EXHIBIT A

SCOPE OF WORK

PI Name: Jason G. Su, Associate Researcher & Principal Investigator

Project Title: Impact of Air Pollution Exposure on Metabolic Outcomes for California Residents

Project Summary/Abstract

Epidemiological studies suggest that short- and long-term exposures to air pollution increase metabolic dysregulation, incidence of diabetes, diabetes mellitus, and diabetes-related deaths. Those studies, however, are largely assessed within a city or metropolitan area, with coarse air pollution resolution or a single long-term (e.g., annual, or multiple years) pollutant exposure linkage. Large population-based studies at state or country level in areas with relatively low exposure using high spatiotemporal resolutions models have not been performed. In this project, the University of California at Berkeley (UCB) will first collaborate with the University of California at Los Angeles (UCLA), and the Methodist Hospital Research Institute to conduct a systematic literature review on impacts of air pollution including criteria pollutants and air toxics on metabolic health endpoints for Type 2 diabetes patients.

This research builds on decades of collaborative work between the UCB and UCLA and now adds the Methodist Hospital Research Institute to form a new research collaboration to conduct California statewide studies. The study will extend Dr. Jason Su's, UCB, completed California Air Resources Board (CARB) contract number 19RD004 that generated statewide daily air pollution surfaces at 100 meters (m) spatial resolution through advanced machine learning approaches for nitrogen dioxide (NO_2), particulate matter of ≤2.5 microns ($PM_{2.5}$) and ozone (0_3) for 2012-2019. This study will also extend the current funded CARB contract 21RD004 (Principal Investigator: UCLA Ritz; Subcontractor: UCB Su) that is enhancing #19RD004 and, generating daily surfaces for the three criteria pollutants and monthly surfaces for the air toxics for 1990-2019. Specifically, this new project will extend the daily air pollution surfaces for the same criteria pollutants and air toxics across all of California to also cover the period of 2020- 2021. Due to the availability of increased spatial resolution (from 12 kilometer [km] to one km) in daily remote sensing data for NO_2 and 0_3 for July 2019 - current (used as a predictor in exposure modeling), UCB will further incorporate the new remote sensing data into UCB's modeling framework to enhance the air pollution surfaces. The newly developed daily air pollution surfaces for the three criteria pollutants and monthly surfaces for the air toxics will be used to identify concentration- response functions for individual pollutants, as well as simultaneous exposures to multiple pollutants. The health outcome analysis will include identification of the impact of air pollution on (a) incidence of diabetes and medication use based on the California Health Interview Survey (CHIS) data, (b) the diabetes mellitus Emergency Department (ED) visits and hospitalizations based on the Department of Health Care Access and Information (HCAI) data, and (c) the diabetes mortality based on the California

Department of Public Health (CDPH) Vital Records data. The study period for the three health outcome datasets will cover 2010-2021 and the datasets will be acquired at individual level for patients of Type 2 diabetes. Further, UCB will develop monetized values (cost of illness and wiliness to pay) for those health impacts and associated uncertainties.

One of the long-term objectives of this research is to enhance the other two CARB funded projects (19RD004 and 21RD004) to provide statewide daily air pollution exposure surfaces for the criteria pollutants and monthly air toxics surfaces for the longest period yet that may also aid future health analyses of other endpoints. The other long-term objective is to supplement literature in concentration-response (C-R) functions for the impact of air pollution on diabetes-related health endpoints and in estimating economic impact from a number of exposures for the statewide California population. This research will allow CARB to assess temporal changes in exposure and/or risk for several adverse health outcomes across California's communities and its diverse population. CARB will also be able to determine the benefits associated with reductions in air pollution and identify where improvements in air quality may still be needed to reduce risk, where economic impacts may be largest, and which deprived communities or groups may be most affected and could possibly be targeted for special interventions.

Statement of Significance

This project will supplement scientific evidence from the impact of PM exposure on cardiopulmonary and respiratory disease-related ED visits, hospitalizations, and death, to adding more air pollutants and health impacts to those currently analyzed. Specifically, UCB, UCLA, and Methodist Hospital Research Institute, herein referred to as the research team, will expand air pollution exposure to include PM, 03), NO_2 and air toxics estimated at high spatiotemporal resolution (daily 100m resolution) across California and assess their impact on metabolic health outcomes in ED visits, hospitalizations, and death for diabetes 2 patients. This project will also add evidence to the impact of air pollution on the increased vulnerability and exposures in impacted communities, including low socio-economic status and racial/ethnic subgroups. The concentration-response (C-R) functions will be estimated not only for the statewide average population, but also for those impacted communities. Further, this project will develop monetized values (cost of illness and wiliness to pay) for those health impacts and associated uncertainties. CARB has advised research efforts to evaluate health benefits more comprehensively. Though metabolic outcomes are currently not listed as a causal relationship based on United States Environmental Protection Agency (U.S. EPA) Integrated Science Assessments categorization, CARB considers diabetes an important health endpoint to investigate. This research will augment the air pollution research and associated health impacts that CARB currently analyzes for its regulations, strategies, and programs.

If Third-Party Confidential Information is to be provided by the State:

☐ Performance of the Scope of Work is anticipated to involve use of third-party Confidential Information and is subject to the terms of this Agreement; OR

☒ A separate CNDA between the University and third-party is required by the third-party and is incorporated in this Agreement as Exhibit A7.

Introduction

Based on the report from the American Diabetes Association,[1] approximately 3,209,418 people in California, or 10.5 percent of the adult population, have diagnosed diabetes. An additional 884,000 people in California have undiagnosed diabetes, which is linked with adverse outcomes. There are 10,320,000 people in California, 33.4 percent of the adult population, who have prediabetes with blood glucose levels that are higher than normal, but not yet high enough to be diagnosed as diabetes.

Every year an estimated 272,814 people in California are diagnosed with diabetes. The total direct medical expenses for diagnosed diabetes in California were estimated at $27 billion in 2017. In addition, another $12.5 billion was spent on indirect costs from lost productivity due to diabetes. Type 1 diabetes is caused by the body's immune system damaging the pancreas and, therefore, no insulin can be manufactured by the body. In Type 1 diabetes, the body fails to make insulin. In Type 2diabetes, the body either develops a resistance to insulin or not enough insulin is produced to lower the blood sugars. The main difference between Type 1 and Type 2 diabetes is that Type 1 diabetes isa genetic condition that often shows up early in life, and Type 2 is mainly lifestyle-related and develops over time. Type 1 affects eight percent of the population with diabetes while Type 2 diabetes affects about 90 percent. Because of the difference in pathophysiology between type 1 and 2 diabetes and the vast majority of diabetes belongs to Type 2, the research team will use Type 2 diabetes patients in this study. It is believed that air pollutants exert their effects, especially on Type 2 diabetes, via impaired endothelial function, elevated systemic inflammation, mitochondrial dysfunction, and oxidative stress.[2] For example, Brook et al.[3] identified low levels of acute ambient $PM_{2.5}$ exposure reduced metabolic insulin sensitivity among 25 rural Michigan patients. Jiang et al.[4]identified that traffic-related $PM_{2.5}$ exposure was associated with metabolic biomarkers of insulin resistance, low-density lipoprotein cholesterol level and high blood pressure. Associations of air pollution exposure with diabetes development[2-5] and diabetes mellitus[6] have also been revealed in scientific research. There is also evidence that long-term air pollution is associated with diabetes mortality.[7] Those studies; however, were either conducted at city or metropolitan level, used coarse air pollution measures or a single fixed long-term surface. Most of the studies used PM2.s as the source of air pollution, although Jerrett et al.[8] reported positive associations between diabetes incidence and 0_3. While most studies detected an increase in the risk of diabetes from air pollution, otherpapers published failed to detect a significant association[9,10]. The statewide population-based study with accurate air pollution exposure from criteria pollutants and air toxics has yet to be seen in literature.

In this project, the UCB will collaborate with the UCLA to develop statewide daily air pollution modelsfor NO_2, fine $PM_{2.5}$ and O_3 at 30 m spatial resolution and generate corresponding daily air pollution surfaces of spatial resolution 100 m for years 2000-2021. Monthly air toxic models for benzene, 1,3-butadiene,total chromium, lead, nickel, and zinc of 30 m spatial resolution will also be developed and corresponding surfaces of 100 m resolution generated for years 2000-2021. Extending air pollution exposure to year 2000 enables us to take into consideration of the potential time-lagged exposure effect for health outcomes measured in years 2010-2021. The UCB will assign daily, or monthly high spatial resolution resolved air pollution exposures to the patients/participants' residentialaddresses and identify the impact of air pollution on incidence of diabetes, diabetes medication use, diabetes mellitus-related ED visits, hospitalizations, and death for years 2010-2021 through time-stratified case-crossover design, cox proportional hazard modeling and other related algorithms.This research will specifically focus on patients with Type 2 diabetes, due to the fact that (a) the vast majority of diabetes are Type 2, with Type 1 diabetes being only about one percent of the diabetes population, and (b) Type 1 and 2 diabetes have very different pathophysiology conditions. Lastly, theresearch team will estimate the economic benefits from diabetes health outcomes due to the decreased air pollution exposure.

