level databases to ascertain outcomes. The results and findings will inform potential opportunities for precision disease prevention and mitigation in Chicago and other urban areas with a diverse population.


**Strengths and limitations of this study:**

- COMPASS has the unique capacity to recruit individuals previously underrepresented in biomedical research.
- Thus far, there is limited variation in socioeconomic strata within race and ethnicity categories as we have oversampled low income households.
- The prospective cohort design reduces possible recall bias and selection bias. In addition, the geographically defined cohort design allows for the exploration of an array of environmental and social exposures potentially affecting health outcomes.
- Participants will continue to be recruited and enrolled at medical centers and aboard mobile units, which will allow for the efficient collection of serial biospecimens and tumor tissue as well as comprehensive health history data from EMRs. We will also be able to capitalize on the latest advances in the understanding of disease etiology, technology for data capture, and analytic tools.

**Registration: NA**

## INTRODUCTION

Despite the tremendous advances in medicine and public health in recent decades, many areas of the United States have seen a growing gap in health outcomes and health equity.[1,2] As one of the most racially segregated cities in the United States[3], Chicago reports alarming health disparities and an unequal burden of cancer and chronic risk and mortality, particularly on the South Side.[3] The South Side of Chicago encompasses roughly 60% of the city's land area. The South Side is one of the nation's largest contiguous urban African-American communities, and is home to a large percentage of the city's 814,500 African American residents.[4] The South Side also includes Chinatown and Bridgeport, the two neighborhoods with the highest concentrations of Asian Americans in Illinois, making Chicago the sixth largest Asian-American city in the United States. The most recent census also illustrated that Hispanics now account for nearly a third of the city's total population of 2.7 million.[4]

Chicago faces significant urban challenges, including concentrated poverty, violence and crime, poor housing and living standards, food deserts, and environmental exposures such as toxins in the air and water due to waste sites, landfills, and abandoned industrial buildings. [5-17] Despite advances in cancer and chronic disease prevention, screening, diagnosis, and treatment that have improved chronic disease risk and outcomes for many Americans, disparities in cancer, cardiovascular disease, diabetes, asthma, and maternal child health persist on Chicago's South Side. Disease related mortality rates continue to climb as well. Health disparities are increasingly understood to be a substantial burden on society in terms of healthcare costs, lost productivity, and general societal wellbeing.[18] According to the National Institute on Minority and Health Disparities, health disparities are multidimensional, complex phenomena that result from the interaction of multiple contributing factors over time.[19] These disparities create trajectories of health vulnerability for certain groups of people that vary from the majority norm.[20] Therefore, there is an urgent need to better understand the influence of factors that drive health disparities, such as health care coverage and low socioeconomic status. It is also important to identify the still-unknown factors behind health disparities in order to mitigate these trends and reduce the unequal burden of disease.

The University of Chicago is home to the Institute for Population and Precision Health (IPPH), as well as an NCI-designated Comprehensive Cancer Center. Given its location in the largest contiguous African-American community in the United States, its extensive community engagement infrastructure and cohort building experience, and the availability of innovative and well-established population science resources, we are in a unique position to focus efforts on understudied populations and to study health disparities on Chicago's South Side. To meet the need for a rigorous study of cancer and chronic disease and disparities in the Chicago metropolitan area, we established the Chicago Multiethnic Prevention and Surveillance Study (COMPASS) in 2013. COMPASS is a population-based longitudinal study of cancer and chronic disease risk and progression with the overall goal of recruiting a diverse study population in Chicago. As those with lowest socioeconomic status and possibly the riskiest health behaviors have historically been the least likely to enroll in biomedical and biobanking research, we have implemented a study design in which recruitment occurs in both the community and clinic, and involves extensive engagement activities. Because of the growing recognition of health disparities in Chicago, COMPASS data collection efforts have focused on potentially important factors for minority health and have emphasized recruitment from minority populations.

## COHORT DESCRIPTION

### Purpose

The aims and primary scientific goals of COMPASS are to focus on the discovery of cancer and chronic disease in general, with the perspective that social, economic, racial, and political marginalization are often primary contributing factors associated with health disparities.[21-31] The COMPASS scientific focus areas are illustrated in **Figure 1**. The primary COMPASS aims driving the design are as follows:

1. To discover novel genetic and epigenetic factors related to the risk of cancers among the urban population through comprehensive yet cost-efficient genome-wide screening of germline DNA for Single Nucleotide Polymorphism (SNP) and methylation variants. This will be accomplished by efficient typing of case-cohort samples using custom gene-chips based on all novel genetic and epigenetic variants identified through next-generation whole genome deep sequencing and methylation profiling of cohort samples supplemented with previously identified and validated variants from 1000 Genome Project. Similar screening of tumor tissues of cancer cases will also be done to distinguish between inherent and acquired genomic alterations.

2. To elucidate and characterize cancer risk defining germline genetic/epigenetic discoveries, we will prospectively screen baseline serum/plasma and RNA samples from representative samples of matched sub-cohorts of all four risk-groups (carriers with cancer, non-carriers with cancer, carriers without cancer and non-carriers without cancer) for proteins/hormones/enzymes and mRNAs/ncRNAs, respectively. Similar screening of tumor tissues of the two cancer groups will also be done to distinguish between inherent and acquired alterations to further elucidate the biology and identification of novel preventive and therapeutic approaches.

3. To integrate the complex interactions of the genomic determinants with non-genomic environmental, nutritional, behavioral, and psychosocial determinants to develop a more complete understanding of cancer risk and causation. This includes consideration of the microbiome. This will be accomplished through tailored statistical and causal modeling approaches developed by statisticians and epidemiological methodologists at the University of Chicago.

4. To discover novel environmental (including air, water and household/workplace pollutants) and dietary (including food additives/toxicants and deleterious food items/nutrients) carcinogens that predispose urban populations to a higher risk for various cancers and chronic disease. These discoveries will be made through interdisciplinary efforts involving environmental sensors using nanotechnology and molecular dosimetry for high-throughput screening for biomarkers of exposure to carcinogens, coupled with sophisticated environmental, statistical, and systems modeling.

5. To identify previously unrecognized factors related to urban lifestyle and behavior (including stress, energy-balance/obesity, sleep, and aging) that increase the risk of various cancers and chronic diseases. These investigations will be conducted through integrated

application of behavioral and psychosocial assessment tools, biomarkers of stress/aging/obesity, and novel multi-level and agent-based statistical modeling.

6. To consider new approaches to prevention that are personalized to one's specific risk profile based on their host, environmental/behavioral, and genomic profiles by exploiting new knowledge about the risk-driving inherent genomic variants, their interactions with environmental, host, behavioral and psychosocial dynamics, and the downstream molecular alterations in serum/plasma as well as tumor tissues that potentially mediate disease risk. COMPASS will function as a platform for focused molecular, mechanistic, and clinical research.

7. To specifically integrate the diverse data sources named here to investigate opportunities for prevention and early diagnosis, with an emphasis on the mitigation of health disparities.

The recruitment of this diverse cohort presents important opportunities for scientific discovery, disease mitigation, and the alleviation of the dramatic disparities observed in Chicago and other urban centers across the United States. According to the National Cancer Institute, overcoming cancer health disparities is one of the best opportunities for lessening the burden of cancer.[32] COMPASS data and biosample collection leverages the advanced population science infrastructure at the University of Chicago and preserves the opportunity to utilize advanced analytic methods and technologies in the future. In the short-term, much will be learned about the health of Chicago communities. In the long-term, the investigation of the aims above will offer insight into the relative contribution of key health drivers to decrease health inequalities, understand disease risk more generally, and elucidate opportunities for prevention and mitigation.

Here, we present this prospective multiethnic cohort study, including a description of the cohort enrolled to date. We also outline the research resources offered by COMPASS and implications for concurrent ancillary studies, as well as future observational and intervention opportunities within the cohort.

**Study Population**

As of 2019, there were an estimated 2.7M people and 1.1M households within the city limits of Chicago. According to a 2018 estimate, the Chicago Metropolitan area is the third largest in the United States, with 9.9M living in city and suburbs combined. More than half the population of the state of Illinois lives in the Chicago metropolitan area. The 2010 United States Census shows the population density of the city itself at 11,842 people per square mile, making it one of the nation's most densely populated cities. The median income for a household in the city as of 2018 was $55,198, with 19.5% of persons in poverty. The goal of COMPASS is to recruit 100,000 participants, oversampling the majority minority population (i.e. more than 45% African Americans and 20% Hispanics) [4, 33].

COMPASS is a longitudinal cohort study including persons recruited using a population-based approach, a community-based recruitment approach, and a hospital/clinic-based recruitment approach. The multiple recruitment modalities were considered in the early phase of this project

in order to capitalize on the strengths and challenges of our Chicago context. As such, here we describe participant recruitment methods using a targeted or community partner approach, as well as at University of Chicago Medicine and on mobile medical units. Parallel data are collected from both community and clinic source populations. These data consist of survey responses, biospecimens including blood, urine, saliva, and stool, clinical measures, electronic health records, and environmental samples from the home.

**Targeted Community-Based Recruitment**

To explore the feasibility of a population-representative sample for COMPASS, we implemented a recruitment plan to target potentially eligible participants using a two-stage cluster sampling approach, first randomizing at the census tract level and then identifying potentially eligible persons from within the tract using commercially available address lists. Field staff would then knock on doors and invite the targeted individuals to participate. Specifically, "cluster sampling" involves dividing the specific population of interest into geographically distinct groups or clusters. Census tract was the primary sampling unit (PSU) for COMPASS, as census tracts are designed to be homogenous with respect to population characteristics. In the City of Chicago, there are 798 census tracts; 169 are identified as majority Hispanic, 239 as majority White, 290 as majority African American, and 6 as majority Asian. We identified 120 (15% of Chicago) study areas of interest using census tracts that would confer the racial/ethnic distribution desired for recruitment. We then used STATA module *gsample*[34] to randomly select approximately 120 tracts (40 original census tracts to which 80 tracts have been added) and addresses within blocks. The number of tracts selected was based on an assumption of a 15% response rate and the assumption that 46% of residents would be 35 years or older. By sampling at the census tract level, study interviewers would recruit from limited geographic areas where many households could be contacted efficiently, minimizing travel time and cost. In particular, we used a probability proportional to size sampling design as this design reduces sampling errors if characters of interest are related to population size. In the first stage, census tracts were randomly selected with probability proportional to the population of census blocks. In the second stage, we randomly chose households within the tract. As two locations within the same census block do not each contribute completely independent information (this is known as the "intra-cluster correlation" or ICC), we selected more clusters rather than more points within any cluster in order to improve precision while maximizing study efficiency and minimizing cost.

The commercially available address lists used for sampling are available from the United States Postal Service's Address Management System and demographic data is available for addresses from commercial sources. The address lists used for COMPASS contained all residential addresses along a mail delivery route, and excluded commercial addresses as well as PO Boxes. The commercial list of addresses provided data for households in the targeted census tracts on age and race and a telephone number (if available). This allowed us to weight the household based on the probability that they would meet eligibility criteria and would reflect the desired racial/ethnic distribution. We excluded all households that did not appear to have a resident 35 or older.

***Engagement for Targeted Community-Based Recruitment***

We used the list of addresses in the chosen census blocks to mail a postcard introducing the study and outlining the inclusion/exclusion criteria. The postcard explained that the resident has been selected to participate in COMPASS, what participation would involve, the time commitment, study eligibility criteria, compensation for participation, and that multiple eligible household members are welcome to participate. The postcard also provided a time frame in which an interviewer would be requesting participation.

After the postcards were mailed to targeted households, the field interviewers went door-to-door on the targeted study census blocks based on a predetermined schedule to request participation. If a subject did not respond, the interviewer revisited the house within the following two weeks. If the interviewer received no response at the third visit, the household would be documented as a non-responder. By "dropping in" on all targeted participant homes, a goal was to recruit as least some portion of those individuals at greatest risk of adverse health outcomes.

If the eligible person/s agreed to participate, they enrolled in the study, completed the interview, and provided biospecimens and physical assessment data onsite in their home.

The results of the door-to-door recruitment response are presented by Press et al.[35] Taken together, the team achieved a much higher response rate in African-American communities than anticipated.

***Untargeted Community-based Recruitment Using Mobile Medical Units***

Community-based COMPASS recruitment is supported by two large mobile medical units (**Figure 2**) and the community engagement infrastructure including partner community-based organizations. The mobile medical units have two fully equipped exam rooms with phlebotomy capabilities, a bathroom to facilitate urine collection, are WiFi-enabled, and include CLIA[36] certified labs with processing/storage capabilities. All study procedures are can be accommodated on-board the mobile medical units.

***Untargeted community-based engagement***

Untargeted community-based recruitment typically occurs both in conjunction with community partners, as well as without partners in communities targeted for participation. Examples of community partners include faith-based organizations, non-profit organizations, retail stores, civic groups, Chicago Park District facilities, and Chicago Public Schools. These partners publicize COMPASS enrollment events and invite participation on days scheduled for mobile unit presence.

