Genetics and the Cognitive Costs of Air Pollution

Vinh Phan*

October 21st, 2026

Abstract

Air pollution is a well-documented threat to physical health, and a growing body of research suggests it may also carry cognitive costs. This paper estimates the causal short-run effect of air pollution on cognitive performance and asks whether genetic propensity for educational attainment shields individuals from this cognitive toll. Using UK Biobank data linked to high-frequency air pollution measures, I study how same-day exposure to PM2.5 affects performance on a fluid intelligence test. To address endogeneity and exposure measurement error, I exploit thermal inversions as plausibly exogenous shocks that trap air pollutants near the ground. Reduced-form estimates show that higher air pollution significantly reduces fluid intelligence. A specification-curve analysis confirms the robustness of the negative effect across a wide range of modeling and data choices. I also find a decline in numeric memory, but no effect on visual memory, prospective memory, or reaction time. I also find find no evidence that genetic propensity for educational attainment moderates the immediate cognitive effects of air pollution. Overall, the results indicate that short-term air pollution exposure impairs cognitive performance broadly, and that this cognitive penalty appears to be shared evenly across genetic endowment for educational attainment.

*Erasmus University Rotterdam & Tinbergen Institute

1 Introduction

Air pollution imposes large social costs through its effects on health. A large and well-established literature has already documented adverse health consequences of exposure to fine particulate matter, including elevated mortality risk, respiratory morbidity, cardiovascular disease, and neurological disorders (e.g., Arceo et al., 2016; Clay et al., 2018; Deryugina et al., 2019; Hoek et al., 2013; Karimi and Samadi, 2024; Xu et al., 2016). While these alone already provide a strong argument for air pollution mitigation, exclusively focusing on health effects risks underestimating the benefits of cleaner air. To the extent that air pollution disrupts cognitive performance, its costs may extend beyond illness and premature death to lower productivity and diminished human capital accumulation.

The effects of air pollution on cognitive performance need not be uniform across individuals, who may differ in their endowments and latent propensities for educational and cognitive performance. A growing literature in economics, epidemiology, and behavioral genetics has shown that genetic endowments can shape how individuals respond to adverse environments, giving rise to gene–environment interactions ($G \times E$) (e.g., Baker et al., 2024; Barcellos et al., 2018, 2021; Muslimova, van Kippersluis, et al., 2023; Virolainen et al., 2023). If pollution interacts with predetermined endowments, then it may not simply lower cognitive performance on average; it may also amplify or compress existing inequalities in cognitive performance. The existing literature does not current offer a clear prediction as to which of these possibilities should prevail. Uncovering these interactions would thus improve our understanding of how the benefits of cleaner air are distributed.

This paper studies whether contemporaneous air pollution, in particular fine particulate matter, affects cognitive performance and whether this effect varies by an individual's endowment, proxied by their genetic propensity towards educational attainment. To do this, I use data from the UK Biobank, which provides detailed information on health, socio-economic status, and genetic background for over 500,000 individuals. My main outcome of interest is fluid intelligence measured through a 13-item cognitive test administered to participants. Air pollution data, specifically daily average concentrations of PM2.5, are obtained from the European Air Quality (EAC4) dataset provided by the Copernicus Atmosphere Monitoring Service, which gives hourly pollution data across the globe. This enables me to explore the effects of local air pollution exposure on cognitive performance in a low-pollution setting like the UK, where pollution levels are generally lower than in many other parts of the world but still exhibit meaningful variability.

Air pollution is not randomly distributed across locations or over time. More polluted areas may also differ in incomes, education, health, or other characteristics that are associated with cognitive performance. Seasonal patterns or trends in participant composition may also be correlated with those in air pollution. Finally, fluctuations in air pollution may also coincide with changes in the weather, e.g., temperature, that could independently affect performance on the cognitive test. Consequently, an observed relationship between pollution and reasoning scores may partly reflect

differences in participants' composition or assessment conditions, rather than the causal effect of pollution itself. To address these concerns, I use two complementary empirical strategies.

