ABSTRACT: Air pollution is a growing public health concern in developing countries and poses a huge epidemiological burden. Despite the growing awareness of all effects of air pollution, the evidence linking air pollution and health effects is sparse. This requires environmental exposure scientist and public health researchers to work more cohesively to generate evidence on health impacts of air pollution in developing countries for policy advocacy. In the Global Environmental and Occupational Health (GEOHealth) Program, we aim to build exposure assessment model to estimate ambient air pollution exposure at a very fine resolution which can be linked with health outcomes leveraging well-phenotyped cohorts which have information on geolocation of households of study participants. We aim to address how air pollution interacts with meteorological and weather parameters and other aspects of the urban environment, occupational classification, and socioeconomic status, to affect cardiometabolic risk factors and disease outcomes. This will help us generate evidence for cardiovascular health impacts of ambient air pollution in India needed for necessary policy advocacy. The other exploratory aims are to explore mediating role of the epigenetic mechanisms (DNA methylation) and vitamin D exposure in determining the association between air pollution exposure and cardiovascular health outcomes. Other components of the GEOHealth program include building capacity and strengthening the skills of public health researchers in India through variety of training programs and international collaborations. This will help generate research capacity to address environmental and occupational health research questions in India. The expertise that we bring together in GEOHealth hub are public health, clinical epidemiology, environmental exposure science, statistical modeling, and policy advocacy.

KEYWORDS: Air pollution, cardiovascular diseases, cohort studies, India, particulate matter

Background
Air pollution is a growing concern in developing countries contributing to >620 000 deaths among Indians annually.1 As per recent estimates from Global Burden of Disease (GBD), 2017, 10.6% of total deaths, 6.4% of total disability-adjusted life years, and 30% of cardiovascular deaths in India are attributable to ambient PM2.5 exposures.2 The annual population-weighted mean exposure to ambient particulate matter PM2.5 in India was 89.9 µg/m3 (95% confidence interval [CI]: 67.0-112.0) in 2017,3 far exceeding the World Health Organization (WHO) target of 10 µg/m3.1,2 Due to the limited spatial coverage of ground-level monitoring stations, there has been an increasing interest in the use of alternative methods to predict air pollution exposures in India. The majority of evidence of health effects of air pollution has come from developed countries which does not account for the complex rapid urbanization happening in developing countries. Moreover, the composition of air pollution is very different in low- and middle-income countries (LMICs), and the health impact is influenced by several other meteorological, demographic, and built environment factors. The health impacts of air pollution are myriad and include respiratory and cardiometabolic diseases to reproductive disorders and infant morbidity.4 After respiratory disorders, cardiometabolic disorders are the most important diseases that are attributable to exposure to ambient air pollution,5 specifically PM2.5.6 Prospective cohort studies are very useful for studying multiple outcomes and are among the strongest designs for evaluating causal effects outside of randomized controlled trials. While several cohort studies exist in India, they have largely focused on pregnancy and birth outcomes. There are very few well-phenotyped adult cohorts focused on cardiometabolic health

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
that are now beginning to study the impact of environmental risk factors on chronic diseases.

Considering that India is home to one-fifth of the world's population, studies specific to the Indian scenario are urgently needed to shift policy discourse around ambient air pollution. Findings from a large multicountry study reported that stroke and ischemic heart disease (IHD) were 2 largest contributors for premature deaths and accounted for 74% of the total premature deaths in South and South-East Asia,7 with India contributing the most premature deaths of any country in the region. To date, only 1 study has evaluated the impact of air pollution on cardiometabolic disease outcomes in India: a time-series study in Varanasi that found that the achievement of the WHO air quality standard would prevent 1900 premature deaths every year.8 With improved air quality, modeled data estimates indicate that 24.0% of IHD and 18.5% of stroke deaths in India could be averted.9

