TITLE: Analyzing Individual-Level Secondary Data with Instrumental Variable Methods is Useful for Studying the Effects of Air Pollution on Dementia

RUNNING TITLE: Secondary Data with Instrumental Variables

Kelly C. Bishop[a], Sehba Husain-Krautter[b], Jonathan D. Ketcham[a], Nicolai V. Kuminoff[a,c], and Corbett Schimming[b,d]

[a] Department of Economics, Arizona State University
[b] Mount Sinai School of Medicine
[c] National Bureau of Economic Research
[d] James J. Peters Veterans Affairs Medical Center

Correspondence to: Nicolai V. Kuminoff: Arizona State University, Department of Economics and NBER, Main Campus, PO BOX 879801, Tempe, AZ 85287-9801, kuminoff@asu.edu.

Abstract: We hypothesize that analyzing individual-level secondary data with instrumental variable (IV) methods can advance knowledge of the long-term effects of air pollution on dementia. We discuss issues in measurement using secondary data and how IV estimation can overcome biases due to measurement error and unmeasured variables. We link air-quality data from the Environmental Protection Agency's monitors with Medicare claims data to illustrate the use of secondary data to document associations. Additionally, we describe results from a previous study that uses an IV for pollution and finds that $PM_{2.5}$'s effects on dementia are larger than non-causal associations.

Introduction

Emerging evidence implicates fine-particulate air pollution smaller than 2.5 microns in diameter ($PM_{2.5}$) as a potential cause of dementia. Although the specific mechanisms remain unknown, $PM_{2.5}$ increases risk for hypertension, hyperlipidemia, atherosclerosis, oxidative stress, insulin resistance, endothelial dysfunction, procoagulant states, inflammation, and stroke, all of which are direct or indirect risk factors for the development of cognitive decline and dementia [1]. The effects of breathing highly concentrated $PM_{2.5}$ over short-term periods on the pathogenesis of cardiovascular events are fairly well understood. These include endothelial barrier dysfunction, pro-inflammatory changes, generation of pro-thrombotic pathways, autonomic dysfunction favoring sympathetic tone, central hypothalamic-pituitary-adrenal axis activation as well as epigenetic changes [2, 3]. However similar short-term studies have limited ability to establish $PM_{2.5}$ as a causative agent in dementia due to the slow and progressive nature of this condition and its underlying pathophysiological changes.

In principle, researchers could conduct a large-scale randomized controlled trial (RCT) that randomly assigns participants at birth to wear masks across their entire lifetimes that mete out different levels and mixes of pollution and regularly implements cognitive assessments, biomarker tests, neuroimaging, and other measures of the onset and progression of dementia. While such a study would be well-designed to evaluate whether $PM_{2.5}$ is a causative agent in dementia, it is neither practical nor ethical to conduct. Instead, to test for the role of $PM_{2.5}$ in dementia, researchers must use other research designs that:

1) Have a large enough sample size to have power to detect or rule out the effects on the clinically relevant endpoints.

2) Have a long enough scope of time to reach the clinically relevant cumulative exposure levels and to observe the clinically relevant endpoints.

3) Account for the fact that people are not randomly assigned to pollution levels so that there are likely to be differences in unmeasured risk factors across pollution levels.

4) Account for the fact that researchers may have error in their measures of dementia.

5) Account for the fact that researchers will have error in their measures of the amount of $PM_{2.5}$ inhaled by each person.

In this article we propose the hypothesis that these requirements can be met by research designs that apply instrumental variables methods to large individual-level secondary data sets. Secondary data are those that are created by someone other than the researchers and exist for reasons other than answering the research question at hand. We also hypothesize that studies that rely on observational research designs, including case-control and cohort studies, are less credible because three types of confounders can make their estimates either overstate or understate any causal role of $PM_{2.5}$ on dementia. These confounders are correlation between individual pollution exposure and individual risk factors for dementia that are unmeasured in the available data, errors in the diagnosis and diagnostic coding of dementia, and errors in measuring the amount of pollution inhaled by people.

In the remainder of this article we first discuss secondary data that can be used to study the effects of long-term exposure to $PM_{2.5}$ on dementia and some of the potential biases that may result from unmeasured factors and from mis-measurement in these data. We then describe how an instrumental variables (IV) research design can be used to overcome these challenges, and the assumptions required to use IV methods for causal inference. Finally, we summarize a recent case study that demonstrates the difference in the estimated effects of long-term exposure to PM2.5 on

dementia from using an IV design versus an observational design. The IV method was first developed by [4] and IV-like methods have long been used in medical contexts to address selection into treatment when RCTs are not feasible [5]. However, the case study we describe is the first to use IV methods to evaluate the role of long-term exposure to $PM_{2.5}$ in dementia. We hypothesize that IV methods offer a promising avenue for further research on the effects of air pollution on dementia.