First, I use a fixed-effects (FE) model that controls for unobserved time-varying and location-specific factors, including location, month, year, and day-of-week fixed effects. These fixed effects control for unobserved heterogeneity across locations and across time periods. However, they are not able to account for location-specific changes over time, which may also simultaneously affect air pollution levels and cognitive performance. For example, changes in traffic intensity or nearby construction activity may affect pollution alongside noise and congestion, which could independently influence participants' cognitive performance.

Therefore, to address potential remaining endogeneity, I use an instrumental variables (IV) approach, where thermal inversion coverage serves as an instrument for air pollution exposure. Thermal inversions, which trap pollutants close to the ground, provide exogenous variation in PM2.5 exposure, isolating the causal effect of pollution on cognitive outcomes.

To examine whether the effects of air pollution vary with one's genetic propensity for educational attainment, I utilize polygenic scores (PGS) for educational attainment (PGS EA). A polygenic score aggregates the effects of many genetic variants associated with a particular trait, weighted by the strength of their individual associations with the trait. PGS provide a genetic risk score that quantifies an individual's genetic propensity to a given outcome, in this case, educational attainment. By including PGS EA in my model, I aim to assess whether genetic factors influence the susceptibility to the cognitive effects of short-term pollution exposure.

The results indicate a significant negative effect of PM2.5 exposure on fluid intelligence scores. An increase in 1 $\mu g/m^3$ of PM2.5 decreases performance on the fluid intelligence test by 0.01 standard deviation. This relationship is robust and consistent across different model specifications. However, I do not find significant heterogeneity by PGS. My findings therefore suggest that the cognitive effects of short-term pollution exposure are similar across individuals, regardless of their genetic propensity.

The study has three primary contributions. First, it contributes to the growing body of research examining the contemporaneous cognitive costs of air pollution. While existing studies often rely on proxies for cognitive function, such as high-stakes exam scores (Cook et al., 2023; Ebenstein et al., 2016; Yao et al., 2023) or labor output in specific industries (Archsmith et al., 2018; X. Chen et al., 2023; Holub and Thies, 2023; Künn et al., 2023), this paper is among the few to utilize direct, standardized cognitive test scores (La Nauze and Severnini, 2025). This paper therefore complements evidence from examination and workplace settings, helping to assess whether the effects documented in these settings extend to cognitive performance more broadly.

Second, this paper brings a contemporaneous perspective to research on the interplay between individual endowments and environmental conditions. Models of human capital formation emphasize how endowments and environmental inputs jointly shape the accumulation of skills (Cunha

and Heckman, 2007; Heckman, 2007), and recent empirical work has begun to investigate these interactions in developmental settings (Muslimova et al., 2026; von Hinke and Sørensen, 2023). At the same time, environmental conditions may also influence how effectively adults draw on the capabilities they have already accumulated. This raises the question of whether endowments associated with stronger cognitive performance also help individuals maintain that performance when conditions deteriorate. By examining pollution exposure on the day of assessment, I investigate both the sensitivity of adult cognitive performance to contemporaneous environmental shocks and whether this sensitivity varies with education-related genetic propensity. The analysis therefore speaks to whether the advantages associated with these endowments extend to resilience against immediate environmental shocks.

Third, the paper examines how estimates of pollution's cognitive effects depend on the construction of environmental exposures and the specification of meteorological controls. It therefore contributes to a growing literature examining how economic estimates depend on the choices researchers make when constructing environmental measures and specifying empirical models. Recent studies show the value of assessing these choices jointly, tracing how estimated relationships change across alternative spatial linkages, weather measures, and regression specifications (Josephson et al., 2026; Yamada, 2026). I bring this approach to the estimation of pollution's cognitive effects, where exposure construction and meteorological adjustment jointly determine the variation used for identification.

