BMJ Public Health

AI4HealthyCities: a protocol for a mixed-method ethnographic study on the social determinants of cardiovascular health in New York City

Anna-Maria Volkmann,[1] Elizabeth Adamson,[2] Emma Boxley ,[1] David Napier,[1] Peter Speyer,[2] Reekarl Pierre,[3] Wenqin Zhang ,[1] Yongkang Zhang[3]

To cite: Volkmann A-M, Adamson E, Boxley E, *et al*. AI4HealthyCities: a protocol for a mixed-method ethnographic study on the social determinants of cardiovascular health in New York City. *BMJ Public Health* 2025;**3**:e002382. doi:10.1136/bmjph-2024-002382

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjph-2024-002382).

Received 21 November 2024
Accepted 2 July 2025

[1]Department of Anthropology, University College London, London, UK
[2]Novartis Foundation, Basel, Switzerland
[3]Department of Population Health Sciences, Weill Cornell Medicine, New York, New York, USA

Correspondence to
Dr Anna-Maria Volkmann;
a.volkmann@ucl.ac.uk

ABSTRACT

Introduction Cardiovascular disease is a leading cause of morbidity and mortality globally and in New York City. Significant disparities in prevalence and risk factors persist across city neighbourhoods and among populations of varying socio-economic status, racial and ethnic backgrounds. These disparities are shaped and sustained by the complex interplay of social determinants of health, including housing, employment, access to healthcare and structural inequities. This study builds on prior quantitative research conducted under the AI4HealthyCities initiative, which applied machine learning to identify spatial clusters of cardiovascular vulnerability and social disadvantage. This research addresses key gaps in that work by generating disaggregated, qualitative data on underrepresented populations. The study also aims to explore the mechanisms through which specific social determinants may contribute to cardiovascular risk, including the role of behavioural and demographic mediators. By combining lived experiences and system-level perspectives, the research will provide contextualised insights to support local stakeholders in designing more effective, equity-oriented interventions.

Methods and analysis This mixed-method ethnographic study will collect data in three phases: expert interviews, community roundtables and vulnerability assessments across three New York City boroughs (Brooklyn, the Bronx and Queens). Qualitative data will be analysed using a combination of deductive and inductive approaches. Thematic synthesis will be used to identify patterns across and within boroughs. Study design and interim findings will be reviewed in collaboration with community stakeholders, a research steering group and the AI4HealthyCities Global Expert Council.

Ethics and dissemination This study has received ethical approval from the Ethics Committee of Weill Cornell Medicine (code number: 23-04025988). The findings of the project will be disseminated via conferences, speaking engagements and peer-reviewed publications.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Social determinants of health play a critical role in shaping cardiovascular outcomes, yet conventional health data often overlook the lived experiences of marginalised and structurally excluded populations.

WHAT THIS STUDY ADDS

⇒ This study will provide context-rich data on how intersecting factors—such as low income, structural marginalisation and time poverty—contribute to elevated cardiovascular risk yet remain underrepresented in existing data systems. Data from roundtables and expert interviews will provide insights into existing services and community needs and priorities, and vulnerability assessments will uncover localised 'hidden' populations and structural inequities.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ By innovatively linking qualitative insights with earlier spatial clustering analysis, this study will generate actionable evidence to support precision population health strategies tailored to the realities of vulnerable communities. Findings are expected to inform urban policy and service design by identifying where and why standard models fail to reach those most at risk—and what can be done differently.

INTRODUCTION

Cardiovascular disease (CVD) remains the leading cause of death in New York City and globally, accounting for approximately 28% of deaths in the city in 2022.[1] Hypertension, the main risk factor for CVD, is prevalent in 30% of adults in New York City, with disproportionately higher rates among African Americans (44%) and Hispanic adults (31%) compared with non-Hispanic White adults (23%).[2] These disparities are shaped by social determinants of health (SDoH)—the conditions in which people are born, grow, live, work and age—that influence health outcomes.[3] These determinants are not randomly distributed; rather, they are structured by the unequal distribution of money, power and resources at global, national and local levels.[4] The U.S. Department of Health and Human Services identifies five key SDoH domains: economic stability, education access and quality, neighbourhood and built environment, social and community context and healthcare access and quality.[5]

Studies consistently have shown that SDoH significantly impact cardiovascular outcomes, particularly in disadvantaged communities, which often experience compounding factors such as environmental stressors, limited access to nutritious foods and healthcare, social isolation and cultural or linguistic barriers—factors that together contribute to elevated CVD risk and worse outcomes among marginalised populations.[6–10] Importantly, these factors interact in complex ways that are difficult to capture through standard data sources and single-method research designs.

Despite extensive research, gaps remain in understanding the direct and indirect effects of SDoH on CVD outcomes, especially in diverse urban areas like New York City, where structural inequities and social heterogeneity are pronounced.[11] Moreover, many studies do not include the perspectives and lived experiences of community members—particularly those from marginalised or 'hidden' populations—who are frequently underrepresented in health system data and, by extension, in policy responses.[12 13]

This study addresses these gaps by combining AI-driven quantitative analysis with ethnographic and qualitative research methods.[14] This mixed-method approach acknowledges the complexity of SDoH and the limitations of any single analytic lens. By engaging directly with local communities and stakeholders, we seek to characterise the complex, multi-layered factors that contribute to CVD disparities and identify actionable interventions tailored to the unique needs of New York City's diverse populations.[15] Our approach enriches the data landscape and ensures that proposed solutions are grounded in the realities of those most affected by health inequities.[16]

AI4HealthyCities[17] is a global initiative led by the Novartis Foundation and partners that aims to understand the true drivers of cardiovascular risk and disease. Since its launch in New York City in September 2022, the initiative has collaborated closely with the New York City Department of Health and Mental Hygiene to align with municipal priorities and support the *HealthyNYC* campaign.