The study will generate exposures for California at the highest temporal (daily) and spatial (100 m) resolution for criteria pollutants and monthly high spatial (100 m) resolution air toxics across all of California to cover the study period of 2000-2021. This will also allow assessment of the impacts of single and multiple pollutants. These exposure surfaces will be used to estimate concentration- response relationships between the highest temporal and spatial resolution air pollution exposure andthe five metabolic health endpoints, all at individual level, including incidence of diabetes, diabetes medication use, diabetes-related ED visits, hospitalizations, and death.

The summary of the tasks in this project are as follows:

1) Generate literature reviews of air pollution for all five health outcomes. UCB and UHCMC the Methodist Hospital Research Institute will conduct a systematic literature review, using peer-reviewed journal papers, to expand the literature citied in the Introduction section of this study on impacts of air pollution on incidence of diabetes, diabetic medication use, diabetes mellitus-related ED visits, hospitalizations, and death (Task 1).

2) Develop daily land use regression (LUR) models through advanced machine learning approaches and generate daily air pollution surfaces with 100 m spatial resolution for criteria pollutants (NO2, PM2.5, and 03) and monthly air pollution surfaces with 100 m spatial resolution for air toxics (benzene, 1,3-butadiene, total chromium, inorganic lead, metallic and inorganic nickel, and zinc) across California for years 2000-2021. With air pollution surfaces extending back ten years farther than the 2010-2021 health outcome data, this research project makes sure that sufficient time-lagged effects can be identified (Tasks 2-3).

3) Based on the new exposure data for criteria pollutants and air toxics, identify exposureconcentration response (C-R) relationships for:

   a) Single and multiple air pollutants (criteria pollutants and air toxics) and the incidence of diabetes and medication use based on the 2010-2021 CHIS data. A cox proportional hazardmodel will be developed to model the impact of air pollution on incidence of diabetes. Mixedeffects logistic regression models will also be developed to identify the impact of air pollution exposure on diabetes medication use.

   b) Single and multiple air pollutants (criteria pollutants and air toxics) and the 2010-2021 diabetes mellitus ED visits and hospitalizations based on the Department of HCAI (formally known as the Office of Statewide Health Planning and Development [OSHPD]) data. A time-stratified case-crossover study design combining with distributed lag nonlinear model will be developed for date of ED visit and for date of hospitalization. For hospital stay, cox proportional hazard models will be developed to identify the impact of short-term (daily and time-lagged) and long-term (seasonal and annual) air pollution exposure on days of hospitalization due to diabetes mellitus.

   c) Single and multiple air pollutants (criteria pollutants and air toxics) and the 2010-2021 diabetes mortality based on the California Department of Public Health (CDPH) Vital Recordsdata. A time- stratified case-crossover study design combining with distributed lag nonlinear model will be developed to identify the impact of air pollution on diabetes-related death.

The above models will also be further stratified by patients' gender, age group, race-ethnicity, and by whether the study subject is in a deprived community based on the Social Deprivation Index (SDI)[11] to identify those at highest risk of air pollutant exposure (Tasks 4-5).

4) Generate estimates of burden of disease and economic impact associated with the five metabolic health endpoints (incidence of diabetes, diabetic medication use, diabetes mellitus-related ED visits, hospitalizations, and death) from different air pollutants, and/or for minority groups and socio-economically deprived communities in California for years 2010-2021 (Task 6).

5) Fulfill project reporting on providing quarterly progress reports, draft reports finalized reports, and research seminar presentations to CARB (Task 7).

This research will provide CARB with statewide daily air pollution exposure surfaces for the criteria pollutants and monthly air toxics surfaces for the longest period yet to aid potential future health analyses of other endpoints. This research will also provide information about concentration-response functions and economic impact for a number of exposures for the statewide California population. CARB will be able to determine the benefits associated with reductions in air pollution and identify where improvements in air quality may still be needed to reduce risk, where economic impacts may be largest, and which deprived communities or groupsmay be most affected and could possibly be targeted for special interventions, in addition to informing stakeholders about the results.

Details of the Project Tasks

Task 1: Literature review of concentration-response functions of air pollution with metabolic outcomes (UCB and the Methodist Hospital Research Institute)

UCB and the Methodist Hospital Research Institute will conduct a systematic literature review, using peer-reviewed journal papers, to expand the literature citied in the introduction section of this study on the concentration- response functions of air pollution and metabolic outcomes, including incidence of diabetes, diabetes medication use, diabetes mellitus-related ED visits, hospitalizations, and death. UCB and the Methodist Hospital Research Institute will also examine the literature on the time- lagged short- and long-term air pollution effects on the planned health end points. The UCB and the Methodist Hospital Research Institute will identify primary lag time/latency period for short term (<one month) and long-term (>one year) exposure windows through systematic review of the literature. These time points will be utilized for the primary analysis.Additional lag and latency windows will be examined as secondary analyses. During the literature review process, UCB and the Methodist Hospital Research Institute will determine inclusion criteria, identify the publication characteristics for review selection, select proper search databases and engines, and decide the search terms and selection process. Specifically, the following inter-connected steps will be used tocomplete the review:

A. Determine inclusion criteria that will include the:

- Study population of children (<18 years), adults (≥18 years), and all ages,
- Study intervention for individuals (1) exposed to air pollution, including NO_2, $PM_{2.5}$, 0_3, and air toxics (including benzene, 1,3-butadiene, total chromium, inorganic lead, metallic and inorganic nickel, and zinc), and
- Study outcome for metabolic health endpoints, such as incidence of diabetes, diabetic medication use, diabetes mellitus-related ED visits, hospitalizations, and death

B. Identify the publications characteristics for studies:

- Published in peer-reviewed journals,

- Published between January 1, 2000, and current, and
- Written in English.

C. Select the proper search databases and engines, including

- PubMed,
- Medline,
- Web of Science Core Collection; and,
- Google Scholar

D. Decide the literature selection process. The following steps will be used to select scientific publications for the literature review:

a) Use one term from each category (A, B, and C) and combine them together(+) to create integrated search terms using the search databases and engines listed above,
b) Merge together the selected publications and remove the duplicates,
c) Obtain abstracts for the remaining publications selected from Step b), screen and remove the publications that are not related to research topic, and
d) Obtain full text for the remaining publications selected from Step C), screen and remove the publications that are not related to the topic.

Deliverable: A copy of the literature review findings on concentration-response functions of air pollution and metabolic outcomes will be submitted to CARB. In addition, a copy of the literature review findings will be included as an Appendix in the draft final report.

Task 2: Develop daily air pollution models and surfaces for criteria pollutants (UCB and UCLA)

Enhancement to the previous daily air pollution modeling of criteria pollutants

Under the CARB funded contract (19RD004, PI Jason Su), UCB identified sources of on-road vehicle emissions and their impacts on respiratory disease symptoms in California. UCB developed high spatiotemporal resolution (daily 30 m spatial resolution for modeling and 100 m resolution for surfaces) LUR air pollution models for criteria pollutants NO2, PM2.5 and 03 for the State of California for years 2012-2019. Under another CARB funded contract (20RD016, PI Jason Su), UCB identified the impacts of train and port pollution on respiratory symptoms and ED visits within socially deprived communities in Southern California. In that project, UCB developed annual emissions grids of NO2 and PM2.5 from train and port operations in Southern California at 30 m spatial resolution for years 2016-2019. Those annual emissions grids were used to (1) enhance the LUR air pollution models and related surfaces for NO2 and PM2.5 for years 2016-2019 and (2) develop a dispersion model (STILT) using emissions sources of NO2 and PM2.5 from port operations and operations of railways in Southern California. Under the third CARB funded project (21RD004, UCLA PI Beate Ritz, and subcontractor Co-PI Jason Su,) UCB is further extending the previously modeled daily air pollution models and surfaces for the three criteria pollutants to years 1990-2019. For this new CARB project, UCB will collaborate with UCLA to further extend the previously generated models and surfaces to include 2020-2021. During all the previous efforts on modeling NO2 and O3, UCB used daily O3 Monitoring Instrument (OMI) data of 25 km spatial resolution as one of the predictors in LUR modeling process. UCB and UCLA will, in this new project, use 2018-2021 Sentinel-SP NRTI (near real-time) NO2 and O3 instruments of daily seven km spatial resolution12 and the 2010-2014 Berkeley High-Resolution (BEHR) instrument of 12 km daily NO2 measures13 to replace OMI data to significantly increase this dataset's spatial resolution. The final generated daily models will have a spatial resolution of 30 m and the corresponding daily surfaces will have a spatial resolution of 100 m, covering the period of 2000-2021. The inclusion of years 2020- 2021 in this new study is critical due to significant temporal variation in air pollution from increasing frequency of wildfires, increasing port and railway goods operations and the impact of COVID19 pandemic that was accompanied by a hypothesized reduction in traffic-related pollution especially early in the pandemic.