The results also carry policy relevance for the valuation of air pollution mitigation. Even in a comparatively low-pollution environment like the UK, PM2.5 significantly impairs cognitive performance, suggesting that further mitigation of air pollution can be beneficial on cognitive grounds alone. Moreover, the cognitive benefits of cleaner air are likely to be distributed relatively evenly across the population, rather than accruing disproportionately to individuals with a particular genetic endowment.

2 Background

2.1 Air pollution and cognitive performance

Cognitive functioning encompasses a broad set of mental processes, including attention, memory, processing speed, executive function, and reasoning. Among these, higher-order cognitive abilities such as reasoning and problem solving are central components of human capital because they shape schooling, productivity, and labor-market performance.

Increasingly, empirical evidence in economics and epidemiology points to a negative link between air pollution and cognitive performance (Bedi et al., 2021; Ebenstein et al., 2016; Sunyer et al., 2015; Weuve et al., 2012). Air pollution refers to multiple toxic components such as carbon

monoxide (CO), sulphur dioxide (SO2), and particulate matter (PM). While the exact mechanisms by which air pollutant affects cognitive performance remains a topic of ongoing research, it is generally thought that air pollution causes an inflammatory response in the body through two possible pathways. The first pathway proposes that particulate matter physically enters brain tissue, either via the olfactory nerve or by crossing into the bloodstream through the lungs and reaching the brain directly, where it triggers local neuroinflammation (Elder et al., 2006; Qi et al., 2022). The second pathway does not require pollutant particles to reach the brain. Instead, it posits that particles deposited in the lung can trigger pulmonary inflammation, and the resulting inflammatory signals spread through systemic circulation to activate neuroinflammatory processes in the brain (Block and Calderón-Garcidueñas, 2009).

Reasoning performance depends on the integrity and functioning of brain networks. Woolgar et al. (2010), for example, shows that performance on a fluid intelligence test was predicted by damage in the frontal and parietal cortex. Recent evidence indicates that air pollutants can already lead to changes in brain activity in a relatively short time. In a randomized study, Gawryluk et al. (2023) shows that same-day exposure to diesel exhaust reduced the coordination between different brain regions, a function that enables the brain to flexibly utilizes its network of connections to handle novel problems (Barbey, 2018; de Souza et al., 2023; Lombardo and Kaufmann, 2023). These changes in the brain can map to functional changes. In a randomized experiment, Faherty et al. (2025) exposed subjects in the treatment group to higher levels of PM2.5 and found that these individuals perform significantly worse on selective attention and emotion recognition when assessed four hours after the exposure compared to the control group. While the sample size of this study, 26 individuals in total, was limited by what was feasible for an interventional human study, it provided some of the first experimental evidence that executive function can decline within hours of a single elevated PM2.5 exposure. In addition to these two pathways, pollution-induced physical discomfort, e.g., respiratory irritation, could also reduce performance on cognitive tests through reduced effort or attention, independent of any direct neurological effect (Frampton et al., 2015; Laumbach et al., 2011).

This pollution-cognition link identified in medical research parallels a growing economics literature showing that transitory, same-day pollution exposure impairs performance on high-stakes cognitive tasks. Ebenstein et al. (2016) find that Israeli students exposed to elevated PM2.5 on their exam day scored worse, with effects persisting into lower postsecondary attainment and earnings a decade later. Similarly, Archsmith et al. (2018) find that baseball umpires make more incorrect calls on high-pollution days, Künn et al. (2023) find that chess players make more and larger errors as indoor PM2.5 rises, and Chang et al. (2019) find that call-center workers complete fewer calls and make more mistakes with higher daily pollution. This literature establishes two facts that motivate our analysis: real-world tasks demanding judgment and sustained attention are vulnerable to acute pollution exposure, even though the coarse, task-specific outcomes used in these studies cannot pin

down which underlying cognitive capacity is responsible; and such short-lived shocks can nonetheless translate into lasting economic consequences, as the exam and productivity settings above make clear. Taken together, these findings underscore both the breadth of pollution's cognitive costs and the need for a more precise account of the specific mental processes involved.