**Hospital/Clinic-based Recruitment**

***Study population***

In conjunction with ongoing epidemiologic efforts at University of Chicago Medicine (UCM) administered by the UChicago IPPH, COMPASS participants are recruited in a UCM research clinic. All UCM patients and their guests ages 35 and older are eligible for participation in COMPASS. In the clinic context, eligibility is not restricted based on place of residence. As such, the eligible study population includes ~60,000 UCM patients per year, over 60% of whom are

African American. However, this population differs from the community-based sampling in some key ways, and will reflect a lower-risk population from the perspective of cancer health disparities.

***Hospital/Clinic-based engagement***

In-person outreach in the medical center is conducted by trained research assistants who approach persons in waiting areas and inform them about COMPASS. Those who meet the eligibility criteria are invited to participate. Digital outreach will be expanded in the near future to use email and text introduction. Participants will be able to self-schedule an enrollment and data collection visit at our IPPH research clinic by following a link in the email or text and completing a REDCap scheduling form.

## DATA COLLECTION

### Overview

Daily enrollment of COMPASS participants was initiated in May of 2013.

***Enrollment of Cohort Members***

To be eligible for participation in COMPASS, participants had to meet the following criteria:

1. Residence in the designated geographic area (census tract)
2. Age 35 and above at time of contact
3. Ability to complete the consent and interview in English or Spanish
4. Willingness to provide blood, urine, and saliva samples
5. Provision of a social security number (for outcomes follow-up)

The COMPASS interview has been revised and is currently divided into three online modules. These modules are administered via the Qualtrics (Qualtrics, Provo, UT) survey platform and have been optimized for mobile devices. During their visit to the UChicago research clinic, one of the mobile data collection units, or during community recruitment, participants complete the consent process and survey module #1 using their own mobile device or an IPPH-provided Apple iPad. Survey modules #2 and #3 can be completed onsite during the same visit, or the participant can elect to receive the survey links via email or SMS to complete offsite. COMPASS staff are available onsite to help participants navigate the online consent and survey modules, and participants who choose to complete survey modules #2 and #3 offsite are encouraged to reach out to staff via phone or email with any questions or to receive technical assistance.

Once the consent and 60-80 minute survey is completed, the following clinical measurements are ascertained: anthropometry, blood pressure, and other hemodynamic measures. Study staff measure weight and body fat percentage using a scale and height, hip, and waist circumference using a measuring device, and they then measure blood pressure and other hemodynamic parameters using a device that monitors pulse (i.e., DynaPulse). An overview of the data collection

is provided in **Table 1**. In **Table 2**, the key topics and measurements in the COMPASS questionnaire are outlined.

The purpose of COMPASS anthropometry is to collect body measurements using standardized examination procedures and calibrated equipment. To ensure the collection of high-quality data, COMPASS staff members are trained to follow standardized examination protocols, calibrate equipment according to a prescribed schedule and method, and precisely measure and record the survey data. Portable digital scales are used in the field on on the mobile units by interviewers to measure weight, as well as percentage body fat and water. Body length measurements are made with waist circumference tape and retractable steel measurement tape.

The DynaPulse Pathway System measures central systolic, diastolic, and mean arterial blood pressures, and the data are also transmitted to the DynaPulse Analysis Center online for an array of hemodynamic profiles/values. Briefly, DynaPulse, is a cuff-sphygmomanometry-based method, which applies Pulse Dynamics waveform analysis principle and has demonstrated and validated its capability of deriving cardiac output (CO) simultaneously with BP, MAP, SVR, systemic vascular compliance (SVC), brachial artery compliance, distensibility and resistance (BAC, BAD and BAR), LV(dP/dt)Max, and other hemodynamic parameters. These hemodynamic values, collectured for a subset participants, obtained simultaneously allow physicians to correlate the dynamic changes of each parameter to evaluate the physiological conditions of a patient's circulatory system.

Following survey data collection and clinical measurements, COMPASS staff members draws a blood sample. For each participant, the blood sample is collected in cryogenic barcode-labeled tubes as follows: (a) Lavender top EDTA vacutainer – 10 ml; (b) Red and black top SST vacutainer – 10 ml; (c) Green top Lithium Heparin vacutainer – 10 ml; (d) Blue top Trace Element K2 EDTA vacutainer – 10 ml; (e) Gold top SST vacutainer – 3.5 ml; (f) Red top serum vacutainer – 5 ml; and (g) cfDNA collection tube – 10 ml. The vacutainers are inverted per protocol for proper mixing, ensuring a source of blood material for hematological, biochemical, hormonal, serological, and other special tests that require high-quality DNA or RNA.

Study participants are asked to provide a buccal cell sample, using an Oragene DNA self-collection kit, as well as a urine sample.

Participants are eligible to receive three $25 Visa gift cards in recognition of their time. The first gift card is distributed when the participant completes the consent process, survey module #1, and contributes biospecimens and physical measurement data. The second gift card is given to the participant upon completion of survey modules #2 and #3. Those who participate in optional home air quality monitoring receive the third gift card when IPPH staff retrieve the air quality devices from their home and the participant completes a short survey on characteristics of their home and environmental exposures.

Study participants are also asked to provide an optional at-home stool sample using a stool sample collection kit. This kit includes instructions to transfer a small swab from toilet paper to a vial. The participant then places the sample vial into a prepaid and addressed envelope and mails it to IPPH.

Biospecimens are stored in medical-grade refrigerators at the IPPH research clinic and aboard the mobile medical units, and are transferred to our laboratory on the University of Chicago campus within 4 to 6 hours of collection.

***Biorepository***

At the UChicago IPPH laboratories, the EDTA-mixed blood sample is centrifuged to separate out the plasma sample and aliquoted in multiple 0.5-ml screw-capped plastic tubes for preservation (to avoid repeated thaw and freeze cycles in future). All components of blood are stored at -80C and urine samples are stored at -20°C with proper inventory. Freezers are connected to emergency power supply line (red line) and a central alarm system. The saliva samples collected using the Oragene DNA self-collection kit are processed by warming in a hot water bath and stored at -20°C in the IPPH laboratory. Stool samples are snap frozen and stored at -20°C.

***Environmental Sample Collection***

Study participants may be asked to allow IPPH staff to collect air quality data from their home using one or more air quality sensors. The sensors provide real-time measures of the presence of particulate matter PM2.5 and PM10 in addition to other contaminants such as CO, VOCs, and radon. Air quality data are collected using the COMPASS best practices protocol. Study participants are provided directions to not disturb the sensors and contact the study team with issues. The IPPH staff member schedules a follow-up visit to retrieve the sensors.

Indoor air quality data will be integrated with ambient levels from the Array of Things (AoT) initiative (https://arrayofthings.github.io/), a joint effort by UChicago and Argonne National Laboratories and supported by the City of Chicago. AoT is currently collecting environmental information on air pollution in addition to light intensity, precipitation, air quality, heat, sound volume, and foot traffic through decorative sensors on traffic poles throughout the City of Chicago. The result is a system that collects data about the city at the micro level, including fine-grained pollution levels in different neighborhoods. The diversity of pollutants measured, the number of monitors that will be installed across the city (>500), and the availability of a complimentary sensor for indoor measures, provides an unprecedented opportunity to more accurately assess exposure to air pollution and understand the relative impact of air pollution in Chicago. AoT is placing devices in communities participating in COMPASS. This presents a significant opportunity to understand the role of air pollution in disease etiology, as well as the potential that findings could inform interventions targeting at-risk Chicago communities.

Details on air quality in-home measurements, ambient concentration modeling, and total exposure estimation can be found in accompanying manuscripts[37].

Interested study participants may be referred to the Chicago Water Quality Study, an initiative of the City of Chicago Department of Water Management. Under this initiative, a water quality kit consisting of sample collection bottles and instructions is mailed to the participant. The participant collects and documents samples per instructions and contacts the Department of Water Management for sample pick up. During the consent process, participants are asked to allow COMPASS to retrieve water quality analysis results from the Department of Water Management.

***Data management***
A secure HIPAA-compliant database was created by the Information Technology Group at Dartmouth for the first four years of the project. Going forward, all study data is being collected and stored in REDCap (Research Electronic Data Capture) and/or Qualtrics (Qualtrics, Provo, UT). In addition, the University of Chicago Medicine Comprehensive Cancer Center (UCCCC) has created an address database to track the response rate as well as details of non-responders and those who opt not to participate. This database is maintained by the COMPASS informatics team.

***Tracking health outcomes***
All participants have provided their social security numbers so that their cancer and vital status can be tracked long-term through the Illinois State Cancer Registry, which is maintained by the Illinois Department of Public Health.

***Electronic Medical Records***
UChicago is a member of the Chicago Area Patient Centered Outcomes Research Network (CAPriCORN) - a platform that integrates EMR across the nine largest academic medical centers in Chicago. COMPASS participants provide consent for access to medical records, which will allow for the efficient collection of comprehensive health history data from EMRs of participants who receive care both at the University of Chicago Medical Centers and outside of our system.

***Follow-up data and biosample collection***
Previously enrolled participants will be re-contacted to schedule in-person visits to our IPPH research clinic and our mobile research units. They will be asked to sign an updated consent form and to complete an online follow-up survey in order to understand changes in health status, health behaviors, and lifestyle. This visit will also include collection of follow up biospecimens and physical measurements. In addition, following COMPASS notification that a participant has had a cancer diagnosis, either from the participants themselves or the Illinois State Cancer Registry, we submit a request to the diagnosing and treatment institutions for medical records such that we may ascertain more detailed tumor characteristic and treatment information.

***Patient and public involvement***
None of the participants was involved in the questionnaire design, biological measurements or outcome measures; they were likewise not involved in the design, recruitment and implementation of the study. Furthermore, all participants were informed of the use of the data for research in this study. There were no plans to disseminate the study results to participants. When genetic results become available, participants will be contacted and asked if they would like to receive them per the protocol/consent documents.

***Research Ethics Approval: Human Participants***
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 Institutional Review Board Committee A (approval IRB12-1660).

**Data Analyses**

Baseline data from the cohort were analyzed for tabular and descriptive statistics (**Table 3)**. Means and standard deviations were calculated for variables with a continuous distribution and proportions were calculated for variables with categorical distributions. Body mass index (BMI) was computed based on weight and height measured by the interviewers. All analyses were performed using STATA 15 (StataCorp, College Station, TX) and GeoDa for Windows[38].

## CHARACTERISTICS OF STUDY PARTICIPANTS RECRUITED TO DATE

Among the 7,728 participants enrolled in the study **(Table 3)**, 80.5% identified as African American, 5.7% as Hispanic, and 47.8% as male. In addition, over 50% reported earning less than $15,000 yearly, 15.2% reported working full-time, 35% were obese, 47.8% were current smokers, and 38% reported having hypertension. High school graduates comprised 19.8% of the cohort, and 17% of the participants were married. Consistent with findings by the Illinois Department of Public Health, the prevalence of diabetes was higher among females than males, and non-Whites compared to Whites.[39] The overall prevalence of Type 2 Diabetes in this study was 10.1%.

The maps in **Figure 3** show the prevalence of the specified variable in the 77 community areas in Chicago. Only community areas with at least 20 observations were included, resulting in 45 communities being analyzed. There is an alarming prevalence of high blood pressure, with the majority of the community areas sampled being classified as having stage 1 or 2 hypertension. In addition, the majority of community areas sample have a smoking prevalence of at least 50%.

## STRENGTHS AND LIMITATIONS

COMPASS efficiently recruits individuals previously underrepresented in biomedical research. The pilot analysis of the first 8,000+ individuals demonstrates our capacity to recruit such a diverse, urban population. Our UCM research space coupled with our mobile research units allow our team to collect and process baseline and follow-up data on campus and in Chicago communities.

Another strength of the study is the prospective cohort design, which reduces possible recall bias and selection bias. Furthermore, the geographically defined cohort design offers a number of advantages for the study of health disparities. The principal reason that we are recruiting a geographic cohort (rather than an occupational cohort, for example) is because it allows for the exploration of an array of exposures potentially affecting health outcomes. In many other cohort designs, subjects are recruited given their unique exposure, such as nurses, farmers, miners, etc. However, the factors that disproportionately affect African Americans' health status and outcomes in Chicago are largely unknown or under-characterized. Similarly, the relevant risk factors in Hispanic and Asian populations are likely unique and remain under-characterized.

COMPASS is also involved in community engagement efforts to address the needs of the South Side community, through collaboration with the UCCCC Office of Community Engagement and Cancer Health Equity (OCECHE). The overall goals of the OCECHE are to: a) establish strong partnerships with underserved communities and organizations; b) support community and local capacity building; and c) establish sustainable educational programs. To ensure both academic and community relevance and to help shape future programming and direction, the OCECHE has both

a Community and Internal Advisory Board comprising key community and institutional stakeholders.

In addition, the University of Chicago is a member of the Chicago Area Patient Centered Outcomes Research Network (CAPriCORN) - a platform that integrates EMR across the nine largest academic medical centers in Chicago. CAPriCORN makes it possible for investigators at its partner institutions in Chicago to meet the growing need for the integration of EMRs in epidemiologic research, including comprehensive multi-institution tracking/follow-up, such that comprehensive EMRs are available for participants who receive care at multiple institutions. The newly recruited participants will be enrolled through integrated health care systems, such as UChicago, which will allow for the efficient collection of serial biospecimens and tumor tissue as well as comprehensive health history data from EMRs. They will be followed longitudinally to study a broad range of risk factors, early markers, and outcomes. We will be able to capitalize on the latest advances in the understanding of cancer etiology, technology for data capture, and analytic tools.