UCB and UCLA will also extend the daily Aerosol Optical Depth (AOD) data from the multi-angle implementation of atmospheric correction (MAIAC) algorithm at one km resolution, every two-weeks of normalized difference vegetation index (NOVI) data at 250 m resolution and the daily four km resolution meteorological data from CARB 19RD004 in 2012-2019 to all the years between 2000- 2021. The land cover, tree canopy, and impervious surfaces data used in CARB 19RD004 project for year 2016 will also be extended to include corresponding for years 2001, 2006, 2011, and 2019.

Daily criteria pollutants modeling and surfaces generation

The relatively simplistic exposure assignment methods are (1) spatial averaging, (2) nearest neighbor, (3) inverse distance weighting, and (4) kriging. All four methods are weighted average methods. Due to simplicity in estimating exposure, they are widely used in literature; however, they do not use environmental characteristics of a location of interest (source or sink of air pollution) and therefore do not reflect the real level of air pollution exposure at the location of interest. By contrast, LUR modeling technique takes into consideration of impact from local environmental characteristics. A typical LUR model uses multiple regression equations to describe the relationship between annual average sample location measurements and their associated annual average environmental variables within buffered distances (e.g., total traffic around a monitoring station of 50 m buffer). It requires some linkage to the environmental characteristics of the location of interest, especially characteristics that influence pollutant emission intensity and dispersion efficiency. LUR modeling technique is more accurate than the weighted average methods in understanding the impact of local environment on air pollution levels and has become a standard way in estimating exposure for locations of interest; however, the vast majority of the current modeling process manually select several buffer distances, which might not be the optimal distances of impact on the monitored data. They also suffer from model overfit due to the use of the modeled data itself to assess modeling accuracy. In the research projects conducted/currently being worked on by UCB (19RD004, 20RD016, and 21RD004), a series of buffer statistics were built around each monitoring station for predictors with high spatial resolution (e.g., land use and land cover data of 30 m spatial resolution) and identify distance decay associations between measured air pollutant concentrations and buffer statistics of predictors. A series of buffer statistics of 50 - 5,000 m at an interval of 50 m are normally created on a predictor of interest, generating 100 covariates for a single predictor. In UCB's LUR modeling, there are normally ~3,000 covariates for a single model. This size of data is impossible for a typical LUR model to handle. To reduce the number of covariates in the modeling process, UCB applies a data reduction strategy to select a subset of covariates that is less spatially correlated (e.g., correlation coefficient < 0.9) while at the same time maintaining the covariates with an optimal distance of impact (i.e., the highest correlation with the measured pollutant concentrations). The same strategy of generating a series of buffer statistics and applying data reduction technique will be applied in this new project through collaborative work between UCB and UCLA.

To avoid model overfit, UCB always develop LUR models and surfaces through the V-fold cross-validation Deletion/Substitution/Addition (D/S/A) algorithm[14-16] after data reduction strategies are applied. The V-fold cross- validation algorithm uses out-of-sample data to validate the model built using training data. Specifically, in D/S/A V-fold cross-validation modeling, the original sample is randomly partitioned into V equal size subsamples. Of the V subsamples, a subsample is retained as the validation data for testing the model, and the remaining V-1 subsamples are used as training data. The cross-validation process is then repeated V times, with each of the V subsamples used exactly once as the validation data. The advantage of this method over the leave-one-out cross-validation technique is that the prediction errors are less impacted by single outliers, and compared to repeated random sub-sampling, all observations in the V-folds are used for both training and validation, and each observation is used for validation once. With each iteration, an independent validation dataset is used to assess the performance of a model built using a training dataset. This technique, therefore, minimizes over-fitting to the data to maximize the probability that the models will predict well at locations that have not been sampled. The D/S/A algorithm can deal with both linear and non-linear associations and the potential for non-linear associations will be explored in the modeling process. The D/S/A model not only generates an optimal model, but also further removing those covariates that have minimal impact.

In the modeling process, UCB and UCLA, referred to as the air pollution modeling team, will incorporate data into a single modeling framework from multiple air pollution measurement instruments , including those from government continuous monitoring across California, the fixed sites saturation monitoring in Los Angeles, Alameda and Sacramento counties, and Google Streetcar mobile saturation monitoring across San Francisco Bay (counties of Alameda, San Francisco, and San Mateo), Los Angeles County, and central valley regions. To integrate three types of air quality measurements into a single modeling framework, the air pollution modeling team will divide each type (e.g., Google Streetcar mobile monitoring) or its sub-type (e.g., Google Streetcar mobile monitoring in Los Angeles) of air quality monitoring data equally into ten-folds and then merge corresponding folds of data into a large ten-fold dataset, with each fold having equal presentation of the three types and corresponding sub-types of air quality monitoring data. The equal presentation of ten-folds data will then be used in av-fold out-of-sample cross-validation technique for LUR modeling.

To make data consistent in spatial resolution, when the spatial resolution of a predictor is finer than 30 m, the surfaces will be resampled through a mean function to have a spatial resolution of 30 m. This included Digital elevation model (DEM), which has an original spatial resolution of 10 m. All the predictors of vector format will be converted to rasters with a spatial resolution of 30 m. For example, the original land use parcel data are vector data and will be converted to raster data of spatial resolution of 30 m. Land cover data are by themselves at 30 m spatial resolution and will remain unchanged. For the predictors of raster with a spatial resolution coarser than 30 m, UCB will treat those pixels under a coarser resolution pixel unchanged in values. This includes, for example, weather data, which has a spatial resolution of four km. The daily weather data includes minimum and maximum relative humidity, minimum and maximum temperature , wind velocity, wind direction, and mean vapor pressure. For a pixel of four km resolution in relative humidity, for example, all the 17,778 sub-pixels of 30 m resolution it covers will have the same value in relative humidity. To make the data consistent in temporal resolution, when the surface of a temporal resolution is coarser than daily, that surface will remain unchanged for all the days of the study period. Due to the large size of a single surface of 30 m spatial resolution for the State of California (~3.9 gigabytes), UCB will generate daily surfaces of 100m across the State, maintaining the spatial resolution (100 m) capable of identifying socially deprived communities in the research.

To assess model performance, models' closeness of fit R^2 and adjusted performance R^2_{adj} will be used. Mean Absolute Error, Root mean squared error and the Akaike's Information Criteria (AIC) will also be employed to assess and compare model accuracies. Similar to R^2_{adj}, AIC penalizes the inclusion of additional variables to a model. AIC adds a penalty that increases the error when including additional terms. The lower the AIC, the better is the model.

For NO_2, the LUR will be developed using the California Environmental Protection Agency (CalEPA) daily NO_2 monitoring data, plus the saturation monitoring data and Google Air data that have been collected as a response variable. The predictors include daily remote sensing pollutant data, daily rush hour traffic flow and speed, daily weather conditions, every two-week NDVI, one-time land use and land cover, distance to shopping centers, distance to freight intermodal facilities, and other traditional geographic features like DEM, distance to highways, major roadways, and ports. To enhance application of remote sensing data as one of the predictors, correlation coefficients between

OMI 25 km data and the Sentinel-SP Near real time (NRTI) $N0_2$ and Berkeley High-Resolution (BEHR) data will be calculated separately for the overlapping days and the adjusted Sentinel-SP NRTI and BEHR $N0_2$ data will replace the OMI 25 km data for years 2010-2014 and 2018-2021 in ther $N0_2$ modeling process. Due to extending the study period to include year's 2020-2021, other related data sources need to be extended as well. The air pollution modeling team will extend the daily AOD data from the MAIAC algorithm at one km resolution, every two-weeks of NOVI data at 250 m resolution and the daily four km resolution meteorological data to cover years 2020-2021. The National Land Cover Database (NLCD) land cover, tree canopy, and impervious surfaces data used for previous projects using NLCD 2016 data will also be extended with inclusion of corresponding data for years 2001-2019. The daily $PM_{2.5}$ and O_3 models will be developed separately; but, in a way, similar to the model used for daily $N0_2$ concentrations, with a mixed effects D/S/A machine learning algorithm to model the daily concentrations at a spatial resolution of 30 m for the study period.