2.2 Genetic endowment and air pollution

Cognitive performance is shaped by both environmental conditions and individual endowments. A growing literature in economics and social-science genetics emphasizes that these forces may interact (Almond et al., 2018; Baker et al., 2024; Barcellos et al., 2018; Muslimova et al., 2026; Pereira et al., 2020). In the context of air pollution, there are several reasons to expect such heterogeneity. Air pollution can impair cognitive performance through physiological channels such as inflammation, oxidative stress, vascular stress, and neurological disruption. If individuals differ in cognitive reserve, health endowments, or biological vulnerability, the same air pollution shock may translate into different cognitive effects.

These differences can be seen more broadly as variation in underlying human-capital endowments. Genetic propensity provide one measurable component of such endowments, capturing differences in cognitive potential, learning ability, and resilience to environmental stressors. From this perspective, air pollution may not only reduce cognitive performance on average but also alter the distribution of outcomes across individuals. If pollution is more harmful for individuals with lower endowments, it may amplify existing inequalities; if it disproportionately affects individuals with higher endowments, it may compress them. Understanding this heterogeneity is therefore central to assessing the distributional consequences of pollution.

An individual's genetic propensity towards a certain phenotype, for example educational attainment, can be quantified using a polygenic score (PGS) (Okbay et al., 2022). Specifically, a polygenic score is a weighted sum of an individual's allele counts across genetic variants known as single nucleotide polymorphisms (SNPs), with weights derived from their estimated associations with the phenotype in a genome-wide association study (GWAS). Formally, a PGS for an individual i is as follows:

$$PGS_i = \sum_{j=1}^{J} w_j a_{ij}$$

where w_j is the weight found through a GWAS, $a_{ij} \in 0, 1, 2$ is the number of effect alleles individual i carries at SNP j, and J is the total number of SNPs included in the score.

Polygenic scores reflect the fact that most complex behavioral and cognitive traits are polygenic, meaning that they are influenced by many genetic variants, each of which has only a very small association with the phenotype. The resulting polygenic score should be interpreted as summary

measure of genetic propensity for a trait and not a deterministic measure of ability. In addition to capturing direct biological effect of genetic variants, PGS may also capture gene–environment correlation. This arises from active and evocative processes in which individuals select into or evoke environments, as well as passive genetic nurture, whereby parental genetics shape the environment in which a child grows up. PGS estimates may also reflect population stratification if genetic ancestry is correlated with social, geographic, or environmental conditions (Biroli et al., 2026). For example, a genetic variant that is more common among people with ancestry from regions with historically better schooling may predict educational attainment partly because it captures differences in educational opportunities rather than a direct genetic effect. As such, gene-environment analyses should be interpreted as documenting systematic differences in the effects of pollution exposure on cognitive performance across individuals with different genetic propensity, rather than as identifying a purely causal biological interaction between genes and environmental shocks.

In this paper, I use the polygenic score for educational attainment as the preferred measure of genetic propensity for cognitive performance. Although a polygenic score for cognitive performance may appear more directly aligned with the fluid intelligence outcome, there are good reasons to prefer the educational-attainment polygenic score. First, educational attainment has been measured in much larger genome-wide association studies, which improves the precision of the estimated genetic variant weights and the predictive power of the resulting score. Second, educational attainment is closely related to cognitive performance and human-capital accumulation, which makes it substantively relevant for a study of pollution and cognition. The educational-attainment PGS (PGS EA) explains substantial variation not only in education, but also in cognitive performance. Lee et al. (2018), for example, construct an educational-attainment PGS from a GWAS of approximately 1.1 million individuals and show that it explains 11–13 percent of the variance in educational attainment and 5-8 percent of the variance in cognitive performance among Add Health and Health and Retirement Study participants.