The collection of environmental data in COMPASS to understand the role of the environment in health disparities is supported by a new NIEHS-sponsored environmental health center in the region- ChicAgo Center for Health and Environment (CACHET). CACHET was established to address EHS-related health disparities as the first center of its kind in the greater Chicago tri-state area (i.e., Illinois, Indiana, and Wisconsin). We have developed an environmental monitoring team (used for many NIEHS P30 investigations), infrastructure, and expertise to explore the impact of urban environmental exposures that are embedded in COMPASS data collection. The CACHET mission is to elucidate the biological and social underpinnings between relevant urban environmental exposures and human disease and translate the findings to reduce health inequities within our communities. COMPASS is a key cohort synergized by the CACHET Integrated Health Sciences Facility Core (IHSFC) at UChicago.

There are some limitations in our study. Thus far, there is limited variation in socioeconomic strata within race and ethnicity categories. Also, the longitudinal nature of the cohort makes it unlikely that we will be able to study cancer-related hypotheses for at least a decade. For rarer diseases, more people and time will be needed. In addition, the oversampling of African Americans and individuals with low socioeconomic status calls into question the representativeness of the study population. However, given the paucity of cohort studies focused on this population, COMPASS provides a unique window into the factors impacting health in Chicago. There is also an oversampling of individuals out of work; however, this is possibly also a strength of the study, as some of the highest risk populations not represented in previous works and studies included.

**Figure 1. COMPASS Scientific Focus Areas**

**Figure 2. An IPPH Mobile Unit Used for COMPASS Enrollment**

**Figure 3. Mapped Prevalence of COMPASS Self-Reported Obesity, Smoking, Type 2 Diabetes, and Hypertension by Chicago Community Area**

## DATA AVAILABILITY AND COLLABORATION

COMPASS will continue to collect a rich set of data on multiple exposure domains and health outcomes. For more information, refer to the web site: compass.uchicago.edu. Researchers interested in collaboration are invited to propose research questions based on the data available within COMPASS or to submit a request for additional data collection. Requests can be submitted electronically on the COMPASS website and will be reviewed by the COMPASS scientific board. The COMPASS study team is particularly interested in collaborations that will enhance research methods for this type of work, assess the impact of environmental exposure, highlight exposures of key significance in urban communities, and address health issues of concern in Chicago and other urban centers.

## CONTRIBUTORS

The study protocol was designed by HA, ML, BAK, ML, KG and MGK and BAK drafted the manuscript. BAK and LS oversaw field operations and KGM and FJ oversaw the biosample processing and biobanking. HA, BAK, and KK oversaw engagement. PZ, AC, and ST performed the data quality control and the statistical analysis. All authors participated in the study design, revised the article and approved the final version.

## COMPETING INTERESTS

None declared.

## FUNDING

This research was supported by funding from the University of Chicago Medicine Comprehensive Cancer Center and the University of Chicago Insisittue for Population and Precision Health.

## ACKNOWLEDGEMENTS

We thank the dedicated COMPASS field staff and community partners for their support of this work.

## AUTHOR AFFILIATIONS

[1] Department of Public Health Sciences, University of Chicago, Chicago, IL, USA.
[2] Institute for Population and Precision Health, University of Chicago, Chicago, IL, USA.
[3] Comprehensive Cancer Center, University of Chicago, Chicago, IL, USA.
[4] Department of Medicine, University of Chicago, Chicago, IL, USA.

## REFERENCES

1. Meyer P, Yoon P, Kaufmann R. Introduction: CDC Health Disparities and Inequalities Report- United States. *Centers for Disease Control and Prevention*. 2013;62:3-5.
2. Baciu A, Negussie Y, Geller A, *et al*. *Communities in Action: Pathways to Health Equity*. Washington DC: National Academies Press, 2017.

3. Orsi J, Margellos-Anast H, Whitman S. Black-White Health Disparities in the United States and Chicago: A 15-Year Progress Analysis. *Am J Public Health*. 2010;100:349–356.
4. United States Census Bureau. U.S. Census Bureau QuickFacts: Chicago city, Illinois. https://www.census.gov/quickfacts/chicagocityillinois. Published 2018. Accessed February 22, 2019.
5. Chicago Department of Public Health. Public Health Statistics – Selected public health indicators by Chicago community area. https://data.cityofchicago.org/Health-Human-Services/Public-Health-Statistics-Selected-public-health-in/iqnk-2tcu/data (Accessed January 23, 2020).
6. Chicago Police Department. Crimes – 2001 to present. https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present-Dashboard/5cd6-ry5g (Accessed January 27, 2020).
7. Kolak, Marynia, et al. Urban foodscape trends: disparities in healthy food access in Chicago, 2007–2014. *Health & Place.* 2018:52: 231-239.
8. United States Commission on Civil Rights. Food Deserts in Chicago: A Report of the Illinois Advisory Committee to the United States Commission on Civil Rights, October 2011. https://www.usccr.gov/pubs/docs/IL-FoodDeserts-2011.pdf (Accessed January 23, 2020.)
9. Shanahan, Caitlin E., et al. Inventory of PCBs in Chicago and opportunities for reduction in airborne emissions and human exposure. *Environmental science & technology* 49.23 (2015): 13878-13888.
10. Jones, Miranda R., et al. Race/ethnicity, residential segregation, and exposure to ambient air pollution: the Multi-Ethnic Study of Atherosclerosis (MESA). *American journal of public health* 104.11 (2014): 2130-2137.
11. United States Environmental Protection Agency. Environmental Issues in Southeast Chicago. https://www.epa.gov/il/environmental-issues-southeast-chicago (Accessed January 23, 2020).
12. Sampson RJ, Winter AS. The racial ecology of lead poisoning: toxic inequality in Chicago neighborhoods, 1995–2013. Du Bois Rev. 2016;13(2):261–283
13. Del Toral MA, Porter A, Schock MR.Detection and evaluation of elevated lead release from service lines: a field study. Environ Sci Technol. 2013 Aug 20;47(16):9300-7. doi: 10.1021/es4003636. Epub 2013 Aug 2.
14. King KE. Chicago Residents' Perceptions of Air Quality: Objective Pollution, the Built Environment, and Neighborhood Stigma Theory. Popul Environ. 2015 Sep;37(1):1-21. Epub 2014 Dec 12.
15. Morenoff JD, Sampson RJ. Violent Crime and the Spatial Dynamics of Neighborhood Transition: Chicago, 1970–1990. Social Forces. 1997;76(1):31–64.
16. Sampson RJ. Chicago and the Enduring Neighborhood Effect. University of Chicago Press; Chicago, IL: 2012.
17. Mohai P, Lantz PM, Morenoff J, House JS, Mero RP. Racial and Socioeconomic Disparities in Residential Proximity to Polluting Industrial Facilities: Evidence From the Americans' Changing Lives Study. American Journal of Public Health. 2009;99(S3):S649–S656.
18. LaVeist TA, Gaskin D, Richard P. Estimating the Economic Burden of Racial Health Inequalities in the United States. *International Journal of Health Services*. 2011;*41*:231–238.
19. Warnecke RB, Oh A, Breen N, *et al*. Approaching health disparities from a population perspective: the National Institutes of Health Centers for Population Health and Health Disparities. *Am J Public Health*. 2008;98:1608-15.
20. Agency for Healthcare Research and Quality. 2018 National Healthcare Quality and Disparities Report. https://www.ahrq.gov/research/findings/nhqrdr/nhqdr18/index.html (Accessed January 27, 2020).
21. Braveman PA, Kumanyika S, Fielding S, *et al*. Health disparities and health equity: the issue is justice. *Am J Public Health*. 2011;101:S149-55.
22. Edberg M, Cleary S, Vyas A. A Trajectory Model for Understanding and Assessing Health Disparities in Immigrant/Refugee Communities. *Journal of Immigrant and Minority Health*. 2010;13:576-84.
23. Edberg M, Cleary S, Collins E, *et al*. The SAFER Latinos Project: Addressing a Community Ecology Underlying Latino Youth Violence. *Journal of Primary Prevention*. 2010;31:247–257.
24. Gaskin DJ, Thorpe RJ, McGinty EE, *et al*. Disparities in diabetes: the nexus of race, poverty, and place. *Am J Public Health*. 2014;104:2147-55.

25. LaVeist TA, Issac LA. Race, Ethnicity, and Health: A Public Health Reader, 2nd Edition. San Francisco, CA: Jossey-Bass 2012:11-41.
26. Lillie-Blanton M, Lewis CB. Policy Challenges and Opportunities in Closing the Racial/Ethnic Divide in Health Care. *Henry J. Kaiser Family Foundation.* 2005.
27. Kawachi I, Daniels N, Robinson DE. Health disparities by race and class: why both matter. *Health Affairs.* 2005;24:343-52.
28. United States Census Bureau. Estimates of Resident Population Change and Rankings: July 1, 2017 to July 1, 2018 - United States - Combined Statistical Area; and for Puerto Rico. https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=PEP_2014_PEPANNCHG.US41PR&prodType=table (Accessed January 27, 2020).
29. Smedley BD, Stith AY, Nelson AR, *et al.* Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. Washington (DC): National Academies Press (US); 2003.
30. Starfield B. Global health, equity, and primary care. *J Am Board Fam Med.* 2007;20:511-3.
31. Margellos H, Silva A, Whitman S. Comparison of health status indicators in Chicago: are Black-White disparities worsening? *Am J Public Health*. 2004;94:116-21.
32. Silva A, Whitman S, Margellos H, Ansell D. Evaluating Chicago's success in reaching the Healthy People 2000 goal of reducing health disparities. *Public Health Rep*. 2001;116:484-94.
33. National Cancer Institute. Cancer Disparities. https://www.cancer.gov/aboutcancer/understanding/disparities. (Accessed March 24, 2019).
34. Jann B. GSAMPLE: Stata module to draw a random sample. *Statistical Software ComponentsS456716, Boston College Department of Economics*, 2005.
35. Press et al, in preparation.
36. U.S. Food & Drug Administration. Clinical Laboratory Improvement Amendments (CLIA). https://www.fda.gov/medicaldevices/deviceregulationandguidance/ivdregulatoryassistance/ucm124105.htm (Accessed February 18, 2019).
37. Aschebrook-Kilfoy, et al, in preparation.
38. Anselin, Luc, Ibnu Syabri and Youngihn Kho. GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis*. 206: 38:1:5-22
39. The Illinois Diabetes Burden 2012, Division of Chronic Disease Prevention and Control, Office of Health Promotion, *Illinois Department of Public Health*, 2012.

**Table 1. An overview of data collection**

| Questionnaire Data | Blood Collection | Other Specimen Data | Clinical Measurements | Other Data Collection |
|---|---|---|---|---|
| Socioeconomic status | EDTA lavender top vacutainer | Urine collected | Anthropometry | Electronic medical records |
| Medical history | SST Red tiger top vacutainer | Saliva collected | Blood pressure | Air quality data |
| Psychosocial | Heparin green top vacutainer | Stool collected | Other hemodynamics | Lead tests |
| Environment | K2 EDTA blue top vacutainer | | | |
| Lifestyle behaviors | SST gold top vacutainer | | | |
| Family cancer history | Serum red top vacutainer | | | |
| Cancer status | cfDNA collection tube | | | |
| Use of medical services | | | | |
| Screening behavior | | | | |
| Medication use | | | | |

**Table 2. Key topics and measurements in the COMPASS questionnaire**

| Construct | Variable |
|---|---|
| Aging | Mother's age at birth; parents' age at death; parents' vital status; osteoporosis |
| Female Medical History | Age at menarche; breastfeeding; fertility interventions; hormonal birth control use; HRT; menopausal status; pregnancy history |
| Lifestyle Behaviors | Alcohol use; cannabis use; e-cigarettes/vaping; heroin use; lifetime passive smoke exposure; physical activity; self-reported health; smoking cessation; tobacco use history |
| Follow-up Data | Address; email address; phone number; contact information of friends or family members; driver's license number; social security number |
| Immune Status | Allergies; blood transfusions; influenza immunization history |
| Medication Use | Antihistamine use; anti-hypertensive use; H2 receptor blocker use; insulin use; NSAID use; opioid use; pain status; proton pump inhibitor use; sleep medication use; statin/cholesterol/lipid lowering drug use |
| Psychological Status | Anxiety; depression; stress |
| Recent Medical Care Utilization | Barriers to health care access; ER/urgent care utilization; health care coverage type and duration; health care utilization; screening activities; usual health care provider/location |
| Significant Medical History | Arthritis; asthma; autoimmune disorders; blood cholesterol; dental history; diabetes; family history of cancer; heart attack/MI; hepatitis B; hepatitis C; HIV/AIDs; hypertension; number of children; number of siblings; other chronic diseases; personal history of cancer; sleep apnea; STDs; surgeries |
| Sociodemographic Status | Age; birth date; household income; household size; education level; marital status; occupation; place of birth; race/ethnicity; sex/gender; work status |
| Social Environment | Caregiver status; community social cohesion; night shift work; religion; self-reported patient experience/satisfaction; sleep habits; stress at home; stress at work |
| Blood Pressure | Resting blood pressure; resting heart rate |
| Body Composition | Height; hip circumference; waist circumference; weight |
| Environmental Exposures | Drinking water sources; lifetime proximity to point sources of pollution; residence during adolescence; self-report crime & violence concerns; self-report environmental concerns; work commute (current & historical) |