In summary, in this new study, UCB and UCLA will develop statewide LUR models that (1) are developed at a high spatial resolution of 30 m at daily level across California (but surfaces at a spatial resolution of 100 m will be generated and used for calculating C-R functions to reduce storage space), (2) can identify optimal distance of impact for a predictor, (3) can reduce model overfit, (4) can deal with multicollinearity among predictors, (5) automatically process potential interactions between predictors, (6) allow potential non-linear associations between a predictor and air pollutant concentrations, (7) avoid excessive number of predictors in the final models, and (8) take into consideration repeated measures for CalEPA air quality monitoring stations across time. Based on current literature available, this has not yet been achieved in the exposure modeling literature. These novel daily air pollution surfaces developed for the State of California will be used to assign daily exposures at the home address or zip code of study subjection.

Deliverable: Daily criteria pollutants surfaces for $N0_2$, $PM_{2.5}$ and O_3 for years 2000-2021 will be submitted to CARB.

Task 3: Develop monthly air pollution models and surfaces for air toxics (UCB and UCLA)

Enhancement to the previous monthly air pollution modeling of air toxics

Under the CARB funded contract #20RD016 (PI Jason Su), UCB developed monthly air toxics models and surfaces for benzene, 1,3-butadiene, total chromium, lead, nickel, and zinc for Southern California for years 2016-2019. Through CARB funding (21RD004), UCB is developing monthly air pollution models and surfaces for the same air toxics across California for years 1990-2019. For this new CARB project, UCB and UCLA will further extend the previously generated models and surfaces for the six air toxics to include years of 2020-2021. In addition, the air pollution modeling team will add point sources of air toxics as a potential predictor in the modeling process using ArcGIS Business Analysts for mining, chemicals and allied products, petroleum refining and related industries, and primary metal industries. The final generated statewide models and surfaces for the six air toxics will cover the period of 2000- 2021.

Monthly air toxics modeling and surfaces generation

Due to the factor that most air toxics are measured on the third or fourth day in a week, this project will develop air toxics at monthly level, including for hazardous air pollutants (1,3-butadiene,total chromium, nickel, lead, and diesel PM) and volatile organic compound and their components, such as benzene, ethylbenzene, formaldehyde. UCB will work together with UCLA to develop monthly air toxics surfaces for six pollutants, including benzene, 1,3-butadiene,total chromium, lead, nickel, and zinc. The number of sites for the six air toxics ranges from 50 to 85 (Table 1), giving the study sufficient air quality measurement data to effectively predict monthly concentrations of an air toxic. The air toxics data will be acquired, respectively, from PM2.5 and VOC speciation.

Table 1. The number of air toxics monitoring sites in the State of California.

Parameter.Code	Parameter.Name	2012	2013	2014	2015	2016	2017	2018	2019	TotalUniqueCount
43218	1,3-Butadiene	40	41	43	40	40	39	42	41	51
45201	Benzene	72	73	76	66	59	56	59	52	85
88112	Chromium PM2.5 LC	45	47	47	42	42	36	36	36	50
88128	Lead PM2.5 LC	45	47	47	42	42	36	36	36	50
88136	Nickel PM2.5 LC	45	47	47	42	42	36	36	36	50
88167	Zinc PM2.5 LC	45	47	47	42	42	36	36	36	50

Air toxics will be modeled in a way similar to the criteria pollutants. Further, the air pollution modeling team will add point sources including those from mining Standard Industrial Classification - (SIC) code 10-14), chemicals and allied products (SIC 28), petroleum refining and related industries (SIC 29), and primary metal industries (SIC 33).

A previous study[17] has shown that 50 sites are sufficient to build an annual LUR model. For air toxics, monthly air pollution surfaces for the 50 sites across California will be built. For those predictors to be used in LUR models, land use, land cover, elevation, and distance to certain air pollution

sources/sinks will largely remain unchanged; however, weather conditions, vegetation greenness (phenology) and AOD data will change throughout the years. Thus, there likely will be sufficient temporal changes to increase variations in land use predictors for the modeling process, i.e., to improve LUR prediction. Emissions inventory data from port operations and operations of vehicles on railways will also be part of the predictors in the LUR modeling process.

Deliverable: Provide monthly air toxics surfaces for benzene, 1,3-butadiene, total chromium, lead, nickel, and zinc for years 2000-2021 to CARB.

Task 4: Data acquisition of human subject's data for 2010-2021, geocoding, and Institutional Review Board (IRB) application (UCB and the Methodist Hospital Research Institute)

This task will comprise the acquisition and coding of health outcomes and complementary measures data from three different entities in the project, including the CHIS data on incidence of diabetes and diabetic medication use, the Department of Health HCAI data on diabetes mellitus ED visits and hospitalizations, and the CDPH Vital Records on diabetes mortality. All the health endpoints data will be acquired at individual level across California, with the CHIS and CDPH Vital Records data acquired with home address as geographic identifier and the HCAI data having five-digit zip code geographic identifier. An IRB will be submitted for approval to cover accessing human subject's data by the UCB and the Methodist Hospital Research Institute employees who will work on the project.

Task 4.1: Acquire and process CHIS data for 2010-2021

CHIS is the largest state health survey in the United States. It is conducted on a continuous basis allowing the survey to generate timely one-year estimates. CHIS provides representative data on all 58 counties in California and provides a detailed picture of the health and health care needs of California's large and diverse population. More than 20,000 Californians, including adults, teenagers, and children are interviewed each year. Participants in the CHIS survey are chosen at random and the sample is extensive enough to be statistically representative of California's diverse population.

The UCB and the Methodist Hospital Research Institute will select survey participants who are diagnosed as diabetic or prediabetic for years 2010-2021. Information collected will include latitude/longitude of home address, age, gender, race-ethnicity, insurance status, body mass index (BMI), smoking status, diabetes medication use (e.g., insulin injections and diabetic pills intake), and hemoglobin A1C checks.

Task 4.2: Acquire and process HCAI data for 2010-2021

The UCB and the Methodist Hospital Research Institute will acquire statewide Type 2 diabetes mellitus (primary diagnosis International Classification of Diseases [ICD]-9 250 and ICD-10 E11) data from HCAI on ED visits and hospitalizations. Other information collected will include patient zip code, date of admission and discharge (and length of stay), age, gender, race-ethnicity, Elixhauser comorbidity index (to becalculated using diagnoses), zip code level median household income, facility number, payer category, preferred language spoken, principal procedure, and care type.

Task 4.3: Acquire and process CDPH mortality data for 2010-2021

UCB and the Methodist Hospital Research Institute will acquire mortality data from diabetes mellitus (underlying or contributory causeof death) through the CDPH Vital Records. Information collected will include residential address, dateof death, age, gender, race-ethnicity, smoking, BMI, and insurance status. UCB and (the Methodist Hospital Research Institute)will additionally acquire the underlying cause of death when diabetes is mentioned as contributory causeof death. Due to IRB restrictions, the metabolic health outcomes data will not be submitted to CARB.

Deliverable: The summary statistics of the metabolic health outcomes data will be submitted to CARB.

Task 5: Identify concentration-response relationships between short-term, intermediate, and long-term air pollution exposures for individual air pollutants and simultaneous exposures to multiple air pollutants (criteria pollutants and air toxics) for the five health outcomes (UCB, UCLA, and the Methodist Hospital Research Institute).

Table 2 lists various exposure types and exposure terms to be modeled in the project with detailed explanations in Tasks 5.1- 5.3.