Our research team brings together investigators from Weill Cornell Medicine (WCM), University College London (UCL) and advisors from the Novartis Foundation. WCM has led the development of the machine learning models used in the earlier phase of the AI4HealthyCities initiative, in collaboration with partners such as Microsoft AI for Good Lab, the New York City Department of Health and Mental Hygiene, and other local stakeholders. The present study builds on that modelling work through three interrelated components—expert interviews, community roundtables and health vulnerability assessments (VAs)—each designed to address distinct research questions and engage different participant groups (figure 1).

To ensure methodological rigour and local relevance, the study will receive inputs from a New York City research steering committee, UCL's internal research

Figure 1 Phased approach to cardiovascular health research: expert, community and ethnographic perspectives. CV, cardiovascular; CVD, cardiovascular disease; EHR, electronic health records.

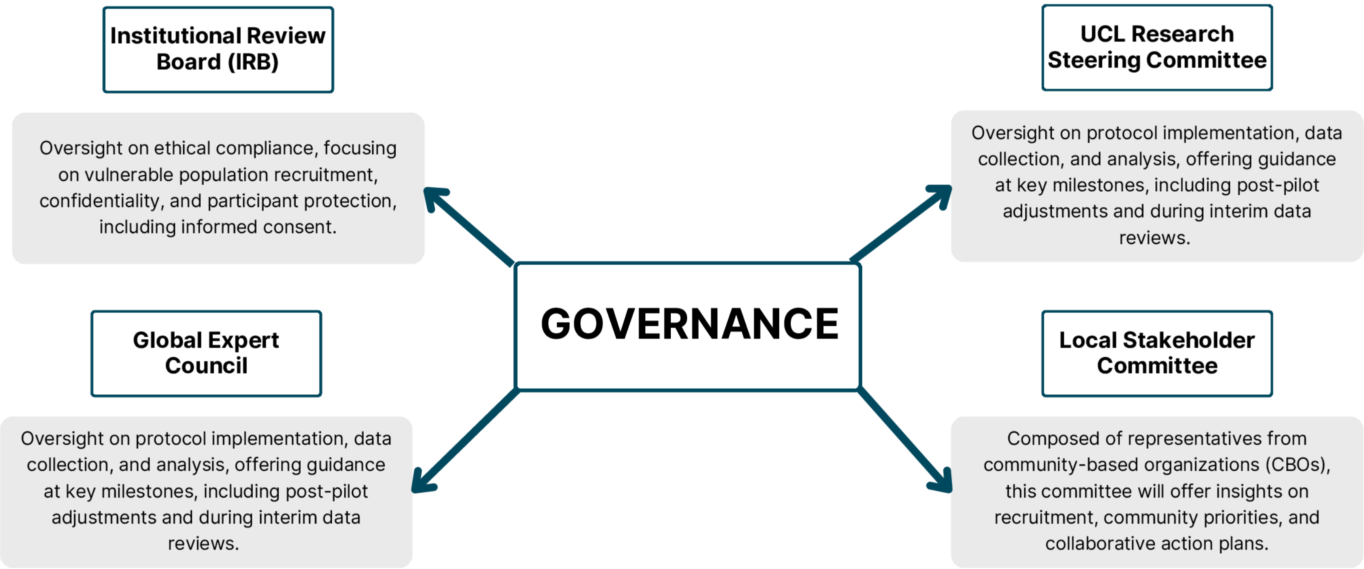

Figure 2 Governance framework for study implementation and data oversight. UCL, University College London.

steering group and the AI4HealthyCities Global Expert Council (figure 2).

Aims and objectives

Building on the quantitative results from AI4HealthyCities, this mixed-method ethnographic study aims to validate prior findings and address potentially critical data gaps, particularly regarding underrepresented populations in the health system data. The study will examine causal pathways of SDoH on CVD disparities and the mediators of cardiovascular outcomes, which are often heterogeneous and compounding, posing significant challenges for quantitative analysis. By capturing nuanced, qualitative insights, the study seeks to elevate the lived experiences of the city's residents, offering a deeper understanding of how these factors interact. Furthermore, it aims to equip local stakeholders with evidence-based, actionable insights to develop tailored interventions that can significantly improve cardiovascular outcomes in their communities.

METHODS AND ANALYSIS

The study employs a mixed-methods, ethnographic and multiphase design to examine the complex interplay of factors such as socioeconomic status, education, neighbourhood environment and access to healthcare (figure 3). While this qualitative study does not directly implement AI or machine learning techniques, it builds on the earlier quantitative phase of the AI4HealthyCities initiative, which applied advanced analytics to a combination of US census-tract level data on SDoH and electronic health records (EHR). The SDoH data were linked to health outcome data from the INSIGHT Clinical Research Network, which aggregates EHR from five major health systems in New York City.[18] In that phase, machine learning models were used to predict disease prevalence and health outcomes, and to identify the key drivers of cardiovascular risk in each census tract.