**Table 3. Characteristics of the Cohort (n= 7,728)**

| | Mean/% | N |
|---|---|---|
| **Demographics** | | |
| **Average Age** | 53.7 | 7,728 |
| **% Male** | 47.8 | 3,693 |
| **% Married** | 17.4 | 1,343 |
| **% African American** | 80.5 | 6,224 |
| **% Hispanic** | 5.7 | 439 |
| **Lifestyle Factors** | | |
| **% Working Full-Time** | 15.2 | 1,174 |
| **% High School Graduate/GED** | 26.3 | 2,030 |
| **% Bachelor's Degree or Higher** | 8.0 | 615 |
| **% Yearly Income Less than $15,000** | 51.4 | 3,975 |
| **% Currently Smoke Marijuana** | 20.2 | 1,559 |
| **% Currently Smoke Cigarettes** | 47.8 | 3,694 |
| **Medical History (Self-Report)** | | |
| **% History of Cancer** | 5.7 | 441 |
| **% High Cholesterol** | 24.2 | 1,874 |
| **% Type 2 Diabetes** | 10.1 | 781 |
| **% Hypertension** | 38.0 | 2,835 |
| **% Health Insurance** | 76.6 | 5,918 |
| **% Heart Attack** | 4.7 | 367 |
| **Observed Values** | | |
| **Average Systolic BP (mmHg)** | 136.0 | 6,646 |
| **Average Diastolic BP (mmHg)** | 84.0 | 6,216 |
| **% Obese** | 35.0 | 2,704 |

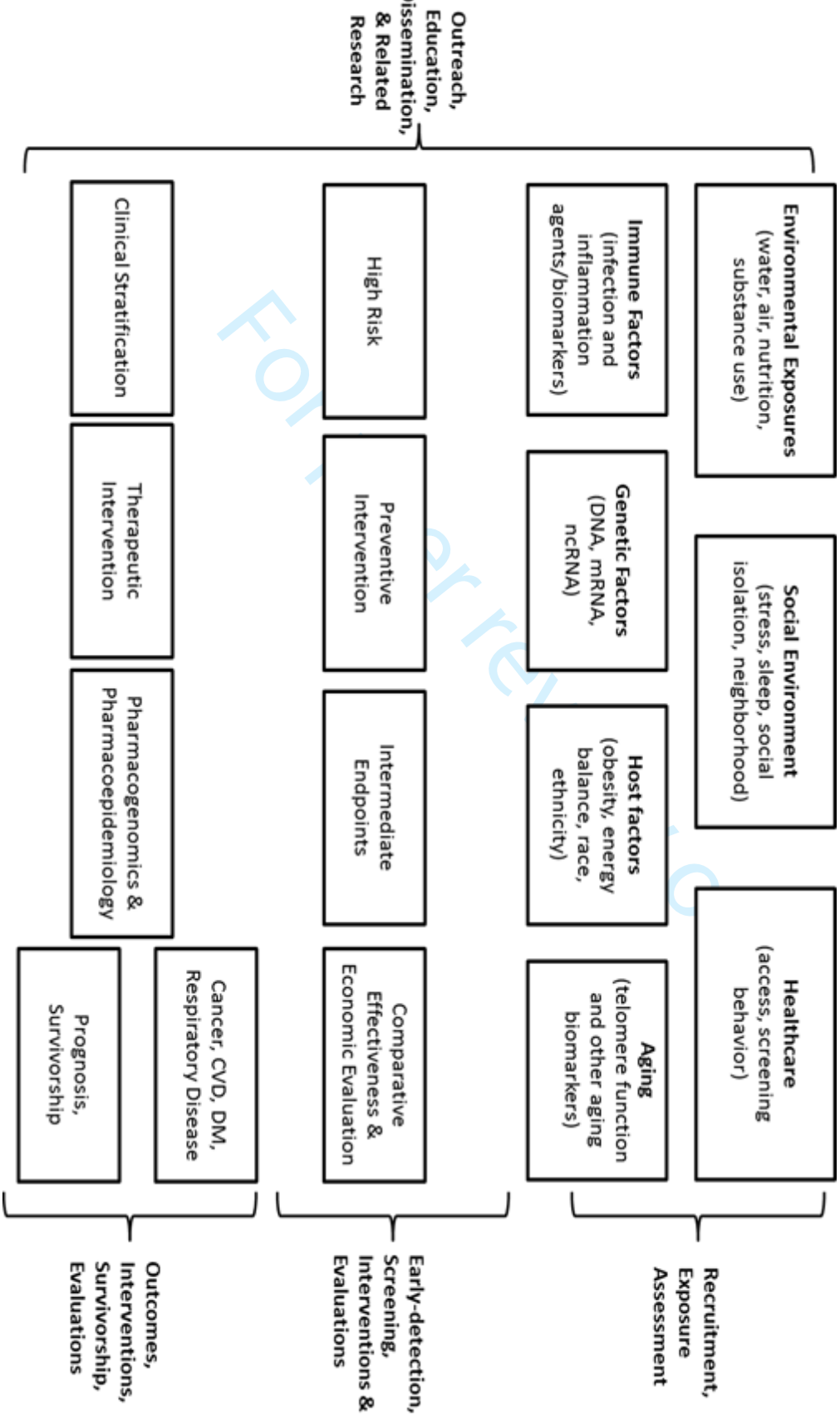



Figure 2. An IPPH Mobile Unit Used for COMPASS Enrollment 2.a Unit Interior

58x74mm (220 x 220 DPI)


Figure 2. An IPPH Mobile Unit Used for COMPASS Enrollment 2.b Unit exterior

285x214mm (72 x 72 DPI)

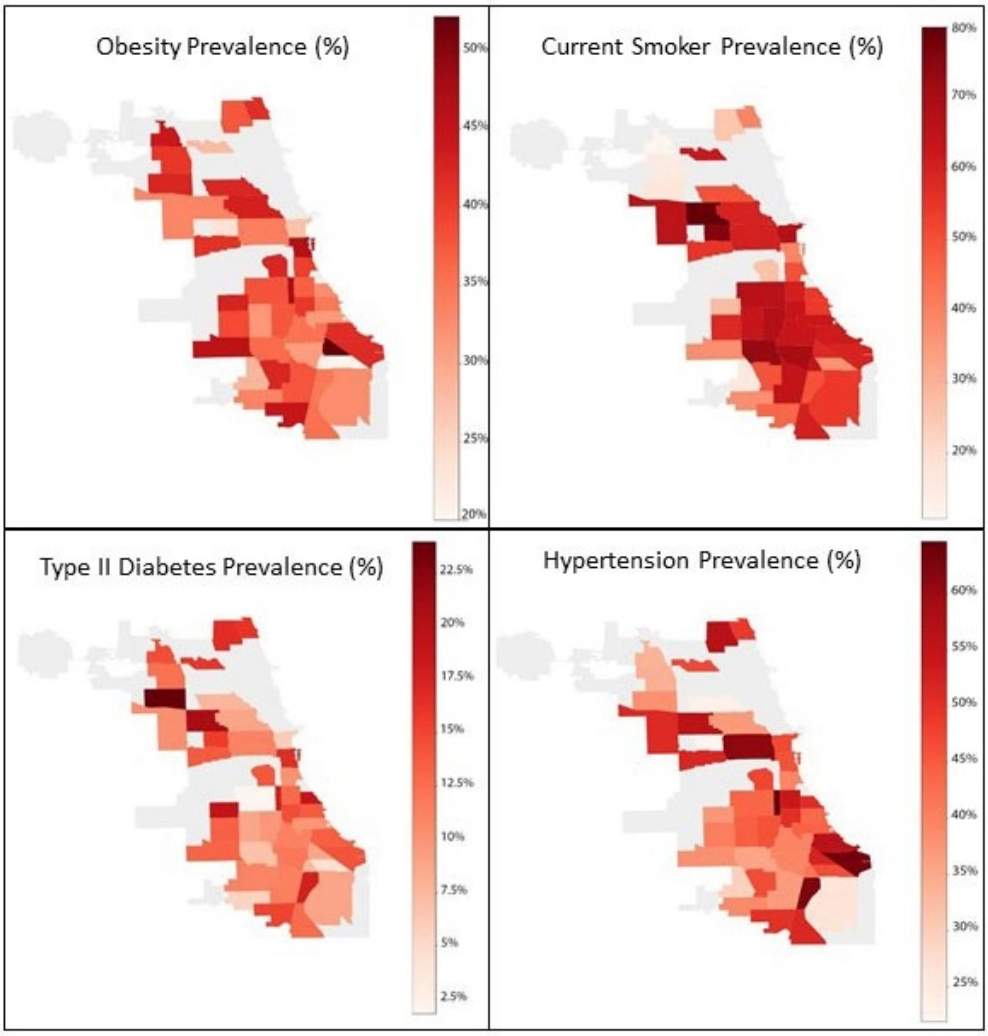



Figure 3. Mapped Prevalence of COMPASS Self-Reported Obesity, Smoking, Type 2 Diabetes, and Hypertension by Chicago Community Area

155x161mm (96 x 96 DPI)

BMJ Open

# Cohort Profile: The ChicagO Multiethnic Prevention and Surveillance Study (COMPASS)

| | |
|---|---|
| Journal: | *BMJ Open* |
| Manuscript ID | bmjopen-2020-038481.R1 |
| Article Type: | Cohort profile |
| Date Submitted by the Author: | 16-Jun-2020 |
| Complete List of Authors: | Aschebrook-Kilfoy, Briseis; University of Chicago, Department of Public Health Science; The University of Chicago, Public Health Sciences<br>Kibriya, Muhammad; The University of Chicago, Public Health Sciences<br>Jasmine, Farzana ; The University of Chicago, Public Health Sciences<br>Stepniak, Liz; The University of Chicago, Public Health Sciences<br>Gopalakrishnan, Rajan; The University of Chicago, Public Health Sciences<br>Craver, Andrew; The University of Chicago, Public Health Sciences<br>Zakin, Paul; The University of Chicago, Public Health Sciences<br>Tasmin, Saira; The University of Chicago, Public Health Sciences<br>Kim, Karen; University of Chicago<br>Goss, Kathleen; The University of Chicago, Public Health Sciences<br>List, Marcy; The University of Chicago, Public Health Sciences<br>LeBeau, Michelle; The University of Chicago, Public Health Sciences<br>Ahsan, Habibul; University of Chicago, Health Studies |
| Primary Subject Heading: | Epidemiology |
| Secondary Subject Heading: | Public health |
| Keywords: | EPIDEMIOLOGY, PREVENTIVE MEDICINE, PUBLIC HEALTH |
| | |

*I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.*

*The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.*

*Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.*

# Cohort Profile: The ChicagO Multiethnic Prevention and Surveillance Study (COMPASS)

Briseis Aschebrook-Kilfoy[1,2], Muhammad G. Kibriya[1,2], Farzana Jasmine[1,2], Liz Stepniak[1,2], Rajan Gopalakrishnan[3], Andrew Craver[2], Paul Zakin[2], Saira Tasmin[1,2], Karen Kim[3;4], Kathleen H. Goss[3], Marcy List[3], Michelle LeBeau[3], Habibul Ahsan[1,2]

[1] Department of Public Health Sciences, University of Chicago, Chicago, IL, USA.
[2] Institute for Population and Precision Health, University of Chicago, Chicago, IL, USA.
[3] Comprehensive Cancer Center, University of Chicago, Chicago, IL, USA.
[4] Department of Medicine, University of Chicago, Chicago, IL, USA.

**Address for correspondence and reprints:**
Dr. Briseis Aschebrook-Kilfoy, Institute for Population and Precision Health, The University of Chicago, 5841 S Maryland Ave, Chicago IL 60636; email: brisa@uchicago.edu

**Running Head**: Cohort Profile for COMPASS

## ABSTRACT

**Purpose:** The Chicago Multiethnic Prevention and Surveillance Study or **"**COMPASS" is a population-based cohort study with a goal to examine the risk and determinants of cancer and chronic disease. COMPASS aims to address factors causing and/or exacerbating health disparities using a precision health approach by recruiting diverse participants in Chicago, with an emphasis on those historically underrepresented in biomedical research.

**Participants:** Nearly 8,000 participants have been recruited from 72 of the 77 Chicago Community Areas. Enrollment entails the completion of an hour-long survey, consenting for past and future medical records from all sources, the collection of clinical and physical measurement data, and the on-site collection of biological samples including blood, urine, and saliva. Indoor air monitoring data and stool samples are being collected from a subset of participants. Upon collection, all biological samples are processed and aliquoted within 24 hours before long-term storage and subsequent analysis.

**Findings to Date:** The cohort reported an average age of 53.7 years, while 80.5% identified as African American, 5.7% as Hispanic, and 47.8% as male. Over 50% reported earning less than $15,000 yearly, 35% were obese, and 47.8% were current smokers. Moreover, 38% self-reported having had a diagnosis of hypertension, while 66.4% were measured as hypertensive at enrollment.