Discussion

Using secondary data to observe associations between ambient $PM_{2.5}$ concentrations and dementia

Recent studies have linked various sources of secondary data to analyze the associations between dementia diagnoses for large cohorts of older adults and inhaled $PM_{2.5}$ as proxied by ambient concentrations [6-8]. In the United States, federal records of health insurance claims can be used to track dementia diagnoses for all Americans over age 65 who are enrolled in "traditional" Medicare (Parts A and B). The US Centers for Medicare and Medicaid Services (CMS) Chronic Condition Warehouse file identifies people with "Alzheimer's disease or related dementia" as those with a single claim that includes one of the relevant diagnosis codes. The supplementary materials for this article provide details. Each person's individual ambient $PM_{2.5}$ concentrations can be approximated by combining information on that person's residential location history with data collected by the Environmental Protection Agency (EPA) air quality monitoring system.

To illustrate this approach, we use administrative data collected by the US Centers for Medicare and Medicaid Services (CMS) for the purposes of processing insurance claims and draw a random sample of 80-year-old traditional Medicare beneficiaries in 2013 (166,059 people). Then we approximate each person's average ambient $PM_{2.5}$ concentrations from 2004-2013 by assigning them the $PM_{2.5}$ levels estimated for their series of residential ZIP+4 centroids using spatial

interpolation of the annual average $PM_{2.5}$ concentrations recorded at EPA air quality monitors. To summarize the association between decadal $PM_{2.5}$ and the prevalence of dementia at end of 2013, we then take state-level averages of these individual data based on where people lived at the end of 2013. Figure 1 shows each state's dementia rate and 10 year average $PM_{2.5}$ concentration ($\mu g/m^3$) for men and women, separately.

Figure 1. Unconditial association between 2013 state-level dementia diagnosis rates and average ambient $PM_{2.5}$ concentrations from 2004-2013 among 80-year-old men and women
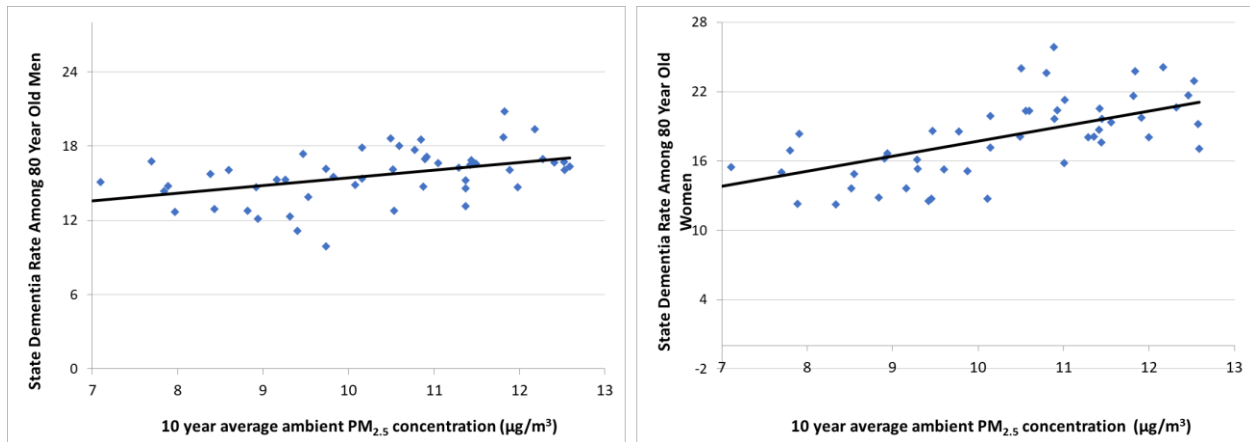

Figure 1 shows that states containing people with higher residential ambient $PM_{2.5}$ concentrations over the prior decade have higher dementia rates. The Pearson correlation coefficient for males is 0.47 ($p<0.001$) and for females it is 0.62 ($p<0.001$). Fitting regression lines to each scatterplot indicates that each additional $\mu g/m^3$ in average $PM_{2.5}$ concentration over the decade is associated with 0.6 percentage point higher rate of dementia for males and 1.3 percentage point higher rate for females. This is consistent with the hypothesis that long-term exposure to $PM_{2.5}$ is a risk factor for dementia, but these associations may overstate or understate any causal effect due to confounding by errors in measuring people's inhalation of $PM_{2.5}$, errors in measuring

dementia, and unmeasured factors that are simultaneously correlated with the observed measures of $PM_{2.5}$ and dementia.