Integrating quantitative analytics with qualitative findings

To ensure that AI-generated insights meaningfully inform public health planning, this study integrates machine learning outputs from the quantitative phase of AI4HealthyCities[19] with in-depth ethnographic research. The goal is to establish an iterative feedback loop between data science, lived experience and stakeholder perspectives—enhancing both explanatory power and local relevance.

During expert interviews and community roundtables (Data Collection Phases I and II), selected findings from the quantitative analysis—such as spatial patterns of SDoH exposure and cardiovascular risk—will be shared with participants to prompt critical reflection. Questions such as 'Do these patterns resonate with your experience?' and 'What might be missing?' will be used to connect data-driven outputs to local realities. Responses will be analysed using a dedicated qualitative codebook, and insights will be synthesised and shared with the quantitative team to inform their ongoing model refinement. We will illustrate areas of alignment and divergence between community perspectives and data-driven outputs.

In data collection phase III (VAs), sampling will be informed by the outputs of an unsupervised machine learning analysis conducted by Yongkang Zhang and colleagues at WCM. This cluster analysis, developed specifically to align with the qualitative phase of the study, identified distinct neighbourhood 'archetypes' across New York City, each defined by a combination of SDoH associated with elevated cardiovascular risk. These profiles will guide purposive sampling by highlighting key characteristics within each cluster—such as low educational attainment, limited English proficiency or housing insecurity. Participants will be recruited to

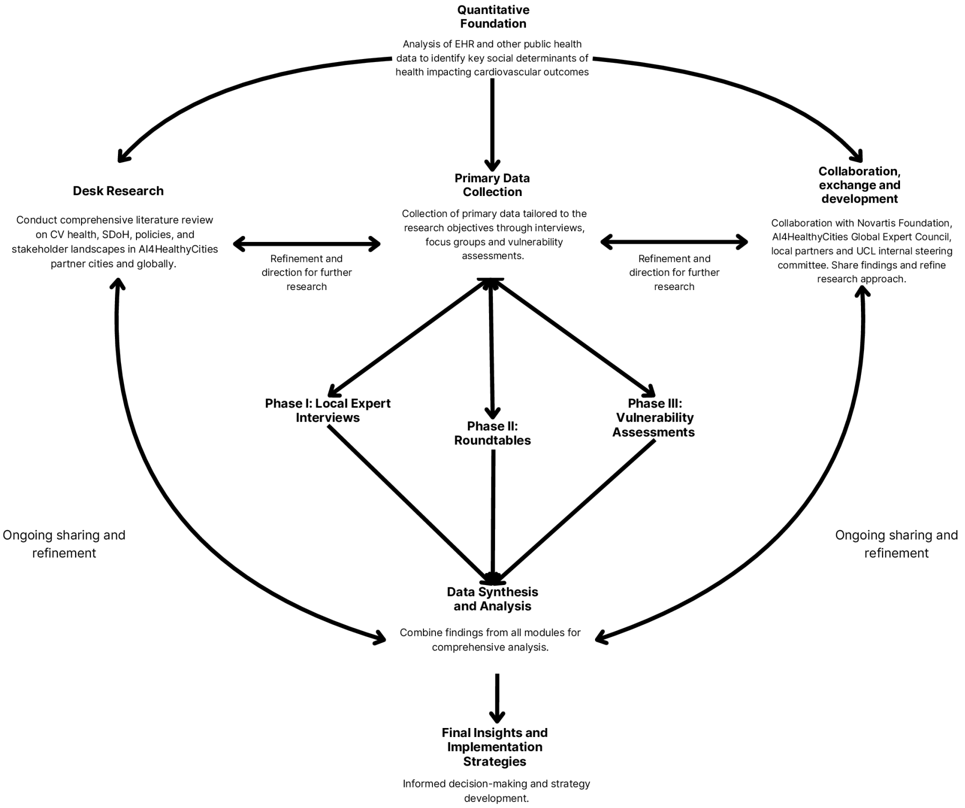

Figure 3 Research protocol framework: data collection, synthesis and implementation process. CV, cardiovascular; EHR, electronic health records; SDoH, social determinants of health; UCL, University College London.

reflect these defining features, enabling the construction of ‘case profiles’ that represent real-world expressions of modelled vulnerability and enrich the interpretive depth of the findings.

Our integrative approach represents a novel contribution to precision public health by embedding ethnographic insight and lived experience into the interpretation and application of AI-informed models—supporting more targeted, equitable and community-responsive interventions.

Data collection phase I: expert interviews

In the first phase, data will be collected through interviews with experts in leadership and decision-making positions and with a wide variety of areas of expertise in New York City. Experts will be identified through a combination of purposive and snowball sampling, drawing on networks within public health departments, healthcare institutions, academic centres and community-based organisations. Selection criteria will include demonstrated leadership in policymaking, programme design or service delivery related to cardiovascular health and SDoH (table 1).

