ARIC Manuscript Proposal #3460

PC Reviewed: 9/10/19 **Status: _____** **Priority: 2**

SC Reviewed: _________ **Status: _____** **Priority: ____**

1.a. Full Title: Accounting for exposure measurement error in assessment of the effects of air pollution on dementia

b. Abbreviated Title (Length 26 characters): Accounting for exposure measurement error

2. Writing Group: Eun Sug Park, Melinda C. Power, Richard Smith, Xiaohui Xu, Eric A. Whitsel, Jay Stewart, Qi Ying, others welcome

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. _____ **[please confirm with your initials electronically or in writing]**

First author: **Eun Sug Park**
Address: Texas A&M Transportation Institute, The Texas A&M University System, 3135 TAMU, College Station, TX 77843-3135

Phone: (979) 317-2466

E-mail: e-park@tamu.edu

ARIC author to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).

Name: Melinda C. Power

Address: 950 New Hampshire Ave NW, Washington DC

Phone: (202) 994-7778

E-mail: power@gwu.edu

3. Timeline: 6-9 months following completion of both air pollution estimation and Visit 7 availability

4. Rationale:

While it is recognized that exposure measurement error is one of major methodological challenges in research estimating the effects of environmental exposures on cognitive decline and dementia incidence (Paul et al. 2019), there has been almost no attempt to explicitly account for exposure measurement error in estimating the effects of air pollution on cognitive decline and dementia incidence. Most prior studies examine the association between air pollution and cognitive test scores by using exposure estimates as an independent variable as if they were true exposure under an implicit assumption that exposure measurement error is negligible. While the direct plug-in of the exposure estimates has been the most common use in a health effects analysis, this approach completely ignores measurement error associated with estimated/predicted exposure. As a matter of fact, all exposure estimates are prone to measurement error regardless of the exposure prediction model and it has been shown that ignoring exposure measurement error can lead to biased health effect estimates as well as incorrect uncertainty estimates (see, e.g., Carroll et al. 1995, Zeger et al. 2000).

There are several potential sources of exposure measurement error depending on available data and exposure estimation methods. If ambient concentrations measured from routine ground-level monitoring are used as exposure, spatial misalignment of monitoring data and study participants becomes one important source of exposure measurement error. Other exposure estimation methods which rely on modelling air pollution (e.g. chemical transport models, dispersion models, land use regression models, spatiotemporal models, etc.) are subject to issues of measurement error caused by uncertainty in prediction models. Temporal misalignment is another source of exposure measurement error. Most previous studies relating air pollution and cognitive changes examine the association between air pollution and cognitive test scores consider the impact of recent (past year) exposures (Power et el. 2016). Such analyses typically assume recent exposures are indicative of exposure over longer periods. Thus, they are subject to significant measurement error resulting from temporal misalignment.

Spatial coverage of ground-level observations of air pollutants is limited for the areas in which ARIC participants live. The only existing study of health effects of long-term air pollution in ARIC participants used a spatiotemporal statistical model (Power et al. 2018) to reduce measurement error from spatial misalignment. However, exposure measurement error was not explicitly taken into account in the final health effects estimation.

The overall goal of this proposal is to develop a method that can account for exposure measurement error in the exposure assessment while quantifying the effects of air pollution on dementia and related outcomes in the ARIC cohort. While we expect that new exposure estimates generated by an ensemble/data-fusing approach in the main project will have significantly reduced spatial and temporal exposure measurement error compared to those generated by the original Community Multiscale Air Quality (CMAQ) model or those using air pollution data measured from only a few monitoring station and/or at a single time point (e.g., in the year prior to cognitive testing), new estimates are still subject to some degree of measurement error including model calibration error (prediction error) as well as remaining spatial misalignment error resulting from using grid level concentration estimates for individual location concentration. As mentioned above, using exposure estimates directly as a covariate in the health model is considered as a plug-in approach which ignores potential exposure measurement error in exposure estimates. Although we plan to quantify measurement error of new exposure estimates using a hold-out monitor in the main project, merely quantifying measurement error does not mean accounting for exposure measurement errors in health effects estimates. Thus, the focus of the present proposal is to explicitly account for measurement error in new exposure estimates in estimation of health effects of air pollution on dementia and assess the impact of measurement error on health effect estimates.

