NBER WORKING PAPER SERIES

DOES EMPLOYMENT SLOW COGNITIVE DECLINE? EVIDENCE FROM LABOR MARKET SHOCKS

Noah Arman Kouchekinia
David Neumark
Tim A. Bruckner

Working Paper 35117
http://www.nber.org/papers/w35117

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2026

We are grateful to the UCI Office of Research Seed Grant program and the Center for Population, Inequality, and Policy at UCI for project development research support. All views expressed are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2026 by Noah Arman Kouchekinia, David Neumark, and Tim A. Bruckner. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Does Employment Slow Cognitive Decline? Evidence from Labor Market Shocks
Noah Arman Kouchekinia, David Neumark, and Tim A. Bruckner
NBER Working Paper No. 35117
April 2026
JEL No. I1, J14

ABSTRACT

With large gains in life expectancy, the population share of disability due to cognitive decline and dementia has substantially increased. Many older adults in the United States leave the workforce well before age 65. Correlational evidence suggests that leaving the workforce before retirement age could accelerate the pace of cognitive decline. We offer causal evidence, using HRS data for the United States, exploiting plausibly exogenous shifts in labor demand in local labor markets as a Bartik instrument for employment variation across these markets. We find substantial declines over time in cognitive scores stemming from negative labor demand shocks. These findings are concentrated among men aged 51 to 64, whose employment decisions and outcomes may be more sensitive to local labor market conditions than are these decisions or outcomes for women or for older men. Our evidence extends past work focusing narrowly on the retirement age window and provides further support to the notion that working to older ages may delay age-related cognitive decline.

Noah Arman Kouchekinia
University of California, Irvine
nkouchek@uci.edu

Tim A. Bruckner
University of California, Irvine
tim.bruckner@uci.edu

David Neumark
University of California, Irvine
Department of Economics
and NBER
dneumark@uci.edu

Introduction

Employment near retirement age appears to reduce the risk of cognitive decline, which is a key precursor to dementia (Coe et al., 2012; Bonsang and Perelman, 2012; Celidoni et al., 2017; Mazzonna and Peracchi, 2017).[1] This potential link between work and forestalling cognitive decline and dementia points to large social benefits from higher employment at older ages; dementia affects an estimated 6 million Americans, and is the leading cause of neurological disability and the third leading cause of overall disability among older adults (>65 years).

Our goal in this paper is to extend research on the causal impact of employment on cognitive decline to employment before the retirement window. The link between employment and cognitive decline at pre-retirement ages is potentially important, because the correspondence of steep cognitive decline with the onset of dementia also arises at younger ages (Deary et al. 2009).[2] However, we know less about whether employment at mid-life (when cognitive decline already starts) delays cognitive decline, aside from research suggesting that enriched opportunities in mid-life – including those associated with work – may reduce risk of dementia in older age by producing resilient cognitive and/or neurobiological architectures (Tucker-Drob and Salthouse, 2011; Kobayashi and Feldman, 2019).

Descriptive work finds that working longer is associated with reduced cognitive decline (e.g., Dufouil et al., 2014; Leist et al., 2013; Lupton et al., 2009; Roberts et al., 2011), but this work does not establish a causal impact of employment in reducing cognitive decline. Unmeasured factors related to mental health may drive decisions about the timing of leaving paid work, or lead to sub-par work performance that can precede leaving a job.[3] Hence, researchers

[1] Dementia is a general term for a group of symptoms that include difficulties with memory, language, problem-solving and other thinking skills.
[2] For more details, see GBD (2016).
[3] The existing work of which we are aware that is closest to this hypothesis studies anxiety and depression as

have used quasi-random variation in employment to isolate the causal effect of employment. Most of these empirical studies with arguably more rigorous causal study designs focus on policy changes to ages at which retirement benefits or Social Security are received, or other exogenous changes affecting employment around the retirement age – including discontinuities created by retirement and pension policies (e.g., Atalay et al., 2019; Bonsang et al., 2012; Celidoni et al., 2017; Coe et al., 2012). This work generally shows (albeit with a few exceptions) that extending employment near retirement ages delays the onset of cognitive decline, consistent with the hypothesis that stimulating cognitive and/or social activities in the workplace may delay cognitive decline (Salthouse, 1991).

Because the studies cited above rely on plausibly exogenous retirement policy changes, they focus on a narrow age range (the mid to late 60s). Thus, it remains an open question whether employment at earlier pre-retirement ages delays cognitive decline. Whereas severe cognitive impairment (and dementia onset) typically occurs after age 65, pathologically meaningful cognitive decline is detectable as early as age 40 (Salthouse, 2009; Singh-Manoux et al., 2012; Yang et al., 2024), and over 32% of adult men aged 51-64 years in the United States are not working.[4] These factors motivate our core question: What is the causal impact of employment on cognitive decline in the years before usual retirement ages – with a particular focus – dictated in part by our data – on ages 51-64?

Our research exploits a different source of exogenous variation in employment than does research using retirement reforms or discontinuities. We use 12 waves of the Health and Retirement Study (HRS) dataset spanning 1996-2018. The HRS has a large number of older

antecedents of non-employment (among those initially employed), rather than cognitive shocks/declines (Dooley et al., 2000; Butterworth et al., 2012; Steele et al., 2013).
[4] Calculated from CPS microdata (Flood et al., 2025).

adults, rich information on employment history and trajectories of cognitive decline, and broad geographic coverage. Our core identification strategy uses County Business Patterns data to construct a Bartik-style shift-share instrumental variable (Bartik, 1991). In particular, we use cross-commuting zone variation in industry composition and national industry-specific growth rates to predict changes in local labor demand. We then use those predicted changes in local labor demand as an instrument for changes in HRS respondent employment rates by commuting zone. This research design is, of course, applicable to ages other than those near retirement.

We believe our work has significant policy implications. In the face of population aging, policymakers are already focused on promoting employment beyond normal retirement ages, motivated by many factors including – in the United States – the solvency of Social Security. But given the high social costs of dementia and cognitive decline, evidence of an employment-cognition link at pre-retirement ages could motivate efforts to increase employment at these ages by, for example, encouraging work among Social Security Disability Insurance (SSDI) recipients (Maestas, 2019), or even, perhaps, broader hiring subsidies.[5] The potential benefits of higher employment at pre-retirement ages could extend beyond improved cognitive functioning, and also include prolonged work, less reliance on SSDI, and enhanced retirement security via later claiming of Social Security benefits and greater accumulation of savings from working longer. In addition, of course, the evidence that exogenous increases in employment around retirement ages helps forestall cognitive decline points to an additional benefit from policies that push back retirement (which are generally being adopted for different reasons – to improve the finances of public pension systems). However, the ability of policy to boost employment at younger, pre-

[5] Because the beneficial effects of employment on cognition may be unknown to individuals, and because a large share of the costs of dementia are likely borne by others, there may be an important externality that could justify such hiring subsidies.

retirement ages, compared to boosting employment at ages very close to retirement, is likely much stronger.

Related Research

Cognitive decline and employment

Small to mild cognitive decline occurs gradually with aging and may not induce disability. Moderate to severe forms of cognitive decline, however, correspond with onset of dementia and result in substantial disability and interruption with daily activities and functioning (Coe et al., 2012). With the large gains in life expectancy in the United States and other high-income countries over the last half-century, the share of the population with disability due to cognitive decline and dementia has substantially increased (Steinmetz et al., 2022). This circumstance has compelled federal research and public health agencies to prioritize the development of interventions to maintain high levels of cognition.[6] A recent Alzheimer's Association Report concludes that, in addition to high levels of cardiovascular functioning and a healthy diet, lifelong learning/cognitive training may reduce the risk of cognitive decline (Baumgart et al., 2015). The notion that cognitive stimulation throughout the life course may promote "cognitive reserve" and stave off neurodegenerative disability is supported by longitudinal associations between employment (as well as job complexity, which we do not explore) and reduced cognitive decline (Kröger et al., 2008; Fisher et al., 2017; Smart et al., 2014; Hanushek et al., 2025).

Numerous descriptive studies show that working longer is associated with reduced cognitive decline. Dufouil et al. (2014) study data on French self-employed workers who were alive and retired as of the end of 2010, merging pension and health databases. They identified

[6] See https://www.nia.nih.gov/report-2020-2021-scientific-advances-prevention-treatment-and-care-dementia/lifestyle-interventions and Anderson et al. (2011).

cases of Alzheimer's or other dementia and found that those who retired later were less likely to have dementia. Leist et al. (2013) use the Survey of Health, Ageing, and Retirement in Europe (SHARE) data to measure cognitive impairment of those aged 50-73 in the mid-2000s, as well as complete employment histories from ages 25-65. Their evidence indicates that employment gaps of more than six months (described as unemployment or sickness by respondents) are associated with higher risk of cognitive impairment.[7] Using data from the United Kingdom's Medical Research Council genetic resource for late onset Alzheimer's, Lupton et al. (2009) document that for males, later retirement is associated with later age at onset.[8] In addition, Roberts et al. (2011) examine data on a prospective study of civil servants in London, which collected a number of cognitive measures (memory, verbal fluency, and vocabulary scale). While scores increased on average over five-year windows measured from ages 38-60 to ages 42-67, the increases were smaller for those who retired.

However, observational studies of employment and cognitive decline or its absence cannot rule out the alternative causal mechanism whereby cognitive decline causes workforce exits (losing a job, choosing not to work, etc.), or the possibility that unmeasured changes drive both cognitive decline and employment decline. For example, in studies that observe a higher incidence of cognitive decline associated with earlier retirement (Dufouil et al., 2014; Leist et al., 2013; Lupton et al., 2009), cognitive decline itself could precede and lead to retirement. And evidence of a correlation based on a longer time span between employment and cognitive decline (Lupton et al., 2009) could reflect earlier causal effects on employment of precursors of the

[7] In contrast – and in relation to the hypothesis that learning can boost cognition, which is not our focus – gaps associated with training are associated with lower risk of cognitive impairment.
[8] They do not find a relationship with education or highest degree/qualification. They also suggest no relationship for "employment," but their variables measure work-related characteristics (foreman, manager, number of people at work, etc.), not employment per se.

measured cognitive decline (or confounding factors that drive both outcomes). For example, cross-sectional studies show correlations between mental disorder (e.g., depression) and cognitive impairment (Vinkers et al., 2004; Migliorelli et al., 1995), and more compelling longitudinal evidence indicates that the onset of depression precedes unemployment and precarious employment (Dooley et al., 2000; Butterworth et al., 2012). These circumstances could induce a positive association between employment and cognition even if there is no causal effect of employment on cognitive decline.

Causal evidence

To address the limitations of correlational evidence, a number of studies have implemented research designs that try to isolate exogenous variation in employment at older ages to estimate causal effects of employment on cognitive decline. Most of these studies use retirement-related policy changes to obtain exogenous changes in employment (Atalay et al., 2019; Bonsang et al., 2012; Celidoni et al., 2017; Coe et al., 2012; Mazzonna and Peracchi, 2012; Rohwedder and Willis, 2010; Rose, 2020; Picchio and van Ours, 2019), and typically find evidence of delayed cognitive decline from extending work around retirement ages (Salthouse, 1991).