**Future Plans:** We plan to expand recruitment up to 100,000 participants from the Chicago metropolitan area in the next decade using a hybrid community and clinic-based recruitment framework that incorporates data collection through mobile medical units. Follow-up data collection from current cohort members will include serial samples, as well as longitudinal health, lifestyle, and behavioral assessment. We will supplement self-reported data with electronic medical records (EMRs), expand the collection of biometrics and biosamples to facilitate increasing digital epidemiologic study designs, and link to state and/or national level databases to ascertain outcomes. The results and findings will inform potential opportunities for precision disease prevention and mitigation in Chicago and other urban areas with a diverse population.


**Strengths and limitations of this study:**

- COMPASS has the unique capacity to recruit individuals previously underrepresented in biomedical research.
- Thus far, there is limited variation in socioeconomic strata within race and ethnicity categories as we have oversampled low income households.
- The prospective cohort design reduces possible recall bias and selection bias.
- The geographically defined cohort design allows for the exploration of an array of environmental and social exposures potentially affecting health outcomes.
- Participants will continue to be recruited and enrolled at medical centers and aboard mobile units, which will allow for the efficient collection of serial biospecimens and tumor tissue as well as comprehensive health history data from EMRs.

**Registration: NA**

## INTRODUCTION

Despite the tremendous advances in medicine and public health in recent decades, many areas of the United States have seen a growing gap in health outcomes and health equity.[1,2] As one of the most racially segregated cities in the United States[3], Chicago reports alarming health disparities and an unequal burden of cancer and chronic risk and mortality, particularly on the South Side.[3] The South Side of Chicago encompasses roughly 60% of the city's land area. The South Side is one of the nation's largest contiguous urban African-American communities, and is home to a large percentage of the city's 814,500 African American residents.[4] The South Side also includes Chinatown and Bridgeport, the two neighborhoods with the highest concentrations of Asian Americans in Illinois, making Chicago the sixth largest Asian-American city in the United States. The most recent census also illustrated that Hispanics now account for nearly a third of the city's total population of 2.7 million.[4]

Chicago faces significant urban challenges, including concentrated poverty, violence and crime, poor housing and living standards, food deserts, and environmental exposures such as toxins in the air and water due to waste sites, landfills, and abandoned industrial buildings. [5-17] Despite advances in cancer and chronic disease prevention, screening, diagnosis, and treatment that have improved chronic disease risk and outcomes for many Americans, disparities in cancer, cardiovascular disease, diabetes, asthma, and maternal child health persist in Chicago[18]. Disease-related mortality rates reflect continued disparity between non-Hispanic Blacks and Chicagoans of other races. [19-22] Health disparities are increasingly understood to be a substantial burden on society in terms of healthcare costs, lost productivity, and general societal wellbeing.[23, 24] According to the National Institute on Minority and Health Disparities, health disparities are multidimensional, complex phenomena that result from the interaction of multiple contributing factors over time.[25] These disparities create trajectories of health vulnerability for certain groups of people that vary from the majority norm.[26-28] Therefore, there is an urgent need to better understand the influence of factors that drive health disparities, such as health care coverage and low socioeconomic status. It is also important to identify the still-unknown factors behind health disparities in order to mitigate these trends and reduce the unequal burden of disease.

The University of Chicago is home to the Institute for Population and Precision Health (IPPH), as well as an NCI-designated Comprehensive Cancer Center. Given its location in the largest contiguous African-American community in the United States, its extensive community engagement infrastructure and cohort building experience, and the availability of innovative and well-established population science resources, we are in a unique position to focus efforts on understudied populations and to study health disparities on Chicago's South Side. To meet the need for a rigorous study of cancer and chronic disease and disparities in the Chicago metropolitan area, we established the Chicago Multiethnic Prevention and Surveillance Study (COMPASS) in 2013. COMPASS is a population-based longitudinal study of cancer and chronic disease risk and progression with the overall goal of recruiting a diverse study population in Chicago. As those with lowest socioeconomic status and possibly the greatest vulnerability to poor health outcomes have historically been the least likely to enroll in biomedical and biobanking research,[29-30] we have implemented a study design in which recruitment occurs in both the community and clinic, and involves extensive engagement activities. Because of the growing recognition of health disparities

in Chicago, COMPASS data collection efforts have focused on potentially important factors for minority health and have emphasized recruitment from minority populations.

## COHORT DESCRIPTION

### Purpose

The aims and primary scientific goals of COMPASS are to focus on the discovery of cancer and chronic disease in general, with the perspective that social, economic, racial, and political marginalization are often primary contributing factors associated with health disparities.[31-41] The COMPASS scientific focus areas are illustrated in **Figure 1**. The primary COMPASS aims driving the design are as follows:

1. To discover novel genetic and epigenetic factors related to the risk of cancers among the urban population through comprehensive yet cost-efficient genome-wide screening of germline DNA for Single Nucleotide Polymorphism (SNP) and methylation variants. This will be accomplished by efficient typing of case-cohort samples using custom gene-chips based on all novel genetic and epigenetic variants identified through next-generation whole genome deep sequencing and methylation profiling of cohort samples supplemented with previously identified and validated variants from 1000 Genome Project. Similar screening of tumor tissues of cancer cases will also be done to distinguish between inherent and acquired genomic alterations.

2. To elucidate and characterize cancer risk defining germline genetic/epigenetic discoveries, we will prospectively screen baseline serum/plasma and RNA samples from representative samples of matched sub-cohorts of all four risk-groups (carriers with cancer, non-carriers with cancer, carriers without cancer and non-carriers without cancer) for proteins/hormones/enzymes and mRNAs/ncRNAs, respectively. Similar screening of tumor tissues of the two cancer groups will also be done to distinguish between inherent and acquired alterations to further elucidate the biology and identification of novel preventive and therapeutic approaches.

3. To integrate the complex interactions of the genomic determinants with non-genomic environmental, nutritional, behavioral, and psychosocial determinants to develop a more complete understanding of cancer risk and causation. This includes consideration of the microbiome. This will be accomplished through tailored statistical and causal modeling approaches developed by statisticians and epidemiological methodologists at the University of Chicago.

4. To discover novel environmental (including air, water and household/workplace pollutants) and dietary (including food additives/toxicants and deleterious food items/nutrients) carcinogens that predispose urban populations to a higher risk for various cancers and chronic disease. These discoveries will be made through interdisciplinary efforts involving environmental sensors using nanotechnology and molecular dosimetry for high-throughput screening for biomarkers of exposure to carcinogens, coupled with sophisticated environmental, statistical, and systems modeling.

5. To identify previously unrecognized factors related to urban lifestyle and behavior (including stress, energy-balance/obesity, sleep, and aging) that increase the risk of various cancers and chronic diseases. These investigations will be conducted through integrated application of behavioral and psychosocial assessment tools, biomarkers of stress/aging/obesity, and novel multi-level and agent-based statistical modeling.

6. To consider new approaches to prevention that are personalized to one's specific risk profile based on their host, environmental/behavioral, and genomic profiles by exploiting new knowledge about the risk-driving inherent genomic variants, their interactions with environmental, host, behavioral and psychosocial dynamics, and the downstream molecular alterations in serum/plasma as well as tumor tissues that potentially mediate disease risk. COMPASS will function as a platform for focused molecular, mechanistic, and clinical research.

7. To specifically integrate the diverse data sources named here to investigate opportunities for prevention and early diagnosis, with an emphasis on the mitigation of health disparities.

The recruitment of this diverse cohort presents important opportunities for scientific discovery, disease mitigation, and the alleviation of the dramatic disparities observed in Chicago and other urban centers across the United States. According to the National Cancer Institute, overcoming cancer health disparities is one of the best opportunities for lessening the burden of cancer.[42] COMPASS data and biosample collection leverages the advanced population science infrastructure at the University of Chicago and preserves the opportunity to utilize advanced analytic methods and technologies in the future. In the short-term, much will be learned about the health of Chicago communities. In the long-term, the investigation of the aims above will offer insight into the relative contribution of key health drivers to decrease health inequalities, understand disease risk more generally, and elucidate opportunities for prevention and mitigation.

Here, we present this prospective multiethnic cohort study, including a description of the cohort enrolled to date. We also outline the research resources offered by COMPASS and implications for concurrent ancillary studies, as well as future observational and intervention opportunities within the cohort.

**Study Population**

As of 2019, there were an estimated 2.7M people and 1.1M households within the city limits of Chicago. According to a 2018 estimate, the Chicago Metropolitan area is the third largest in the United States, with 9.9M living in city and suburbs combined. More than half the population of the state of Illinois lives in the Chicago metropolitan area. The 2010 United States Census shows the population density of the city itself at 11,842 people per square mile, making it one of the nation's most densely populated cities. The median income for a household in the city as of 2018 was $55,198, with 19.5% of persons in poverty. The goal of COMPASS is to recruit 100,000 participants, oversampling the majority minority population (i.e. more than 45% African Americans and 20% Hispanics) [4, 43].

COMPASS is a longitudinal cohort study including persons recruited using a population-based approach, a community-based recruitment approach, and a hospital/clinic-based recruitment approach. The multiple recruitment modalities were considered in the early phase of this project in order to capitalize on the strengths and challenges of our Chicago context. As such, here we describe participant recruitment methods using a targeted or community partner approach, as well as at University of Chicago Medicine and on mobile medical units. Parallel data are collected from both community and clinic source populations. These data consist of survey responses, biospecimens including blood, urine, saliva, and stool, clinical measures, electronic health records, and environmental samples from the home.

To be eligible for participation in COMPASS, participants had to meet the following criteria:

1. Residence in the designated geographic area (census tract)
2. Age 35 and above at time of contact (to enhance efficiency to obtain cancer outcomes)
3. Ability to complete the consent and interview in English or Spanish
4. Willingness to provide blood, urine, and saliva samples
5. Provision of a social security number (for outcomes follow-up)

**Targeted Community-Based Recruitment**

To explore the feasibility of a population-representative sample for COMPASS, we implemented a recruitment plan to target potentially eligible participants using a two-stage cluster sampling approach, first randomizing at the census tract level and then identifying potentially eligible persons from within the tract using commercially available address lists. Field staff would then knock on doors and invite the targeted individuals to participate. Specifically, "cluster sampling" involves dividing the specific population of interest into geographically distinct groups or clusters. Census tract was the primary sampling unit (PSU) for COMPASS, as census tracts are designed to be homogenous with respect to population characteristics. In the City of Chicago, there are 798 census tracts; 169 are identified as majority Hispanic, 239 as majority White, 290 as majority African American, and 6 as majority Asian. We identified 120 (15% of Chicago) study areas of interest using census tracts that would confer the racial/ethnic distribution desired for recruitment. We then used STATA module *gsample*[44] to randomly select approximately 120 tracts (40 original census tracts to which 80 tracts have been added) and addresses within blocks. The number of tracts selected was based on an assumption of a 15% response rate and the assumption that 46% of residents would be 35 years or older. By sampling at the census tract level, study interviewers would recruit from limited geographic areas where many households could be contacted efficiently, minimizing travel time and cost. In particular, we used a probability proportional to size sampling design as this design reduces sampling errors if characters of interest are related to population size. In the first stage, census tracts were randomly selected with probability proportional to the population of census blocks. In the second stage, we randomly chose households within the tract. As two locations within the same census block do not each contribute completely independent information (this is known as the "intra-cluster correlation" or ICC), we

selected more clusters rather than more points within any cluster in order to improve precision while maximizing study efficiency and minimizing cost.

The commercially available address lists used for sampling are available from the United States Postal Service's Address Management System and demographic data is available for addresses from commercial sources. The address lists used for COMPASS contained all residential addresses along a mail delivery route, and excluded commercial addresses as well as PO Boxes. The commercial list of addresses provided data for households in the targeted census tracts on age and race and a telephone number (if available). This allowed us to weight the household based on the probability that they would meet eligibility criteria and would reflect the desired racial/ethnic distribution. We excluded all households that did not appear to have a resident 35 or older.

***Engagement for Targeted Community-Based Recruitment***

We used the list of addresses in the chosen census blocks to mail a postcard introducing the study and outlining the inclusion/exclusion criteria. The postcard explained that the resident has been selected to participate in COMPASS, what participation would involve, the time commitment, study eligibility criteria, compensation for participation, and that multiple eligible household members are welcome to participate. The postcard also provided a time frame in which an interviewer would be requesting participation.

After the postcards were mailed to targeted households, the field interviewers went door-to-door on the targeted study census blocks based on a predetermined schedule to request participation. If a subject did not respond, the interviewer revisited the house within the following two weeks. If the interviewer received no response at the third visit, the household would be documented as a non-responder. By "dropping in" on all targeted participant homes, a goal was to recruit as least some portion of those individuals at greatest risk of adverse health outcomes.

If the eligible person/s agreed to participate, they enrolled in the study, completed the interview, and provided biospecimens and physical assessment data onsite in their home.

The results of the door-to-door recruitment response are presented by Press et al.[45] Taken together, the team achieved a much higher response rate in African-American communities than anticipated.

