**Request for Approval under the "Generic Clearance for Citizen Science and Crowdsourcing Projects" (OMB Control Number: 2080-0083)**

**TITLE OF INFORMATION COLLECTION:** Smoke Sense

**PURPOSE:**

The health effects of air pollution range from low severity outcomes that impact large segments of populations (respiratory symptoms), to the most severe ones including hospitalization and death. Wildfires are a significant source of air pollution, however less is known about wildfire specific impacts on health. Better insight into the magnitude and scope of health impacts and improved health risk communication tools are key elements to improving health outcomes in the affected communities. We are proposing a specific research activity that integrates environmental models of smoke with EPA's current communication strategy in the context of wildfire smoke to characterize the impacts on human health and provide an understanding on how the public responds to existing health risk communication strategy.

The purpose of this research is to quantify health impacts and examine the ability of environmental models to provide timely information on when and where smoky conditions are likely to occur and improve health outcomes.

**NEED AND AUTHORITY FOR COLLECTION:**

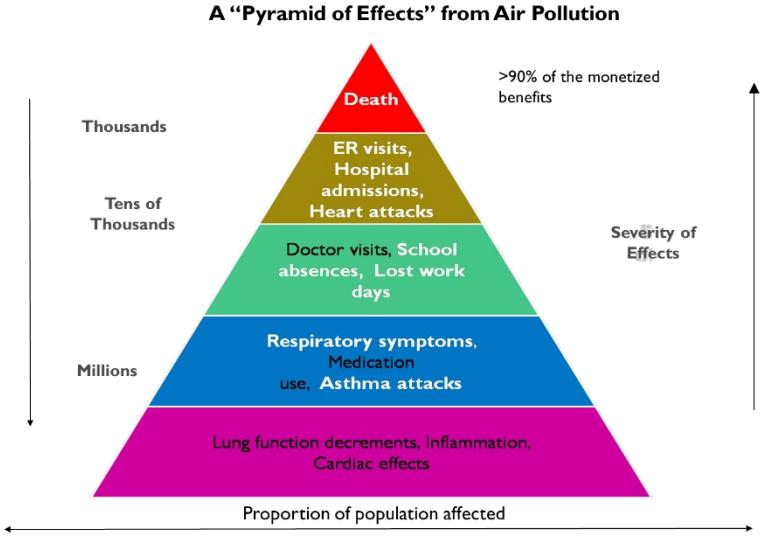
The information is needed to: quantify health impacts, examine the ability of environmental models to provide timely information on when and where hazardous conditions are likely to present, and to quantify how individuals respond to that information.

The health impacts from air pollution exposures range from low severity outcomes that impact large segments of populations such as increased incidence of respiratory symptoms, to the most severe ones including hospitalizations and death. The impacts of air pollution on the most severe outcomes are established based on hospital records and large administrative databases and are well documented. However, a number of questions remain unanswered; at what concentrations do impacts occur? Who is the most susceptible? What pre-existing health conditions increase susceptibility? Is the susceptible segment of the population aware of their risks?

By comparison, but less understood, is the magnitude of the health burden and loss of productivity due to less severe outcomes. This is due in part to the difficulty in measuring the health burden using traditional research methods. Low levels of smoke can be intangible and often invisible, resulting in the public's perception of risk that is lower than actual risk. It is therefore critical that we improve health risk messaging surrounding wildfire smoke, and provide people with tools that allow them to make informed decisions on how to reduce exposures and protect their health.

Figure 1 summarizes the magnitude of air pollution-related health effects, along with the severity of effects and the proportion of population affected. The largest portion of population is affected by less severe effects such as lung function decrements and airway inflammation. Symptoms stemming from these effects are also the hardest to measure. At the top of the pyramid is death, the most severe effect, impacting the smallest population, while accounting for 90% of the monetized benefits of air pollution regulation and the most straight-forward and easiest to measure.

In this study we are particularly interested in the frequency of symptoms including low severity symptoms (e.g., coughing, headache, etc.) that impact the population at large but are not reported officially.




This information will be collected under the legal authority of the Clean Air Act § 103, 42 U.S.C. § 7403, the National Environmental Education Act, § 4, 20 U.S.C. § 5503, and OMB Memo M-15-16. The Clean Air Act authorizes research into techniques for monitoring and controlling air pollution, including the health effects of air pollution. This study is concerned with low severity health effects of wildfire smoke, a growing contributor to particulate matter air pollution. The National Environmental Education Act, § 4, 20 U.S.C. § 5503 authorizes EPA to develop and support programs to increase environmental literacy. The EPA provides information on air quality and wildfire smoke, but how individuals respond to that information is generally not well known. This study has the potential to improve EPA's communication on wildfire smoke through better quantification of responses to air quality messaging. OMB Memo M-15-16 encourages agencies to use approaches such as citizen science, which is a cornerstone of this study.