We discuss three examples. Atalay et al. (2019) study a number of cognition measures in the Household, Income and Labour Dynamics in Australia dataset. They use an instrumental variables (IV) estimate, instrumenting for retirement with a policy reform that affected retirement incentives. For men, they find a negative effect of retirement duration on word reading (interpreted as "crystallized ability") but not on working memory or speed of information processing. They find no significant impacts for women. The study by Bonsang et al. (2012) has many parallels, estimating the effect of retirement on episodic memory in the HRS, using the age

of eligibility for Social Security as an IV (which creates spikes in retirement at age 62 – the age at which one can first draw benefits – and 65 – when "full" benefits are available). They find a significant negative effect of retirement on their cognitive functioning measure. Mazzonna and Peracchi (2012) use a similar strategy with the SHARE data, studying effects of retirement on orientation in time, memory, verbal fluency, and numeracy. OLS estimates indicate a significant negative relationship between years since retirement (zero for the non-retired) and each cognitive domain, for both men and women. Using an IV for the early and normal eligibility ages for retirement, they tend to also find significant negative effects of retirement on cognition. Their IV estimates for women tend to be larger (negative) than the OLS estimates.[9]

We have identified one paper that uses a research design not focused on retirement ages and retirement policies, although it still focuses on ages near retirement. In particular, Mosca and Wright (2018) study Ireland, using as an IV the "marriage bar" that required women to leave paid employment upon marriage, and was in effect from the 1930s and only abolished in the 1970s. They use data from the Irish Longitudinal Study of Aging, including information on retirement and tests of processing speed and mental switching. Their results indicate a small negative effect of retirement duration on cognitive functioning in OLS estimates. In the IV estimates – where the abolition of the marriage bar serves as the IV – the estimates are larger in absolute value although not statistically significant, but there is no statistical evidence that retirement is endogenous (i.e., no reverse causal impact of cognitive functioning on retirement). Mosca and Wright's study parallels ours in that the exogenous variation in when people worked occurs earlier than traditional retirement ages.[10]

[9] Some papers that use a similar approach reach more mixed or heterogeneous results (Celidoni et al. 2017; Coe et al., 2012).

[10] In their study, the marriage bar led to long spells of non-employment at much younger ages. As the authors put it, "Many of these women spent a significant proportion of their lives after getting married in retirement" (p. 1320),

Data

We use the HRS, which consists of a nationally representative, longitudinal survey of approximately 40,000 older adults. Respondents are interviewed biennially, with 16 waves of interviews completed to date. The original HRS cohort was interviewed in 1992. We use the 1996-2018 waves because from 1996 onward a consistent battery of cognitive questions assessing recall, mental processing, and working memory was asked of participants of any age.[11] The HRS target population aims to represent the non-institutionalized U.S. population aged 51 or older but also interviews spouses or partners of respondents. The survey records employment history as respondents enter the study and documents changes in employment status (and other work information) in each wave. We obtained restricted-use data on county of residence and industry of employment (RDA #2023-095). The conditions of using these data impose restrictions on the geographic granularity at which we can report statistics.

Upon joining the HRS, each respondent (both those dwelling in the community, and nursing home residents beginning in 2000) receives a detailed survey of several cognitive domains.[12] Consistent with prior research, we aggregate scores from surveyed cognitive tasks (a four-item sum of Immediate Word Recall, Delayed Recall, Serial 7s, and Backwards Counting index), to derive a "global cognitive status" score (referred to as the "Langa-Weir" score).[13] This score has been evaluated against a subsample of HRS participants who received a detailed

although this is not "retirement" in the conventional sense.

[11] We forego using 2020 and 2022 data because of the potential influence of COVID on the results.

[12] There is a potential argument for using data from 2000 on, given that consistent cognitive screening began for nursing home residents only in in 2000 (Farina et al., 2020). We present results indicating that our results are robust to using this subperiod. This is the same sub-period for which we check robustness to opioid epidemic exposure. Table 3, column (3), Panel B reports our core analysis restricted to the 2000-2018 period.

[13] The HRS collects some additional measures, but we do not use them because they are not routinely collected from participants under 65, and they were not assessed as part of the Langa-Weir score in Crimmins et al. (2011); additionally, some of the items have a potential direct connection to employment independent of cognitive ability (e.g., naming the day of the week).

neuropsychiatric assessment (Crimmins et al., 2011), and can be used to categorize participants as likely cognitively impaired (7-11) or likely demented (1-6). In our broad sample, very few respondents fall into each of these categories. We instead rely on variation in the continuous score, much of remains in the cognitively normal (12-27) range. We use as the key dependent variable the within-person difference in global cognitive scores across waves.

The HRS data contain individual-level responses about contemporaneous employment and years since last worked. From these self-reported data, we construct employment measures aggregated to the commuting zone (CZ) and year level. We aggregate (in almost all of our analyses) for reasons discussed below – but the key reason is because this is the level at which our Bartik IV is constructed. Economists view CZs as an appropriate measure of a local labor market, as CZs capture geographic areas defined by a large share of workers who both reside and work within the CZ.[14] Thus, the Bartik IV at the CZ level characterizes changes in job opportunities facing residents of a CZ. Counties are grouped into CZs using Journey to Work data by the U.S. Department of Agriculture . We use the USDA's 1990 vintage of CZ groupings throughout our analysis.

We use data from County Business Patterns (CBP) to construct our IV used in our core empirical strategy.[15] The CBP data are based on the Census Business Register, a database of all known establishments derived from firm survey data and administrative records. The data sources include the Economic Censuses and annual data collected by the Census Bureau, as well as administrative records from the Internal Revenue Service, the Social Security Administration, and the Bureau of Labor Statistics, providing accurate employment information. The CPB reports employment counts for each county by detailed industry. Our primary analysis uses 83

[14] See the discussion in Jha et al. (forthcoming).
[15] See U.S. Census Bureau, County Business Patterns, https://www.census.gov/programs-surveys/cbp.html.

major industries based on the 1987 SIC to characterize a CZ's initial industry shares. Although later HRS waves switch to NAICS classifications, we are able to map these codes to the 1987 SIC major industries. At this level of aggregation, suppression of small county-by-industry cells does not occur. As a robustness exercise in light of recent concerns raised about Bartik IVs (Borusyak et al., 2022) we use finer industry categorizations, up to 6-digit industries in the NAICS coding scheme. With detailed industry categories, small county-by-industry cells are suppressed due to confidentiality concerns. This suppression is overcome using Eckert et al.'s (2020) imputed values for suppressed cells.

Finally, as a potentially important confounding variable – which we discuss in the next section – we measure local exposure to the opioid crisis using the CDC's underlying cause of death statistics. We measure the share of deaths attributable to drug overdose using the CDC's standard definition of drug overdose mortality (Jalal et al., 2018).[16] Mortality statistics are reported at the county-year level, but small counties are suppressed for confidentiality and data quality reasons. We calculate CZ-level level drug overdose rates in two different ways to overcome this suppression. First, we calculate rates simply excluding small counties, as long as those counties represent the minority of a CZ's population. CZs with most of their population residing in suppressed counties are simply excluded. Second, we assign to suppressed counties their state's drug overdose rate net of unsuppressed counties before aggregating up to the CZ level.

Drug mortality data are available in a consistent form starting in 2000, so we lack overdose rates for the earliest period of our analysis. To control for opioid epidemic exposure without sacrificing earlier waves of data, we use an alternative approach in which we assign CZs

[16] Drug overdose deaths are identified using ICD–10 underlying cause-of-death codes: X40–X44, X60– X64, X85, and Y10–Y1. See cdc.gov/nchs/nvss/vsrr/drug-ovrdose-data.htm.

to exposure intensity categories based on drug overdoses from 2000 onwards. Using a K-means clustering algorithm, we find four exposure intensity groups, whose drug overdose rates over time are displayed in Figure 1, Panel A, and which are mapped in Panel B. Over our time period, exposure increased approximately linearly within each group. Thus, in this alternative approach using all the years of data, we account for variation in opioid epidemic exposure by controlling for separate linear trends by exposure group in our models.[17]

Empirical Approach

We begin by estimating an individual-level regression:

$$\Delta C_{i,l,t}^{HRS} = \alpha + \beta \Delta E_{i,l,t}^{HRS} + \Delta X_{i,l,t}^{HRS} \gamma + \Delta AGE_{i,l,t}^{HRS} \delta + \varepsilon_{i,l,t}. \quad (1)$$

In this equation, C is the global cognitive score, E is a dummy variable for employed,[18] X is a vector of demographic dummies (for gender, race/ethnicity, and education), $AGE_{i,l,t}^{HRS}$ is a vector of 3-year age bin dummies, and $\varepsilon_{i,l,t}$ is the error term. The *HRS* superscript indicates the data are computed from HRS respondents. The i, l, and t subscripts denote individual, location (CZ), and year. Fixed observable and unobservable characteristics of individuals are differenced out by the first-difference design (denoted by Δ). We also do not include fixed effects for CZs or year. CZ dummy variables, at least to the extent people do not move, drop out of the first differences as long as there are not CZ-specific trends. And time effects are inappropriate in our context: aggregate economic shocks common to CZs provide valuable identifying information, and there is no reason to expect aggregate time effects in cognition. The demographic controls and age dummies allow for variation in the trajectories of cognitive scores across different

[17] Given that we use a model in first differences, as explained in the next section, we include dummy variables for the exposure groups, which correspond to linear trends in a model in levels.
[18] This indicator is derived from the question: "Are you currently working for pay?" In some waves a response of working but not for pay is allowed. This is treated as not employed.

demographic groups and ages.

As discussed above, this naive approach is subject to bias from reverse causality or common unobservables that affect both cognition and employment. Bias can also arise from measurement error in the employment variable.[19] Aggregating across local areas can potentially mitigate some of the biases, to the extent that the variation that creates the bias is idiosyncratic at the individual level but averages out across individuals in local areas. We use the same first-difference approach, augmented by controls for changes in respondent composition. This aggregated regression model is:

$$\Delta \bar{C}_{l,t}^{HRS} = \alpha + \beta \Delta \overline{EP}_{l,t}^{HRS} + \Delta \bar{X}_{l,t}^{HRS} \gamma + \Delta \overline{AGE}_{l,t}^{HRS} + \varepsilon_{l,t}. \quad (2)$$

In this equation, the bars over the variables represent CZ-year averages of the same variables defined above. $\overline{EP}$ is now the employment-to-population ratio, and $\overline{AGE}$ is a vector of dummies for binned average age of CZ residents. The vector of demographic controls $\overline{X}$ now represents the shares of HRS participants by demographic group (all of which are dummy variables).