The overall goal of the present GEOHealth (Global Environmental and Occupational Health) India Hub program is to leverage an ongoing cohort study in India, the Centre for cArdiometabolic Risk Reduction in South-Asia (CARRS) surveillance study,10 to evaluate the prospective effects of ambient air pollution on cardiometabolic health outcomes and associated traits. The study also leverages publicly available information on air quality, meteorological variables, and other environmental factors (like land use and emission inventories) for the analysis of complex spatiotemporal data and multipollutant exposures, and can serve as a proof of concept for other ongoing cohort studies in India and around the world. The analysis is facilitated by previously developed and validated prediction models for PM2.5 in the United States that combine land-use regression (LUR) with satellite-derived aerosol optical depth (AOD) data to estimate particle exposure.11

In addition, we are building on existing laboratory capacity in India to explore factors that may mediate the association between air pollution and cardiometabolic disease such as DNA methylation and vitamin D levels. Methylomics provides a unique opportunity to reconstruct past exposures, particularly those such as air pollution and other exposures that augment oxidative stress, to which the methylome is exquisitely sensitive.12 Air pollution in particular is known to alter DNA methylation that in turn is known to be associated with cardiometabolic health outcomes.13-15 Similarly, there is strong evidence on effect of Vitamin D levels and deficiency with various cardiometabolic outcomes.16-21 Particulate air pollution reduces effective UV-B exposure, which is critical for producing the biologically active form of vitamin D, and contributes to variation in the UV index, along with other factors such as latitude, season, skin pigmentation, and the use of sun-protective wear.22,23

Similar to this context, the GBD estimates emerge from a low-resolution chemical transport model that estimates particulate matter levels with considerable error and exposure-response functions based largely on research from low- and mid-level exposure settings.24 Most of the air pollution studies

in LMICs (largely represented by China) are time-series/ecological study design with a short observation period which often has no spatiotemporal resolution of pollution parameters and focuses only on short-term health outcomes.25 Hence, this research will provide crucial, country-specific evidence of the health effects of air pollution to provide evidence for appropriate policy reforms. This study will also demonstrate how existing cohorts with longitudinal information on cardiometabolic health can be used to understand emerging risk factors and provide timely scientific data to inform cardiometabolic disease prevention and air pollution mitigation policies.

**CARRS Cohort**

The CARRS surveillance study is a hybrid cohort-modeled cross-sectional study involving a baseline survey followed by repeat surveys carried out in subsequent years with a response rate of approximately 85%. The CARRS participants were recruited at baseline in 2010-2012 (Cohort 1) from 3 urban sites, Delhi and Chennai in India and Karachi in Pakistan. Thereafter, more participants were recruited from these cities in 2014-2016 (Cohort 2) to achieve larger sample size to understand the incidence of cardiometabolic risk factors, diseases, comorbidities, and mortality10 in this south-east Asian region.

Households were selected in each of the 3 cities using a multistage cluster random sampling technique from each ward and census enumeration blocks. Two participants, 1 male and 1 female, aged 20 years or older and permanently residing in the household, were selected from each household using “Kish method” used in the WHO’s (World Health Organisation stepwise Approach to Surveillance) (STEPS) surveys. Pregnant women and bed-ridden individuals were excluded from the study, and information on basic demographic details of these excluded individuals was recorded along with non-participating eligible participants. To provide consistency and reproducibility of the results across multiple sites and across different follow-ups, comprehensive and uniform data collection instruments were used to capture measurements. The details of all the data collection and study procedures have been described previously.10

The CARRS participants (Table 1) were phenotyped for a range of cardiovascular disease (CVD) risk factors at baseline. Thereafter, every year these participants are being followed for CVD events and additionally for lifestyle factors, physical examinations, and biological samples as well in every alternate year (Table 2). This intense phenotype and built environment data are integrated into a Geographical Information System (GIS)–linked database. The data on geocoded residence of the participants and how long they have lived at their present location provide an excellent opportunity to estimate air pollution exposure levels. As far as possible, we are also trying to gather information on migration and also geocode current residence of the participants, if they have migrated within the cities. Written informed consent was obtained from CARRS
participants to utilize their de-identified phenotype data and stored de-identified biological samples for future cardiovascular research.