Through the expert interviews, perspectives will be gathered on how this project can inform health policy and care delivery to mitigate the adverse effects of SDoH on cardiovascular health, with a special focus on the feasibility and impact of potential interventions or policies. Additionally, advice will be sought on identifying neighbourhoods, communities and populations of concern for the next phases of the project.

Research questions for Phase I:

1. What are the local experts’ views on the impact of SDoH and health behaviours on cardiovascular health in New York City? Which do they perceive to be most relevant, which might be modifiable?
2. What does the current initiative, programme or policy landscape look like regarding SDoH and cardiovascular health? Where are geographical and population ‘hot spots’ for prevalence of CVD and/or poor health outcomes?
3. How do local experts view AI-driven data insights generated in the quantitative phase, and how do they/would they use such insights?

Procedure

Expert interviews will be conducted remotely via telephone or online platforms. The interview will be audio-recorded, conducted by two researchers and follow a detailed interview guide (online supplemental Appendix 1) that captures information across the following broad domains of interest (table 2).

Sample size and sampling

Participants will be purposefully selected from stakeholders with relevant expertise or prior collaboration with the AI4HealthyCities initiative. The research team will also engage stakeholders who, while not directly interviewed, may help identify additional informants. Based on prior studies and feasibility in this setting, we aim to initiate Phase I with approximately 10–12 stakeholders.[20]

Table 1 Expert selection framework

Experts category	Stakeholder example
Policy makers	Municipal departments
Hospital leadership	Senior leadership in a municipal hospital
Insurance provider	New York City-based nonprofit health plan
Health providers	Medical doctors, community health workers
Researcher	Universities, research institutions
Community-based organisation leadership	Community clinics, philanthropy foundations
Specialities of the experts	
Primary healthcare services Health financing Population health and health equity in general Epidemiology Workforce Built environment Transportation Environment and pollution Education and English proficiency Housing SNAP benefits, social support Food security Racial and ethnic disparities Immigration Ageing/elderly (population over 65) Gender	

SNAP, Supplemental Nutrition Assistance Program.

The sample will then be expanded through participant referrals (snowball sampling) until sufficient thematic breadth has been achieved, guided by the Expert Selection Framework (table 1). This framework draws on key determinants identified in the AI4HealthyCities quantitative phase as well as insights from exploratory fieldwork conducted in June 2024. The research team will assess data sufficiency collaboratively, using practical and capacity considerations alongside the diminishing emergence of new insights as indicators of saturation.

Phase II: community roundtables

In Phase II, roundtables will be conducted with community representatives in three New York City boroughs: Brooklyn, Queens and the Bronx. These boroughs were selected based on findings from the AI4HealthyCities quantitative phase, which identified a high prevalence of CVD, elevated risk factors and distinct SDoH profiles compared with other boroughs.

Community representatives will be identified through established partnerships with local institutions: Elmhurst Hospital (Queens), Montefiore Medical Centre (Bronx) and EmblemHealth Neighbourhood Care (Brooklyn). These institutions will also serve as community-based sponsors and host the roundtables in their local centres. Community representatives include staff and leaders from community-based organisations, faith-based groups and local service providers with direct knowledge of neighbourhood health challenges and social vulnerabilities.

The roundtable discussions will help ensure that local priorities and lived experiences inform the research design and sampling for Phase III. The process is guided by UNICEF's Minimum Quality Standards and Indicators for Community Engagement.[21]

Research questions for Phase II:

1. What are the local community's priorities, needs and opportunities regarding SDoH and cardiovascular health?
2. What community-identified initiatives could positively impact cardiovascular health?
3. How do community representatives view and potentially use AI-driven data insights generated in the quantitative phase?

Procedure

The roundtable format was chosen over focus groups to foster community based organisation (CBO)

Table 2 Expert interview outline

I. Background	I.1. Role and experience
II. General SDoH	II.1. Understanding of and perspective on SDoH
	II.2. Populations and neighbourhoods
	II.3. COVID-19 and SDoH
	II.4. Presenting the quantitative results for validation
III. Expert knowledge	III.1. Specific questions around the expertise and experience of the participant relevant to the interview topic
	III.2. Relevant policies, initiatives and interventions
IV. Community and collaborative efforts	IV.1. Community engagement
	IV.2. Collaboration and partnerships
V. Future directions	V.1. Vision for the future

SDoH, social determinants of health.

collaboration and shared understanding of local SDoH challenges, pooling knowledge and resources.[22] Unlike focus groups, which typically focus on gathering data from individuals in a structured manner, roundtables provide a more open, participatory platform that promotes dialogue and inter-organisation cooperation. By focusing on CBO collaboration, the roundtable format ensures that CBOs can contribute to later translations of insights into action and collectively address SDoH in a way that is tailored to shared goals and community-specific needs.

A roundtable discussion guide (online supplemental Appendix 2) will be adapted collectively by the study leads, their research team and the community-based sponsor to capture the communities' perspectives on SDoH priorities, challenges and opportunities, existing resources and co-develop an action plan to potentially address the identified SDoH.

Sample size and sampling

The roundtable participants will be selected through desk research and recommendations from Phase I. In-depth desk research focusing on the Bronx, Brooklyn and Queens will create a detailed 'map' of the CBO landscape in each site. This CBO 'map' will include a longlist of potential roundtable participants, based on SDoH-relevant focus areas, geographical locations, as well as reach and impact assessments where available. Phase I will also provide input into the selection criteria through recommendations from experts and stakeholders in the AI4HealthyCities initiative. The research team and the community-based sponsor will then select participants for invitation to the roundtables, maintaining an even representation of gender, race and ethnicity.