To see how aggregation might help reduce bias, first suppose that, in addition to equation (1) – our core regression at the individual level – there is a second equation that captures the endogeneity of employment (*E)* in equation (1):

$$\Delta E_{i,l,t}^{HRS} = \alpha' + \beta' \Delta C_{i,l,t}^{HRS} + \Delta X_{i,l,t}^{HRS} \gamma' + \Delta AGE_{i,l,t}^{HRS} \delta + \varepsilon'_{i,l,t}. \quad (3)$$

Suppose the error terms in equations (1) and (3) are i.i.d. across individuals – i.e., they have no common element at the *l* level. However, suppose they both reflect a common individual component, i.e.:

[19] This has been pointed out by other researchers, including, for example, Atalay et al. (2019), Bonsang et al. (2012), and Mosca and Wright (2018).

$$\varepsilon_{i,l,t} = \tau_{i,l} + \nu_{i,l,t} \quad (4)$$

$$\varepsilon'_{i,l,t} = \tau_{i,l} + \nu'_{i,l,t} \quad (5)$$

In this case, $\varepsilon_{i,l,t}$ is correlated with $\Delta E_{i,l,t}^{HRS}$ in equation (1), creating endogeneity bias. If $\beta' > 0$ (e.g., an adverse cognitive shock reduces employment), then $Corr(\varepsilon_{i,l,t}, \Delta E_{i,l,t}^{HRS}) > 0$, and the OLS estimate of β is upward biased. If $\beta' < 0$ (e.g., an adverse cognitive shock reduces wealth and increases employment), the bias is in the opposite direction. We can also get the latter result from measurement error in $\Delta E_{i,l,t}^{HRS}$. Suppose $\Delta E_{i,l,t}$ is true employment, and $\mu_{i,l,t}$ is classical measurement error ($E\left(\mu_{i,l,t}\middle|\Delta E_{i,l,t}\right) = 0$), and i.i.d. across individuals, so that

$$\Delta E_{i,l,t}^{HRS} = \Delta E_{i,l,t} + \mu_{i,l,t} \quad (6)$$

Then our equation (1) becomes

$$\Delta C_{i,l,t}^{HRS} = \alpha + \beta \Delta E_{i,l,t}^{HRS} + \Delta X_{i,l,t}^{HRS} \gamma + \varepsilon_{i,l,t} - \mu_{i,l,t} \quad (7)$$

Because $Corr(\Delta E_{i,l,t}^{HRS}, -\mu_{i,l,t}) < 0$, the OLS estimate of β is biased downward.

Suppose we now aggregate across observations in a CZ and year. Then the error term in equation (1) in the endogeneity case becomes the average of $\tau_{i,l} + \nu_{i,l,t}$ across each *i* within *l*, so the problematic part of the error term, for each observation on *l* and *t*, is $\sum_i \nu_{i,l,t}/N_{l,t}$, where $N_{l,t}$ is the number of observations for the CZ and year. Similarly, in the case of measurement error, the problematic part of the error term is for each observation on *l* and *t* is $\frac{\sum_i \mu_{i,l,t}}{N_{l,t}}$. The law of large numbers implies that the probability limit of these sums equals 0, given that they are i.i.d. within each CZ and year, so the bias from these sources is eliminated (in the limit), implying consistent estimates. Of course, if these components of the error have a group-specific structure or are correlated within CZ-year pairs, then the aggregation does not lead to consistent estimates.[20]

[20] Two of the earliest sources on aggregation bias are Theil (1954) and Griliches (1957). For a discussion of similar

That said, there is a potential problem of aggregation bias from estimating regression relationships on averages, one source of which is omitting individual-level variation that drives the behavioral parameter of interest (sometimes called the "ecological fallacy"), and another of which is heterogeneity of true coefficients across individuals. In our case, however, it seems the fundamental issues posited in this literature are the effects of cognitive shocks that themselves influence employment, and measurement error, so we believe the aggregated estimates are likely to be more reliable.

Still, we believe there remains a need to use an IV for employment at the CZ level because of omitted community factors that affect both cognition and employment, which are not fully accounted for in our controls. If the sources of bias are not only at the individual level but at the community level, so that averaging respondent cognition and employment within CZ simply aggregates up the individual biases, then instrumenting is still necessary. In addition, the possibility of measurement error in employment remains, in part because we still have sampling error at the CZ level, and in part because the employment measure we use (and average) may not perfectly measure all the dimensions of the experience of employment that potentially affect cognition (in which case there is common measurement error at the CZ level).

There is substantial geographic variation in CZ level employment that supports the utility of local area shocks as an instrument. Panels A-D of Figure 2 show substantial geographic variation in employment change by commuting zone (CZ), as well as large reductions in employment in the 51-64 age range in many areas of the southeast (which appears to be driven entirely by men, in Panel C). Of course, the variation documented in Figure 2 could be caused by local variation in cognitive changes, or associated with common factors driving both,

references to and uses of aggregation in other research contexts, see Appendix B.

necessitating use of an IV for local employment change.

Our approach is to instrument for local variation in employment with a Bartik-style shift-share instrument (Bartik, 1991). This instrument predicts local labor market conditions using local industry composition at "baseline" and national industry growth, isolating variation in local employment due to national trends in labor market demand that have different impacts depending on the local industry mix. This Bartik IV uses employment in the CZ in the base year to construct the industry shares of local employment in that CZ and year. National trends in employment by the same industries are then applied to the base-year industry weights for each CZ. So, for example, if a CZ has a heavy representation of automotive industry employment, and the automotive industry has declined nationally, such a CZ will have a more negative (or less positive) employment shock, *ceteris paribus*, than a CZ with a lower baseline auto industry share. We aim to predict employment-to-population ratios, so predicted employment is then scaled by baseline period population.

In particular, let $E_{l,j,b}^{CBP}$ denote industry employment (E) by local labor market (l) and industry (j), at the baseline year (b), as measured in the CBP. Denote national employment as lacking labor market subscripts (e.g. $E_{j,b}^{CBP}$). Then the predicted one-period change in the employment-to-population ratio in l at time t is:

$$\widehat{\Delta EP_{lt}} = \frac{1}{Pop_{lb}} \sum_i \frac{E_{j,t}^{CBP} - E_{j,t-1}^{CBP}}{E_{j,B}^{CBP}} E_{j,l,B}^{CBP} \quad (8)$$

$\widehat{\Delta EP_{lt}}$ is the Bartik IV. It is a valid IV under two conditions. First, it must predict changes in local employment-to-population ratios, which we can test directly by examining the strength of the first stage. Second, it must be excludable from the regression above in that it should have no independent effect on $\Delta \bar{C}_{l,t}^{HRS}$ conditional on $\Delta \overline{EP}_{l,t}^{HRS}$ and the other controls in equation (2).

Throughout our main results, we provide evidence related to the first condition –

predicting changes in local employment-to-population ratios, or "relevance." First, we present first-stage F-statistics, which measure the strength of the relationship between the instrument and the employment change measure. A conventional threshold for a strong instrument is an F-statistic of at least 10. Second, we present weak-instrument robust Anderson-Rubin (AR) tests of significance and confidence intervals, which are robust to instruments with low relevance. The test statistics can be inverted to produce weak-IV robust confidence intervals. These confidence intervals are calculated numerically using a grid-search algorithm. This can result in unbounded intervals when identification is weak, and in cases with a very weak IV, it can even result in a disjoint confidence interval.[21] In practice, disjoint AR confidence intervals only arise for subsamples in which our instrument performs very poorly, and these are cases where we do not attempt to interpret the estimated coefficients.

Whereas the exclusion restriction seems defensible on a priori grounds, recent work has sharpened our understanding of this identifying assumption and how to assess it further when using a Bartik shift-share IV (Borusyak et al., 2022, 2025; Goldsmith-Pinkham et al., 2020). We adopt analyses suggested by this recent work in our empirical analysis below. In our view, these analyses generally bolster our confidence in the IV and hence the credibility of our findings.

Nonetheless, we also focus on one specific factor for which the exclusion restriction could be violated owing to correlation with predicted employment growth in an area (based on the baseline industry composition and national industry employment growth), and an independent effect on cognition (ΔC) separate from the direct effect on employment for the CZ

[21] This occurs when the sign of the first-stage relationship between the instrument and employment is uncertain, but the reduced-form relationship between the instrument and cognition is significant. Since the IV is the reduced-form estimate divided by the first-stage estimate, if the sign of the first stage is unclear we may reject a region around zero but not two symmetric negative and positive regions. The unboundedness could potentially be resolved with an expanded grid search, but regardless, the confidence intervals would be very large and hence finding the bound(s) would not affect the conclusions. (We use the *weakiv* ado file in Stata.)

for which $\Delta\bar{C}_{l,t}^{HRS}$ is measured. In particular, we consider opioid use and abuse. It is conceivable that opioid use is correlated with the job composition of an area – for example, because of injury rates,[22] and its use and abuse could also spill over to other persons in the local area. Given that opioid use and abuse could influence both employment, and cognition net of its effect on employment, in local areas, the Bartik IV estimate could be biased if we do not control for it. Hence, we test the robustness of our results to the presence of controlling for local patterns in opioid use and abuse.

Results

Descriptive statistics

Among the HRS participants in our sample, cognitive performance gradually decreases with age, although there is significant variation. In Figure 3, we display the trajectory of the global (Langa-Weir) cognitive score. Panel A displays the average score and standard deviations at each age. The mean score decreases by about 3 points (or 17%) from age in the mid-50s to the mid-70s, and the standard deviation is consistently around 4.[23] The average decline between survey waves for an individual respondent is on average 0.18 points, with a standard deviation of the change of about 3.5, implying significant within-person variation in the cognition measure.[24] Across demographic groups, cognitive performance varies in levels but maintains a roughly parallel trajectory. Figure 3, Panel B displays cognitive trajectories separately for men and women. A similar pattern holds by educational attainment and race (see Figure A2).

[22] For example, Asfaw and Boden (2020) document the strong relationship between workplace injuries and opioid-related morbidity (abuse, dependence, adverse effects, or poisoning). Lundstrom (2025) and Metcalf and Wang (2020) document links between opioids, industry, and occupation.

[23] Figure 3 uses all of the data, not a balanced panel of the same people at all ages, because the age composition of the sample changes a good deal over time.

[24] Figure A1 displays the distribution of these scores by gender and age group, along with the Langa-Weir classification of cognitive function, which divides the score range into dementia probable (1-6), cognitive impairment but not dementia probable (7-11), and cognitively normal (12-27) (Crimmins et al. 2011).

There are marked differences in both level and trajectory by economic geography, displayed in Figure 3, Panel C. Respondents in the CZs in the lowest quartile of employment growth start their early 50s with scores at the same level as the other three quartiles. However, a gap emerges over time, particularly as respondents enter potential retirement years. Relative to the top quartile, there is a full one-point gap for participants in their 70s between the bottom and top quartiles of CZ employment growth. This suggests a role for work – related to local variation in employment – in slowing cognitive decline.

The variation in employment is significant across space, as we already discussed with regard to Figure 2, and this is especially true for men at pre-retirement ages. Figure 2 maps changes in the employment-to-population ratio for older adults from shortly before the HRS commenced (1990) to just before the COVID-19 pandemic (2019). Employment-to-population ratios among older adults (aged 51-75) and adults before retirement (aged 51-64) – without regard to gender (Panel A) – increased near universally but somewhat unevenly, with the lowest gains in rural parts of East North Central and East South Central census divisions. This masks significant heterogeneity by gender. Employment-to-population ratios for pre-retirement aged men (51-64) declined in much of the country during this period, while women saw gains in every CZ. As noted earlier, the last panel of Figure 2 displays the shift-share predicted change in the employment-to-population ratio by CZ. Note that there is some correspondence with the geography of realized changes in employment, particularly for pre-retirement men; this will be more apparent when we discuss the first-stage estimates.