**USES OF RESULTING DATA:**

The information collected in this study will provide knowledge about the effectiveness of wildfire communications, the effects of wildfire smoke on daily activities and low severity health symptoms, and the engagement of participants in citizen science. This knowledge will allow for better communication planning for public health messages regarding wildfire smoke. This information is not for regulatory use.

**DATA COLLECTION METHODS**:

Data will be collected through the "Smoke Sense" mobile application. The Smoke Sense App was developed by Sonoma Technologies, Inc. (STI) 1450 N. McDowell Blvd., Suite 200, Petaluma, CA 94954-6515. STI has developed air pollution and health web-based studies including the EPA Village Green. The Smoke Sense application is currently being updated and streamlined by The EPA's Environmental Modeling and Visualization Laboratory (EMVL, https://www.epa.gov/hesc/environmental-modeling-and-visualization-laboratory). The Smoke Sense App will be run May - Oct 2018 for users to download and use, and data will be collected for 16 weeks

thereafter. Each data file shall be accompanied with clear documentation of both data management programs and the data fields in hardcopy and in electronic format.

Data management: Data will be backed up daily. Study documents and data will be kept on a secure workstation in the EPA Human Studies Facility. Secondary data files, computer code, and documentation will reside on the Smoke Sense shared network directory administered by the data custodian.

This project falls under record schedule 1035(b), which has a 20-year retention requirement. Records will be kept in accordance with ORD PPMs 13.2 and 13.4 which can be found here: http://intranet.ord.epa.gov/about-ord/chapter-13-quality-assurance. Study materials that are not Federal records according to EPA Records Schedule 0008: Non-records (http://intranet.epa.gov/records/schedule/final/0008.html), including copies, intermediate data files, computer programs under development, output not used for any manuscripts or reports, prior versions of analytic software, working papers, draft manuscripts, and technical reference materials, are disposable and may be destroyed when obsolete, superseded or no longer needed for reference.

Smoke Sense study protocol has been reviewed and approved by the Office of Human Research Ethics University of North Carolina-Chapel Hill Institutional Review Board and was determined to be exempt from further review according to the regulatory category 2. Survey, interview, public observation under 45 CFR 46.101(b). Study protocol does not use personally identifiable data such as phone numbers, names or addresses however, for research purposes we have to collect data longitudinally and need to be able to track responses by user. For this purpose we will use “vendor/advertiser mobile device ID”. Research data will be stripped of “vendor/advertiser ID” and only an assigned participant ID will be saved for research purposes. We will use advertiser/vendor ID of device because these identifiers are “non-permanent and non-personal” mobile device identifiers that cannot be traced to individuals. (https://developer.apple.com/reference/uikit/uidevice/1620059-identifierforvendor). We will not use GPS to track individuals, only their preferred locations established in the Profile Survey will be saved and used to establish exposure levels retrospectively. Other information collected, such as age group, race, and sex are not considered Personally identifiable information (PII).

The probability of harm from a potential data breach is very low, as the only way to identify individuals is through advertising and vendor identifiers of the app. The magnitude of the possible risk of harm is also very low because disclosure of the general information (i.e., demographic, general health, activity level and symptoms) collected would not reasonably place the participants at risk of criminal or civil liability or be damaging to the participants’ financial standing, employability, or reputation. Moreover, unlike most survey-based research, participation will provide benefits to participants. Participants will be provided with sophisticated environmental model predictions of smoke, as well as EPA-approved and recommended measures to reduce health impacts as benefits of participating in the study. It is important to state that all recommendations and health risk communications in this study will be based on the same data that federal agencies provide to citizens through AirNow (https://www.airnow.gov/), the most widely used air quality and health risk communication tool.

**PARTICIPANT UNIVERSE:**

**Burden based on 2017 pilot study**

| Category of Respondent | No. of Respondents | Number of responses per respondent | Total Number of Responses | Participation Time per response | Burden Hours |
|---|---|---|---|---|---|
| Participant registration | 4,508 | 1 | 4,508 | 0.1 hours | 451 |
| Health and behavior surveys | 4,508 | Average 2.5 | 11,270 | ~0.25 hours | 2,818 |
| **Total** | **15,778** | | **15,778** | | **3,269** |