***Untargeted Community-based Recruitment Using Mobile Medical Units***

Community-based COMPASS recruitment is supported by two large mobile medical units (**Figure 2**) and the community engagement infrastructure including partner community-based organizations. The mobile medical units have two fully equipped exam rooms with phlebotomy capabilities, a bathroom to facilitate urine collection, are WiFi-enabled, and include CLIA[46] certified labs with processing/storage capabilities. All study procedures are can be accommodated on-board the mobile medical units.

***Untargeted community-based engagement***

Untargeted community-based recruitment typically occurs both in conjunction with community partners, as well as without partners in communities targeted for participation. Examples of

community partners include faith-based organizations, non-profit organizations, retail stores, civic groups, Chicago Park District facilities, and Chicago Public Schools. These partners publicize COMPASS enrollment events and invite participation on days scheduled for mobile unit presence.

**Hospital/Clinic-based Recruitment**

***Study population***

In conjunction with ongoing epidemiologic efforts at University of Chicago Medicine (UCM) administered by the UChicago IPPH, COMPASS participants are recruited in a UCM research clinic. All UCM patients and their guests ages 35 and older are eligible for participation in COMPASS. In the clinic context, eligibility is not restricted based on place of residence. As such, the eligible study population includes ~60,000 UCM patients per year, over 60% of whom are African American. However, this population differs from the community-based sampling in some key ways, and will reflect a lower-risk population from the perspective of cancer health disparities.

***Hospital/Clinic-based engagement***

In-person outreach in the medical center is conducted by trained research assistants who approach persons in waiting areas and inform them about COMPASS. Those who meet the eligibility criteria are invited to participate. Digital outreach will be expanded in the near future to use email and text introduction. Participants will be able to self-schedule an enrollment and data collection visit at our IPPH research clinic by following a link in the email or text and completing a REDCap scheduling form.

**DATA COLLECTION**

**Overview**

Daily enrollment of COMPASS participants was initiated in May of 2013.

***Enrollment of Cohort Members***

The COMPASS interview has been revised and is currently divided into three online modules. These modules are administered via the Qualtrics (Qualtrics, Provo, UT) survey platform and have been optimized for mobile devices. During their visit to the UChicago research clinic, one of the mobile data collection units, or during community recruitment, participants complete the consent process and survey module #1 using their own mobile device or an IPPH-provided Apple iPad. Survey modules #2 and #3 can be completed onsite during the same visit, or the participant can elect to receive the survey links via email or SMS to complete offsite. COMPASS staff are available onsite to help participants navigate the online consent and survey modules, and participants who choose to complete survey modules #2 and #3 offsite are encouraged to reach out to staff via phone or email with any questions or to receive technical assistance.

Once the consent and 60-80 minute survey is completed, the following clinical measurements are ascertained: anthropometry, blood pressure, and other hemodynamic measures. Study staff measure weight and body fat percentage using a scale and height, hip, and waist circumference

using a measuring device, and they then measure blood pressure and other hemodynamic parameters using a device that monitors pulse (i.e., DynaPulse). An overview of the data collection is provided in **Table 1**. In **Table 2**, the key topics and measurements in the COMPASS questionnaire are outlined. We conducted an extensive review of questionnaires used in other large cancer studies, including the American Cancer Societies Cancer Prevention Study and National Cancer Institute cohorts (including PLCO, NIH-AARP, and the Agricultural Health Study) in addition to the NHANES questionnaires, such that we would be able to harmonize our data in cohort consortium collaboration and maximize our ability to enhance study power for key diseases and populations. Our data dictionary can be found at compass.uchicago.edu.

The purpose of COMPASS anthropometry is to collect body measurements using standardized examination procedures and calibrated equipment. To ensure the collection of high-quality data, COMPASS staff members are trained to follow standardized examination protocols, calibrate equipment according to a prescribed schedule and method, and precisely measure and record the survey data. Portable digital scales are used in the field on on the mobile units by interviewers to measure weight, as well as percentage body fat and water. Body length measurements are made with waist circumference tape and retractable steel measurement tape.

The DynaPulse Pathway System measures central systolic, diastolic, and mean arterial blood pressures, and the data are also transmitted to the DynaPulse Analysis Center online for an array of hemodynamic profiles/values. Briefly, DynaPulse, is a cuff-sphygmomanometry-based method, which applies Pulse Dynamics waveform analysis principle and has demonstrated and validated its capability of deriving cardiac output (CO) simultaneously with BP, MAP, SVR, systemic vascular compliance (SVC), brachial artery compliance, distensibility and resistance (BAC, BAD and BAR), LV(dP/dt)Max, and other hemodynamic parameters. These hemodynamic values, collectured for a subset participants, obtained simultaneously allow physicians to correlate the dynamic changes of each parameter to evaluate the physiological conditions of a patient's circulatory system.

Following survey data collection and clinical measurements, COMPASS staff members draws a blood sample. For each participant, the blood sample is collected in cryogenic barcode-labeled tubes as follows: (a) Lavender top EDTA vacutainer – 10 ml; (b) Red and black top SST vacutainer – 10 ml; (c) Green top Lithium Heparin vacutainer – 10 ml; (d) Blue top Trace Element K2 EDTA vacutainer – 10 ml; (e) Gold top SST vacutainer – 3.5 ml; (f) Red top serum vacutainer – 5 ml; and (g) cfDNA collection tube – 10 ml. The vacutainers are inverted per protocol for proper mixing, ensuring a source of blood material for hematological, biochemical, hormonal, serological, and other special tests that require high-quality DNA or RNA.

The blood collection approach was conceived to confer the ability to implement a variety of future analyses relevant for cancer, environmental, and population and precision health research. Specifically, the lavender top collection can accommodate assays of whole blood DNA, PBMC – RNA, lymphocyte markers, cytokines, cardiac/CVD markers, metabolic panels, cancer panels, and investigation of viral & bacterial load (PCR). The green top collection can accommodate assays including thyroid parameters, drug screening, and other biochemical tests. The gold top collection accommodates a variety of hormone assays, measurement of lithium, iron, vitamin B12, folate,

immunoglobulins, autoantibodies and antibiotic assays. The blue top tubes can accommodate a variety of environmental measurements, including the measurement of some metals.

Study participants are asked to provide a buccal cell sample, using an Oragene DNA self-collection kit, as well as a urine sample.

Participants are eligible to receive three $25 Visa gift cards in recognition of their time. The first gift card is distributed when the participant completes the consent process, survey module #1, and contributes biospecimens and physical measurement data. The second gift card is given to the participant upon completion of survey modules #2 and #3. Those who participate in optional home air quality monitoring receive the third gift card when IPPH staff retrieve the air quality devices from their home and the participant completes a short survey on characteristics of their home and environmental exposures.

Study participants are also asked to provide an optional at-home stool sample using a stool sample collection kit. This kit includes instructions to transfer a small swab from toilet paper to a vial. The participant then places the sample vial into a prepaid and addressed envelope and mails it to IPPH.

Biospecimens are stored in medical-grade refrigerators at the IPPH research clinic and aboard the mobile medical units, and are transferred to our laboratory on the University of Chicago campus within 4 to 6 hours of collection.

***Biorepository***

At the UChicago IPPH laboratories, the EDTA-mixed blood sample is centrifuged to separate out the plasma sample and aliquoted in multiple 0.5-ml screw-capped plastic tubes for preservation (to avoid repeated thaw and freeze cycles in future). All components of blood are stored at -80C and urine samples are stored at -20°C with proper inventory. Freezers are connected to emergency power supply line (red line) and a central alarm system. The saliva samples collected using the Oragene DNA self-collection kit are processed by warming in a hot water bath and stored at -20°C in the IPPH laboratory. Stool samples are snap frozen and stored at -20°C.

***Environmental Sample Collection***

Study participants may be asked to allow IPPH staff to collect air quality data from their home using one or more air quality sensors. The sensors provide real-time measures of the presence of particulate matter PM2.5 and PM10 in addition to other contaminants such as CO, VOCs, and radon. Air quality data are collected using the COMPASS best practices protocol. Study participants are provided directions to not disturb the sensors and contact the study team with issues. The IPPH staff member schedules a follow-up visit to retrieve the sensors.

Indoor air quality data will be integrated with ambient levels from the Array of Things (AoT) initiative (https://arrayofthings.github.io/), a joint effort by UChicago and Argonne National Laboratories and supported by the City of Chicago. AoT is currently collecting environmental information on air pollution in addition to light intensity, precipitation, air quality, heat, sound volume, and foot traffic through decorative sensors on traffic poles throughout the City of Chicago. The result is a system that collects data about the city at the micro level, including fine-grained pollution levels in different neighborhoods. The diversity of pollutants measured, the number of

monitors that will be installed across the city (>500), and the availability of a complimentary sensor for indoor measures, provides an unprecedented opportunity to more accurately assess exposure to air pollution and understand the relative impact of air pollution in Chicago. AoT is placing devices in communities participating in COMPASS. This presents a significant opportunity to understand the role of air pollution in disease etiology, as well as the potential that findings could inform interventions targeting at-risk Chicago communities.

Details on air quality in-home measurements, ambient concentration modeling, and total exposure estimation can be found in accompanying manuscripts[47].

Interested study participants may be referred to the Chicago Water Quality Study, an initiative of the City of Chicago Department of Water Management. Under this initiative, a water quality kit consisting of sample collection bottles and instructions is mailed to the participant. The participant collects and documents samples per instructions and contacts the Department of Water Management for sample pick up. During the consent process, participants are asked to allow COMPASS to retrieve water quality analysis results from the Department of Water Management.

***Data management***

A secure HIPAA-compliant database was created by the Information Technology Group at Dartmouth for the first four years of the project. Going forward, all study data is being collected and stored in REDCap (Research Electronic Data Capture) and/or Qualtrics (Qualtrics, Provo, UT). In addition, the University of Chicago Medicine Comprehensive Cancer Center (UCCCC) has created an address database to track the response rate as well as details of non-responders and those who opt not to participate. This database is maintained by the COMPASS informatics team.

***Tracking health outcomes***

All participants have provided their social security numbers so that their cancer and vital status can be tracked long-term through the Illinois State Cancer Registry, which is maintained by the Illinois Department of Public Health.

***Electronic Medical Records***

UChicago is a member of the Chicago Area Patient Centered Outcomes Research Network (CAPriCORN) - a platform that integrates EMR across the nine largest academic medical centers in Chicago. COMPASS participants provide consent for access to medical records, which will allow for the efficient collection of comprehensive health history data from EMRs of participants who receive care both at the University of Chicago Medical Centers and outside of our system. The collection of medical records will be completed continuously, when available.

***Follow-up data and biosample collection***

Previously enrolled participants will be re-contacted to schedule in-person visits to our IPPH research clinic and our mobile research units. They will be asked to sign an updated consent form and to complete an online follow-up survey in order to understand changes in health status, health behaviors, and lifestyle. This visit will also include collection of follow up biospecimens and physical measurements. Follow-up survey data collection will be requested annually and

frequency of in-person biospecimen collection will be funding dependent. In addition, following COMPASS notification that a participant has had a cancer diagnosis, either from the participants themselves or the Illinois State Cancer Registry, we submit a request to the diagnosing and treatment institutions for medical records such that we may ascertain more detailed tumor characteristic and treatment information. We note in the consent that we plan to follow study participants for at least 10 years. In a recent effort to collect follow-up data, we were able to reach 25% of participants by phone. In the future, we will reach out via text and phone to enhance response rates.

***Patient and public involvement***

None of the participants was involved in the questionnaire design, biological measurements or outcome measures; they were likewise not involved in the design, recruitment and implementation of the study. Furthermore, all participants were informed of the use of the data for research in this study. There were no plans to disseminate the study results to participants. When genetic results become available, participants will be contacted and asked if they would like to receive them per the protocol/consent documents.

***Research Ethics Approval: Human Participants***

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 Institutional Review Board Committee A (approval IRB12-1660).

**Data Analyses**

Baseline data from the cohort were analyzed for tabular and descriptive statistics (**Table 3)**. Means and standard deviations were calculated for variables with a continuous distribution and proportions were calculated for variables with categorical distributions. Body mass index (BMI) was computed based on weight and height measured by the interviewers. All analyses were performed using STATA 15 (StataCorp, College Station, TX) and GeoDa for Windows[48].

## CHARACTERISTICS OF STUDY PARTICIPANTS RECRUITED TO DATE

Among the 7,728 participants enrolled in the study **(Table 3)**, 80.5% identified as African American, 5.7% as Hispanic, and 47.8% as male. In addition, over 50% reported earning less than $15,000 yearly, 15.2% reported working full-time, 35% were obese, 47.8% were current smokers, and 38% reported having hypertension. High school graduates comprised 19.8% of the cohort, and 17% of the participants were married. Consistent with findings by the Illinois Department of Public Health, the prevalence of diabetes was higher among females than males, and non-Whites compared to Whites.[49] The overall prevalence of Type 2 Diabetes in this study was 10.1%.

The maps in **Figure 3** show the prevalence of the specified variable in the 77 community areas in Chicago. Only community areas with at least 20 observations were included, resulting in 45 communities being analyzed. There is an alarming prevalence of high blood pressure, with the majority of the community areas sampled being classified as having stage 1 or 2 hypertension. In addition, the majority of community areas sample have a smoking prevalence of at least 50%.