Table 2. Project health endpoints, exposure types, and exposure terms

Endpoint	**Data Source**	**Time Period and Frequency**	**Details of Measure**
Incidence of diabetes	CHIS	2011 to 2021 annual	Year of diabetes II identification
Diabetic medication use	CHIS	2011 to 2021 annual	Whether diabetes medication in a year
Diabetes-related ED visits	HCAI	2000 to 2021 event day	Date of ED visit
Diabetes-related hospitalizations	HCAI	2000 to 2021 event day	Date of hospitalization
Diabetes-related death	CDPH Vital Records	2000 to 2021 event day	Date of death
Effect of Measure	**Short-term Outcome**	**Intermediate Outcome**	**Long-term Outcome**
Chronic	1-2 years	3-5 years	Over 5 years
Chronic	1-2 years	3-5 years	Over 5 years
Acute	Daily and time-lagged (<7 days)	Monthly and seasonal	Annual to multiple years
Acute	Daily and time-lagged (<7 days)	Monthly and seasonal	Annual to multiple years
Acute	Daily and time-lagged (<7 days)	Monthly and seasonal	Annual to multiple years

CHIS: California Health Interview Survey; HCAI : Department of Health Care Access and Information; CDPH: California Department of Public Health

Task 5.1: Modelling incidence of diabetes and diabetes medication use using CHIS data

Daily air pollution surfaces will be assigned to the home addresses of the survey participants for years 2000-2021. The daily air pollution exposures will be aggregated to monthly, seasonal, and annual exposures through an average function. The research team will use a participant's age at the year of survey and the response to the survey question on "how old were you when a doctor first told you that you have diabetes" to identify the incidence year of developing diabetes (if the year of incidence is< 2000, the data will be removed from analysis). Because every year more than 270,00 people in California are newly diagnosed with diabetes and the annual CHIS survey sampled more than 20,000 participants, it is expected that about 2,000 CHIS participants were diagnosed with diabetes across the 2010-2021 period (assuming the same participant not sampled more than once). The C-R functions of incidence of diabetes will be generated for:

a) different lengths of exposure (with different lagging times according to windows of interests , e.g., 1, 3, or 5 years before index date),
b) different air pollutant exposures (criteria pollutants and air toxics) based on the newly generated pollution surfaces with high spatiotemporal resolution, shorter term (1-2 years), intermediate (3-5) and long-term (>5 years) air pollutant exposure effects, and
c) different subgroups defined by age group, race-ethnicity, and SDI at census tract level as sample size allows.

SDI is a composite measure of area level deprivation (e.g., at census tract or zip code) based on sevendemographic characteristics collected in the American Community Survey, including percent living in poverty, percent with less than 12 years of education, percent single-parent household , percent living in the rented housing unit, percent living in the overcrowded housing unit, percent of households without a car, and percent non-employed adults under 65 years of age.11 The composite measure value rangesfrom 0 to 1. The research team will treat those census tracts with a composite score over 0.75 as socially deprived communities.

The C-R functions will be generated through longitudinal models, such as generalized linear mixed effects accounting for time and study design.[18,19] Cox proportional hazards regression models will also be used, with calendar time (5 years maximum traceback) as the underlying time scale to assess the impact of air pollutants exposures on incident diabetes. The analyses will use age, gender, and race-ethnicity (also potentially social deprivation) as stratification variables and, insurance status, BMI and smoking status will be used as confounding factors. Latency windows approach for time to event will beused, i.e., average exposure effects prior to diagnosis/event will be assessed either by extending the averaging period consecutively from 1 to 3 to 5 years prior to event or by using a fixed time window (such as 1, 3 or 5 years) and move this window back in time prior to event, such as by one, three, and five years prior to event. Analyses will be stratified by gender, age group, race-ethnicity, and neighborhood social deprivation.

Co-pollutants will also be assessed in the same model (e.g., two pollutants at a time) as long as they are not strongly correlated, and for correlated pollutants the best pollutant will either be selected to model the common source or if deemed appropriate, mixture models will be employed, such as Bayesian Kernel Machine Regressions[20] or hierarchical cluster analysis.[21]

Based on UCB's latest CARB research funding (contract #'s 19RD004 and 20RD016), the UCB research team found that $N0_2$ had the greatest spatial variability, with the most vulnerable communities having the highest air pollution exposure. The spatial distribution of $PM_{2.5}$ had a similar pattern to $N0_2$, but the degree of spatial variability decreased due to the secondary $PM_{2.5}$ formation. The spatial distribution of 0_3 was found totally opposite to $N0_2$, with those locations of high $N0_2$ concentrations being low in 0_3. The research team expects the combination of $N0_2/0_3$ or $PM_{2.5}/0_3$ will generate the most significant interaction effect due to the greater difN)Nference in their spatial distribution. The research team expects the combination of $N0_2$ and $PM_{2.5}$ will also generate, though in a smaller scale, significant interaction effect because their spatial distributions are still different. Due to the difference in spatial distribution of the three criteria pollutants, the research team expects the simultaneous exposure to those three will significantly exert health impacts, as has shown in the UCB-led CARB contract #19RD004.

In California, 10.5 percent of the adult population have diabetes, the annual CHIS survey of more than 20,000 participants makes it possible to have about 20,000 CHIS participants being diabetic for the 2010-2021 period. Mixed effects logistic regression models will be developed to identify the impact of air pollution exposure on insulin or diabetic pills use for those 20,000 CHIS participants. Similar to model incidence of diabetes, the C-R functions of insulin or diabetic pills use will be generated for different lengths exposures, different air pollutant exposures, shorter term, as well as long-term air pollutant exposure effects, and different subgroups by socio-economic status as sample size allows. Analyses will be stratified by gender, age group, race-ethnicity, and neighborhood socioeconomic level, whenever possible.

Task 5.2: Modelling diabetes mellitus ED visits and hospitalizations using HCAI data

Due to the finest geographic information in the HCAI data for an individual is at zip code level, the individual level daily air pollution exposure will be assigned using zip code weighted mean for residential land use within that zip code. A time-stratified case-crossover study design combining with distributed lag nonlinear model will be developed for date of ED visit and for date of hospitalization. A case-crossover design utilizes an individual as a stratum, comparing individuals to themselves at different times.[22 23] In air pollution research, it compares exposure level in the day when the health event occurs (case day) with the levels in nearby days (control days).

The control days represent the counterfactual exposure experience of each case, independently of the exposure on case day. Given that diabetes is affected by multiple biologics, behavioral, social, and genetic factors, the case-crossover design is ideal to minimize these confounding factors due to intra-individual comparisons, while minimizing inter-individual comparisons. Due to time trends and seasonal patterns in exposure or health events and non-independent selection of control days, the research team will use the time-stratified case-crossover design to control those potential biases.[24, 25] Specifically, the research team will select, for each case day, same week/weekend days within each month of each year for matching. Each case will have three or four matched control days (before and/or after the case day in the same month). With this design, the long-term trend and seasonality of unmeasured variables are controlled for.

For hospital stay, the research team will also develop cox proportional hazard models to identify the impact of short-term (daily and time-lagged), intermediate (monthly and seasonal), and long-term (annual and longer years) air pollution exposure on days of hospitalization due to diabetes mellitus. The research team will identify primary lag time/latency period for short term (<1 week) and long-term (>one year) exposure windows through systematic review of the literature (in Task 1). These time points will be utilized for the primary analysis. Additional lag and latency windows will be examined as secondary analyses. The C-R functions will be generated for:

a) different lengths exposures (Task 1- guided primary analysis, and secondary analysis by daily and time-lagged in a week, monthly, seasonal, annual, or over a year),
b) different air pollutant exposures (criteria pollutants and air toxics) based on the newly generated pollution surfaces of high spatiotemporal resolution (guided by systematic review of literature in Task 1),
c) different subgroups by race/ethnicity and socio-economic status as sample size allows.

Given that some factors may have differential effects on the C-R function due to susceptibility factors, the research team will include stratification variables for more robust modelling, including age group (e.g., <vs 65 years) , gender, race/ethnicity, and social deprivation status.[11] This will allow generation of flexible C-R functions accounting for susceptibility. These variables were selected a priori for stratification (rather than for confounding control) based on the available health effects literature on susceptibility to air pollution. It is also true that comorbidity index, facility number, payer category, preferred language spoken, principal procedure and type of care will be used for confounding control.

The stratification procedure also allows for analysis of effect modification of each stratification variable on the C-R function. The research team will evaluate effect modification if there is significant interaction between strata and exposure. Selection of age group, gender, race/ethnicity, and social deprivation as stratification also meets CARB's requirement in identifying C-R functions for socially deprived communities or groups may be most affected.