Core results

Our core OLS and IV results are reported in Table 1. Here we use a large age range: 51-74. Column (1) shows the change in cognitive score as a function of changes in an individual's

employment, conditional on the demographic and three-year age bin controls.[25] The estimated effect of individual employment changes is positive but statistically insignificant. Given that the standard deviation of the cognitive score across individuals is 4.16, the estimated coefficient implies that becoming non-employed would reduce the cognitive score by 0.0013 of a standard deviation, which seems very small (even if it were significant). However, as discussed above, this estimate could be biased in either direction relative to the true effect.

In column (2) we aggregate to the CZ level and, again, estimate by OLS. The estimated magnitude is positive, much larger, and statistically significant. The positive sign is consistent with employment slowing cognitive decline (or conversely leaving employment leading to cognitive decline). With the employment rate measured on a 0-1 scale, the magnitude implies that a 10-percentage point decrease in the share employed decreases the cognitive score by 0.0579, or 0.014 of a standard deviation. Thus, in this specification, although the implied effect is roughly 10 times larger than what we find in column (1), it is still arguably quite small. As discussed above, this aggregated estimate could still be biased by reverse causality, common influences on employment and cognition (such as opioid use) at the CZ rather than the individual level, or CZ-level measurement error, and hence may not be causal.

In column (3) we instrument for the change in a CZ's employment-to-population ratio with the predicted change in the employment-to-population ratio based on the Bartik instrument. The estimated effect is significantly larger than the OLS estimate. The magnitude implies that for a 10-percentage point decrease in the employment-to-population ratio among 51-75 year olds, the cognitive score declines by 0.19 standard deviations. This is a sizable but not unreasonably large estimate, given that a 10-percentage point change in the employment-to-population ratios is

[25] Descriptive statistics at the individual level are reported in Table A1. The table lists the demographic controls and other information.

a large change.

The change in the estimated effect to a more positive estimate of employment on cognition, or equivalently a larger effect of employment decline on cognitive decline, runs counter to what seems like the most intuitive direction of bias – reverse causality whereby adverse cognitive changes reduce employment, which would imply positive rather than negative bias.[26] In fact, this is not surprising in light of other research literature (focused on retirement ages). For example, in studying the effects of retirement, Atalay et al. (2019) find a larger impact in their IV estimate on word reading for men than in their first-difference estimate. The same is true in Bonsang et al. (2012), and in Mazzonna and Peracchi (2012) for women. Bonsang et al. suggest that measurement error in the timing of retirement could cause attenuation bias towards zero, which is ameliorated by the IV estimation. We find this plausible, as it is not clear, a priori, exactly what dimension of employment change, including its persistence, is most relevant to its impact on cognition. Another possibility that would lead to bias in the same direction – which we have not seen discussed in the literature – is that adverse cognitive changes lower income and wealth, with the income/wealth effects leading to higher labor supply rather than lower labor supply, implying a bias against finding a positive impact in regressions of cognitive changes on employment. Finally, Bonsang et al. also suggest the stronger IV finding could arise from heterogeneous treatment effects, which could be larger for those whose retirement is induced by the IV (per the Local Average Treatment Effect interpretation; Imbens and Angrist, 1994). This could arise if those whose behavior is changed by the IV make sharper reductions in hours than those who transition to retirement more slowly and are less affected by Social Security retirement incentives.

[26] The estimate becomes more positive from aggregation (column (2)), and more so from instrumenting (column (3)).

One might also be concerned that the IV estimate is unreliable because of a weak IV. However, as reported in the lower panel of Table 1, the IV is relatively strong. The F-statistic from the first stage is near 10. We also report the Anderson-Rubin weak instrument robust test of significance and confidence interval. We continue to reject the null hypothesis of no effect, and the confidence interval excludes negative (or small positive) effects.

Table 2 reports alternative versions of the IV estimates in column (3) for variation in the age range, and disaggregates estimates by gender. After repeating the last estimates from Table 1, columns (2) and (3) split the data by age: 51-64 and 65-75. In the first-stage estimates, as well as in the results from the Anderson-Rubin test and confidence intervals, the IV works well for the younger group but fails for the older group. This is perhaps not surprising, because at pre-retirement ages local economic conditions have much more influence on employment than at post-retirement ages (where retirement incentives and norms may predominate). We thus put store only in the estimates for 51-64 year-olds. For them, we find a larger impact (9.877 vs. 7.749, or about 27% larger).

Next, we focus on the younger age range, but disaggregate by gender (columns (4) and (5)). Paralleling the results for the younger vs. older individuals, we find that the IV works well for men aged 51-64, but not for women – again, based on the first-stage estimate and the other diagnostics. For women, the first-stage has almost no power and the IV estimate is implausible (but also essentially invalid). Again, this might not be surprising, if for women in their 50s and 60s, in particular, employment variation is driven more by other factors than by local employment shocks, perhaps because these women are more likely to work in sectors not affected by economic shocks, such as teaching, health care, and the public sector. Moreover, we show below that the variation in the IV is largely driven by blue-collar industries, which are

more important predictors of employment for men than for women. For men alone, the estimated impact is smaller (4.44) which implies that a 10-percentage point decline in the employment rate lowers cognitive scores by 0.11 standard deviations.

Accounting for opioid exposure

We next turn to an assessment of whether the results are robust to controlling for exposure to the opioid epidemic, which could have affected both local employment and local cognition. If exposure to the opioid epidemic is correlated with a CZ's baseline (1990) industrial composition, we may have a violation of the exclusion restriction required by the IV.[27]

In Table 3, we control for exposure to the opioid epidemic based on the data on drug overdose deaths. Across columns (1)-(3), we take different approaches to deal with suppressed data in small counties. In column (1), if a CZ contains a county with suppressed data, it is omitted from the sample. This eliminates the vast majority of CZs (as is apparent from the sample sizes in the last row of the table) and hence does not lead to evidence we regard as very reliable – reflected also in the IV diagnostics. In column (2), we only omit CZs if counties with missing data contain more than 50% of the CZ's population. In column (3), missing counties are assigned their state's opioid death rate net of unsuppressed counties before being aggregated to the commuting zone level. Even in this case, the sample is about 15% smaller than our base case (3,595 observations), because CDC death rate data are not available before 2000, precluding using the 1996 and 1998 HRS survey waves. In each of these columns, Panel A shows the IV estimate with the opioid exposure control, while Panel B shows the estimate without the control, for the same sample.

To use the full sample period back to 1996, but still control for opioid exposure, we

[27] This framing of the validity of the IV in terms of a correlation between baseline industry shares and omitted variables that affect the outcome aligns with Goldsmith-Pinkham et al.'s (2020) interpretation of the Bartik IV.

include in the model indicators for three of the four groups capturing variation in exposure to the opioid epidemic during the years for which we have mortality data. This amounts to adding linear trends in cognition by opioid exposure category. (As we noted earlier, Figure 1 indicates that drug death rate by year for each of these groups increases close to linearly, so controlling for opioid exposure in this way seems reasonable, including the imputation to the two earlier waves not covered by the drug death data.)

Overall, in Table 3, for all but the most conservative approach (which yields a small sample), we find an effect of employment on cognition that survives the inclusion of opioid epidemic exposure controls, and does not change appreciably when adding the control. Most important, there is not clear evidence that the causal effect weakens when this control is included, which would be the expectation if higher opioid exposure lowered both employment and cognition at the local labor market level, and hence was an important confounder generating a positive correlation between changes in the two variables.

Assessing the Bartik IV

Finally, we turn our attention to evidence on the validity of the Bartik shift-share IV. As explained earlier, this IV is the cross-product of industry-level growth rates and local baseline industry shares. A good instrumental variable must be relevant – which we have already addressed in the preceding analyses. It must also satisfy the exclusion restriction, which in our setting means that the Bartik shift-share must affect cognition only through employment.

With regard to the exclusion restriction, recent econometrics research has suggested that shift-share IVs, while very widely used, are somewhat of a "black box" with regard to the underlying source of exogenous variation that provides identification, and as a result it can be a bit difficult to think about the validity of the identifying assumption. This research offers two

different interpretations of the "experiment" provided by the shift-share IV. These two interpretations help clarify the assumptions that are required for the exclusion restriction to hold, and point to some useful diagnostics for assessing the validity of the IV.

In Borusyak et al.'s (2022) interpretation, the shift-share IV captures exogenous shocks to employment in many industries, which are allocated differentially to exposed areas according to the baseline industry shares, without making any assumption about the exogeneity of these baseline shares. The exclusion restriction then relies on the nationwide changes in employment by industry only affecting cognition through employment. Borusyak et al. rely on a law of large numbers argument with regard to industry shocks. In other words, exogeneity of the common shocks is more plausible when there is a large number of industries, because if the shocks are random, when there is only a small number of them there is more likely to be a correlation between industry shocks and unobservables just by chance.

We use this idea to provide some evidence on the validity of the exclusion restriction by assessing the robustness of the results to using finer levels of industry disaggregation. Per Borusyak et al.'s argument, we might interpret robustness of the findings to the finer industry classification as providing assurance about the validity of the instrument, because with a finer classification, the law of large numbers is more likely to hold and random variation in national industry growth rates is less likely to be correlated with unobservables.

For our main analysis, we use 83 major industries from the 1987 Standard Industry Classification (SIC). However, crosswalks to map between detailed SIC codes and detailed NAICS codes are imprecise because the detailed codes often do not map one to one. As a consequence, we disaggregate more finely only for the years in which NAICS codes are used in the HRS. Thus, disaggregating more finely by industry entails a cost – excluding the earliest

observations from the data. In Table 4 we rerun our core results with increasingly fine industry classifications as defined in the 2012 North American Industry Classification System (NAICS). We find that the estimated impact of employment changes on cognitive changes is very robust to the level of industry disaggregation. We believe this evidence bolsters the exclusion restriction in Borusyak et al.'s framing of the Bartik IV. Moreover, on a priori grounds, we believe that the assumption that nationwide changes in employment by industry only affect cognition through employment is reasonable.

Goldsmith-Pinkham et al. (2020) provide a different interpretation of the shift-share instrument, as a weighted aggregation of IV estimates using the shift in employment in each industry separately. In their interpretation of the shift-share IV, shares provide the exogenous variation, and shifts weight the shares' importance. The exclusion restriction is satisfied if industry shares in the initial period affect the outcome of interest only through the potentially endogenous variable (employment change, in our case). In this case, we require that the 1990 industry composition of a CZ affects changes in cognition only through changes in employment – which to us also seems a reasonable assumption.

Nonetheless, we also explore Goldsmith-Pinkham et al.'s (2020) recommendation for evaluating this assumption. They suggest computing Rotemberg weights, which estimate the relative importance of specific industries in the shift-share IV, and can provide useful insight into identification. For example, we could find that a particular industry plays a very large role in identifying the effect using the Bartik IV, and we could then think about reasons why local variation in the baseline share of that industry might or might not affect local cognitive changes net of local employment variation.

We display a summary of the Rotemberg weights for our instrument in Table 5.