Potential confounder: unmeasured risk factors for dementia

The first type of potential confounder is the set of risk factors for dementia that may be correlated with measures of $PM_{2.5}$ exposure but unmeasured in the available secondary data. The fact that people are not randomly assigned to their different levels of pollution will cause observational studies to overstate (understate) the true effect of $PM_{2.5}$ on dementia if people with other unmeasured risk factors for dementia have higher (lower) levels of exposure on average [9]. This makes it important to consider what can and cannot typically be observed about dementia risk factors in administrative data sets.

Administrative data often enable researchers to observe some but not all risk factors for dementia. For example, the Medicare claims data include each person's age, race, gender, prescription drug use, medical expenditures, and timing of diagnoses of numerous chronic diseases including several that are common co-morbidities with dementia such as hypertension, diabetes, ischemic heart disease, heart failure, and stroke. Further, administrative data on each person's residential location can be matched to neighborhood-level data on socioeconomic status from the Census Bureau (e.g. education, income). Unmeasured in these data are potential risk factors such as the presence or absence of the APOE4 allele and other genetic mutations that increase the likelihood of dementia, socioemotional engagement, and history of diet, exercise, smoking, and toxic exposures apart from measures of $PM_{2.5}$. The nature of the correlation between these factors and $PM_{2.5}$ measures is currently unknown.

Potential confounder: errors in measuring dementia

Secondary data on dementia are also likely to embed some errors that may confound causal interpretations. The direction of any bias depends on whether the errors in dementia diagnosis are positively or negatively correlated with $PM_{2.5}$ levels. This makes it important to consider the nature of potential errors in measuring dementia in secondary data sets.

The available secondary data differ in their size, scope, sample selection and measures of dementia and its risk factors. Small-scale studies of the prevalence of dementia have relied on diagnoses based on cognitive assessments at a single point in time in surveys such as the Health and Retirement Study [10, 11]. Individual-level longitudinal data for specialized cohorts are also available from a number of sources, such as the National Alzheimer's Coordinating Center's Uniform Data Set (NACC UDS) [12] and the Nurses' Health Study Cognitive Cohort [13]. These types of data offer a range of measures of cognition and cognitive impairment, including subclinical symptoms. However, their small size and scope limit their usefulness for investigating the causes of dementia. While the NACC UDS is larger in size and scope, it also has limited usefulness for investigating the causes of dementia because its sample is defined largely as those with cognitive impairment. See [1] for a literature review of articles that have used secondary data to investigate the link between air pollution and dementia.

Existing administrative data have potential to meet each of the criteria listed in the introduction. These include electronic medical records (EMR), particularly in relatively closed systems such as the Veteran's Administration and Kaiser Permanente. EMRs provide individuals' diagnostic test scores for tests conducted over time, as well as the presence of diagnostic codes. Insurance claims data likewise include diagnosis codes developed by the World Health Organization for mild cognitive impairment, Alzheimer's disease, and dementia currently denoted as "ICD-10".

Definitions for the relevant codes are provided in the appendix. By tracking the appearance of these ICD-10 codes for individual claims over time, researchers can observe when an individual's symptoms rise to the level of clinical attention and subsequently progress and worsen.

A primary concern for using administrative data is error in measuring dementia. Because claims are used for billing and not clinical practice, clinicians may either under-report or over-report their dementia diagnoses. Furthermore, coding and diagnostic practices may differ across contexts and geographies. For example, academic medical centers may be more systematic or rigorous than primary care physicians' office, although some research has found a high degree of consistency between diagnostic codes on Medicare claims and diagnoses by clinicians at the University of Southern California Alzheimer Disease Research Center [14]. Other research concluded that Medicare claims, compared to the Health and Retirement Study, have a sensitivity and specificity of 0.85 and 0.89 for dementia but lower for Alzheimer's disease specifically [15]. However, the relationship between any errors in diagnoses in claims data and any cause of dementia being investigated remains unknown. As a result, research designs should account for the possibility that administrative data on dementia onset may embed errors that could be correlated with measures of $PM_{2.5}$ exposure, resulting in biased estimators.

Potential confounder: errors in measuring people's inhaled air pollution

Errors in measuring the amount and types of air pollution inhaled by each person are unavoidable with current technology. Outside of controlled chamber studies, researchers cannot directly observe what people inhale, nor can they observe what people have inhaled over their lifetimes. The direction and magnitude of confounding from measurement error of $PM_{2.5}$ depends on whether it is unrelated, positively related, or negatively related to the level of $PM_{2.5}$. If the size of the measurement error in individual pollution exposure is unrelated to the measured level of

exposure, then the estimated relationship between $PM_{2.5}$ and dementia will be biased toward zero due to statistical noise, e.g. the lines in Figure 1 would be flatter than the true relationship. Similarly, the lines would be flatter if the measurement error in $PM_{2.5}$ is negatively correlated with $PM_{2.5}$ levels, while they would be steeper and hence overstate the true relationship between $PM_{2.5}$ and dementia if the error is positively correlated with $PM_{2.5}$ levels. Because of this ambiguity and the fact that the nature of this relationship cannot be observed, research designs should rely on methods to account for the presence of any such errors.