The GEOHealth program is utilizing information from the Delhi and Chennai sites having a very different cardiometabolic profile along with different geospatial determinants and air pollution levels and composition. We are restricting our GEOHealth proposed objectives to only Indian cities for logistic and feasibility purposes around estimating ambient air pollution exposure levels.

**Research Aims**

*Aim 1: Estimate air pollution exposure in Chennai and Delhi at fine spatiotemporal resolution*

We will develop and validate exposure models to estimate daily exposure to fine particulate matter (PM$_{2.5}$) at a 1 km $\times$ 1 km spatial resolution from 2010 to 2016.$^{26}$ The predicted concentrations will be used to assign ambient air pollution exposure values to >15 000 CARRS households in Chennai and New Delhi. The prediction models are based on machine learning methodologies.
and ensemble averaging while using ground monitoring data, satellite measurements, meteorological data, land-use variables, and emission inventories. The major advantage of this modeling exercise is that it enables us to obtain neighborhood-level ambient concentrations irrespective of the presence or absence of the monitoring network. In addition, the fine spatiotemporal resolution of the exposure enables us to estimate effects on health outcomes at an individual level over time. The future aim is to extend this model across all of India as well as over longer periods of time and also for other pollutants, including NO₂ and ozone.

**Aim 2: Estimate the association between exposure to air pollution, temperature, cardiometabolic disease risk factors and diseases, and risk factors, and identify potential susceptible subpopulations**

We aim to estimate that the association of ambient air pollution exposure from Aim 1 is within the CARRS cohort. In addition to estimating main effects, we will evaluate effect modification by population subgroups, based on their socioeconomic status, built environment, occupational status, and nutritional status to identify those most susceptible groups. The minimum detectable extreme quartile relative risks for 80% power with a 5% Type I error rate were calculated for the CARRS-1 cohort in Delhi and Chennai (n = 12 271) using the observed 2- and 3-year follow-up rates. To assess power of this study to detect correlations in prospective changes in markers of cardiometabolic (CM) risk such as HBA1c, lipid profiles, serum creatinine, and blood pressure, we find that we will have 80% power to detect correlations as low as 2% to 3% longitudinally as well as cross-sectionally with baseline air pollution (AP) constituents, given the baseline sample size and observed follow-up rates at 2 and 3 years.

**Aim 3: Characterize DNA methylation patterns—associated cardiovascular events and explore whether DNA methylation mediates the association between air pollution exposure and cardiovascular outcomes**

Given the limited sample size and budget, the methylationomics aim will focus on cardiometabolic outcomes through a nested case-control design (approximately n = 192 cardiovascular events and controls [myocardial infarction/strokes or CVD deaths] to explore whether methylation patterns mediate the effect of PM₂.₅ exposures on CVD events using mediation analyses; 96 cases and controls will provide 98% power to detect 5% methylation difference between the groups assuming a conservative standard deviation (SD) of 5% in each group at \( P = 1.1 \times 10^{-6} \).

**Aim 4: Explore the association between ambient exposure to air pollution and blood vitamin D levels**

We will explore associations between air pollution exposure and blood vitamin D levels (measured as 25-OH-D levels), and we will examine whether vitamin D levels are a mediator of the association between air pollution and cardiometabolic outcomes using causal mediation analysis. We will randomly sample 600 CARRS participants from Delhi and Chennai who will provide 80% power or more to detect correlations as low as 2% to 3% longitudinally as well as cross-sectionally at baseline.

The detailed analysis approach for all the 4 research aims is described in Table 3.