Goldsmith-Pinkham et al.'s result that the shift-share instrument is a weighted aggregation of IV estimates using the shift in employment in each industry separately implies that, in the presence of heterogeneous treatment effects, a shift-share estimate can be interpreted as a weighted average of heterogeneous effects. With positive weights, this average will be a convex combination of the heterogeneous effects, while with negative weights it is not easily interpretable. Panel A of Table 5 shows that most weights are positive, and Panel C shows that weights are consistently positive over the whole sample period; both results indicate that negative weights are not problematic. Panel C, however, shows that identification is disproportionately driven by the 2002-2004 period. We do not have any particular concern about some years being more important than others for the validity of the exclusion restriction, however.[28]

Most important, perhaps, Panel D displays the industries receiving the most weight, and the most negative weight.[29] Identification is driven relatively strongly by equipment manufacturing and construction. However, no one industry appears to receive an excessive share of the weight, reassuring us that our identification strategy is based on broad-based changes to industrial composition that make exogeneity of the IV more plausible (although even a large share on an industry would not necessarily imply that the IV is invalid).[30]

Conclusions

With large gains in life expectancy in the United States (and other high-income countries) over the last half-century, the population share of disability due to cognitive decline and

[28] One reason for the large influence of the 2002-2004 data is that the Early Baby Boomers were added to the HRS in 2004 (at ages 56-61). Our earlier results indicated that the Bartik IV has its strongest influence in approximately this age range.
[29] Panel B shows the correlations between the industry weights, industry growth rates, and the estimates obtained using one industry for the IV.
[30] The distribution of Rotemberg weights by industry is presented in Figure A3. Aside from the top five industries listed, about a dozen others have large positive weights.

dementia has substantially increased (GBD, 2016). Employment near retirement age appears to reduce the risk of cognitive decline, which can in turn forestall the onset of dementia. However, many older adults in the United States leave the workforce well before age 65, and it is an open question whether leaving the workforce before retirement age also accelerates the pace of cognitive decline. Our contribution is to provide causal evidence on the effects of work – or alternatively, leaving employment – on cognitive decline in years preceding retirement. We use HRS, focusing on pre-retirement ages above 50, and leverage exogenous shifts in labor demand in local labor markets as a source of exogenous changes in employment.

Our main IV specification, using a Bartik shift-share instrument for employment changes, finds substantial declines over time in cognitive scores following exposure to negative labor market shocks. We also find an effect of the same sign, although much smaller in magnitude, using data that aggregates from the individual to the local labor market level as a strategy to reduce biases in individual-level OLS estimates. Our IV findings of beneficial effects of employment on forestalling cognitive decline are concentrated among men aged 51 to 64, which coheres with the notion that employment in this subgroup (compared to women, or persons of retirement age) would respond most strongly to local area labor market shocks. Our evidence extends past work focusing narrowly on the retirement age window (some of which also uses rigorous research designs), and provides further support to the notion that working to older ages may delay age-related cognitive decline.

Limitations of our work include the inability to determine which aspects of work, or of employment loss, causally affect cognition. We suspect that the relation between work and cognition at pre-retirement ages is heterogeneous in ways that depend on both the nature of work and the characteristics of the people studied. Nevertheless, our group-average findings, if

replicated by others, would hold clear policy implications. Federal efforts to promote work at pre-retirement ages would not only reduce reliance on SSDI and enhance retirement security, but would also promote healthy aging through delaying cognitive decline.

References

Anderson, L. A., Day, K. L., Vandenberg, A. E. 2011. Using a concept map as a tool for strategic planning: The Healthy Brain Initiative. *Preventing Chronic Disease* 8, A117.

Asfaw, A., Boden, L.I. 2020. Impact of workplace injury on opioid dependence, abuse, illicit use and overdose: A 36-month retrospective study of insurance claims. *Occupational and Environmental Medicine* 77, 648-653.

Atalay, K., Barrett, G. F., Staneva, A. 2019. The effect of retirement on elderly cognitive functioning. *Journal of Health Economics* 66, 37-53.

Bartik, T. J. 1991. *Who Benefits from State and Local Economic Development Policies?* (Kalamazoo, MI: W.E. Upjohn Institute).

Baumgart, M. et al. 2015. Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective. *Alzheimer's & Dementia Journal* 11, 718-726.

Bonsang, E., Adam, S., Perelman, S. 2012. Does retirement affect cognitive functioning? *Journal of Health Economics* 31, 490-501.

Borusyak, K., Hull, P., Jaravel, X. 2025. A practical guide to shift-share instruments. *Journal of Economic Perspectives* 39, 181-204.

Borusyak, K., Hull, P., Jaravel, X. 2022. Quasi-experimental shift-share research designs. *Review of Economic Studies* 89, 181-213.

Butterworth P., Leach L. S., Pirkis J., Kelaher, M.. 2012. Poor mental health influences risk and duration of unemployment: A prospective study. *Social psychiatry and psychiatric epidemiology* 47, 1013-1021.

Celidoni, M., Dal Bianco, C., Weber, G. 2017. Retirement and cognitive decline. A longitudinal analysis using SHARE data. *Journal of Health Economics* 56, 113-125.

Coe, N. B., von Gaudecker, H.-M., Lindeboom, M., Maurer, J. 2012. The effect of retirement on cognitive functioning. *Health Economics* 21, 913-927.

Crimmins, E. M., Kim, J. K., Langa, K. M., Weir, D. R. 2011. Assessment of cognition using surveys and neuropsychological assessment: The Health and Retirement Study and the Aging, Demographics, and Memory Study. *Journals of Gerontology, Series B, Psychological Sciences and Social Sciences* 66, i162-i171.

Deary, I. J. et al. 2009. Age-associated cognitive decline. *British Medical Bulletin* 92, 135-152.

Dooley D., Prause, J., Ham-Rowbottom K. A. 2000. Underemployment and depression: Longitudinal relationships. *Journal of Health and Social Behavior* 41, 421-436.

Dufouil, C. 2014. Older age at retirement is associated with decreased risk of dementia. *European Journal of Epidemiology* 29, 353-361.

Eckert, F., Fort, T. C., Schott, P. K., Yang, N. J. 2020. Imputing missing values in the US Census Bureau's county business patterns. NBER Working Paper 26632.

Farina, M. P., Hayward, M. D., Kim, J. K., Crimmins, E. M. 2020. Racial and educational

disparities in dementia and dementia-free life expectancy. *Journals of Gerontology, Series B, Psychological Sciences and Social Sciences* 75, e105-e112.

Fisher, G. G., Chaffee, D.S., Tetrick, L. E., Davalos, D. B., Potter, G. G. 2017. Cognitive functioning, aging, and work: A review and recommendations for research and practice. *Journal of Occupational Health Psychology* 22, 314-336.

Flood, S. et al. 2025. IPUMS CPS: Version 13.0 [dataset]. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D030.V13.0.

GBD. 2016. Neurology collaborators. Global, regional, and national burden of neurological disorders, 1990-2016: A systematic analysis for the Global Burden of Disease Study. *Lancet Neurology* 18, 459-480.

Goldsmith-Pinkham, P., Sorkin, I., Swift, H. 2020. Bartik instruments: What, when, why, and how. *American Economic Review* 110, 2586-2624.

Griliches, Z. 1957. Specification bias in estimates of production functions. *Journal of Farm Economics* 39, 8-20.

Hanushek, E. A., Kinne, L., Witthöft, F., Woessmann, L. 2025. Age and cognitive skills: Use it or lose it. *Science Advances* 11(10).

Hanushek, E. A., Rivkin, S. W., and Taylor, L. R. 1996. Aggregation and the estimated effects of school resources. *Review of Economics and Statistics* 78, 611-627.

Imbens, G. W., Angrist, J. D. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62, 467-475.

Jalal H. et al. 2018. Changing dynamics of the drug overdose epidemic in the United States from 1979 through 2016. *Science*. Sept. 21; 361(6408).

Jha, P., Neumark, D., Rodriguez-Lopez, A. Forthcoming. What's across the border? Re-evaluating the cross-border evidence on minimum wage effects. *Journal of Political Economy: Microeconomics*.

Kobayashi, L. C., Feldman, J. M. 2019. Employment trajectories in midlife and cognitive performance in later life: Longitudinal study of older American men and women. *Journal of Epidemiology and Community Health* 73, 232-238.

Kramer, G. H. 1983. The ecological fallacy revisited: Aggregate- versus individual-level findings on economics and elections, and sociotropic voting. *American Political Science Review* 77, 92-111.

Kröger, E. et al. 2008. Is complexity of work associated with risk of dementia? The Canadian Study of Health and Aging. *American Journal of Epidemiology* 167, 820-830.

Leist, A. K., Glymour, M. M., Mackenbach, J. P., van Lenthe, F. J, Avendano, M. 2013. Time away from work predicts later cognitive function: Differences by activity during leave. *Annals of Epidemiology* 23, 455-462.

Lundstrom, E.W. 2025. Synthetic opioid and stimulant co-involved overdose deaths by occupation and industry—United States, 2022. MMWR. Morbidity and Mortality Weekly Report, 74.

Lupton *et al.* 2009. Education, occupation and retirement age effects on the age of onset of

Alzheimer's disease. *International Journal of Geriatric Psychiatry* 25, 30-36.

Maestas, N. 2019. Identifying work capacity and promoting work: A strategy for modernizing the SSDI program. *Annals of the American Academy of Political and Social Science* 686, 93-120.

Mazzonna, F., Peracchi, F. 2017. Unhealthy retirement? *Journal of Human Resources* 52, 128-151.

Mazzonna, F., Peracchi, F. 2012. Ageing, cognitive abilities and retirement. *European Economic Review* 56, 691-710.

Metcalf, G.F., Wang, Q. 2020. Abandoned by coal, swallowed by opioids. *Milken Institute Review*, https://www.milkenreview.org/articles/abandoned-by-coal-swallowed-by-opioids.

Migliorelli R. et al. 1995. Prevalence and correlates of dysthymia and major depression among patients with Alzheimer's disease. *American Journal of Psychiatry* 152, 37-44.

Mosca, I., Wright, R. E. 2018. Effect of retirement on cognition: Evidence from the Irish marriage bar. *Demography* 55, 1317-1341.

Picchio, M., van Ours, J. 2019. The mental health effects of retirement. SSRN Scholarly Paper, https://doi.org/10.2139/ssrn.3490031.

Roberts, B.A., Fuhrer, R., Marmot, M., Richards, M. 2011. Does retirement influence cognitive performance? The Whitehall II study. *Journal of Epidemiology and Community Health* 65, 958-963.

Rohwedder, S., Willis, R. J. 2010. Mental retirement. *Journal of Economic Perspectives* 24, 119-138.

Rose, L. 2020. Retirement and health: Evidence from England. *Journal of Health Economics* 73(102352)

Salthouse, T. A. 2009. When does age-related cognitive decline begin? *Neurobiology of Aging* 30, 507-514.

Salthouse, T. A. 1991. *Theoretical perspectives on cognitive aging* (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.).

Singh-Manoux A. et al. 2012. Timing of onset of cognitive decline: Results from Whitehall II prospective cohort study. *BMJ*. Jan. 5;344.

Smart, E. L., Gow, A. J., Deary, I. J. 2014. Occupational complexity and lifetime cognitive abilities. *Neurology* 83, 2285-2291.

Steele, F., French, R., Bartley, M. 2013. Adjusting for selection bias in longitudinal analyses using simultaneous equations modeling: The relationship between employment transitions and mental health. *Epidemiology* 24, 703-11.

Steinmetz, N.E. et al. 2022. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: An analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health* 7: e105-25.

Theil, H. 1954. *Linear Aggregation of Economic Relations* (Amsterdam, Netherlands: North-Holland).