5. Main Hypothesis/Study Questions:

This project supports the aims of a funded grant to investigate the association of ambient exposure to air pollutants with cognitive decline, incident cognitive impairment, and neuroimaging markers of dementia-related pathology in the ARIC study. Our objective in this proposal is to develop an enhanced approach that can account for exposure measurement error (resulting from uncertainty in exposure estimates from the ensemble/data-fusing approach as well as remaining spatial misalignment error) in health effects evaluation. Our specific aims are:

Aim 1. Develop a method that can account for exposure measurement error in estimation of health effects of air pollution on cognitive and neuroimaging outcomes in the ARIC study.

Aim 2. Compare health effects estimates and the associated uncertainty estimates obtained by approaches with and without accounting for exposure measurement errors based on the ARIC data.

Aim 3. Quantify influence of accounting for exposure measurement error in health effects evaluation based on the ARIC data.

6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

The datasets used in the previous proposal by Power (ARIC MP#3348) will also be used in this proposal. Exclusions will be non-white/non-black, black in MN or MD, or missing relevant exposure or covariate data. We will also use monthly (or yearly) ambient air pollution exposure estimates for $PM_{2.5}$, PM_{10}, and O_3 obtained at the grid cell level (e.g., 1 km by 1 km grid squares) by an ensemble/data-fusing approach (combining two chemical transport models (CMAQ, UCD/CTI), two emission inventories (NEI and EDGAR), and monitor data. We will, however, explicitly recognize that the exposure estimates obtained from an ensemble/data-fusing approach are not true exposure (but surrogates of true exposures) and are still subject to measurement error caused by uncertainty in prediction process as well as spatial misalignment. Dependent variables will be the ADRD-related outcome variables. Covariates will include demographics (age, sex, race), personal SES (education, pre-retirement health insurance status), self-rated health (alcohol use, smoking status, leisure time physical activity), and area-level SES (% of residential census tract below the poverty line, a summary measure of neighborhood wealth/income, education and occupation). Effect modifiers will include APOE e4, sex, smoking status, physical activity, obesity, SES, and co-exposures.

We define true exposure as unobserved true (monthly or yearly) ambient outdoor pollutant (primarily $PM_{2.5}$ or PM constituents) concentrations at each participant's residential address over the exposure time window. We admit that true monthly ambient pollutant (e.g., $PM_{2.5}$) concentrations at each participant's residential address during the study period are unknown and exposure estimates are subject to exposure measurement error. To address the issue of measurement errors and uncertainties associated with use of predicted exposure from an ensemble/data-fusing approach, we will employ Bayesian hierarchical modeling to estimate the impact of air pollution exposure on our cognitive and neuroimaging outcomes. Bayesian hierarchical models have been successfully used as a way of accounting for uncertainty in exposure assessment in health effects evaluation (see, e.g., Holloman et al. 2004; Molitor et al. 2006; Calder et al. 2008; Park et al. 2014, 2015), but have not been used in assessing the long-term effects of air pollution on the ADRD-related outcomes. Here, our Bayesian model consists of two levels: (Level 1) a measurement model that links predicted air pollution concentrations with the unobserved true ambient exposure, and (Level 2) a health model that links the ADRD-related outcomes with true ambient exposure.

For a health model, we will use an appropriate generalized linear model for modelling health effects depending on the ADRD-related outcome (e.g., linear mixed models for DSST, DWRT, WFT cognitive test scores or brain volumes, Cox proportional hazard models for incident dementia, or logistic regression for cerebrovascular findings). Specifically, we will link the unobserved true ambient exposure at the geocoded residential address for each participant to health outcome variable (e.g., cognitive test score) along with relevant covariates and random effects to handle the within person correlations of health outcome arising from the repeated measures across time (when needed). We will assume the predicted pollutant concentrations and the health outcomes are conditionally independent given the unobserved true exposures and model covariates.