### Table 3. Proposed analysis approach for the research aims.

| RESEARCH AIMS | ANALYSIS APPROACH |
|---------------|-------------------|
| Aim 1: Estimation of air pollution exposure | A 3-stage hybrid model utilizing machine learning algorithms, ensemble averaging, and tensor product smoothing |
| Aim 2: Estimation of the association between air pollution exposure and cardiometabolic disease risk factors and diseases | • Longitudinal mixed effects modeling  
• Time to event modeling for incidence data |
| Aim 3: Characterization of DNA methylation patterns associated cardiovascular health effects and cardiometabolic (CM) outcomes | • Linear regression analyses to identify the DNA methylation profile associated with cardiovascular disease (CVD) events  
• Mediation analyses to explore whether these associations are mediated by PM₂.₅ levels |
| Aim 4: Estimation of the association between air pollution exposure and blood vitamin D levels | Mediation analyses to examine the mediating effect of vitamin D in association between PM₂.₅ and cardiovascular disease (CVD) |

**Capacity Building**

One of the major goals of this study is to build training and research capacity to address environmental and occupational health research questions beyond the specific aims of this grant. We have laid out multiple different ways to achieve this: faculty from the Harvard T.H. Chan School of Public Health (HSPH) will collaborate and train investigators from Centre for Chronic Disease Control (CCDC) and Public Health Foundation of India (PHFI):

- Mentored training program wherein the researchers from PHFI and CCDC will work and learn along with an identified mentor at HSPH.
- Summer exchange visits at HSPH, to further strengthen capacity of the researchers from PHFI and CCDC.
• Master’s training program (fully sponsored MPH or MSc in environmental health at HSPH).
• PHFI and CCDC are running 5-day short courses to complement research activities. Courses on introduction to environmental health, research ethics in environmental health, air pollution epidemiology, food and the environment, principles of toxicology, environmental exposure assessment, occupational health and medicine, causal modeling, and air pollution, climate, and health: modeling and methods have been conducted. The design of the courses is such that in the first couple of years, HSPH faculty will lead the course and PHFI and CCDC faculty can take lead thereafter.
• HSPH along with faculty from PHFI and CCDC are working together in developing curriculum for MPH in environmental health track in India.

Innovation
The innovation of our study lies in the methodology of exposure assessment and the estimation of health effects in a cohort study with longitudinally measured health outcomes in 2 major Indian cities. To date, ambient air pollution exposure assessment in India has been reliant on source apportionment, emission inventories, satellite remote sensing, and LUR techniques. Due to inherent limitations of each methodology, the exposure estimates are often coarse in spatial resolution and/or fail to capture temporal variability. The methodology used in this study incorporates the strengths of multiple machine learning techniques along with the most relevant sources of data, thus providing high resolution on both spatial and temporal scales. To the best of our knowledge, the GEOHealth study is the first to assess the effects of ambient air pollution on multiple incident cardiometabolic disease and associated risk factors in India with one of the highest ambient levels of PM$_{2.5}$ in the world. Our results will also help in understanding the complex interplay of the role of the air pollution, the built environment, occupational exposure, and sociodemographic factors on cardiometabolic risk factors in India which is facing major developmental and epidemiological transitions.

Air pollution modeling work that will be undertaken in this project will improve upon the GBD estimates. This exposure assessment approach is much more rigorous and comprehensive, with a fine spatiotemporal resolution of 1 km by 1 km compared with GBD resolution of 11 km by 11 km. This can serve as a resource even for other health outcomes in that space-time boundary. In addition, we have individual-level information on health and other variables that help in providing more reliable exposure-response curves.

Impact
Through this GEOHealth Hub, we are enhancing research activities and providing scientific infrastructure, training, and capacity building to characterize the relationship between air pollution and cardiometabolic risk factors and diseases in India. This is the largest and most extensive effort to address this issue in India, an LMIC with very high air pollution levels and prevalence of CM risk factors. The study is expected to produce results that will (1) advance the science regarding exposure assessment and effects of air pollution on CM risk factors; (2) inform urban planning and transportation planning policies designed to improve health in India, while taking into account air pollution exposures; (3) contribute important information to the gap in knowledge on the environmental contributions to CM risk factors and how effects are mediated by vitamin D levels and epigenetic mechanisms; and (4) inform development and implementation of targeted regulations, policies, and interventions to promote healthier living in India. In addition, this research can serve as a template for developing national-level pollution models, which can be further used to study the effects of pollution on diverse health outcomes.