Data on ambient $PM_{2.5}$ levels exist for the US beginning in 1999 due to the EPA establishing a national network of air quality monitors. Initially, 900 monitors were placed around the US, tracking $PM_{2.5}$ and other regulated air pollutants, and this number has grown to 1,800. With additional information and assumptions, researchers can use these data to develop proxy measures for each person's inhaled $PM_{2.5}$ from 1999-present.

Several methods exist for using EPA monitor data to measure ambient pollution levels at each geographic location in the US. First, researchers can spatially interpolate $PM_{2.5}$ from the air quality monitor readings. Second, $PM_{2.5}$ concentrations can be predicted by air dispersion models that use additional data on emissions. Third, satellite images of aerosol optical depth (AOD) can be translated into $PM_{2.5}$ by calibrating AOD to data from ground level monitors [16, 17]. All three of these methods yield positive associations between $PM_{2.5}$ and mortality from circulatory system diseases, but the choice of exposure measure affects the estimated sizes of the relative risks [18].

These three methods embed tradeoffs for studying the effects of air pollution on health [19]. Using calibrated AOD data can reduce measurement error for places that are far from EPA monitors. But, because the calibration relies on the EPA monitors, these data add the assumption that the calibration method that works best at locales with EPA monitors transfers equally well to

areas far from monitors. Recent research has revealed problems with this assumption and demonstrated that the ability of AOD to predict near-surface $PM_{2.5}$ that people breathe varies across geographic areas due differences in topography, weather patterns, and pollution sources and speciation [20]. As a result, use of this method may introduce systematic but unknown measurement error and subsequent bias in the estimated relationship between $PM_{2.5}$ and dementia.

Errors in measuring individuals' inhaled pollution would persist even if researchers could perfectly observe outdoor $PM_{2.5}$ concentrations at every geographic location. This is because researchers cannot observe exactly where each person spent their time, the quality of building sealing and filtration systems, and individuals' respiration levels. Perhaps the most significant confounder in measuring exposure to air pollution is the need for longer-term air quality data than exists currently. Given the current models of pathogenesis, potential risk factors for dementia need to be studied for many years prior to the onset of symptoms and clinical syndromes. Studying only proximate air pollution data may confound assessment of risk if vulnerability or resilience vary across the lifespan. Overall, errors in measuring individuals' inhaled pollution can be expected to confound estimated associations between $PM_{2.5}$ and dementia in secondary data, but the direction and magnitude of confounding are unknown.

In summary, the inability to include every known risk factor for dementia, errors in measuring inhaled $PM_{2.5}$, and errors in measuring dementia may confound efforts to make causal inferences about $PM_{2.5}$ and dementia from secondary data. Furthermore, the net bias from these confounders on the observed estimates has uncertain direction and magnitude.

Using instrumental variables to mitigate confounding

Empirical research in many disciplines has relied on a range of designs to draw causal inferences from observational data. In addition to IVs, these include regression discontinuity,

difference-in-difference, and matching methods such as propensity scoring and synthetic controls. Each of these methods differ in the set of assumptions required to make valid causal inferences.

Drawing causal inferences from IV methods requires two key assumptions: 1) the IV must be statistically associated with the hypothesized causative agent (i.e. the explanatory variable of interest) after controlling for other factors but 2) otherwise unrelated to the outcome of interest. RCTs satisfy these assumptions by randomizing people to treatment conditions, where the treatment is the explanatory variable of interest. In fact, RCTs have been used as IVs in prior medical research (e.g. Baicker et al. 2013). To analyze the health benefits of having Medicaid coverage rather than being uninsured, researchers used the Oregon Medicaid RCT that randomly assigned people to become eligible for Medicaid as an IV to study the health effects of having Medicaid coverage relative to being uninsured. Even within this RCT an IV was valuable because some people who were randomly chosen to be eligible could still choose not to enroll for reasons unmeasured by the researchers.

When RCTs are not available researchers often search for "quasi-natural experiments" to apply to secondary data. In a quasi-natural experiment, some intervention outside of the researchers' control affects some people differently than others, where the assignment of the intervention to people is as good as random once other factors are accounted for. If this intervention affects the hypothesized causative agent but does not otherwise potentially affect the outcome of interest, then it meets the assumptions needed to use it as an IV and make causal inferences. As an example in the medical literature, researchers used IV methods to evaluate the survival benefits of more intensive treatment for acute myocardial infarction relative to less intensive treatment [5]. Their IV was defined as the differential distance that a person lived from a high-treatment-intensity hospital to a low- treatment-intensity hospital. The idea is that distance to one hospital relative to

another hospital is a strong predictor of which hospital the patient will be admitted to. As a result, patients who live relatively closer to high intensity hospitals will be more likely to be treated intensively due to the hospital's treatment styles, technology, and clinical expertise (e.g. the presence of a coronary catheterization lab or a cardiac surgery service line) rather than due only to differences in the patients' risk factors and severity of illness. The aim of the IV is to capitalize on this quasi-random assignment of patients to treatment, with the assumption that the difference in the probability of intensive treatment due to the differential distance measure is unrelated to any omitted aspects of the severity of the risk factors.