The study aims to include 10–15 participants per roundtable, ensuring representation from organisations that address key SDoH identified in the AI4HealthyCities quantitative phase, including housing insecurity, food access, transportation, healthcare access and socioeconomic inequality. Participating organisations may include local government agencies, faith-based organisations and CBOs working across domains like maternal and child health, migrant health and community care. Efforts will also be made to include organisations serving historically marginalised populations, such as individuals experiencing homelessness or those with histories of incarceration and to ensure racial, ethnic and faith-based diversity across roundtable participants.

Phase III: local ethnographic research

This phase will focus on New York City populations that meet specific health vulnerability criteria such as being exposed to the five SDoH domains identified by the U.S. Department of Health and Human Services,[5] living in one of the three high-risk boroughs, or experiencing poor health outcomes, as identified in the AI4HealthyCities quantitative phase. The criteria will be further refined following Phases I and II results. The team will conduct mixed-method ethnographic research using an adapted VA instrument.[23]

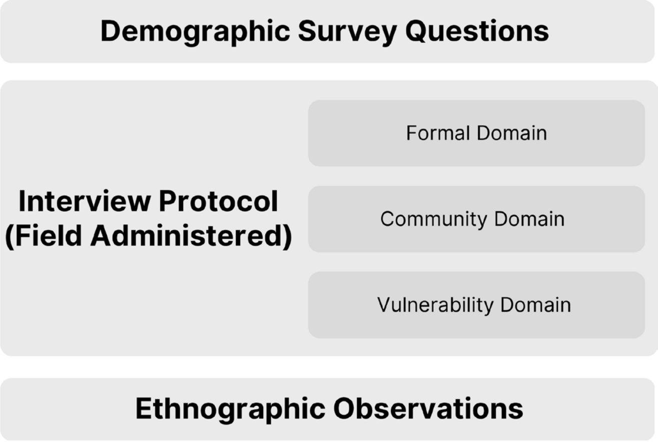

Figure 4 Data collection domains for vulnerability assessment.

The VA instrument is a data collection tool designed to understand the daily experiences and social environments that influence health, moving beyond individual health concerns. Used for over a decade, the instrument will explore the impact of SDoH on cardiovascular health, non-communicable disease risk and broader themes of resilience, equity and inclusivity in New York City, with a particular focus on marginalised and 'hidden' populations (for this study, the instrument will be referred to as CVD-VA). The CVD-VA data collection for each participant will include a semi-structured interview, a questionnaire and an ethnographic observation (figure 4).

Central to the VA approach is the use of semi-structured interviews. For the implementation of the CVD-VA, New York City residents will be engaged in conversations about health beliefs, available resources and services and community experiences. These interviews will explore the social, cultural and environmental dimensions of their neighbourhoods, covering issues such as community challenges, health equity and inclusivity. By focusing on these broader factors, the study aims to uncover how they shape health behaviours and outcomes.

Before each interview, a questionnaire will be administered to gather demographic and contextual information (e.g., gender, education, language and household income), offering a foundation for understanding how personal circumstances influence access to health resources and services.

Ethnographic observation will provide context for analysis, as researchers document observations of the neighbourhood environment, participant interactions and interview dynamics.

Research questions for Phase III:

1. What are the health and SDoH characteristics of 'hidden' populations in New York City (populations often omitted from surveys and other data sets)?
2. What are residents' priorities, needs and opportunities regarding SDoH and cardiovascular health?

3. What initiatives are likely to be effective and have a positive impact on cardiovascular health locally?
4. What are the ‘unknown unknowns’ regarding health vulnerability and resilience in New York City?

Procedure

Before commencing the main study, the research team will pilot the assessment protocol with 2–3 participants per cluster, recruited through community sponsors and CBOs from Phase II. The pilots will validate the data collection instruments and support researcher training.

Following the pilots, the research team will make needed adjustments to the instruments. Minor adjustments post-pilot will allow retention of pilot data for full analysis, while major adjustments will require separate documentation, reported to the UCL steering committee and exclusion of pilot data from further analysis.

The assessments will be conducted in the local community centres and will be audio-recorded. After each assessment, researchers will complete a one-page assessment summary, flagging adverse events or incidents that might require immediate follow-up, as well as key insights and contact referrals to other potential participants. Summaries will be uploaded daily to a shared server monitored by the study lead, supporting progress tracking, prioritisation of transcription and coding.

At the halfway point of data collection, after approximately 15–18 assessments, an in-depth review will be held to ensure demographic and characteristic variation targets are met, with protocol adjustments made if needed. The steering committee and local stakeholder committee will be consulted for further input, and the AI4HealthyCities Global Expert Council will receive a two-page update for any mid-study adjustments.