Limitations and potential solutions:

Of the 1,751 five-digit zip codes in California, the average area is 37,002 has (SD: 91,407 ha). This is far coarser than the 100 m spatial resolution air pollution surfaces generated for the State. Due to the finest geographic identifier in the HCAI data for an individual is at zip code level, the individual level daily air pollution exposure will be assigned using zip code weighted mean for residential land use within that zip code. Further, the research team will calculate standard deviations (SD) of daily air pollutant concentrations for residential area within a zip code and generate subsets of zip codes with concentration SDs being less than 0.5, 1, 2 and 3, respectively, for the health outcome analysis. The smaller the SDs in pollutant concentration in a zip code, the more homogenous is the air pollution level in that zip and subsequently the smaller the exposure misclassification would occur. The research team will strive to reduce air pollution exposure misclassification, but at the same time maintain the sample size for the analysis. The research team will additionally perform sensitivity analyses by including/excluding the largest five percent of ZIP codes by area to evaluate the impact on C-R function. The research team will additionally test the impact of using the upper and lower limit of exposure estimates in the largest five percent of ZIP codes on the C-R function to account for the uncertainty in exposure assignment.

Task 5.3: Modelling diabetes mortality using CDPH Vital Records data

Mortality data from death certificate will be used in this study. Daily air pollution surfaces will be assigned to the home addresses of the patients. The research team will apply a time-stratified case-crossover study design combining with distributed lag nonlinear model in a way similar to Task 5.2 to identify the impact of air pollution on diabetes-related death. The research team will identify short-term (daily and time-lagged), intermediate (monthly and seasonal) and long-term (annual and longer years) exposure windows through systematic review of the literature in Task 1. These time points will be utilized for the primary analysis. Additional lag and latency windows will be examined as secondary analyses. The C-R functions will be generated for:

a) Different lengths exposures (Task 1- guided primary analysis, and secondary analysis by daily and time-lagged in a week, monthly, seasonal, annual, or over a year),
b) Different air pollutant exposures (criteria pollutants and air toxics) based on the newly generated pollution surfaces of high spatiotemporal resolution (guided by systematic review of literature in Task 1, and
c) Different subgroups by race/ethnicity and socio-economic status (SES) as sample size allows.

Similar to Task 5.2, the research team will include stratification variables for more robust modelling, including age group (e.g., <vs ≥ 65 years), gender, race/ethnicity, and social deprivation status. Effect modification will also be analyzed through testing of interaction. Confounding factors will be identified including smoking, BMI, and insurance status.

Task 5.4: Sensitivity Analysis

In addition to identifying whether the above-mentioned modeling techniques generate valid results, the research team will apply machine learning algorithms including random forest and ensemble machine learning algorithms to confirm the main modeling results. Briefly, random forest is composed of a large number of decision trees and allows for regression analysis by averaging the output of these trees.[26] As such, random forest methods allow for complex interactions between input variables as it does not assume a linear relationship. Random forest allows for evaluation of combination of variables (exposure, SES, etc) on the outcome (diabetes), and can help identify the percentage of variation in outcome that can be explained by the input variables. Random forest also contains variable ranking process that allows for identification of the most important determinants of outcome, from a list of input variables. Random forests are increasingly utilized in exposure studies given complexity of interactions between input variables, which are not well handled by traditional statistical methods.[27-29] Those machine learning algorithms have the potential advantage over traditional linear models for several reasons: (1) lack of underlying data structure assumptions, (2) resistance to model overfitting, (3) ability to handle multicollinearity among predictors, and (4) increased accuracy in prediction.[30] Using the advanced machine learning modeling framework may allow identifying the impacts of individual pollutants on outcomes and to compare results generated by advanced linear mixed models. This includes exposure C-R relationships overall and for subgroups, such as for deprived vs. non-deprived communities due to the inherent de-correlated nature of the decision trees. Differences seen for different C-R functions can be compared and impacts from different pollutants to pollutant mixes can be assessed.

Develop concentration-response relationships of short- and long-term between air pollution and (1) incidence of diabetes, (2) diabetes medication use, (3) diabetes-related ED visits, hospitalizations, and hospital length of stay, and (4) diabetes-related death. The identified air pollution dose effect will include those of (1) single air pollutant exposure, (2) simultaneous criteria pollutants exposure, (3) simultaneous air toxics exposure, and (4) simultaneous criteria pollutants and air toxics exposures. Time lagged models will be developed with a maximum of 10 years . All the models will be developed with individual level data.

Deliverable: The summary model(s) for concentration-response relationships between air pollution exposure and metabolic health endpoints and a draft summary of the methods and findings will be submitted to CARB.

Task 6: Estimate economic benefits from reducing air pollution exposures on metabolic health outcomes, stratified by race/ethnicity, and neighborhood deprivation status in California (UCB, UCLA, and the Methodist Hospital Research Institute).

Task 6.1: Identification of healthcare costs for metabolic endpoints

Cost of illness studies broadly aim to identify and measure the total costs attributable to a particular disease and estimate its economic burden. The first step is to obtain estimates of the cost per case for each of the outcomes, including direct and indirect costs. Direct costs are generally costs associated with medical procedures and treatment while indirect costs are the loss of money because of disease-related disability or mortality. Previous studies have analyzed the cost of illness associated with air pollution on other outcomes.[31-33] Annual direct and indirect cost estimates will be generated for 2010-2021 through a literature review and separately for incidence of diabetes, its medication use, ED visits, hospitalizations , and death. To account for uncertainty, the components of costs will be evaluated based on lists of items previous studies relied on for each of the outcomes.[34-35] The list could include : (i) medical and healthcare service costs, (ii) therapeutic costs, (iii) (special) education costs, (iv) costs of production loss for adults, (v) costs of informal care and lost productivity for family/caregivers, and (vi) costs of accommodation, respite care, and out-of-pocket expenses. A minimum and maximum range of cost per case will be generated. As suggested in a previous study, medical costs will then be *(vii)* indexed using the Bureau of Economic Analysis' Personal Consumption Expenditures health price index and for indirect economic losses, UCB will use the Personal Consumption Expenditures Price Index.[37] The medical expenses per person (direct and indirect) from metabolic health outcome j in yeark is represented here as $c_{j,k}$,

Task 6.2: Identification of health impact from air pollution exposure

Using year 2010 air pollution exposure as baseline, the research team will identify improvements in air pollution exposure from year 2010 to year 2021 for three types of datasets. Through the concentration-response functions identified from this research (Task 5), the research team will identify health impact separately for each of the metabolic endpoints (including incidence of diabetes, medication use, ED visits, hospitalizations, and death) from improvements in air pollution exposure:

$$H_{i,j,k} = f_{i,j} * P_k * D_k \quad (1)$$

where $H_{i,j,k}$ and $f_{i,j}$ are respectively total health impact and population attribution factor of pollutant *i*, metabolic health outcome *j* at year *k* (year 2010 as reference). P_k is the statewide total population of

type 2 diabetes in year *k*. D_k = 1 for annual long-term effect and D_k = number of days in a year for daily short-term effect. The form of a population attribution factor $(f_{i,j})$ decides how a health impact is calculated, with for example, linear, logistic and log-linear models being estimated separately by:

linear models: $\beta_{i,j} * Y_{0,j} * \Delta E_{i,k}$ (2)

logistic models: $Y_{0,j}\left(1 - \frac{1}{\exp(\beta_{i,j} * \Delta E_{i,k})}\right)$, (3)

and log-linear models: $Y_{0,j}\left(1 - \left(\frac{1}{(1 - Y_{0,j}) * \exp(\beta_{i,j} * \Delta E_{i,k}) + Y_{0,j}}\right)\right)$ (4)

where $\beta_{i,j}$ is the model coefficient for metabolic health outcome *j* from exposure to pollutant *i*. $\Delta E_{i,k}$ is improvements in air pollution exposure to pollutant *i* at year *k* (year 2010 as reference). $Y_{0,j}$ is an health endpoint specific baseline rate, including the rate of incidence of diabetes, medication use, ED visits, hospitalizations and death in 2010. The U.S. EPA BenMAP-CE [38] provides details on the selection of health impact functions for evaluation of health benefits from changes in exposure. The research team will identify the baseline rate for the general population, but also for the race-ethnicity and socioeconomic subgroups. The research team will also create interactions between race-ethnicity/socioeconomic status and air pollution exposure to create minority racial-ethnic and socio-economically deprivation specific $\beta_{i,j}$.