Tucker-Drob, E. M., Salthouse, T. A. 2011. Individual differences in cognitive aging. In

Chamorro-Premuzic, von Stumm, and Furnham, eds. *The Wiley-Blackwell Handbook of Individual Differences* (Hoboken, NJ: Wiley Blackwell), 242-267

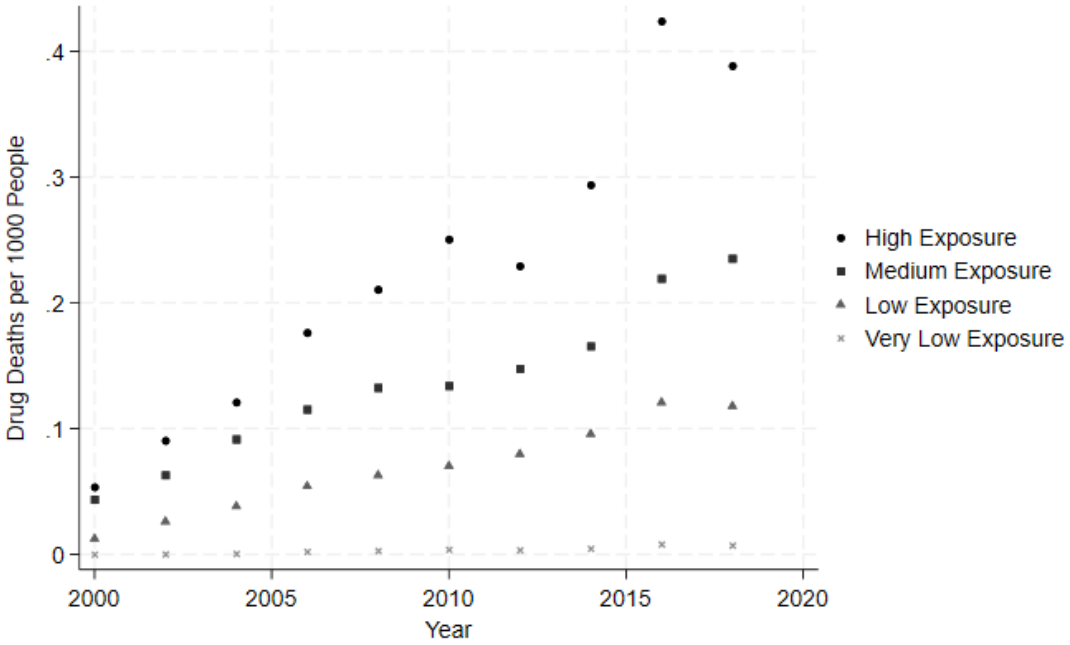
U.S. Department of Agriculture, Economic Research Service. 2012. 1980 and 1990 commuting zones and labor market areas. https://www.ers.usda.gov/data-products/commuting-zones-and-labor-market-areas.

Vinkers, D. J. et al. 2004. Temporal relation between depression and cognitive impairment in old age: Prospective population based study. *BMJ.* Oct 14;329.

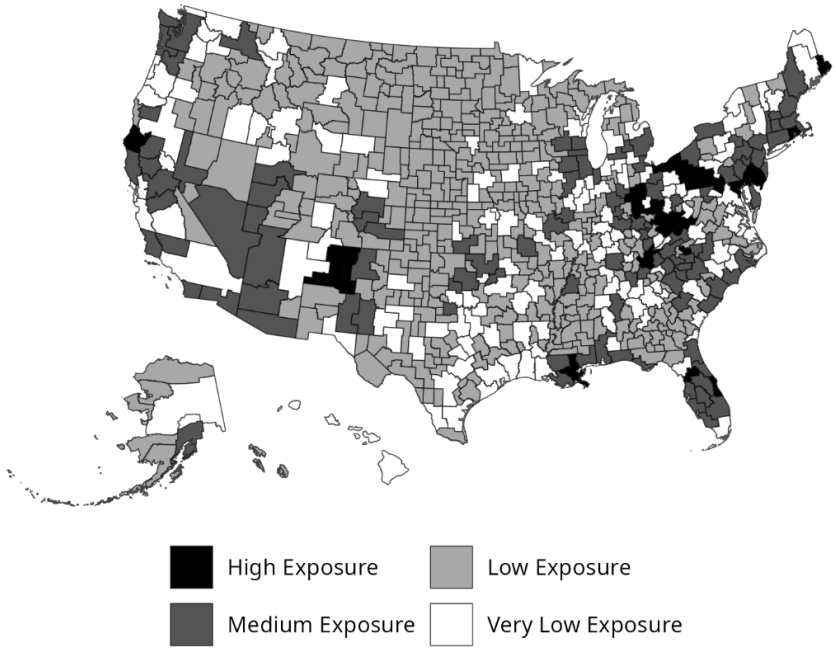
Yang, Y. C. et al. 2024. An early and unequal decline: Life course trajectories of cognitive aging in the United States. *Journal of Aging and Health* 36(3-4), 230-45.

Figure 1: Opioid Deaths by CZ Type

Panel A: Changes in Drug Deaths by Exposure Intensity Group

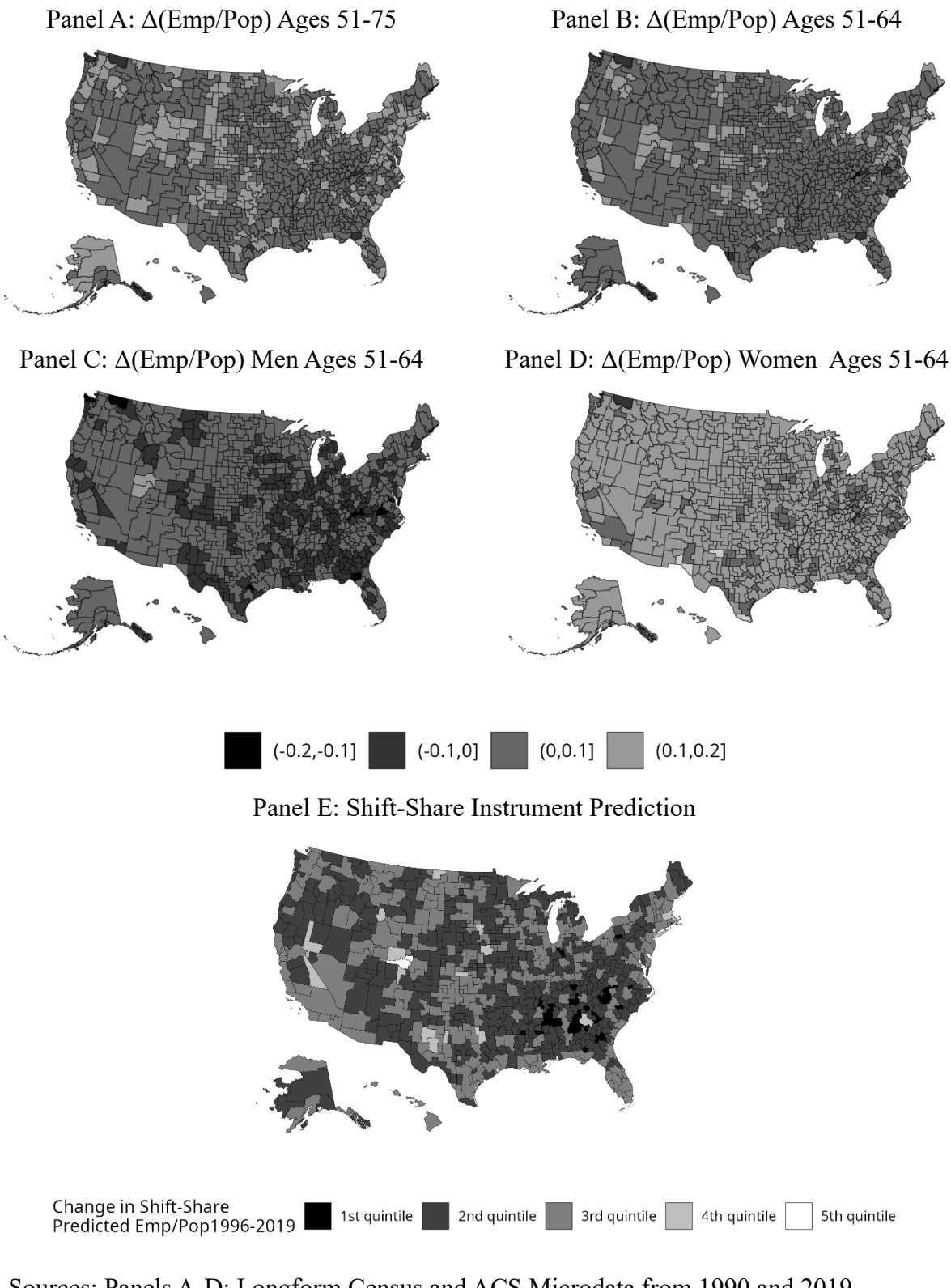


Panel B: Map of Exposure Intensity Groups



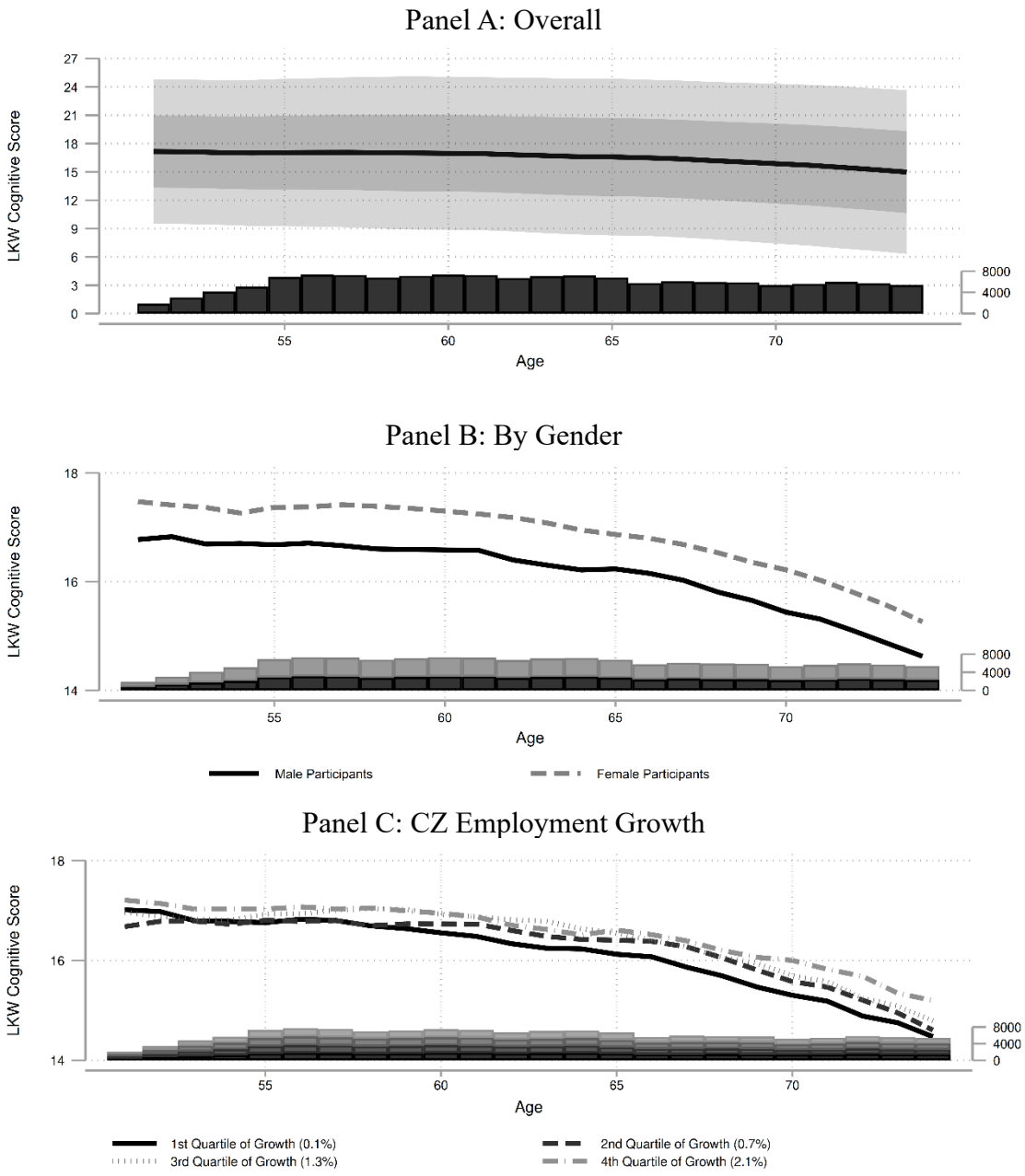
Notes: Data sourced from the Centers for Disease Control and Prevention's National Center for Health Statistic: National Vital Statistics System, Mortality 1999-2020 on CDC WONDER Online Database, released in 2021. Data are from the Multiple Cause of Death Files, 1999-2020. Drug overdose deaths are identified using ICD–10 underlying cause-of-death codes: X40–X44, X60– X64, X85, and Y10–Y1. See cdc.gov/nchs/nvss/vsrr/drug-ovrdose-data.htm. CZ types derived K-means clustering algorithm applied to each CZs overdose rate history.