Sample size and sampling

Residents aged 18 and older from Brooklyn, Queens and the Bronx who are able to give informed consent will be invited to participate in the VA. Sampling will be informed by a cluster analysis conducted by Yongkang Zhang *et al* as part of the AI4HealthyCities quantitative phase. This analysis has to date identified five neighbourhood-level clusters, each defined by a distinct combination of SDoH associated with elevated cardiovascular risk. While full results are forthcoming, anticipated sampling characteristics include lower levels of educational attainment, limited English proficiency, housing insecurity, long commute times, economic instability and limited food security—appearing in different combinations across clusters. Sampling will be purposive and guided by these cluster profiles, with further refinement based on insights generated through Phases I and II of this study. This approach ensures that participants reflect the complexity and heterogeneity of local vulnerability contexts.

Recruitment will be conducted in partnership with CBO partners and supplemented by snowball sampling to reach individuals who may be less connected to formal service systems. Participants engaged through snowball sampling will be encouraged to refer others in their communities who may face similar or greater social and structural barriers.

The study aims to recruit approximately 90–100 participants, with an even distribution across the identified clusters. Particular attention will be given to including individuals who are less likely to appear in EHR based on healthcare utilisation, such as those with limited access due to lack of insurance or transport options. The final sample size will be adjusted as needed based on ongoing review of data richness and representativeness. This sample size target aligns with guidance from the *Cities Changing Diabetes Vulnerability Assessment: How-To Guide*,[24] as well as established qualitative research practice, which recommends adapting sample size to study design, analytic aims and population diversity.[25]

Recruitment progress and data quality will be monitored throughout, and the sampling approach may be modified to address gaps in representation. VAs will be conducted in trusted community settings, with remote participation options available to accommodate accessibility and participant preference.

Researcher training

Fieldworkers involved in data collection—experienced researchers trained in qualitative methods—will complete a 2 day training programme delivered by A-MV prior to the start of fieldwork. The training will cover core topics outlined in table 3 and is designed to ensure consistent, ethical and culturally responsive research practices across study sites. It emphasises relational ethics, reflexivity and community engagement, aligning with the study’s commitment to equity and the respectful inclusion of marginalised populations.

During data collection, the research team in the field will meet weekly—or more frequently if needed—to exchange insights, discuss challenges and ensure consistency across locations. A-MV will provide ongoing feedback, guidance and methodological support throughout, helping to maintain quality and steer the study in response to emerging findings.

Data analysis—all phases

To ensure participant privacy, all data will be anonymised before analysis. Personal identifiers such as names, organisational affiliations and contact details will be removed from all transcripts and datasets. Each participant will be assigned a unique identifier, and only anonymised data will be used in the analysis software NVivo.[26] Audio and visual recordings will be securely stored at WCM and destroyed after transcription. Access to data will be restricted to the research team, and all sensitive data will be stored on encrypted servers at WCM to ensure confidentiality. For each interview and roundtable, case files will be developed to include participants’ professional or demographic characteristics and location.

Table 3 Cardiovascular disease-vulnerability assessment researcher training outline

	Session
Day 1	Overview of AI4HealthyCities, SDoH and cardiovascular health in New York City
	Introduction to method and best ethnographic practice
	Working with, and for, 'vulnerable' populations
	Review and adaptation of the draft demographic questionnaire and interview guide, familiarisation with data collection instruments
Day 2	Familiarisation with technical equipment, adverse event reporting and other relevant reporting procedures
	Risk mitigation and safety practices, researcher well-being and support
	Practicum (mock assessments)

SDoH, social determinants of health.

A consistent data analysis strategy will be employed across all three data collection phases. Each transcript will be reviewed and coded independently by two researchers. Codes will be assigned to text segments that represent distinct ideas or concepts and subsequently grouped into broader themes through iterative discussion. NVivo's functionalities, such as node creation, case development and matrix queries, will facilitate the organisation and visualisation of the data, allowing for a comprehensive understanding of the participants' perspectives on SDoH, behaviour and CVD in New York City. Discrepancies will be resolved through consensus-building approaches, such as the Nominal Group Technique.[27]

A thematic analysis approach will be used to identify patterns and themes relevant to the study's research questions.[28] Coding will be guided by an initial codebook, developed in line with study aims and insights from earlier phases. This codebook will be refined iteratively to incorporate emerging themes and unanticipated patterns of meaning.

The analysis will attend closely to how participants' lived experiences relate to specific SDoH dimensions. Informed by the cluster analysis used for sampling, the research team will also generate a set of analytic exemplars—individual cases selected to illustrate key themes and configurations of SDoH across the five identified clusters. These cases will offer rich, context-specific insight into how structural and behavioural factors intersect in the lived experience of cardiovascular risk. The use of analytic exemplars follows established qualitative traditions of selecting illustrative cases to convey key configurations and deepen interpretive insight.[29]

Patient and public involvement

Stakeholder partnerships and community engagement are central to the study's design, implementation and policy relevance. The AI4HealthyCities initiative in New York City has built relationships with numerous local partners, with Elmhurst Hospital, Montefiore Medical Centre and EmblemHealth Neighbourhood Care formally supporting this study through outreach and engagement.

Community roundtables and expert interviews serve not only as data collection activities but also as platforms for identifying priorities and expanding the stakeholder network. Regular result-sharing sessions will enable community members to review emerging findings and shape how insights are interpreted and applied. This approach helps ensure that the study remains grounded in lived experience and informs context-specific, equity-focused interventions.