Task 6.3: Identification of economic benefits from reduced air pollution exposure:

The economic benefits of improved metabolic endpoints from reduced air pollution could be calculated as:

$$E_{i,j,k} = H_{i,j,k} * C_{j,k} \quad (5)$$

where $E_{i,jk,}$ is the overall economic benefits from reduced exposure to pollutant i for metabolic health outcome j in year k. $c_{j,k}$ is the medical expenses (direct and indirect) from metabolic health outcome $j$in year k (year 2010 as reference). The direct medical expenditures include hospital ED visits and hospitalizations, prescription medications to treat complications of diabetes, anti-diabetic agents and diabetes supplies, and physician office visits. The indirect costs include increased absenteeism at school for students, reduced productivity while at work for those employed or inability to work as a result of disease-related disability. $H_{i,j,k}$ has the same definition as in eq (1). Because $H_{i,j,k}$ includes health impact for the general population and for the minority racial-ethnic and socio-economically deprived subgroups, the economic benefits estimated will also be stratified by (1) minority racial- ethnic groups vs. White, and (2) socio-economically deprived communities vs. non-deprived ones.

Deliverable: The draft final report should include a description of the numerical estimates of economic benefits from reducing air pollution exposures on metabolic health outcomes in California.

Task 7: Project Reporting (UCB)

UCB will meet with CARS staff quarterly and submit quarterly progress reports using the CARB- designated template, and in invoice for the same period will accompany each progress report. Six months prior to the end of the study, UCB will submit a draft final report (DFR), which will include theresults of the study and the additional deliverables identified in Exhibit A1, Schedule of Deliverables.The DFR will be submitted in accordance with the Final Report format and will be reviewed by CARSstaff. CARB's comments will be sent to UCB and after receiving the reviewer's comments, UCB shallmodify and resubmit the modified DFR to the CARB Project Manager. The modified DFR will be subject to formal review by the Research Screening Committee (RSC). Once accepted by the RSC, UCB will revise the modified draft final report addressing the RSC comments and any remaining concerns from CARB staff and will submit the revised final report to CARB. If CARB has additional comments on the report, the UCB will be notified so appropriate changes can be made; otherwise, CARS will accept the revised final report as the final. The UCB will submit the final report in an Americans with Disabilities compliant format. A notation in the Final Report task should denote that the University will incorporate a one-page Public Outreach Document into the Final Report, that will be widely used to communicate, in clear and direct terms, the key research findings from the study to the public. The format for the Public Outreach Document is outlined in Exhibit A1, Section 2.

In addition, UCB will present study findings at a CARB research seminar and provide electronic dataand analyses to CARS.

Project Schedule

The project will be completed in 24 months from the start date and the timeline is listed below.

	Months																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Task 1	■	■	■	■	■	■	■	■	■	■	■	■												
Task 2	■	■	■	■	■	■																		
Task 3							■	■	■	■	■	■												
Task 4	■	■	■	■	■	■	■	■	■	■	■	■												
Task 5													■	■	■	■	■	■						
Task 6															■	■	■	■						
Task 7	m		mp			mp			mp			mp			mp			md			m			fr
Note:	m = meeting with CARB staff; p = quarterly progress report; d = deliver draft final report; f = deliver final report; r = research seminar																							

Meetings

A. Initial meeting. Before work on the contract begins, the Principal Investigator and key personnel will meet with the CARS Contract Project Manager and other staff to discuss the overall plan, details of performing the tasks, the project schedule, items related to personnel or changes in personnel, and any issues that may need to be resolved before work can begin.

B. Progress review meetings. The Principal Investigator and appropriate members of his or her staff will meet with CARB's Contract Project Manager at quarterly intervals to discuss the progress of the project. This meeting may be conducted by phone.

C. Technical Seminar. The Contractor will present the results of the project to CARB staff and a possible webcast at a seminar at CARS facilities in Sacramento or El Monte.

HEALTH AND SAFETY

Contractors are required to, at their own expense, comply with all applicable health and safety laws and regulations. Upon notice, Contractors are also required to comply with the state agency's specific health and safety requirements and policies. Contractors agree to include in any subcontract related to performance of this Agreement, a requirement that the subcontractor comply with all applicable health and safety laws and regulations, and upon notice, the state agency's specific health and safety requirements and policies.

CONFIDENTIAL HEALTH DATA AND PERSONAL INFORMATION

CARB will be provided access to and will not receive any confidential health data or other confidential personal information under this contract. Further, CARB will have no ownership of confidential health data or other confidential personal information used in connection with this contract. The entities conducting the research in this contract will follow all applicable regulations regarding access to and the use of confidential data, including the Health Insurance Portability and Accountability Act (HIPAA) and requirements related to the Institutional Review Board (IRB) process. CARB will not be a listed entity with authorized access to confidential information pursuant to the IRB process for this contract.

Project Management Plan

The project will be led by UCB, in collaboration with the Methodist Hospital Research Institute and UCLA for a 24-month period and the project management plan is as follows (see diagram below):

- Dr. Jason Su, Principal Investigator, will act as project director and UCB-CARB contact and will be responsible for carrying out all the seven tasks. Dr. Su will be working with the Berkeley Undergraduate Research Apprentice Program (URAP) undergraduates to complete those seven tasks.

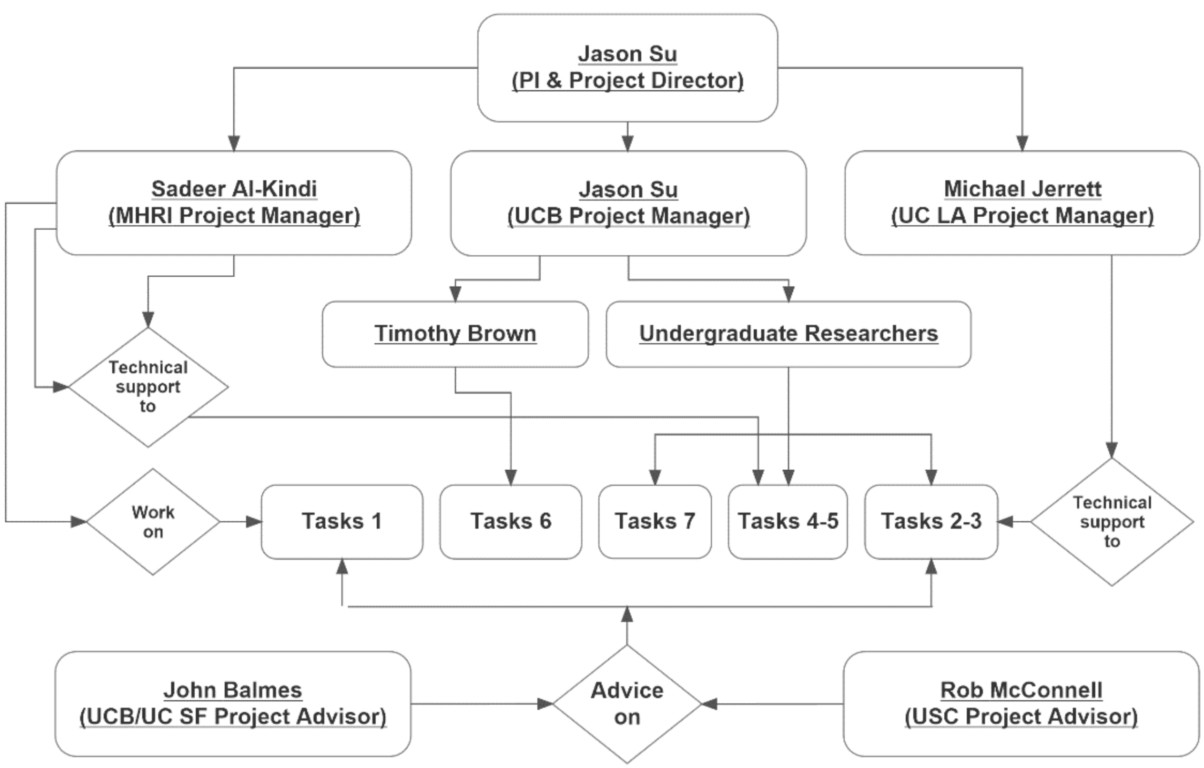
- **Dr. Timothy Brown, Co-I will focus on economic benefits estimation from reducing air pollution exposures on metabolic health outcomes (Task 6).**

- Dr. Michael Jerrett from UCLA will focus on providing expertise support for exposure assessment (Tasks 2-3), and will also provide advice on health outcome (Task 5) and economic benefits analyses (Task 6).

- Dr. Sadeer Al-Kindi from the Methodist Hospital Research Institute will focus on providing expertise support for literature review (Task 1), and health outcome data acquisition and analysis (Tasks 4-5). Dr. Al-Kindi will also provide technical support to economic benefits analyses
(Task 6)

- Drs. John Balmes (UCB/University of California, San Francisco) and Rob McConnell (University of Southern California) will provide advice on all aspects of the project, especially on the health outcome and economic benefits analyses. They are not budgeted on the project and will participate in the project as volunteers and senior advisors.