Figure 2: Δ(Employment/Population) by Commuting Zone



Sources: Panels A-D: Longform Census and ACS Microdata from 1990 and 2019 respectively. Panel E: County Business Patterns and 1990 Full Count Census. Note that shift-shares for some Alaskan commuting zones were not calculated due to data suppression.

Figure 3: Cognitive Trajectories (Global Cognitive Score)



Notes: 3-year rolling average applied to cognitive score series. Dark and light shaded regions in Panel A display one and two standard deviations from the mean at each age. Bars at bottom of each panel show sample size by age and category. Sample size bar heights correspond with the axis labels on the right.

Table 1 : Employment and Cognition, Full Sample, Ages 51-74

	(1) Individual OLS	(2) CZ OLS	(3) IV
		Δ Cognitive Score	
ΔEmployment	0.0523		
	(0.0424)		
Δ(Emp/Pop)		0.579**	7.749**
		(0.236)	(3.947)
ΔDemographic Controls	N	Y	Y
Age Controls (3-Year Bins)	Y	Y	Y
N	114726	3595	3595
First Stage F- Stat.			9.36
AR Test Statistic			5.351
AR P-value			0.0207
AR Weak-IV Robust CI			[1.3415, ...]

Notes: Changes in cognitive score, the employment-to-population ratio, and demographic controls are defined relative to one HRS wave prior. HRS waves are two years apart. Standard errors clustered by individual participant in column (1). Standard errors clustered by commuting zone in columns (2) and (3). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Employment and Cognition, IV Estimates, Heterogeneity by Age and Gender

	(1)	(2)	(3)	(4)	(5)
			Δ Cognitive Score		
Δ(Emp/Pop)	7.749**	9.877***	15.62	4.440**	129.4
	(3.947)	(3.485)	(13.84)	(1.737)	(642.1)
ΔDemographic Controls	Y	Y	Y	Y	Y
Age Controls (3-Year Bins)	Y	Y	Y	Y	Y
Sample Restriction	Age 51-75	Age 51-64	Age 65-75	Men 51-64	Women 51-64
N	3595	2713	2712	1942	2401
First Stage F- Stat.	9.36	12.60	1.16	31.38	0.04
AR Test Statistic	5.351	12.87	3.922	6.362	7.366
AR P-value	0.0207	0.000334	0.0477	0.0117	0.00665
AR Weak-IV Robust CI	[1.34, ...]	[4.77, ...]	[... ,-5.75] U [.823, ...]	[1.21, 8.77]	[... ,-48.59] U [53.10, ...]

Notes: Changes in cognitive score, the employment-to-population ratio, and demographic controls are defined relative to one HRS wave prior. HRS waves are two years apart. Standard errors clustered by commuting zone. * p < 0.10, **p< 0.05,p< 0.01***

Table 3: Robustness to Opioid Exposure Controls , IV, Men and Women Aged 51-75

	(1)	(2)	(3)	(4)
	ΔCognitive Score			
	Panel A: Drug Mortality Controls			
Δ(Emp/Pop)	6.239	7.782*	8.905**	7.851*
	(5.643)	(4.113)	(3.559)	(4.096)
ΔDrug Mortality	-0.555	0.356	0.466	
	(0.856)	(0.600)	(0.403)	
Drug Mortality Category				Y
First Stage F-stat	5.15	8.07	12.32	8.64
AR Test Statistic	1.16	4.60	9.20	5.31
AR P-value	0.2044	0.0320	0.0024	0.0212
AR Weak-IV Robust CI	[-5.156, …]	[0.779, …]	[3.409, 21.730]	[1.526, …]
	Panel B: No Drug Mortality Controls, Same Sample			
ΔEmp/Pop	4.484	8.768	10.07**	7.749**
	(5.867)	(5.224)	(4.683)	(3.947)
AR Test Statistic	3.13	5.57	7.95	9.36
AR P-value	0.67	4.13	7.91	5.35
AR Weak-IV Robust CI	0.4124	0.0421	0.0049	0.0207
	[… ,…]	[0.7004, …]	[3.2067, …]	[1.3415, …]
ΔDemographic Controls	Y	Y	Y	Y
Age Controls (3-Year Bins)	Y	Y	Y	Y
Imputation method for counties with missing or suppressed drug mortality data	No Imputation, CZ dropped	Missing counties omitted if <50% CZ pop	Net-of-state used for missing counties	Net-of-state used for missing counties
N	241	1,676	3,080	3,595

Notes: Changes in cognitive score, the employment-to-population ratio, and demographic controls are defined relative to one HRS wave prior. In column (4) we include dummy variables for the opioid exposure categories (Figure 1); coefficient estimates are not reported for confidentiality reasons. HRS waves are two years apart. Standard errors clustered by commuting zone. Broad sample of men and Women age 51-75 is used in all models. * p < 0.10, **p< 0.05,p< 0.01***

Table 4: Employment and Cognition, Robustness to Industry Classification Resolution, IV, Men and Women Aged 51-75

	(1)	(2)	(3)	(4)	(5)
			ΔCognitive Score		
Δ(Emp/Pop)	7.570*	8.101*	7.393*	6.773*	7.110*
	(4.330)	(4.244)	(4.113)	(3.828)	(4.040)
ΔDemographic Controls	Y	Y	Y	Y	Y
Age Controls (3-Year Bins)	Y	Y	Y	Y	Y
Industry Classification	SIC		NAICS		
	Major Industry	Subsector	Industry Group	5-Digit Industry	6-Digit Industry
Industry Count	83	96	308	689	1012
N	3370	3370	3370	3370	3370
First Stage F- Stat.	9.48	8.29	8.70	10.05	9.45
AR Test Statistic	4.600	5.715	4.649	4.137	4.136
AR P-value	0.0320	0.0168	0.0311	0.0420	0.0420
AR Weak-IV Robust CI	[.88, ...]	[1.55, ...]	[.72, ...]	[.56, ...]	[.55, ...]

Notes: Changes in cognitive score, the employment-to-population ratio, and demographic controls are defined relative to one HRS wave prior. HRS-waves occur two years apart. Standard errors clustered by commuting zone. * p < 0.10, **p< 0.05,p< 0.01***

Table 5: Summary of Rotemberg Weights

Panel A: Negative and Positive Weights

	Sum	Mean	Share
Negative	-0.249	-0.011	0.166
Positive	1.249	0.029	0.834

Panel B: Correlations

	α_k	g_k	β_k
Rotemberg Weights by Industry (α_k)	1		
Growth in Employment by Industry (g_k)	0.089	1	
Single Industry Instrument Estimate (β_k)	-0.006	-0.136	1

Panel C: Distribution of Rotemberg Weights across HRS Waves

	Sum		Sum
Δ1996-1998	0.027	Δ2008-2010	0.006
Δ1998-2000	0.015	Δ2010-2012	0.040
Δ2000-2002	0.052	Δ2012-2014	0.021
Δ2002-2004	0.508	Δ2014-2016	0.066
Δ2004-2006	0.110	Δ2016-2018	0.016
Δ2006-2008	0.139		

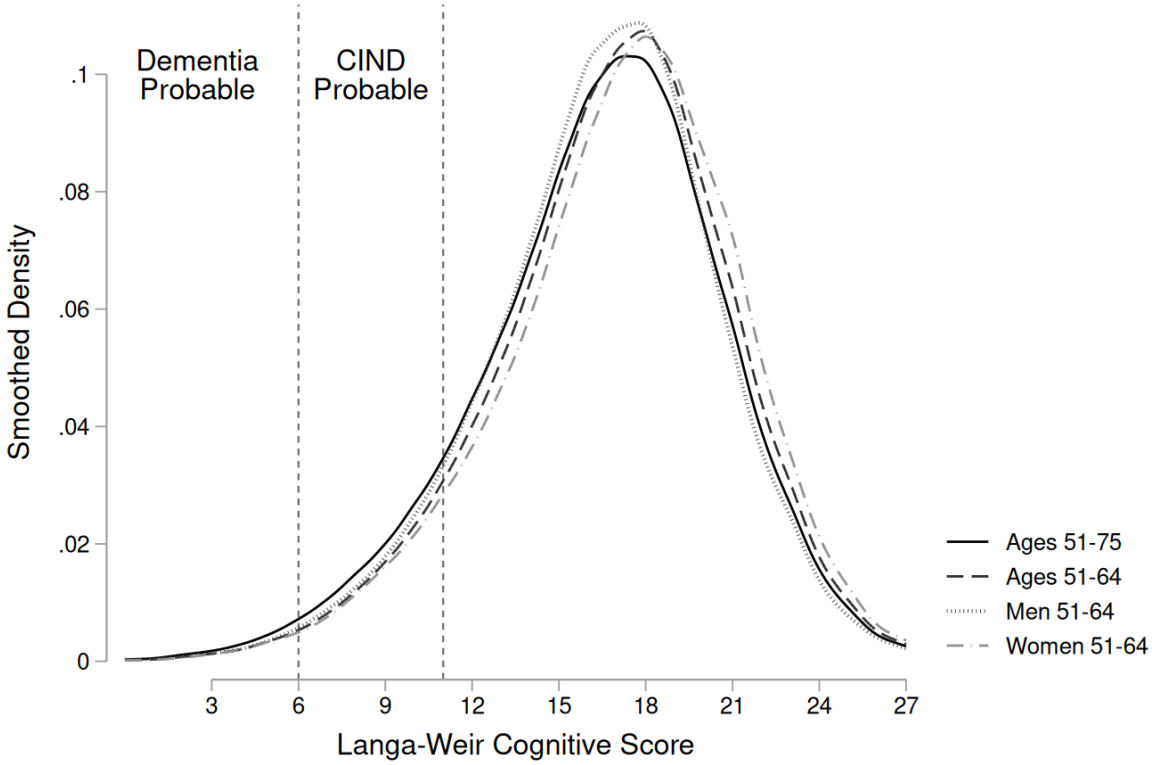
Panel D: Outlier Rotemberg Weight Industries

Rank	SIC Major Industry Group	α_k
1	Transportation Equipment Manufacturing	0.145
2	Ind./Comm. Machinery and Computer Equipment Manufacturing	0.124
3	Construction Special Trade Contractors	0.113
4	Automotive Dealers and Gasoline Service Stations	0.093
5	Eating and Drinking Retail	0.082
…		
83	Personal Services	-0.016
84	Fabricated Metal Product Manufacturing	-0.016
85	Engineering, Accounting, Research, and Management Services	-0.023
86	Food Store Retail	-0.050
87	Wholesale Trade of Durable Goods	-0.051

Notes: Rotemberg weights calculated for the instrumental variable estimation in Table 2, column (3). Panel A separately aggregates positive and negative Rotemberg weights. Panel B shows the correlation across industries in their Rotemberg weight, growth rate, and single instrument estimate. Panel C shows the sum of Rotemberg weights in each first difference used in our estimation strategy. Panel D lists the industries with the largest and smallest (most negative) Rotemberg weights.