Ethics and dissemination

A participant information document will be shared with all participants before the interview or roundtable. Oral informed consent will be taken from each participant in Phase I and a written informed consent form will be obtained from each participant in Phase II and III. This study has been approved by the Ethics Committee of WCM (code number: 23–04025988).

The research team will synthesise findings from the preceding quantitative study and all research phases into a comprehensive report. A key outcome will be a characterisation of SDoH clusters and citizen exemplars—individuals who have elevated CVD risk and/or poor health outcomes because of exposure to certain or multiple SDoH. The identification of different clusters of SDoH across neighbourhoods will help the development of targeted interventions to improve cardiovascular health in the city population at large.

Findings will be disseminated through conferences, speaking engagements and peer-reviewed publications.

DISCUSSION

This study protocol is aimed at addressing persistent CVD disparities in New York City by examining how SDoH shape risk across underserved groups and those that remain hidden from traditional quantitative data evaluations. As part of the AI4HealthyCities initiative, the study will complement earlier machine learning analysis that mapped spatial patterns of cardiovascular vulnerability and extend this work by integrating ethnographic and community-based methods to validate, contextualise and deepen those insights.

By focusing on historically underrepresented urban populations, the study will address critical gaps in both

Table 4 Proposed outputs and their intended outcomes

Study output	Source phase(s)	Intended outcome/use
Expert interview report	Phase I (expert interview)	Guide borough-level prioritisation of interventions and inform resource allocation, validate and provide feedback on the quantitative results
Community insights on barriers and local priorities	Phase II (community roundtables)	Ensure action plans are grounded in lived experience and reflect community needs
Exemplars of vulnerable populations	Phase III (vulnerability assessments)	Inform equity-focused policies and reach underserved or 'hidden' populations
Policy briefs	Synthesised across Phases I–III	Communicate actionable, place-based recommendations to the municipal health department, WCM and CBOs
Stakeholder workshops, webinars and visual dashboards	Dissemination phase	Co-develop and refine localised action plans targeting key SDoH such as housing, transport and healthcare access

CBO, Community-based organization; SDoH, social determinants of health; WCM, Weill Cornell Medicine.

predictive modelling and public health literature. Its mixed-method design facilitates a nuanced understanding of the systemic barriers and social contexts that shape health vulnerability—factors typically omitted from EHR and other administrative datasets. Each phase of the study offers a distinct perspective: expert interviews (Phase I) provide systems-level insights; roundtables (Phase II) elicit priorities and perceived service gaps in local communities; and VAs (Phase III) explore the lived realities of residents situated within distinct SDoH-defined clusters.

A key innovation of our protocol is the use of purposive sampling guided by machine learning cluster analysis from the AI4HealthyCities quantitative phase. This approach enables the development of analytic exemplars—case profiles that illustrate typical patterns of social and structural disadvantage within each cluster. These exemplars are intended to complement predictive models and inform the design of more precise, equity-oriented public health interventions that address both individual-level risk and systemic inequities.

The participatory design of the study enhances its relevance and potential for impact. Insights from all three phases will be synthesised and translated into borough-specific outputs—such as policy briefs, stakeholder workshops, webinars and visual dashboards—co-developed with local partners such as the WCM and CBOs. These outputs (table 4) will inform the design of locally tailored action plans to address priority SDoH domains, including housing insecurity, transportation barriers and limited access to care.

While rooted in the New York City context, this study offers a transferable model for integrating lived experience into precision public health approaches. By combining advanced analytics with community-based methods, it contributes a novel framework for generating actionable, socially grounded evidence to inform both local intervention design and broader systems change.

Strengths and limitations

A key strength of this study lies in its integrative and participatory design. Collaboration with local stakeholders, CBOs and global experts enhances the study's policy relevance, credibility and potential for local impact.

The relatively small qualitative sample enables in-depth exploration of lived experience and the development of rich, analytically grounded case profiles. However, this also introduces limitations: findings may risk being episodic or not generalisable beyond the sampled clusters and boroughs. To mitigate this, sampling is carefully structured to reflect SDoH cluster profiles identified in the prior quantitative phase and insights from the qualitative data will be mapped onto broader datasets and analytic models developed through the AI4HealthyCities initiative. This layered approach enhances interpretive depth while supporting relevance to population-level planning.

Other potential limitations include reliance on self-reported data, which may be affected by recall or social desirability bias, and the challenges of maintaining consistency across an iterative, multi-phase study design. These will be addressed through structured training, ongoing team debriefs and shared analytic protocols to ensure rigour and coherence across phases.

Acknowledgements We would like to thank Sri Lekha Tummalapalli at Weill Cornell Medicine for her valuable feedback on the tool development and support in data collection and Lillian Kerns for her editing support. We also acknowledge our CBO hosts as research partners, in particular Terri Rhymes-Lowery and Nazneen Rahman at Emblem Health, Kevin Fiori and Renee Whiskey-LaLanne at Montefiore, Helen Arteaga-Landaverde at Elmhurst, and other partner organisations that are yet to be determined.

Contributors A-MV and YZ conceptualised the study. A-MV, EB and WZ drafted a first version of the protocol. All authors (A-MV, EA, EB, DN, PS, RP, WZ and YZ) participated in the revisions of the manuscript and have given final approval of the version to be published. A-MV is the guarantor.

