INITIAL APPLICATION PART 2

1. Abstract:

Wildfires are increasing in size and intensity. The smoke plumes generated by fires travel thousands of miles and affect communities even when they are distant from actual fire activity. We will measure the impact of wildfire exposures (both smoke and exposure to local fire activity) on health, health service utilization, and costs. Wildfire data will come from a variety of sources that include satellite measures of smoke plumes, modelled estimates of air pollution, and data on fire perimeters available from the federal government. These data will be merged (by geography and time) to administrative health care records from the Colorado All Payer Claims Data (APCD). The data have insurance enrollment and medical claims for Medicaid and a subset of commercial plans in Colorado. Data will be limited to people 0-64 years old. The data contain information on diagnoses, services, and costs. These claims data will also be linked with birth certificate records for 0-18 year olds which provide higher quality demographic variables than is available in the claims and they allow us to measure health at birth for people we only start to observe in the APCD after their birth. They will also be merged with hospital discharge and ER admission files that cover the entire population. Those data allow us to track hospitlizations for people we initially observe in the APCD, but who then exit the data either because they become uninsured or because they obtain health insurance that is not part of the APCD. None of the merging conducted in the project will be done to identify respondents and the act of merging will not allow the research team to identify individuals. These secondary data will be de-identified, but will include some "limited" data elements (such as date of birth, resident location, service dates, etc). Data will stored in a highly secure research environment overseen by the PI in conjunction with IT professionals in the School of Public Health.

2. Subject Selection:

a. **Recruitment:** Subjects will be obtained from existing administrative data. No subjects will be pro-actively recruited.

b. **Eligibility Criteria:** All 0-64 year olds with Medicaid and commercial plan enrollment covered by the Colorado APCD.

c. **Rationale:** Wildfire exposures are known to affect a range of health outcomes across the life course. We must limit to Medicaid and a subset of selected commercial plans because that is all that is available.

d. **Enrollment Numbers:** The data vendor (CIVHC) provides only limited information on the number of unique people. They report 3.7 unique people by year. We will obtain data for 2016-2024, however many people will be observed in multiple years. We expect to enroll no more than 6 million unique individuals.

e. **Rationale for Enrollment Numbers**: Our enrollment size is determined by the data that is available. However, because we are using administrative data on all people within a defined target population, we will have large sample sizes that ensure both scientific benefit and minimize disclosure risks.

3. **Procedures:**

   The data are restricted use. They require separate applications to the owners of the APCD, the birth file, and the hospital/er files.

   As part of the application process the PI needs to attend a data privacy review meeting overssen by state officials in Colorado. Our IRB application must reviewed and a decision letter issued prior to the meeting. The date of the meeting is May 7th. While the data owners do not require "expedited review", having the IRB decision before May 7th would facilitate an orderly application process without delays.

   The merging of the CO APCD, birth records, all payer hospital discharge data, and wildfire exposures will provide a rich set of demographic, health care utilization, health, and wildfire exposure data. The analytic data will contain hundreds of variables. They include information on mothers race, nativity, age, and education, date of birth of infant, gestational age, birth weight, and income ranges (from the birth certificate), patient age and sex, medical diagnoses and procedures, health care costs, dates of services, provider type, insurance type and enrollment dates (from the APCD), hospital and emergency department events (including diagnoses, procedures, and costs) from the hospital/ER linkage, and wildfire exposures (exposure to smoke plumes, exposure to air pollutants such as PM2.5, and exposure to local fire activity) from our wildfire data.

   The merge itself will be facilitated by the data owners. The data owners will create anonymous IDs that are common across each data source and then remove all real personal identifiers from the analytic files they deliver to us. We will use these anonymous identifiers to merge the files.

   After all data elements have been merged, we will conduct several statistical analyses. These analyses will estimate the association between wildfire exposure levels and outcomes (health service use, health measured via diagnoses, and health care costs) in a multiple regression framework. Exposures of interest include smoke plumes, wildfire attributable air pollution, and local fire activity. The regressions relate changes in exposures within a geographic area to changes in outcomes within that geographic unit, using other geographic units observed at the same time that do not have exposure changes as a comparison group. Economists refer to this kind of regression as a two-way fixed effects model. The geographic units of interest will vary based on the exact regression, but will include Census Tracts or Counties. Time will also vary based on the regression and will include individual dates or months. We will a number of different health outcomes (e.g. asthma, cardiovascular conditions). We will focus on three primary patient subpopulaitons of interest: infants, children, maternal patients (those about to or recently given birth), and non-elderly adults (outside of the maternal patient population).

4. **Risks:**

   We expect minimal risks to this secondary data study. These patients data have already been collected and thus our use of the data poses no additional risks or burdens

associated with data collection. We will not observe patient identity. Colorado state agencies and the data vendor will create masked subject ID to facilitate merging between the APCD, birth files, and hospital/er files. We will not have access to the cross-walk between actual patient id and this masked identifier. We will take a number of steps to further protect against indirect identification (e.g. we will apply CMS safe harbor rules and will store and analyze data in a highly protected research environment).

5. Benefits:

Benefits will be indirect. Our analysis will yield valuable insights about the harms caused by wildfire exposure and the current ability of the health system to adequately treat relevant conditions. Our results can be used by public health officials and health care providers to improve prevention and treatment efforts.

6. Confidentiality:

This secondary data study will use de-indentified data to protect the confidentiality of the subjects. For some cases, date of birth is being requested in order to accurately identify wildfire smoke exposures during the pre-natal period. Smoke exposures can be short in duration and without date of birth we would likely misclassify prenatal smoke exposure.

Data will be stored on a research server that is only accessible through a virtual machine. Access to the data and computing enviorment is safe guarded by UMDs two-factor CAS system. Only the PI in collaboration with IT staff in the school of public health will be able to grant permissions to this environment. The environment is both physically and virtually secure. All data will be destroyed after the project is complete.

7. Consent Process:

We seek a waiver of written consent.

(1) The research poses no more than minimal risks (see section 4).
(2) The study is based on solely on secondary data; no interaction with participants will occur.
(3) We will not have access to any contact information in this data set to be able to contact participants to obtain consent
(4) It is not applicable to provide additional information after participation because we will not have contact information for participants.

8. Conflict of Interest:

We have no conflicts

9. HIPAA Compliance:

This project does not use data covered by HIPAA.

10. Research Outside of the United States:

No research will be conducted outside of the United States

11. Research Involving Prisoners:

This project will not involve prisoners.

12. SUPPORTING DOCUMENTS

Your Initial Application must include a **completed Initial Application Part 1 (On-Line Document),** the information required in items 1-11 above, and all relevant supporting documents including: consent forms, letters sent to recruit participants, questionnaires completed by participants, and any other material that will be presented, viewed or read to human subject participants.

The consent forms in your approved <u>IRBNet PACKAGE</u> must be used. When creating or editing your consent form, please provide the most recent IRBNet package number at the bottom, right corner of the consent form. This ensures you are using the most "up-to-date" version of the form.

To find your IRBNet package number, go to the MY PROJECTS tab and click on the title of your project. In the PROJECT OVERVIEW page, your IRBNet package number will be listed at the top, next to your project title.