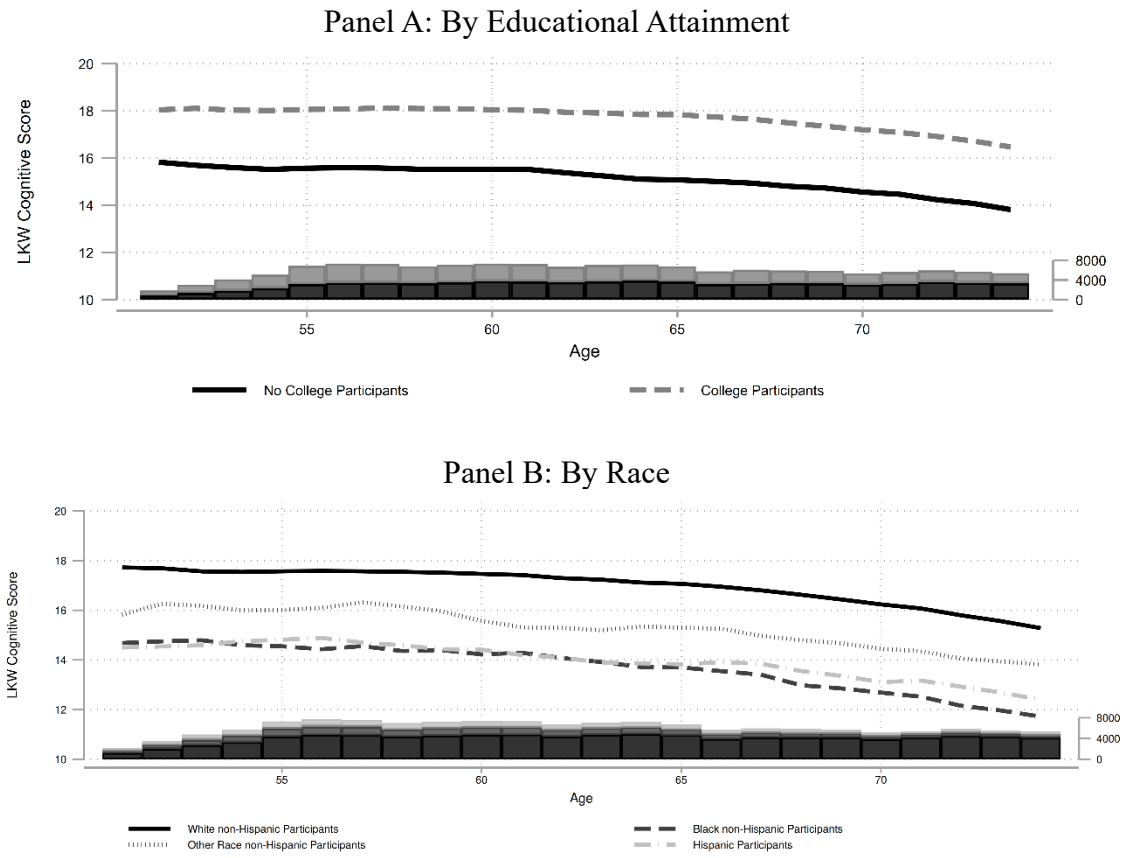
Appendix A: Additional Figures and Tables

Figure A1: Distribution of Global (Langa-Weir) Cognitive Score with Verified Cognitive Thresholds



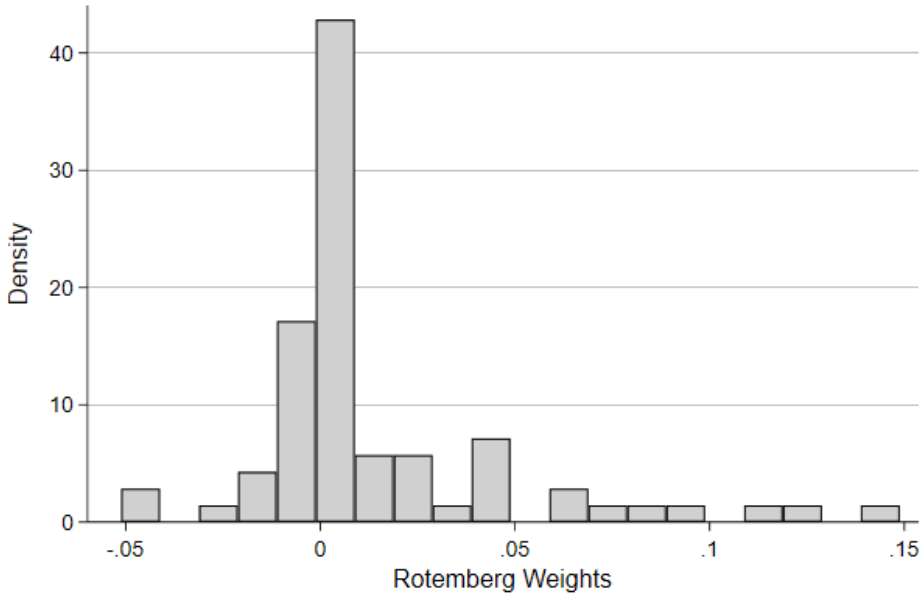
Notes: Calculated from HRS microdata.

Figure A2: Additional Cognitive Trajectories



Notes: 3-year rolling average applied to cognitive score series. Bars at bottom of each panel show sample size by age and category. Sample size bar heights correspond with the axis labels on the right.

Figure A3: Distribution of Rotemberg Weights



Notes: Rotemberg weights calculated for the model in Table 2, column (3).

Table A1: Summary of Employment, Cognition, and Demographics by Sample, Individual-Level Data

	Broad Age Range (51-75)	Pre-Retirement Ages (51-64)		
		All	Men	Women
Employed	52.57%	66.00%	71.56%	61.30%
Became Employed	3.28%	3.27%	2.90%	3.53%
Became Non-employed	7.86%	7.38%	7.10%	7.50%
Langa-Weir Index (0-27)	16.41 (4.16)	16.81 (3.99)	16.46 (3.96)	17.11 (4.09)
Immediate Word Recall (0-10)	5.83 (1.61)	6.00 (1.57)	5.77 (1.56)	6.21 (1.54)
Delayed Word Recall (0-10)	4.85 (1.92)	5.03 (1.88)	4.73 (1.83)	5.29 (1.88)
Serial 7s Subtraction (0-5)	3.76 (1.56)	3.83 (1.52)	4.01 (1.42)	3.67 (1.59)
Backwards Counting (0-2)	1.91 (0.41)	1.92 (0.40)	1.92 (0.39)	1.92 (0.40)
ΔLanga-Weir Index per wave	-0.15 (3.63)	-0.06 (3.59)	-0.74 (3.55)	-0.04 (3.63)
Age	62.00 (6.60)	57.89 (3.66)	57.79 (3.71)	57.83 (3.73)
Share non-Hisp. White (%)	77.77%	75.83%	76.58%	75.17%
Share non-Hisp. Black (%)	10.19%	10.78%	9.94%	11.57%
Share Hispanic (%)	8.37%	9.04%	8.88%	9.13%
Share College Educated (%)	53.45%	56.77%	59.10%	54.79%
Obs. per Commuting Zone	36.86 (81.07)	27.03 (56.53)	15.05 (27.37)	18.21 (34.95)
Waves per Participant	4.76 (3.05)	3.53 (1.89)	3.36 (1.84)	3.74 (2.00)
Total Observations	150,306	88,067	37,149	51,638

Notes: Means reported, with standard deviations in parentheses. All summary statistics besides observation counts are weighted by HRS sampling weights. College Educated is defined as holding a Bachelor's degree.

Appendix B: Related Work on Aggregation and Bias

Our argument about the impact of aggregation is reflected, in reverse, in earlier research on aggregation bias. For example, Hanushek et al. (1996) study the issue of bias in regressions of student outcomes on earlier test scores, individual and family characteristics, community variables, and school characteristics. They are motivated by differences in the results depending on the level of aggregation, arguing that a positive relationship between school resources and student outcomes tends to arise in more aggregated regressions, as opposed to disaggregated regressions. They point out that the direction of bias from aggregation is unclear. Like the argument in the main text, "aggregation of the relationships may actually have beneficial effects through reduction in measurement error or the bias introduced by the endogeneity of school and residential location choices …" (p. 612), but aggregation can increase omitted variable bias when aggregation is at the level of the omitted factors. For example, "state-level resources are correlated with a variety of important state influences on school performance …" in which case "aggregation increases the magnitude of omitted variable biases" (p. 612). Indeed, in the context of measurement error, they note that "The strategy of aggregating within defined groups … and replacing the microdata with the group average is a variant of instrumental variables, in which the group average is used as the instrument. Group averaging reduces bias if it raises the signal to noise" (p. 618), and they note that his happens when observations are "grouped on the basis of a characteristic that is correlated with the true underlying values of the regressor" (p. 618). This holds in the case of grouping at the CZ and year level, which is associated with $\Delta E_{i,l,t}$.

A similar argument is made in the context of studying the effects of macroeconomic conditions on voting. Kramer (1983) considers the evidence that aggregate time-series regressions of voting behavior on economic conditions reveal that better economic conditions

increase votes for the incumbents. But individual-level regressions of voting behavior on individual economic variables generally fail to find such a relationship. The argument is the same. Individual-level variation in economic conditions can reflect aggregate variation, but it will also reflect sources of idiosyncratic variation, including variation that may happen to be correlated with party affiliation (and perhaps differently in different elections). In the aggregate regressions the variation in the independent economic variable(s) is more reflective of aggregate conditions and economic policy, because this individual-level variation may average out. In the individual-level data, however, the idiosyncratic variation dominates; indeed, in data from one election, the role of macroeconomic conditions washes out completely. Thus, the aggregate regression can provide more reliable evidence on how macroeconomic conditions affect voting.