**Burden estimates for 2018**

| Category of Respondent | No. of Respondents | Number of responses per respondent | Total Number of Responses | Participation Time per response | Burden Hours |
|---|---|---|---|---|---|
| Participant registration | 19,000* | 1 | 19,000 | 0.1 hours | 1,900 |
| Health and behavior surveys | 19,000 | 16 maximum | 304,000 | ~0.20 hours** | 60,800 |
| **Total** | **323,000** | | **323,000** | | **62,700** |
| **Total Respondents and Burden Hours, 2017 and 2018** | **338,778** | | **338,778** | | **65,969** |

*expectation of increased word of mouth engagement and partner outreach post-pilot study
** questions have been updated and streamlined for better interaction

Burden estimates are based on the previous and expected number of Smoke Sense app users per day (number of respondents), actual average participation for the 2017 pilot or full participation for the 16-week study period, and the expectation that participants will take no more than 10 minutes (or less) to respond to each week's surveys. The respondent universe may vary among the activities listed because not all respondents must complete each activity.

There will not be fees associated with participation in the data collections. Participants will not be required to purchase any equipment to collect data, as all data collection will be through applications on personal smartphones.

**AGENCY COST:** The estimated annual cost to the Federal government is $106.779. The cost estimate accounts for 10% GS 14 at $56.49/hour for 30 weeks = $6,779 and $100,000 in contract and internal support.

**STATISTICAL ANALYSIS:**

Once the study period is complete, data will be analyzed to determine onset and incidence rates of health impacts in association with wildfire events/PM.

Analyses will be performed in standard statistical software, such as SAS, R, or Stata. Descriptive analyses using univariate and bivariate distributions and frequencies will be used to characterize and describe the data. Poisson regression and case crossover analysis are the main statistical techniques that will be employed to estimate effect measures and 95% confidence intervals. For a Poisson regression analysis (time-series), health events will be aggregated into time-series of Poisson counts by date of event within each category and metropolitan region, matched with data on environmental stressors by date and locality, and analyzed using either a generalized additive model or a mixed-effects model to estimate the association between time-varying health event counts and time-varying environmental measure(s) while adjusting for time-varying confounders such as trend, season, and other environmental stressors. For a case-crossover analysis, the date of each health event (case-period) is used to identify one or more randomly selected dates within the same calendar month (reference-periods). Each case-period and reference-period is then matched with environmental stress and meteorological data by date and region. Finally, conditional logistic regression is used to analyze the relationship between case status and an indicator of environmental stress while adjusting for time-varying confounders such as meteorology. In this analytic approach, confounding by trend and season are controlled by design.

Survey responses will be tested for representativeness of underlying population and appropriate statistical methods will be used to adjust sampled populations. Incidence rates will be compared to those obtained in the health care institution based surveillance system.

The anticipated results will satisfy the study objectives of quantifying low severity health outcome response to wildfire smoke exposure, and provide information on behavioral response to wildfire smoke exposure. This information may then be used to improve messaging/communication in regard to wildfire smoke hazards.

**DATA QUALITY ASSESSMENT PROCEDURES:**

After data is received by EPA investigators, will perform cleaning and quality checks on the data. Within the study design, we incorporated a number of measures to reduce the chance of erroneous data. For example, because responses are constrained by the nature of the electronic survey (i.e., answers to questions are restricted to pull down or checkbox responses. Therefore, no out of range or inappropriate responses will be allowed.), out of range values are not expected to be of concern; data will be examined to ensure these criteria is met. This will be done through frequency and distribution statistics. Data will also be examined for potential outliers. While at present potential outliers are unknown, an example might be a person who fills out experiencing all symptoms for every day of every week during the study period, as this is an unlikely scenario. Any observed outliers will be examined for influence on the data using sensitivity analyses with and without their inclusion. In this way impact of outlier removal can be ascertained. Any outlier removal from further analyses will be documented, in both study procedures and in any resulting study publications. Due to the nature of citizen science, missingness of survey responses is a potential issue. Participants are neither required to answer the surveys, or to answer particular questions in the surveys. Data will therefore be examined for patterns of missing data, in particular data missing at random (MAR) or data missing not at random (MNAR). Established statistical techniques for handling MAR and MNAR data will then be used. For MAR data, these include adjusting for variables

which are known to influence missingness. For MNAR, possible techniques include complete case analysis that is then specified to only apply to certain subpopulations within the study, imputation of missing data which depends on completeness of non-missing data, stratification, and others (Little and Rubin 2014; Rubin 1976).

**ADMINISTRATION OF THE INSTRUMENT:** (Check all that apply)

[X] Web-based or Social Media
[ ] Telephone
[ ] In-person
[ ] Mail
[ ] Other, Explain

**INSTRUMENT:** Append a copy of the questionnaire or a screen shot of the website or app that includes the information collection.

**CONTACT NAME:** Ana Rappold **EMAIL:** Rappold.Ana@epa.gov