IV estimation can be viewed as a two-stage approach. In the first stage, the hypothesized causative agent is regressed on the IV and other factors that may affect the outcome of interest. The residuals (i.e. the empirical "error terms") from this regression absorb all unexplained variation in the causative agent, including variation that may be correlated with measurement errors and unmeasured variables. In the second stage the relationship between the outcome of interest and the hypothesized causative agent is estimated, conditional on the same set of controls from the first stage and the first stage residuals. Intuitively, this two-stage process uses the IV to break the hypothesized causative agent into two parts, one part that is correlated with the IV and as good as random and another part (the residual) that is correlated with potential confounders. Controlling for this residual in the second stage avoids confounding. The supplemental materials for this article contain a Monte Carlo simulation of an instrumental variables model that we designed to demonstrate how the method can overcome confounding from measurement error and unmeasured variables.

Inverse F-statistics are commonly used to test whether an IV is correlated with the hypothesized causative agent conditional on other factors, and whether that correlation is strong

enough to sufficiently mitigate statistical biases that arise from working with finite samples [21]. In contrast, there is no direct test of the second key IV assumption that the IV is uncorrelated with measurement errors and unmeasured variables. Instead, researchers often test this assumption in informal ways such as documenting how the IV generates "as good as random" variation in exposure to the hypothesized causative agent, comparing mean characteristics of groups that are exposed to different levels of the IV, and repeating estimation after replacing the outcome measure with a placebo that should be unaffected by the hypothesized causative agent [22].

Researchers routinely use IVs to measure how air pollution affects people's health. Examples of instruments for daily variation in $PM_{2.5}$ include measures of wind direction [23] and traffic [24]. In the case of longer-term variation, such as changes in average in $PM_{2.5}$ concentrations over years or decades, instruments are often constructed from public policies that create spatial and temporal discontinuities in regulatory intensity [8, 25, 26]. All of these studies find that using IVs to address confounding from measurement error and unmeasured variables *increases* estimates for $PM_{2.5}$'s negative effects on morbidity and mortality. That is, they suggest that failing to control for measurement errors and unmeasured variables tends to attenuate estimates for $PM_{2.5}$'s effect on human health in secondary data. We briefly summarize [8] analysis of $PM_{2.5}$ and dementia as a case study in the use of IVs.

Case Study: Using instrumental variables to estimate the effect of $PM_{2.5}$ on dementia

The study described in [8][henceforth BKK] uses the IV methodology to estimate how long-term $PM_{2.5}$ exposure affects dementia diagnoses in administrative data. They link Medicare records for 2.5 million Americans who were over age 65 in 2004 and not diagnosed with dementia at that point to their $PM_{2.5}$ exposure from 2004-2013. The outcome measure is an indicator for whether each person had been diagnosed with Alzheimer's Disease or related dementias by the end of the

calendar year. Each person's annual average air pollution exposure is estimated by spatial interpolation from EPA monitoring stations. Because the relevant exposure duration is unknown, BKK estimate models for two years to ten years of exposure (i.e., from 2004-2005, 2004-2006,…, 2004-2013). The models control for observable differences in individual risk factors such as each person's integer age, gender, race, gross medical expenditures in 2004, and indicators for the full set of possible combinations of past diagnoses of hypertension, diabetes, ischemic heart disease, heart failure and stroke. The models also include neighborhood-level measures of socioeconomic status including Census block group measures of income, house value, and educational attainment. Further, the models use indicators for the core-based statistical area (CBSA) in which people live to control for spatial variation in diagnostic procedures and toxic exposures apart from $PM_{2.5}$. Finally, the models control for annual average $PM_{2.5}$ concentration at each person's residential location from 2001-2003.