For a measurement error model, we will consider estimated exposure by an ensemble/data-fusing approach as surrogated measures for true exposure. Note that while health outcome data are measured at the individual level, estimated exposure from an ensemble/data-fusing approach is obtained at the grid cell level (e.g., 1 km by 1 km grid squares). We will assume that estimated exposure by an ensemble/data-fusing approach predicts the true ambient surface at the centroid of the grid cell with measurement error modeled by a Gaussian process and construct a continuous spatial process for exposure by the discrete process convolution approach of Higdon (1998), or other more recent approaches to modeling nonstationary spatial processes. Higdon's approach creates a continuous spatial process by convolving a discrete white noise process with a smoothing kernel. Using this approach, the exposure estimate at the geocoded residential address for each participant can easily be obtained and thus further reducing spatial misalignment error. Estimation of measurement error variance will be based on the available monitoring data (held-out sample) not used in the ensemble/data-fusing step. For the prior distribution for measurement error variance, we will assume a vague but proper inverse-gamma distribution, and the posterior distribution will be calculated based on the observed prediction errors (i.e., the residuals computed as the difference between the exposure estimates and the monitoring data from the held-out sample). Although we will initially assume that measurement errors are spatially independent and the measurement error distribution (Gaussian with constant mean and variance) does not change over space because of limited monitoring data (e.g., for $PM_{2.5}$, there are only between 1 and 3 monitoring stations at each of 3 ARIC sites (NC, MD, and MS) and between 2 and 7 monitoring stations at MN site in any given year during 1999-2017), we will check whether the constant mean and variance assumption for measurement errors is approximately satisfied whenever the external data from at least two held-out monitors (which are not used in any steps of the ensemble/data-fusing approach) are available to ensure that there is no residual spatial structure in the estimated exposure. It is possible that the size of measurement error may differ by each site, which will facilitate investigation of influence of measurement error of different sizes in health effects analysis. Depending on data availability, we will also explore additional models for true exposure incorporating land-use regression or GIS-based covariates that may help with modeling fine-scale spatial variation of exposure, and assess how incorporating those variables will affect prediction accuracy of the exposure at a held-out monitor (not used in an ensemble/data-fusing process or in estimating measurement error variance) and health effect estimates.

We will consider a 2-stage Bayesian approach as well as a fully Bayesian approach modeling exposure and health outcome jointly. In the 2-stage Bayesian approach, we will first fit the measurement model and then use the interim posterior for true exposure from the measurement model as a prior distribution for true exposure in the health model. An advantage of the 2-stage approach is that the exposure model does not need to be refitted when running multiple health effect analyses, although it does not cut feedback between the health observations and the exposure estimates as noted in Gryparis et al. (2009).

As an alternative to Bayesian approaches, we should note that non-Bayesian approaches to measurement error regression estimation have also been derived, e.g. Szpiro et al. (2011), Szpiro and Paciorek (2013). These approaches focus particularly on the spatial misalignment problem (monitors in different locations from participants) and implicitly assume that monitors are the principal source of air pollution data. Szpiro and Paciorek (2013, discussion section) note that their approach does not apply to the case where air pollution models are the principal source of exposure information. Nevertheless, we plan to explore the possibility of likelihood-based approaches as well, with simulation as a means to assess their frequentist properties.

We will compare health effects estimates from a plug-in approach (not accounting for measurement error), a 2-stage Bayesian approach accounting for measurement error, and a fully Bayesian approach accounting for measurement error to assess the impact of accounting for exposure measurement error in health effects estimation by different pollutants and ARIC sites. We will also assess the effect of the size of measurement error (quantified to some degree by the estimate of measurement error variance) on the health effect estimates and provide guidance for future studies.