Funding The project is funded by the Novartis Foundation (Grant Number: 107119). Colleagues from the Novartis Foundation contribute actively to the project by making available contacts in Novartis Foundation stakeholder groups, liaising with local government and other key institutions, and supporting the implementation of actionable insights post-study. Results and findings will be generated independently. The funder aids any enquiry, audit or investigation related to the funded work.

Competing interests None declared.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by the Ethics Committee of Weill Cornell Medicine (code number: 23-04025988). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information. This is a study protocol; no participant datasets have yet been generated. The authors confirm that the research materials supporting this study, including the semi-structured interview guides, are available within the article and its supplementary materials

ORCID iDs
Emma Boxley http://orcid.org/0000-0001-6276-758X
Wenqin Zhang http://orcid.org/0000-0002-2922-218X

REFERENCES

1 Li W, Castro A, Gurung S, *et al*. *Summary of Vital Statistics, 2022*. New York, NY: Bureau of Vital Statistics, New York City Department of Health and Mental Hygiene, 2023.
2 New York City Department of Health and Mental Hygiene. *Prevalence of Hypertension, Awareness, Treatment and Control in New York City. Epi Data Brief, No. 135*. New York, NY: New York City Department of Health and Mental Hygiene, 2023.
3 World Health Organization. About social determinants of health. 2020.
4 Marmot M, Wilkinson R, eds. *Social Determinants of Health*. Oxford: Oxford University Press, 2006.
5 Healthy New York City. New york city department of health and mental hygiene. New York City Health; 2020. Available: https://www.nyc.gov/site/doh/about/about-doh/healthyNewYorkCity.page
6 Hayes RB, Lim C, Zhang Y, *et al*. PM2.5 air pollution and cause-specific cardiovascular disease mortality. *Int J Epidemiol* 2020;49:25–35.
7 Brook RD, Newby DE, Rajagopalan S. Air Pollution and Cardiometabolic Disease: An Update and Call for Clinical Trials. *Am J Hypertens* 2017;31:1–10.
8 Berkman LF, Glass T. Social integration, social networks, social support, and health. In: Berkman LF, Kawachi I, eds. *Social Epidemiology*. New York: Oxford University Press, 2000: 158–62.
9 Kawachi I, Subramanian SV, Almeida-Filho N. A glossary for health inequalities. *J Epidemiol Community Health* 2002;56:647–52.
10 Heiman HJ, Artiga S. Beyond health care: The role of social determinants in promoting health and health equity. *Health Equity Series* 2015;1–10.
11 Diaz CL, Shah NS, Lloyd-Jones DM, *et al*. State of the Nation’s Cardiovascular Health and Targeting Health Equity in the United States: A Narrative Review. *JAMA Cardiol* 2021;6:963–70.
12 Williams DR, Lawrence JA, Davis BA. Racism and Health: Evidence and Needed Research. *Annu Rev Public Health* 2019;40:105–25.
13 Braveman P, Egerter S, Williams DR. The social determinants of health: coming of age. *Annu Rev Public Health* 2011;32:381–98.
14 Creswell JW, Plano Clark VL. *Designing and Conducting Mixed Methods Research*. Sage Publications, 2018.
15 Berkman LF, Kawachi I, Glymour MM, eds. In: *Social Epidemiology*. Oxford University Press, 2014.
16 Marmot M, Friel S, Bell R, *et al*. Closing the gap in a generation: health equity through action on the social determinants of health. *Lancet* 2008;372:1661–9.
17 AI4HealthyCities. AI4HealthyCities – a global initiative. n.d. Available: https://www.novartisfoundation.org/ai4healthycities
18 INSIGHT Clinical Research Network. Data overview. INSIGHT Clinical Research Network; 2023. Available: https://insightcrn.org/data-overview/
19 Tanner D, Zhang Y, Chang JE, *et al*. Machine learning to evaluate the relationship between social determinants and diabetes prevalence in New York City. *bmjph* 2024;2:e001394.
20 Guest G, Bunce A, Johnson L. How many interviews are enough? An experiment with data saturation and variability. *Field Methods* 2006;18:59–82.
21 UNICEF. Minimum quality standards and indicators for community engagement. UNICEF Middle East and North Africa; 2020. Available: https://www.unicef.org/mena/reports/community-engagement-standards
22 Pennel CL, Carpender SK, Quiram BJ. Rural health roundtables: a strategy for collaborative engagement in and between rural communities. *Rural Remote Health* 2008;8:1054.
23 Napier AD. The rapid assessment of vulnerable populations: a “barefoot” manual. Sonar-Global; 2020.
24 Napier AD. *Diabetes Vulnerability Assessment: How-To Guide*. London: University College London & Cities Changing Diabetes, 2017.
25 Hennink M, Kaiser BN. Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Soc Sci Med* 2022;292:114523.
26 QSR international pty ltd. Nvivo (version 12) [software]. 2018.
27 Potter M, Gordon S, Hamer P. The nominal group technique: A useful consensus methodology in physiotherapy research. *N Z J Physiother* 2004;32:126–30.
28 Joffe H. *Thematic Analysis*. Chichester: Wiley-Blackwell, 2012:209–23.
29 Patton MQ. *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*. Thousand Oaks (CA): SAGE Publications, 2015.