Despite a rich set of controls, ordinary least squares (OLS) or probit estimation of BKK's model is vulnerable to potential confounding from measurement error in $PM_{2.5}$, measurement error in dementia, and unmeasured risk factors for dementia such as genetics and exposure to $PM_{2.5}$ before they turned 65 and joined Medicare. BKK address this concern by constructing an IV for long-term $PM_{2.5}$ exposure based on the EPA's initial implementation of national air quality standards for $PM_{2.5}$. In 2004 the EPA identified 132 counties containing about 27% of the US population as exceeding the EPA's 15.05 $\mu g/m^3$ standard for the maximum allowable annual average $PM_{2.5}$ concentrations based on the county's highest monitor reading. These counties were designated as being "non-attainment" (NA) and required to take actions to reduce their pollution levels. Another 528 counties containing 43% of the US population were classified as "attainment" (A). The remaining counties that were not designated as NA or A lacked monitoring data and were

designated "unclassifiable" and not subjected to additional regulation. As in other research on the health effects of pollution, BKK exclude them from the analysis because they are likely to violate the condition that the instrument is uncorrelated with measurement error of $PM_{2.5}$. BKK show that average $PM_{2.5}$ concentrations in NA counties were reduced by the regulations by an average of $0.97\ \mu g/m^3$ across the subsequent decade.

BKK use a county's nonattainment status in 2004 to construct a binary IV for the change in pollution exposure after 2004 (NA=1, A=0). To test whether this variable satisfies the underlying IV assumptions BKK first show that the IV has a strong negative correlation with $PM_{2.5}$ exposure after 2004 conditional on the controls. The first-stage F-statistic on the instrument interacted with baseline exposure is 637, indicating that the IV is strongly correlated with air pollution exposure. Intuitively, the IV compares people who were treated differently by EPA regulations but were identical in terms of all of the measured individual risk factors, lived in the same CBSA in 2004, and lived in neighborhoods that had similar socioeconomic characteristics and $PM_{2.5}$ levels prior to regulation. Some of these neighborhoods were in newly regulated counties; others were not. People who happened to be living in the newly regulated counties experienced significantly larger reductions in $PM_{2.5}$ after regulation. This setup attempts to mimic the design of an RCT in secondary data. BKK provide indirect evidence that the IV is unlikely to be correlated with unmeasured risk factors and measurement errors by showing that trends in dementia diagnosis rates across regulated and unregulated areas were nearly identical prior to regulation. BKK also show that their IV analysis yields null effects of $PM_{2.5}$ on placebo outcomes that are not known or suspected to be caused by air pollution such as glaucoma, fibromyalgia, breast cancer, prostate cancer, and peripheral vascular disease.

BKK's research design is similar to prior studies of health effects of long-term air pollution exposure that constructed IVs from spatial and temporal discontinuities in regulatory intensity [21, 22]. Changes in regulatory intensity are commonly thought to provide the most effective IVs for long-term air pollution exposures because of the way they create unexpected differences in exposure among otherwise similar groups of people with similar baseline exposures. By contrast, candidate IVs that could be derived by summing over short-term random variation in wind direction [23] and traffic [24] could easily be correlated with unmeasured risk factors for dementia in the long run. One problem is that it is relatively easy for people to observe whether neighborhoods tend to be upwind or downwind of major pollution sources such as factories or freeways. If people who are less likely to be diagnosed with dementia are also more likely to pay to live in less polluted neighborhoods at baseline, then an IV based on long-term wind or traffic patterns would suffer from confounding. Further, such instruments would not provide a viable approach to separating the effects of the observed exposure period from the effects of prior exposure if exposure levels are correlated throughout life.

Figure 2 presents point estimates and 95% confidence intervals from BKK's models estimated with and without using county nonattainment status as an IV for average $PM_{2.5}$ concentrations across the decade from 2004-2013. The underlying standard errors are clustered at the Census block group level to account for the possibility that the model's error terms are not independent across observations within block groups. The IV estimates increase steadily with measured duration, with the confidence intervals excluding zero beginning in the eighth year (2011) and persisting through the tenth year. In contrast, the OLS estimates are smaller, their confidence intervals typically include zero, and the estimated effect does not increase with the measured duration. In both sets of models, the estimator is comparing the difference in the probability of a

dementia diagnosis due to a one µg/m^3 higher average PM$_{2.5}$ concentration but identical on all of the other observed risk factors, including the interactions among diagnoses and the baseline levels of PM$_{2.5}$, as described above. The single difference between the estimates is whether the model uses the IV to address biases due to measurement error and unmeasured risk factors. The systematically larger estimates from the IV models indicate that the net effect of these confounders is to bias the OLS estimates downward.

The estimates from this case study are best interpreted as local average treatment effects (often referred to as LATEs). That is, they provide the average treatment effects (ATEs) of decadal PM$_{2.5}$ concentrations on the probability of receiving a dementia diagnosis among the group of people whose decadal PM$_{2.5}$ concentration was affected by the assignment rule of whether or not their county was classified as being in attainment, all else equal. The 1.68 percentage-point effect estimated for this group represents an average over individuals with different decadal PM$_{2.5}$ concentrations, demographics, and baseline health conditions. Investigating how PM$_{2.5}$'s effects on dementia vary across these dimensions and extend to other subpopulations is an important area for further research.