7.a. Will the data be used for non-CVD analysis in this manuscript? __x__ Yes ____ No

b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = "CVD Research" for non-DNA analysis, and for DNA analysis RES_DNA = "CVD Research" would be used? __x_ Yes ____ No

(This file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript? _x__ Yes ____ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER03 must be used to exclude those with value RES_DNA = "No use/storage DNA"? __x_ Yes ____ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status. ARIC Investigators have access to

the publications lists under the Study Members Area of the web site at:
http://www.cscc.unc.edu/aric/mantrack/maintain/search/dtSearch.html

___X___ Yes ________ No

10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

2412; “Association of particulate matter air pollution with MRI outcomes” (Power)

2288; “Associations of Brain Imaging with Cognitive Change over 20 Years” (Knopman)

2248: “Ambient air pollution and late-life cognition and dementia” (Power)

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? __X__ Yes ____ No

11.b. If yes, is the proposal

_x_ A. primarily the result of an ancillary study (list number* 2016.20

___ B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* __________ __________ __________)

*ancillary studies are listed by number at https://www2.cscc.unc.edu/aric/approved-ancillary-studies

12a. Manuscript preparation is expected to be completed in one to three years. If a manuscript is not submitted for ARIC review at the end of the 3-years from the date of the approval, the manuscript proposal will expire.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is **your responsibility to upload manuscripts to PubMed Central** whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.cscc.unc.edu/aric/index.php, under Publications, Policies & Forms. http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

References:

Calder CA, Holloman CH, Bortnick SM, Strauss W, Morara M. 2008. Relating ambient particulate matter concentration levels to mortality using an exposure simulator.

Carroll RJ, Ruppert D, Stefanski LA. 1995. Measurement Error in Nonlinear Models. London: Chapman & Hall.

Gryparis A, Paciorek CJ, Zeka A, Schwartz J, Coull BA. Measurement error caused by spatial misalignment in environmental epidemiology. Biostatistics. 2009;10(2):258-74. Epub 2008/10/18. doi: kxn033 [pii]

Higdon, D. (1998), "A Process-Convolution Approach to Modeling Temperatures in the North Atlantic Ocean," *Environmental and Ecological Statistics,* 5, 173–190.

Holloman CH, Bortnick SM, Morara M, Strauss WJ, Calder CA. A Bayesian Hierarchical Approach for Relating $PM_{2.5}$ Exposure to Cardiovascular Mortality in North Carolina. *Environmental Health Perspectives*. 2004;112(13):1282-1288. doi:10.1289/ehp.6980.

Molitor J, Molitor NT, Jerrett M, McConnell R, Gauderman J, Berhane K and Thomas D. 2006. Bayesian modelling of air pollution health effects with missing exposure data. Am J Epidemiol 164: 69-76.

Park ES, Hopke PK, Oh MS, Han D, Symanski E, Spiegelman CH. *Assessment of source specific health effects associated with an unknown number of major sources of multiple air pollutants: A unified Bayesian approach*. Biostatistics 2014; 15:484-497.

Park ES, Symanski E, Han D, Spiegelman C. *Part 2. Development of Enhanced Statistical Methods for Assessing Health Effects Associated with an Unknown Number of Major Sources of Multiple Air Pollutants.* In: Development of Statistical Methods for Multipollutant Research. Research Report 183, 2015, Health Effects Institute.

Paul KC, Haan M, Mayeda ER, Ritz BR. Ambient Air Pollution, Noise, and Late-Life Cognitive Decline and Dementia Risk. *Annu Rev Public Health*. 2019;40:203–220. doi:10.1146/annurev-publhealth-040218-044058

Power MC, Adar SD, Yanosky JD, Weuve J. Exposure to air pollution as a potential contributor to cognitive function, cognitive decline, brain imaging, and dementia: A systematic review of epidemiologic research. Neurotoxicology. 2016;56:235-53. Epub 2016/06/23. doi: 10.1016/j.neuro.2016.06.004. PubMed PMID: 27328897; PMCID: PMC5048530.

Power MC, Lamichhane AP, Liao D, Xu X, Jack CR, et al. 2018. The association of long-term exposure to particulate matter air pollution with brain MRI findings: The ARIC study. *Environ. Health Perspect* 126(2):027009.

Szpiro, A.A. and Paciorek, C.J. (2013), Measurement error in two-stage analyses, with application to air pollution epidemiology. *Environmetrics* 24, 501-517.

Szpiro, A.A., Sheppard, L. and Lumley, T. (2011), Efficient measurement error correction with spatially misaligned data*. Biostatistics* 12 (4), 610-623.

Zeger SL, Thomas D, Dominici F, et al. Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environ Health Perspect*. 2000;108(5):419–426. doi:10.1289/ehp.00108419.