## STRENGTHS AND LIMITATIONS

COMPASS efficiently recruits individuals previously underrepresented in biomedical research. The pilot analysis of the first 8,000+ individuals demonstrates our capacity to recruit such a diverse, urban population. Our UCM research space coupled with our mobile research units allow our team to collect and process baseline and follow-up data on campus and in Chicago communities.

Another strength of the study is the prospective cohort design, which reduces possible recall bias and selection bias. Furthermore, the geographically defined cohort design offers a number of advantages for the study of health disparities. The principal reason that we are recruiting a geographic cohort (rather than an occupational cohort, for example) is because it allows for the exploration of an array of exposures potentially affecting health outcomes. In many other cohort designs, subjects are recruited given their unique exposure, such as nurses, farmers, miners, etc. However, the factors that disproportionately affect African Americans' health status and outcomes in Chicago are largely unknown or under-characterized. Similarly, the relevant risk factors in Hispanic and Asian populations are likely unique and remain under-characterized.

COMPASS is also involved in community engagement efforts to address the needs of the South Side community, through collaboration with the UCCCC Office of Community Engagement and Cancer Health Equity (OCECHE). The overall goals of the OCECHE are to: a) establish strong partnerships with underserved communities and organizations; b) support community and local capacity building; and c) establish sustainable educational programs. To ensure both academic and community relevance and to help shape future programming and direction, the OCECHE has both a Community and Internal Advisory Board comprising key community and institutional stakeholders.

In addition, the University of Chicago is a member of the Chicago Area Patient Centered Outcomes Research Network (CAPriCORN) - a platform that integrates EMR across the nine largest academic medical centers in Chicago. CAPriCORN makes it possible for investigators at its partner institutions in Chicago to meet the growing need for the integration of EMRs in epidemiologic research, including comprehensive multi-institution tracking/follow-up, such that comprehensive EMRs are available for participants who receive care at multiple institutions. The newly recruited participants will be enrolled through integrated health care systems, such as UChicago, which will allow for the efficient collection of serial biospecimens and tumor tissue as well as comprehensive health history data from EMRs. They will be followed longitudinally to study a broad range of risk factors, early markers, and outcomes. We will be able to capitalize on the latest advances in the understanding of cancer etiology, technology for data capture, and analytic tools.

The collection of environmental data in COMPASS to understand the role of the environment in health disparities is supported by a new NIEHS-sponsored environmental health center in the region- ChicAgo Center for Health and Environment (CACHET). CACHET was established to address EHS-related health disparities as the first center of its kind in the greater Chicago tri-state area (i.e., Illinois, Indiana, and Wisconsin). We have developed an environmental monitoring team (used for many NIEHS P30 investigations), infrastructure, and expertise to explore the impact of

urban environmental exposures that are embedded in COMPASS data collection. The CACHET mission is to elucidate the biological and social underpinnings between relevant urban environmental exposures and human disease and translate the findings to reduce health inequities within our communities. COMPASS is a key cohort synergized by the CACHET Integrated Health Sciences Facility Core (IHSFC) at UChicago.

There are some limitations in our study. Thus far, there is limited variation in socioeconomic strata within race and ethnicity categories. Also, the longitudinal nature of the cohort makes it unlikely that we will be able to study cancer-related hypotheses for at least a decade. For rarer diseases, more people and time will be needed. In addition, the oversampling of African Americans and individuals with low socioeconomic status calls into question the representativeness of the study population. However, given the paucity of cohort studies focused on this population, COMPASS provides a unique window into the factors impacting health in Chicago. There is also an oversampling of individuals out of work; however, this is possibly also a strength of the study, as some of the highest risk populations not represented in previous works and studies included.

**Figure 1. COMPASS Scientific Focus Areas**

**Figure 2. An IPPH Mobile Unit Used for COMPASS Enrollment**

**Figure 3. Mapped Prevalence of COMPASS Self-Reported Obesity, Smoking, Type 2 Diabetes, and Hypertension by Chicago Community Area**

## DATA AVAILABILITY AND COLLABORATION

COMPASS will continue to collect a rich set of data on multiple exposure domains and health outcomes. For more information, refer to the web site: compass.uchicago.edu. Researchers interested in collaboration are invited to propose research questions based on the data available within COMPASS or to submit a request for additional data collection. Requests can be submitted electronically on the COMPASS website and will be reviewed by the COMPASS scientific board. The COMPASS study team is particularly interested in collaborations that will enhance research methods for this type of work, assess the impact of environmental exposure, highlight exposures of key significance in urban communities, and address health issues of concern in Chicago and other urban centers.

## CONTRIBUTORS

The study protocol was designed by HA, ML, BAK, ML, KG and MGK and BAK drafted the manuscript. BAK and LS oversaw field operations and MGK and FJ oversaw the biosample processing and biobanking. HA, BAK, and KK oversaw engagement. PZ, RG, AC, and ST performed the data quality control and the statistical analysis. All authors participated in the study design, revised the article and approved the final version.

## COMPETING INTERESTS

None declared.

## FUNDING

This research was supported by funding from the University of Chicago Medicine Comprehensive Cancer Center and the University of Chicago Insisittue for Population and Precision Health.

## ACKNOWLEDGEMENTS

We thank the dedicated COMPASS field staff and community partners for their support of this work.

## AUTHOR AFFILIATIONS

[1] Department of Public Health Sciences, University of Chicago, Chicago, IL, USA.
[2] Institute for Population and Precision Health, University of Chicago, Chicago, IL, USA.
[3] Comprehensive Cancer Center, University of Chicago, Chicago, IL, USA.
[4] Department of Medicine, University of Chicago, Chicago, IL, USA.

## REFERENCES

1. Meyer P, Yoon P, Kaufmann R. Introduction: CDC Health Disparities and Inequalities Report- United States. *Centers for Disease Control and Prevention*. 2013;62:3-5.
2. Baciu A, Negussie Y, Geller A, *et al*. *Communities in Action: Pathways to Health Equity*. Washington DC: National Academies Press, 2017
3. Orsi J, Margellos-Anast H, Whitman S. Black-White Health Disparities in the United States and Chicago: A 15-Year Progress Analysis. *Am J Public Health*. 2010;100:349–356.
4. United States Census Bureau. U.S. Census Bureau QuickFacts: Chicago city, Illinois. https://www.census.gov/quickfacts/chicagocityillinois. Published 2018. Accessed February 22, 2019.
5. Chicago Department of Public Health. Public Health Statistics – Selected public health indicators by Chicago community area. https://data.cityofchicago.org/Health-Human-Services/Public-Health-Statistics-Selected-public-health-in/iqnk-2tcu/data (Accessed January 23, 2020).
6. Chicago Police Department. Crimes – 2001 to present. https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present-Dashboard/5cd6-ry5g (Accessed January 27, 2020).
7. Kolak, Marynia, et al. Urban foodscape trends: disparities in healthy food access in Chicago, 2007–2014. *Health & Place.* 2018:52: 231-239.
8. United States Commission on Civil Rights. Food Deserts in Chicago: A Report of the Illinois Advisory Committee to the United States Commission on Civil Rights, October 2011. https://www.usccr.gov/pubs/docs/IL-FoodDeserts-2011.pdf (Accessed January 23, 2020.)
9. Shanahan, Caitlin E., et al. Inventory of PCBs in Chicago and opportunities for reduction in airborne emissions and human exposure. *Environmental science & technology* 49.23 (2015): 13878-13888.
10. Jones, Miranda R., et al. Race/ethnicity, residential segregation, and exposure to ambient air pollution: the Multi-Ethnic Study of Atherosclerosis (MESA). *American journal of public health* 104.11 (2014): 2130-2137.
11. United States Environmental Protection Agency. Environmental Issues in Southeast Chicago. https://www.epa.gov/il/environmental-issues-southeast-chicago (Accessed January 23, 2020).
12. Sampson RJ, Winter AS. The racial ecology of lead poisoning: toxic inequality in Chicago neighborhoods, 1995–2013. Du Bois Rev. 2016;13(2):261–283
13. Del Toral MA, Porter A, Schock MR. Detection and evaluation of elevated lead release from service lines: a field study. Environ Sci Technol. 2013 Aug 20;47(16):9300-7. doi: 10.1021/es4003636. Epub 2013 Aug 2.

14. King KE. Chicago Residents' Perceptions of Air Quality: Objective Pollution, the Built Environment, and Neighborhood Stigma Theory. Popul Environ. 2015 Sep;37(1):1-21. Epub 2014 Dec 12.
15. Morenoff JD, Sampson RJ. Violent Crime and the Spatial Dynamics of Neighborhood Transition: Chicago, 1970–1990. Social Forces. 1997;76(1):31–64.
16. Sampson RJ. Chicago and the Enduring Neighborhood Effect. University of Chicago Press; Chicago, IL: 2012.
17. Mohai P, Lantz PM, Morenoff J, House JS, Mero RP. Racial and Socioeconomic Disparities in Residential Proximity to Polluting Industrial Facilities: Evidence From the Americans' Changing Lives Study. American Journal of Public Health. 2009;99(S3):S649–S656.
18. Chicago Department of Public Health. Healthy Chicago 2.0 Community Health Assessment. https://www.chicago.gov/content/dam/city/depts/cdph/CDPH/Healthy%20Chicago/HealthyChicago_CHA_4102017.pdf (Accessed June 11, 2020)
19. Chicago Health Atlas. Cancer Deaths. https://www.chicagohealthatlas.org/indicators/cancer-deaths (Accessed June 11, 2020)
20. Chicago Health Atlas. Heart Disease Deaths. https://www.chicagohealthatlas.org/indicators/heart-disease-deaths (Accessed June 11, 2020)
21. Chicago Health Atlas. Diabetes - Related Deaths. https://www.chicagohealthatlas.org/indicators/diabetes-related-deaths (Accessed June 11, 2020)
22. Chicago Health Atlas. Stroke Deaths. https://www.chicagohealthatlas.org/indicators/stroke-deaths (Accessed June 11, 2020)
23. LaVeist TA, Gaskin D, Richard P. Estimating the Economic Burden of Racial Health Inequalities in the United States. *International Journal of Health Services*. 2011;*41*:231–238.
24. Thornton, Rachel L. J., et al. "Evaluating Strategies For Reducing Health Disparities By Addressing The Social Determinants Of Health." Health Affairs, vol. 35, no. 8, Aug. 2016, pp. 1416–23.
25. Warnecke RB, Oh A, Breen N, *et al*. Approaching health disparities from a population perspective: the National Institutes of Health Centers for Population Health and Health Disparities. *Am J Public Health*. 2008;98:1608-15.
26. Agency for Healthcare Research and Quality. 2018 National Healthcare Quality and Disparities Report. https://www.ahrq.gov/research/findings/nhqrdr/nhqdr18/index.html (Accessed January 27, 2020).
27. Adler, Nancy E., and Judith Stewart. "Health Disparities across the Lifespan: Meaning, Methods, and Mechanisms: Health Disparities across the Lifespan." Annals of the New York Academy of Sciences, vol. 1186, no. 1, Feb. 2010, pp. 5–23.
28. Landrine, Hope, and Irma Corral. "Separate and unequal: residential segregation and black health disparities." Ethnicity & disease vol. 19,2 (2009): 179-84.
29. George, Sheba, et al. "A Systematic Review of Barriers and Facilitators to Minority Research Participation Among African Americans, Latinos, Asian Americans, and Pacific Islanders." American Journal of Public Health, vol. 104, no. 2, Feb. 2014, pp. e16–31.
30. McDonald, Jasmine A., et al. "Understanding Participation by African Americans in Cancer Genetics Research." Journal of the National Medical Association, vol. 104, no. 7–8, July 2012, pp. 324–30.
31. Braveman PA, Kumanyika S, Fielding S, *et al*. Health disparities and health equity: the issue is justice. *Am J Public Health*. 2011;101:S149-55.
32. Edberg M, Cleary S, Vyas A. A Trajectory Model for Understanding and Assessing Health Disparities in Immigrant/Refugee Communities. *Journal of Immigrant and Minority Health*. 2010;13:576-84.
33. Edberg M, Cleary S, Collins E, *et al*. The SAFER Latinos Project: Addressing a Community Ecology Underlying Latino Youth Violence. *Journal of Primary Prevention*. 2010;31:247–257.
34. Gaskin DJ, Thorpe RJ, McGinty EE, *et al*. Disparities in diabetes: the nexus of race, poverty, and place. *Am J Public Health*. 2014;104:2147-55.
35. LaVeist TA, Issac LA. Race, Ethnicity, and Health: A Public Health Reader, 2nd Edition. San Francisco, CA: Jossey-Bass 2012:11-41.
36. Lillie-Blanton M, Lewis CB. Policy Challenges and Opportunities in Closing the Racial/Ethnic Divide in Health Care. *Henry J. Kaiser Family Foundation.* 2005.