Acknowledgements
We are thankful to the entire India CARRS (Centre for cardioMetabolic Risk Reduction in South-Asia) Delhi and Chennai teams for their cooperation in GEOHealth (Global Environmental and Occupational Health) study. We would also like to acknowledge all the members of the GEOHealth Team for their contributions.

Author Contributions
G.K.W. wrote the first draft of the manuscript. D.P., K.S.R., J.S., R.A.C., G.K.W., P.K.D., B.K., and M.S.M. conceptualized the study. G.K.W., S.J., S.M., L.M.J., N.L.S., M.S.M., K.K.M., and D.K. are implementing and managing the study. All the authors provided comments and finalized the manuscript. Apart from the listed authors, the people mentioned in the “Acknowledgements” section under India GEOHealth (Global Environmental and Occupational Health) Team are contributing in different components of the study.

Ethical Approval
The GEOHealth (Global Environmental and Occupational Health) Hub Grant was reviewed and cleared by the Institutional Ethics Committees of Public Health Foundation of India (PHFI; Ref. No. TRC-IEC.264.2/15), Centre for Chronic Disease Control (CCDC; Ref. No. CCDC_IEC_10_2017), and Madras Diabetes Research Foundation of India (MDRF) and All India Institute of Medical Sciences (AIIMS) (Ref. No. IEC/NP-401/09.10.2015). The GEOHealth research grant is nested within the CARRS (cArdiometabolic Risk Reduction in South-Asia) cohort where participants provided informed written consent to utilize their de-identified phenotype data and biological samples for future studies and publish the research findings.

The Global Environmental and Occupational Health (GEOHealth) Team
Principal Investigators: D Prabhakaran, K Srinath Reddy, Joel Schwartz, and Richard A Cash.
Global Environmental and Occupational Health Team (GEOHealth) Core Team (Co-Investigators and Investigators): Gagandeep Kaur Walia, Siddhartha Mandal, Suganthi Jaganathan, Poornima Prabhakaran, Saishef Mohan, Melina S Magsumbol, Kishore K Madhipatla, Preet K Dhillon, Bhargav Krishna, Dimple Kondal, Safraj S Hameed, Roopa Shivasankar, Lindsay M Jaacks, and Nancy L Sieber.

Other GEOHealth Project Members: Jyothi S Menon, Shivam Pandey, Kalpana Singh, Garima Rautela, Ruby Gupta, Naveen Kaushik, and Praggya Sharma.

Co-opted members including other HSPH and CARRS Team: Francesca Dominici, Douglas Dockery, Petros Koutrakis, David Christiani, Nagarjun Konduru, Gary Adamkiewicz, Kaushik, and Praggya Sharma.