Figure 2. OLS and IV Estimates for the Effects of PM$_{2.5}$ on Incident Dementia Diagnosis by Exposure Duration from a prior case study [8]
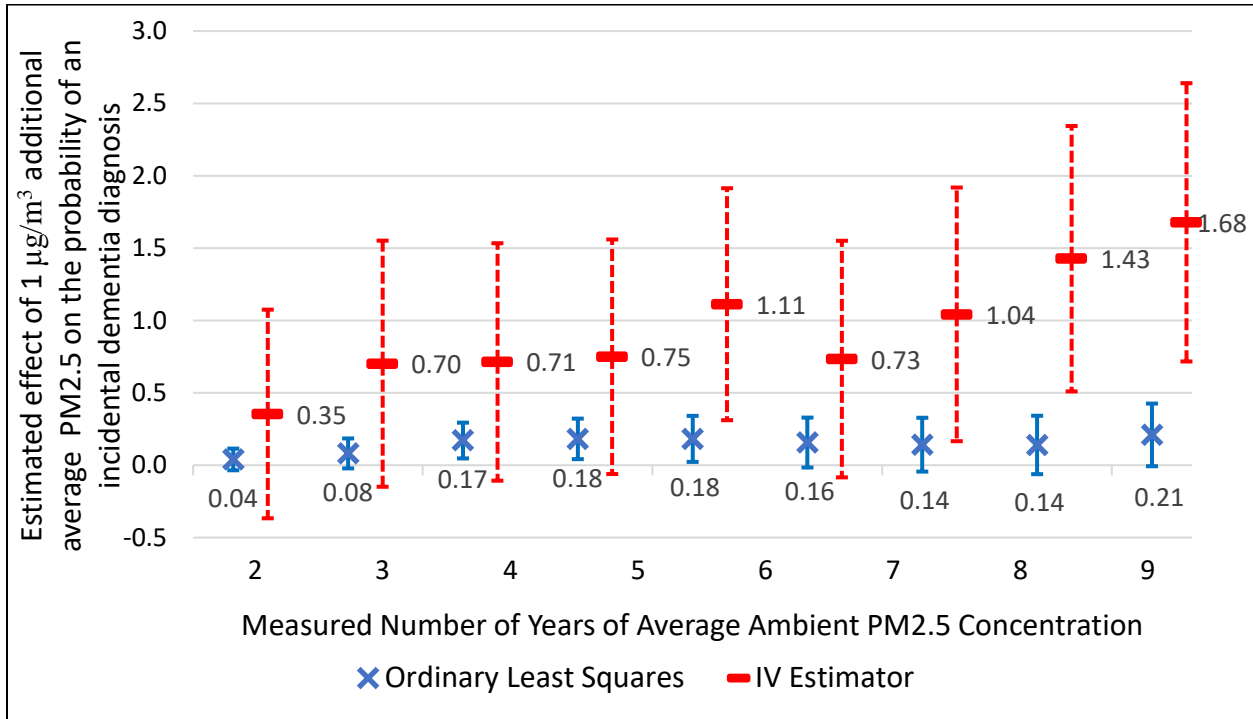
Conclusions

We propose that secondary data can be analyzed with instrumental variables methods to credibly test whether PM$_{2.5}$ acts as a causative agent in dementia. We support this hypothesis with results from a prior case study that demonstrates using an instrumental variable method to overcome biases due to mismeasured and unmeasured variables leads to substantially larger effects of PM$_{2.5}$ on the probability of an individual receiving a dementia diagnosis than estimated by observational studies. This is consistent with results from studies of the effects of PM$_{2.5}$ on other outcomes.

Even if a body of research convincingly establishes PM$_{2.5}$ as a causative agent in dementia, clinicians and medical researchers seeking to prevent, treat, or cure dementia would benefit from understanding the precise pathophysiological changes precipitated by long-term inhalation of

PM$_{2.5}$. Likewise policy makers would benefit from understanding these mechanisms so that the precise sources and types of PM$_{2.5}$ can be targeted. Further, the existing evidence indicates that the role of PM$_{2.5}$ is probabilistic rather than deterministic, raising questions about the dose-response function and the importance of various potential moderating and mediating factors such as obesity, diabetes mellitus and hyperlipidemia. Such insights may be gained by research that applies causal inference methods to secondary data on ambient PM$_{2.5}$ concentrations linked with longitudinal individual-level data that includes risk factors and neuroimaging biomarkers and post-mortem neuropathological measures, including increased production of Aβ peptides and activation of microglia.

References

[1]   Peters R, Ee N, Peters J, Booth A, Mudway I, Anstey KJ. Air Pollution and Dementia: A Systematic Review. Journal of Alzheimer's disease : JAD, 2019;70:S145-S63.

[2]   Rajagopalan S, Al-Kindi SG, Brook RD. Air Pollution and Cardiovascular Disease: JACC State-of-the-Art Review. Journal of the American College of Cardiology, 2018;72:2054-70.