- Dr. Su and his URAP students will also draft quarterly progress reports and the final report.
The project team members, except Drs. Balmes and McConnell will have weekly conference calls to discuss issues associated with implementation of the project tasks, solutions for the tasks, and progress of the project. Drs. Balmes and McConnell will attend the meetings at least once per month or at critical moments when advice is required for moving the project forward. Dr. Su will track project progress as it relates to budgeted amounts in order to ensure that the project meets the targets.



BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. DO NOT EXCEED FIVE PAGES.

NAME: Brown, Timothy Tyler

eRA COMMONS USER NAME (credential, e.g., agency login): TIMOTHYBROWN

POSITION TITLE: Associate Professor

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
California State University, East Bay	B.S.	12/87	Business Administration
California State University, East Bay	M.A.	12/91	Economics
University of California, Berkeley	Ph.D.	12/99	Health Services and Policy Analysis

A. Personal Statement

I am Associate Professor of Health Economics in the Division of Health Policy & Management at the School of Public Health, University of California Berkeley, where I am also Associate Director for Research of the Berkeley Center for Health Technology. I am a health economist who teaches econometrics and health economics at the graduate level, including cost-effectiveness analysis.

Ongoing and recently completed projects include the following:

Ongoing Research Support

Commonwealth Fund Rodriguez (PI) 9/1/2022 – 8/31/2024
Early Evaluation of the Blue Cross Blue Shield of Massachusetts Multilevel Strategy to Advance Health Equity
Role: Co-Investigator

California Department of Health Care Services Rodriguez (PI) 7/01/2022 - 6/30/2024
Medical Interpreter Pilot Project Evaluation
Role: Co-Investigator

National Institute for Health Care Management Foundation Xiao (PI) 1/20/2023 – 1/19/2024
Do Hospital Partnerships with Community-Based Organizations that Address Social Determinants of Health Reduce Hospital Readmission and Mortality?
Role: Co-Investigator

Monterey County Behavioral Health Winston (PI) 3/01/2023 - 10/31/2023
Help@Hand Mental Health Evaluation
Role: Co-Investigator

Blue Shield of California Foundation Brewster (PI) 10/01/2022 – 9/30/2024
Evaluation of Contra Costa County Behavioral Health Crisis Initiative for Domestic Violence Prevention.
Role: Co-Investigator.

Completed Research Support

National Institute for Health Care Management Foundation Brown (PI) 5/1/2021 - 4/30/2022
The Causal Effect of High-Quality Physician-Patient Relationships on Healthcare Costs and Outcomes: Differences by Race/Ethnicity and the Effect of Racial/Ethnic Concordance
Role: Principal Investigator

Peter G. Peterson Foundation Brown (PI)
7/1/2020 – 3/31/2021
Health System ACO Evaluation
Role: Principal Investigator

B. Positions, Scientific Appointments, and Honors

Positions and Scientific Appointments

2014 – Present Associate Director for Research, Berkeley Center for Health Technology; Associate Professor of Health Economics, School of Public Health, University of California, Berkeley.
2011 – 2014 Associate Director for Research, Berkeley Center for Health Technology; Assistant Professor of Health Economics, School of Public Health, University of California, Berkeley.
2008 – 2011 Associate Director for Research and Training, Nicholas C. Petris Center on Health Care Markets and Consumer Welfare; Assistant Professor of Health Economics, School of Public Health, University of California, Berkeley
2007 – 2008 Associate Director for Research and Training, Nicholas C. Petris Center on Health Care Markets and Consumer Welfare, School of Public Health, University of California, Berkeley.
2004 – 2007 Associate Director for Research, Nicholas C. Petris Center on Health Care Markets and Consumer Welfare, School of Public Health, University of California, Berkeley.
2003 – 2004 Senior Researcher and Acting Director, Nicholas C. Petris Center on Health Care Markets and Consumer Welfare, School of Public Health, University of California, Berkeley.
2002 – 2003 Senior Researcher, Nicholas C. Petris Center on Health Care Markets and Consumer Welfare, School of Public Health, University of California, Berkeley.

Honors

2016 American Journal of Public Health Editor's Choice Award for Paper of the Year
2014 PHSR (Public Health Services Research) Article of the Year: AcademyHealth
2014 – 2015 School of Public Health Award for Teaching Excellence (Spring 2014, Spring 2015)
2003 Recognition and Reward Award, School of Public Health, University of California at Berkeley.

C. Contributions to Science

Valuation of Public Health Activities

1. I published the first set of studies that rigorously estimate the return on investment for the public health system in California. This was based on determining the causal relationship between public health expenditures and health status, the causal relationship between public health expenditures and mortality, and performing a rigorous valuation of health status. This information was combined to determine return on investment. My most recent work in this area shows that for every $1 spent on public health, Medicaid expenditures are reduced by approximately $3 (public health pays for itself by lowering Medicaid expenditures). The techniques used here can be used in many settings.

   a. Brown T. T., Murthy V. 2020. Do Public Health Activities Pay for Themselves? The Effect of County-Level Public Health Expenditures on County-Level Public Assistance Medical Care Benefits in California. *Health Economics* 29(10): 1220-1230.
   b. Brown, T. T. 2016. "Returns on Investment in California County Departments of Public Health." *American Journal of Public Health*, 106(8): 1477–1482.

c. Brown, T. T., M. S. Martinez-Gutierrez, and B. Navab. 2014. “The Impact of Changes in County Public Health Expenditures on General Health in the Population.” *Health Economics, Policy, and Law*, 9(3): 251–269.
d. Brown, T. T. 2014. “How Effective Are Public Health Departments at Preventing Mortality?” *Economics and Human Biology*, 13: 34–45.

Valuation of Non-Market Goods

2. I have published studies valuing non-market goods in various contexts. These techniques can be used in many settings.

a. Brown, T. T. 2015. “The Subjective Well-Being Method of Valuation: An Application to General Health Status.” *Health Services Research*, 50(6): 1996–2018.
b. Brown, T. T. 2013. “A Monetary Valuation of Individual Religious Behaviour: The Case of Prayer.” *Applied Economics*, 45(15): 2031–2037.

Conjoint/Discrete-Choice Valuation Studies

3. I have applied conjoint/discrete-choice approaches that can be used to value telemedicine and comprehensive navigation systems in healthcare settings. These techniques can be used in many settings.

a. Hague E., Brown T.T. “Care Concierge: A Discrete Choice Experiment to Quantify Value to Health Plan Members of a Patient Navigation and Advocacy Feature.” *Under Review*
b. Tierney A.A., Brown T.T., Aguilera A., Shortell S.M., Rodriguez H.P. “Conjoint Analysis of Telemedicine Preferences for Hypertension Management Among Adult Patients.” *Under Review*

List of Peer-Reviewed Work: https://www.ncbi.nlm.nih.gov/myncbi/timothy.brown.3/bibliography/public/

EXHIBIT A6

CURRENT & PENDING SUPPORT

University will provide current & pending support information for Key Personnel identified in Exhibit A2 at time of proposal and upon request from State agency. The "Proposed Project" is this application that is submitted to the State. Add pages as needed.

Brown, Timothy					
Status (currently active or pending approval)	**Award #** (if available)	**Source** (name of the sponsor)	**Project Title**	**Start Date**	**End Date**
CURRENTLY ACTIVE	**21RD010**	**California Air Resources Board (CARB)**	**Impact of Air Pollution Exposure on Metabolic Health Outcomes for California Residents**	**3/1/2023**	**2/27/2026**
Currently Active		**Commonwealth Fund**	**Early Evaluation of the Blue Cross Blue Shield of Massachusetts Multilevel Strategy to Advance Health Equity**	**09/01/2022**	**08/31/2024**
Currently Active		**California Department of Health Care Services**	**Medical Interpreter Pilot Project Evaluation**	**07/01/2022**	**06/30/2024**
Currently Active		**National Institute for Health Care Management Foundation**	**Do Hospital Partnerships with Community-Based Organizations that Address Social Determinants of Health Reduce Hospital Readmission and Mortality?**	**01/20/2023**	**01/19/2024**
Currently Active		**Monterey County Behavioral Health**	**Help@Hand Mental Health Evaluation**	**03/01/2023**	**10/31/2023**
Currently Active		**Blue Shield of California Foundation**	**Evaluation of Contra Costa County Behavioral Health Crisis Initiative for Domestic Violence Prevention**	**10/01/2022**	**09/30/2024**