Specific Aims. Ambient air pollution is a major environmental threat. In 2015, exposure to $PM_{2.5}$ (particles with diameter $\leq$ 2.5 μm) was the fifth-ranking mortality risk factor, causing 4.2 million deaths globally,[1] and the sixth-ranking factor for disability-adjusted life-years.[2] Research groups are developing distinct spatio-temporal models (*exposure models*) to predict ambient air pollution exposures of study participants even in areas where air pollution monitors are sparse.[3–13] Use of these exposure models has led to strong evidence of air pollution-related adverse health effects. In the US, two recent nationwide studies[14,15] using such a model[4] to assign exposures and data from the entire Medicare population ($>$60 million) confirmed previous findings[16–20] and reported increased mortality risks even at $PM_{2.5}$ levels below the current national standards. A **significant limitation** is that health effect estimates from all these studies and their statistical uncertainty are based on the *very strong assumption that a single exposure model is correct*. Even when distinct exposure models predict similarly well, the spatio-temporal patterns of their prediction errors might differ, potentially leading to misleading evidence of health effect heterogeneity across sub-populations, which in turn could be wrongly attributed to other factors that vary in a similar manner either in space (e.g. rural vs. urban) or time (e.g. by season). Additionally, even very small exposure errors can lead to incorrect inferences, especially when estimating health effects at low levels of air pollution. To date, we lack statistical methods to **comprehensively quantify the spatio-temporal uncertainty associated with model selection and exposure assessment, and propagate it into the estimated health effects**. With the proposed work we will overcome this limitation (**Aim 1**) and, furthermore, fully characterize exposure uncertainty across multiple pollutants simultaneously (**Aim 2**). Overcoming these two methodological gaps is critical for (1) guiding the placement of new monitoring sites in areas of high exposure uncertainty; (2) propagating exposure uncertainty when estimating health effects in single- and multi-pollutant settings; and (3) avoiding misleading evidence of heterogeneity when characterizing vulnerable sub-populations. Our goal is to develop innovative methods to overcome these challenges and leverage nationwide data at an unprecedented scale (millions of Medicare and Medicaid enrollees—the two largest health insurance providers in the US—as well as more than 25 existing air pollution prediction models) to provide policy-relevant evidence on air pollution health effects (**Aim 3**), within an open source platform (**Aim 4**). We have assembled a highly interdisciplinary team of biostatisticians, data scientists, epidemiologists, atmospheric chemists, and engineers to achieve the following specific aims:

1. **Develop an innovative ensemble model framework for air pollution exposure prediction and spatio-temporal uncertainty characterization.** We will combine information across multiple exposure prediction models, weighing each model by its predictive accuracy at each space and time point.
   (a) We will build and validate nationwide ensemble models and characterize the uncertainty in exposure prediction spatially and temporally (e.g. daily vs. annual predicted concentrations).
   (b) We will apply the newly developed framework to global air pollution models to identify those areas with highest uncertainty in $PM_{2.5}$ estimated concentrations to inform placement of future monitoring sites.
2. **Develop a multi-pollutant ensemble model framework.** We will develop a framework for joint prediction of multiple pollutants and characterize their spatio-temporal uncertainty. We will build and validate a nationwide multi-pollutant ensemble model for $PM_{2.5}$, O_3 and NO_2.
3. **Estimate health effects that fully incorporate exposure uncertainty.** We will apply our methods to: a) the whole **Medicare** and **Medicaid** populations, with information available at zip-codes; b) all hospitalizations in **New York State** with information available at residential addresses. We will estimate health effects associated with short- and long-term exposure to $PM_{2.5}$, O_3 and NO_2, fully accounting for the spatio-temporal uncertainty in the exposure assessment. The NY data will be used to validate our nationwide findings and to quantify the uncertainty related to outcome spatial resolution (zip-code vs. residential address).
4. **Create an open science portal to allow data and resource sharing for exposure prediction and uncertainty characterization and propagation**. Develop peer-reviewed software and visualization tools for communication of the spatio-temporal uncertainty of air pollution exposures and health effect estimates.

Impact. The current administration has either reversed or has proposed to reverse numerous air pollution regulations which will lead to an estimated increase of $>$80,000 extra deaths per decade.[21] Therefore, rigorous evidence of air pollution health effects that fully accounts for exposure uncertainty, including at low concentrations, is of **utmost importance**. The proposed ensemble modeling framework will shift the research paradigm, and the methods developed can be applied to any spatio-temporally varying exposure, beyond air pollution, locally or globally.