REFERENCES

1. India State-Level Disease Burden Initiative Air Pollution Collaborators. The impact of air pollution on deaths, disease burden, and life expectancy across the states of India: the Global Burden of Disease Study 2017. Lancet Planet Health. 2019;3:e226-e39.
2. Institute for Health Metrics and Evaluation (IHME). GBD Compare Data Visualization. Seattle, WA: IHME, University of Washington; 2016. http://vizhub.healthdata.org/gbd-compare.
3. World Health Organization (WHO). Ambient air pollution: a global assessment of exposure and burden of disease. 2016. https://apps.who.int/iris/bitstream/handle/10665/230514/9789241511353-eng.pdf?sequence=1.
4. Thurston GD, Kipen H, Annesi-Maesano I, et al. A joint ERS/ATS policy statement: what constitutes an adverse health effect of air pollution? An analytical framework. Eur Respir J. 2017;49:1600419.
5. Brook RD, Rajagopalan S. Air pollution and cardiometabolic disease: an update and call for clinical trials. Am J Hypertens. 2017;31:1-10.
6. Brook RD, Rajagopalan S, Pope CA 3rd, et al. Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association. Circulation. 2010;121:2371-2378.
7. Shi Y, Matsunaga T, Yamaguchi Y, Zhao A, Li Z, Gu X. Long-term trends and spatial patterns of PM2.5-induced premature mortality in South and Southeast Asia from 1999 to 2014. Sci Total Environ. 2018;631-632:1504-1514.
8. Jain Y, Dey S, Chowdhury S. Ambient PM2.5 exposure and premature mortality burden in the holy city Varanasi, India. Environ Pollut. 2017;226:182-189.
9. Chowdhury S, Dey S. Cause-specific premature death from ambient PM2.5 exposure in India: estimate adjusted for baseline mortality. Environ Int. 2016;91:283-290.
10. Nair M, Ali MK, Ajay VS, et al. CARRS Surveillance study: design and methods to assess burdens from multiple perspectives. BMC Public Health. 2012;12:268.
11. Correia AW, Pope CA 3rd, Dockery DW, Wang Y, Ezzati M, Dominici F. Effect of air pollution control on life expectancy in the United States: an analysis of 545 U.S. counties for the period from 2000 to 2007. Epidemiology. 2013;24:23-31.
12. Bollati V, Baccarelli A. Environmental epigenetics. Heredity (Edinb). 2010;105:105-112.
13. Plusquin M, Guida F, Poldoro S, et al. DNA methylation and exposure to ambient air pollution in two prospective cohorts. Environ Int. 2017;108:127-136.
14. Vick AD, Burris HH. Epigenetics and health disparities. Curr Epidemiol Rep. 2017;4:31-37.
15. Rider CF, Carlsten C. Air pollution and DNA methylation: effects of exposure in humans. Clin Epigenetics. 2019;11:131.
16. Wang TJ, Pencina MJ, Booth SL, et al. Vitamin D deficiency and risk of cardiovascular disease. Circulation. 2008;117:503-511.
17. Anderson JL, May HT; Horne BD, et al. Relation of vitamin D deficiency to cardiometabolic risk factors, disease status, and incident events in a general healthcare population. Am J Cardiol. 2010;106:963-968.
18. Pile S, Dohnig H, Fischer JE, et al. Low vitamin d levels predict stroke in patients referred to coronary angiography. Stroke. 2008;39:2611-2613.
19. Deleskog A, Hilding A, Brismar K, Hamsten A, Efendic S, Ostenson CG. Low serum 25-hydroxyvitamin D level predicts progression to type 2 diabetes in individuals with prediabetes but not with normal glucose tolerance. Diabetologia. 2012;55:1668-1678.
20. Holick MF. Environmental factors that influence the cutaneous production of vitamin D. Spec Biol Med. 1995;61:638S-645S.
21. Parker J, Hashmi O, Dutton D, et al. Levels of vitamin D and cardiometabolic disorders: systematic review and meta-analysis. Maturitas. 2010;65:225-236.
22. Holick MF. Environmental factors that influence the cutaneous production of vitamin D. Am J Clin Nutr. 1995;61:638S-645S.
23. Glorup H, Mikkelsen K, Poulsen L, et al. Commonly recommended daily intake of vitamin D is not sufficient if sunlight exposure is limited. J Intern Med. 2000;247:260-268.
24. Cohen AJ, Brauer M, Burnett R, et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. Lancet. 2017;389:1907-1918.
25. Jaganathan S, Jaacks LM, Magsumbol M, et al. Association of long-term exposure to fine particulate matter and cardiometabolic diseases in low- and middle-income countries: a systematic review. Int J Environ Res Public Health. 2019;16:E2351.
26. Mandal S, Madhipatla KK, Gutrikunda S, Klooq I, Prabhakaran D, Schwartz JD. Ensemble averaging based assessment of spatiotemporal variations in ambient PM2.5 concentrations over Delhi, India, during 2010-2016. Atmospheric Environment. 2020;224:117309.
27. Kraemer HC, Thiemann S. How Many Subjects? Newbury Park, CA: SAGE; 1987.