37. Kawachi I, Daniels N, Robinson DE. Health disparities by race and class: why both matter. *Health Affairs.* 2005;24:343-52.
38. United States Census Bureau. Estimates of Resident Population Change and Rankings: July 1, 2017 to July 1, 2018 - United States - Combined Statistical Area; and for Puerto Rico. https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=PEP_2014_PEPANNCHG.US41PR&prodType=table (Accessed January 27, 2020).
39. Smedley BD, Stith AY, Nelson AR, *et al.* Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. Washington (DC): National Academies Press (US); 2003.
40. Starfield B. Global health, equity, and primary care. *J Am Board Fam Med.* 2007;20:511-3.
41. Margellos H, Silva A, Whitman S. Comparison of health status indicators in Chicago: are Black-White disparities worsening? *Am J Public Health*. 2004;94:116-21.
42. Silva A, Whitman S, Margellos H, Ansell D. Evaluating Chicago's success in reaching the Healthy People 2000 goal of reducing health disparities. *Public Health Rep*. 2001;116:484-94.
43. National Cancer Institute. Cancer Disparities. https://www.cancer.gov/aboutcancer/understanding/disparities. (Accessed March 24, 2019).
44. Jann B. GSAMPLE: Stata module to draw a random sample. Statistical Software ComponentsS456716, Boston College Department of Economics, 2005.
45. Press et al, in preparation.
46. U.S. Food & Drug Administration. Clinical Laboratory Improvement Amendments (CLIA). https://www.fda.gov/medicaldevices/deviceregulationandguidance/ivdregulatoryassistance/ucm124105.htm (Accessed February 18, 2019).
47. Aschebrook-Kilfoy, et al, in preparation.
48. Anselin, Luc, Ibnu Syabri and Youngihn Kho. GeoDa: An Introduction to Spatial Data Analysis. Geographical Analysis. 206: 38:1:5-22
49. The Illinois Diabetes Burden 2012, Division of Chronic Disease Prevention and Control, Office of Health Promotion, Illinois Department of Public Health, 2012.

**Table 1. An overview of data collection**

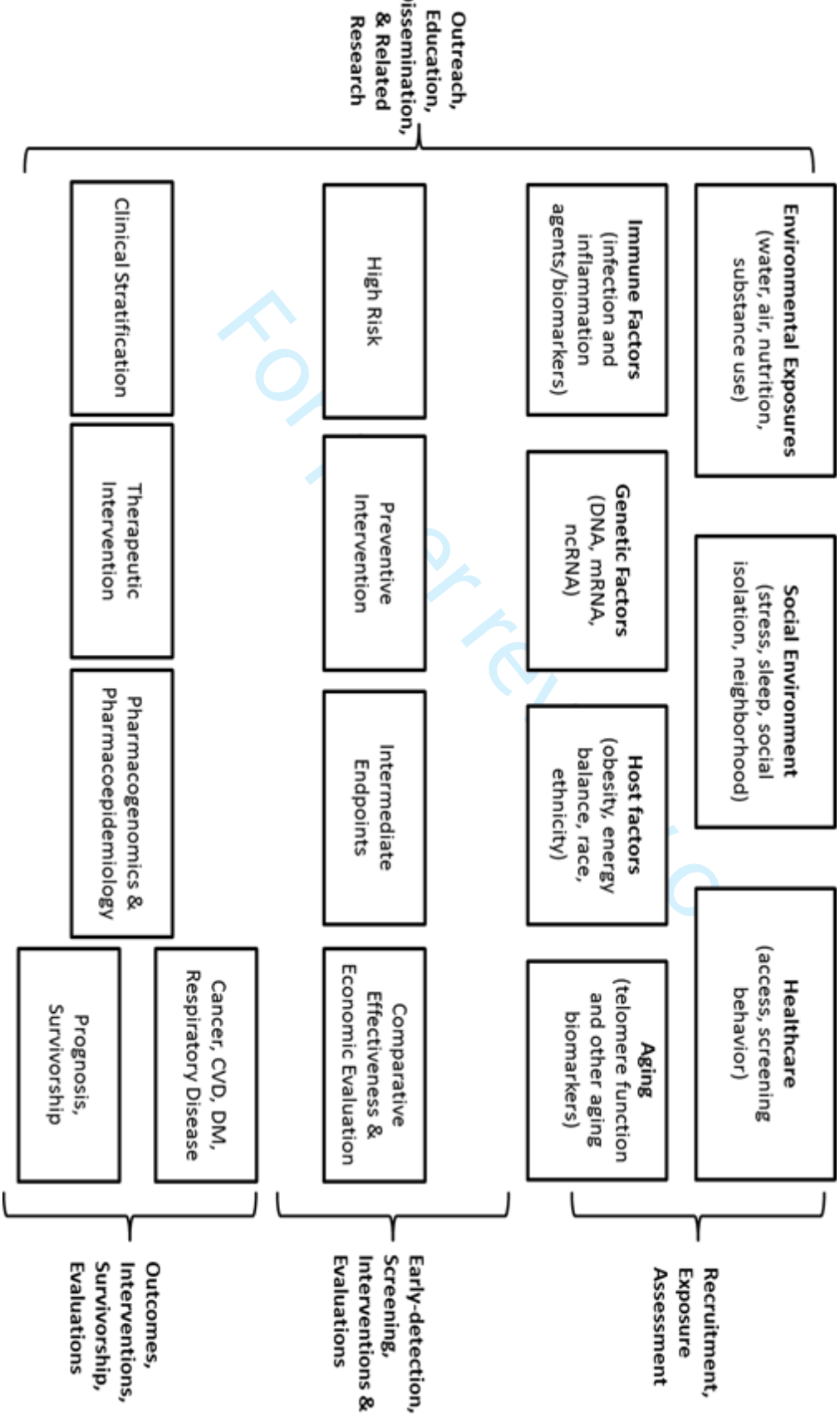
| Questionnaire Data | Blood Collection | Other Specimen Data | Clinical Measurements | Other Data Collection |
|---|---|---|---|---|
| Socioeconomic status | EDTA lavender top vacutainer | Urine collected | Anthropometry | Electronic medical records |
| Medical history | SST Red tiger top vacutainer | Saliva collected | Blood pressure | Air quality data |
| Psychosocial | Heparin green top vacutainer | Stool collected | Other hemodynamics | Lead tests |
| Environment | K2 EDTA blue top vacutainer | | | |
| Lifestyle behaviors | SST gold top vacutainer | | | |
| Family cancer history | Serum red top vacutainer | | | |
| Cancer status | cfDNA collection tube | | | |
| Use of medical services | | | | |
| Screening behavior | | | | |
| Medication use | | | | |

**Table 2. Key topics and measurements in the COMPASS questionnaire**

| Construct | Variable |
|---|---|
| Aging | Mother's age at birth; parents' age at death; parents' vital status; osteoporosis |
| Female Medical History | Age at menarche; breastfeeding; fertility interventions; hormonal birth control use; HRT; menopausal status; pregnancy history |
| Lifestyle Behaviors | Alcohol use; cannabis use; e-cigarettes/vaping; heroin use; lifetime passive smoke exposure; physical activity; self-reported health; smoking cessation; tobacco use history |
| Follow-up Data | Address; email address; phone number; contact information of friends or family members; driver's license number; social security number |
| Immune Status | Allergies; blood transfusions; influenza immunization history |
| Medication Use | Antihistamine use; anti-hypertensive use; H2 receptor blocker use; insulin use; NSAID use; opioid use; pain status; proton pump inhibitor use; sleep medication use; statin/cholesterol/lipid lowering drug use |
| Psychological Status | Anxiety; depression; stress |
| Recent Medical Care Utilization | Barriers to health care access; ER/urgent care utilization; health care coverage type and duration; health care utilization; screening activities; usual health care provider/location |
| Significant Medical History | Arthritis; asthma; autoimmune disorders; blood cholesterol; dental history; diabetes; family history of cancer; heart attack/MI; hepatitis B; hepatitis C; HIV/AIDs; hypertension; number of children; number of siblings; other chronic diseases; personal history of cancer; sleep apnea; STDs; surgeries |
| Sociodemographic Status | Age; birth date; household income; household size; education level; marital status; occupation; place of birth; race/ethnicity; sex/gender; work status |
| Social Environment | Caregiver status; community social cohesion; night shift work; religion; self-reported patient experience/satisfaction; sleep habits; stress at home; stress at work |
| Blood Pressure | Resting blood pressure; resting heart rate |
| Body Composition | Height; hip circumference; waist circumference; weight |
| Environmental Exposures | Drinking water sources; lifetime proximity to point sources of pollution; residence during adolescence; self-report crime & violence concerns; self-report environmental concerns; work commute (current & historical) |

**Table 3. Characteristics of the Cohort (n= 7,728)**

| | Mean/% | N |
|---|---|---|
| **Demographics** | | |
| **Average Age** | 53.7 | 7,728 |
| **% Male** | 47.8 | 3,693 |
| **% Married** | 17.4 | 1,343 |
| **% African American** | 80.5 | 6,224 |
| **% Hispanic** | 5.7 | 439 |
| **Lifestyle Factors** | | |
| **% Working Full-Time** | 15.2 | 1,174 |
| **% High School Graduate/GED** | 26.3 | 2,030 |
| **% Bachelor's Degree or Higher** | 8.0 | 615 |
| **% Yearly Income Less than $15,000** | 51.4 | 3,975 |
| **% Currently Smoke Marijuana** | 20.2 | 1,559 |
| **% Currently Smoke Cigarettes** | 47.8 | 3,694 |
| **Medical History (Self-Report)** | | |
| **% History of Cancer** | 5.7 | 441 |
| **% High Cholesterol** | 24.2 | 1,874 |
| **% Type 2 Diabetes** | 10.1 | 781 |
| **% Hypertension** | 38.0 | 2,835 |
| **% Health Insurance** | 76.6 | 5,918 |
| **% Heart Attack** | 4.7 | 367 |
| **Observed Values** | | |
| **Average Systolic BP (mmHg)** | 136.0 | 6,646 |
| **Average Diastolic BP (mmHg)** | 84.0 | 6,216 |
| **% Obese** | 35.0 | 2,704 |




An IPPH Mobile Unit Used for COMPASS Enrollment

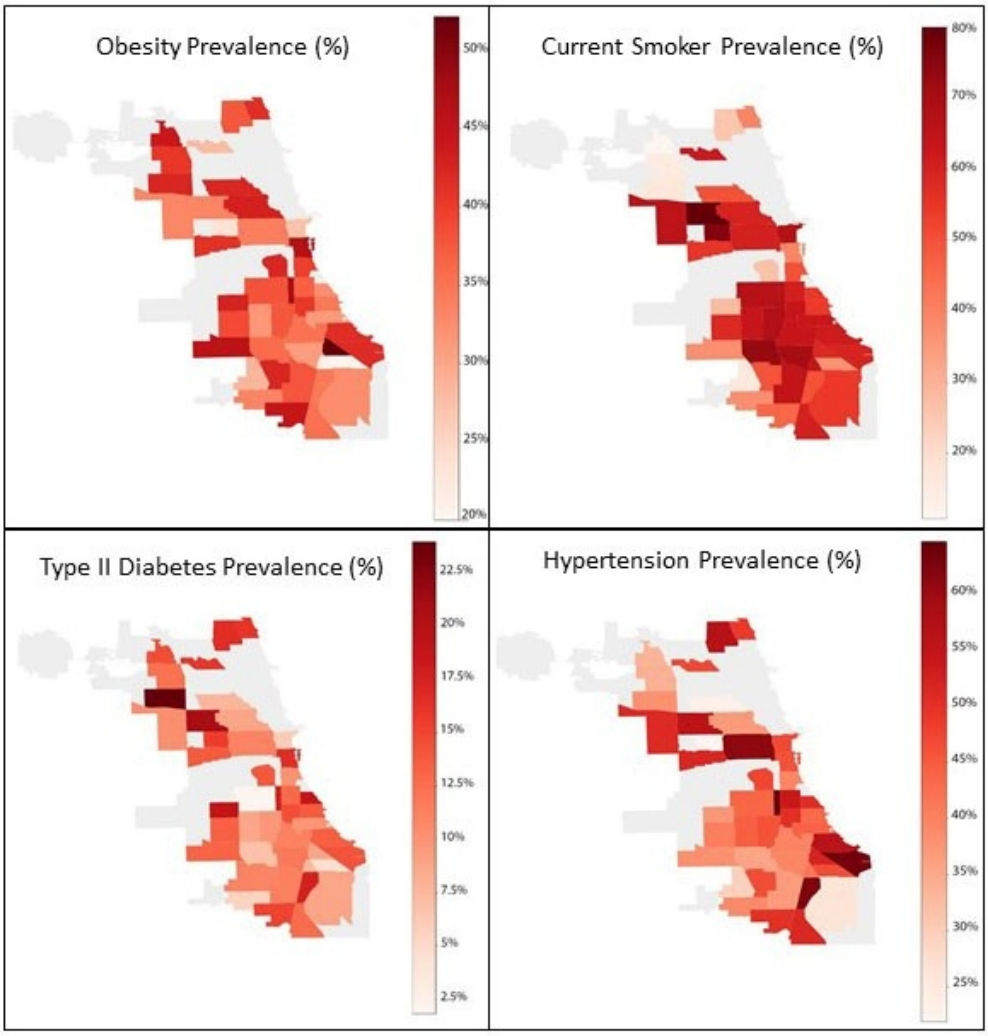



Figure 3. Mapped Prevalence of COMPASS Self-Reported Obesity, Smoking, Type 2 Diabetes, and Hypertension by Chicago Community Area

155x161mm (96 x 96 DPI)