[3]   Miller MR, Raftis JB, Langrish JP, McLean SG, Samutrtai P, Connell SP *et al.* Inhaled Nanoparticles Accumulate at Sites of Vascular Disease. ACS nano, 2017;11:4542-52.

[4]   Wright PG. The tariff on animal and vegetable oils. The Macmillan company, 1928.

[5]   McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. Jama, 1994;272:859-66.

[6]   Chen H, Kwong JC, Copes R, Tu K, Villeneuve PJ, van Donkelaar A *et al.* Living near major roads and the incidence of dementia, Parkinson's disease, and multiple sclerosis: a population-based cohort study. Lancet, 2017;389:718-26.

[7]   Carey IM, Anderson HR, Atkinson RW, Beevers SD, Cook DG, Strachan DP *et al.* Are noise and air pollution related to the incidence of dementia? A cohort study in London, England. BMJ open, 2018;8:e022404.

[8]   Bishop KC, Ketcham JD, NV. K. Hazed and Confused: The Effect of Air Pollution on Dementia., NBER Working Paper; 2019.

[9]   Dominici F, Greenstone M, Sunstein CR. Science and regulation. Particulate matter matters. Science, 2014;344:257-9.

[10]   Hudomiet P, Hurd MD, Rohwedder S. Dementia Prevalence in the United States in 2000 and 2012: Estimates Based on a Nationally Representative Study. The journals of gerontology Series B, Psychological sciences and social sciences, 2018;73:S10-S9.

[11]   Fishman E. Risk of Developing Dementia at Older Ages in the United States. Demography, 2017;54:1897-919.

[12]   Spackman DE, Kadiyala S, Neumann PJ, Veenstra DL, Sullivan SD. Measuring Alzheimer disease progression with transition probabilities: estimates from NACC-UDS. Current Alzheimer research, 2012;9:1050-8.

[13]   Weuve J, Puett RC, Schwartz J, Yanosky JD, Laden F, Grodstein F. Exposure to particulate air pollution and cognitive decline in older women. Archives of internal medicine, 2012;172:219-27.

[14]   Lee E, Gatz M, Tseng C, Schneider LS, Pawluczyk S, Wu AH *et al.* Evaluation of Medicare Claims Data as a Tool to Identify Dementia. Journal of Alzheimer's disease : JAD, 2019;67:769-78.

[15]   Taylor DH, Jr., Ostbye T, Langa KM, Weir D, Plassman BL. The accuracy of Medicare claims as an epidemiological tool: the case of dementia revisited. Journal of Alzheimer's disease : JAD, 2009;17:807-15.

[16]   Di Q, Kloog I, Koutrakis P, Lyapustin A, Wang Y, Schwartz J. Assessing PM2.5 Exposures with High Spatiotemporal Resolution across the Continental United States. Environmental science & technology, 2016;50:4712-21.

[17]   van Donkelaar A, Martin RV, Brauer M, Hsu NC, Kahn RA, Levy RC *et al.* Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors. Environmental science & technology, 2016;50:3762-72.

[18]   Jerrett M, Turner MC, Beckerman BS, Pope CA, van Donkelaar A, Martin RV *et al.* Comparing the Health Effects of Ambient Particulate Matter Estimated Using Ground-Based versus Remote Sensing Exposure Estimates. Environmental health perspectives, 2017;125:552-9.

[19]   Fowlie M, Rubin E, Walker R. Bringing Satellite-Based Air Quality Estimates Down to Earth. AEA Papers and Proceedings, 2019;109**:**283-8.

[20]   Jin QC, P.; Pryor, S. C. Spatial characteristics and temporal evolution of the relationship between PM2.5 and aerosol optical depth over the eastern USA during 2003–2017. Atmospheric Environment, 2020;239.

[21]   Staiger DS, J. Instrumental Variables Regression with Weak Instruments. Econometrica, 1997;65**:**557-86.

[22]   Angrist JDPJ-K. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, 2008.

[23]   Deryugina T, Heutel G, Miller NH, Molitor D, Reif J. The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction. The American economic review, 2019;109**:**4178-219.

[24]   Schlenker W, WR. W. Airports, Air Pollution, and Contemporaneous Health: Review of Economic Studies. Review of Economic Studies; 2016, p. 768-809.

[25]   Ebenstein A, Fan M, Greenstone M, He G, Zhou M. New evidence on the impact of sustained exposure to air pollution on life expectancy from China's Huai River Policy. Proceedings of the National Academy of Sciences of the United States of America, 2017;114**:**10384-9.

[26]   Deschênes O, Greenstone M, Shapiro JS. Defensive Investments and the Demand for Air Quality: Evidence from the NOx Budget Program. American Economic Review, 2017;107**:**2958-89.