Epidemiological studies have begun to document gene-by-pollution interactions in cognitive outcomes. For instance, Franz et al. (2023) found that PM2.5 and NO2 exposure are more strongly associated with contemporaneous declines in executive function and episodic memory among APOE $\varepsilon 4$ carriers than non-carriers. However, these studies rely on residential or modeled ambient pollution exposure rather than exogenous variation, leaving estimates vulnerable to confounding from neighborhood sorting and other unobserved factors correlated with both pollution exposure and cognitive outcomes. Economics research, by contrast, has used exogenous sources of pollution variation to study gene-environment interaction, but has so far focused exclusively on long-run outcomes measured decades after exposure: von Hinke and Sørensen (2023) exploit the 1952 London smog to study its effects on cognition and health in later adulthood, and Fukushima et al. (2024) similarly study the long-term human capital and health impacts of a historical pollution reduction program. Neither literature therefore speaks directly to whether genetic propensity moderates the contem-

poraneous cognitive effects of pollution - the immediate, same-day physiological response that this paper is designed to isolate.

3 Data

3.1 Cognitive assessment and genetics data

I obtain data on cognitive performance and genetics from the UK Biobank. The UK Biobank collected data on approximately 500,000 individuals aged 40-69 between 2006 and 2010 throughout the UK (Sudlow et al., 2015). At the initial assessment, participants provided information on their socioeconomic and family backgrounds, health, and behaviors through questionnaires, interviews, and physical measurements. Biological samples were also collected, with DNA extracted from blood samples for genotyping (Bycroft et al., 2018). Subsequent data collection includes a repeat assessment of approximately 20,000 participants in 2012–2013, additional online questionnaires, and ongoing follow-up through linked health records (Sudlow et al., 2015). This study uses cognitive performance measured at the initial assessment, linked to participants' genetic data.

During the initial assessment, participants were also asked to complete a battery of validated cognitive tests (Fawns-Ritchie and Deary, 2020). These include a reaction time test, a visual memory test, a numeric memory test, a prospective memory test, and a fluid intelligence test. I focus on fluid intelligence as the primary outcome and additionally examine reaction time and memory to assess whether the effects extend to other cognitive domains. A thorough description of how these assessments are implemented in the UK Biobank is detailed in section B.2 of the Appendix.

Fluid intelligence scores in the UK Biobank were derived from a 13-item test administered to participants during the baseline assessment interview. The test includes tasks such as pattern recognition, abstract reasoning, and logical problem solving, which require participants to process information in real-time without relying on prior knowledge. Section B.1 of the Appendix lists all 13 items in this test. The fluid intelligence (FI) score is then based on the number of correct answers participants provide. Importantly, the fluid intelligence section of the baseline assessment was only introduced towards the end of the baseline and is not mandatory, i.e., participants can choose whether or not they do this test. This could complicate the interpretation of the estimated effect of air pollution on fluid intelligence scores, specifically if air pollution also affects the probability of taking the test. About one third of the baseline participants were presented with this test module, which was entirely voluntary and could have been skipped if the participants chose to do so. The vast majority of these participants chose to participate in this test. In section 5.2, I show that air pollution does not predict whether someone participates in the test.

Genetic data in the UK Biobank were obtained from biological samples collected at the baseline assessment visit, from which DNA was extracted and genotyped. Standard quality control proce-

dures were applied by the UK Biobank prior to data release. Using this genetic dataset, Muslimova, Dias Pereira, et al. (2023) constructed polygenic scores (PGS) that summarize participants' genetic propensity toward educational attainment based on the EA3 GWAS from Lee et al. (2018). In independent European-ancestry validation samples, Lee et al. (2018) report that their education-based PGS explains an additional 12.7% of the variance in years of schooling in the National Longitudinal Study of Adolescent to Adult Health and 10.6% in the Health and Retirement Study, beyond demographic and genetic ancestry controls. The score also predicts cognitive performance: in the Health and Retirement Study, it explains an additional 2.7% of the variance in total cognition and 4.7% in verbal cognition (Lee et al., 2018).

Figure 1 shows the distribution of fluid intelligence scores stratified by whether individuals fall below or above the median of the PGS EA distribution. There exists substantial variation in the FI scores within both groups, with most individuals answering five to nine out of 13 questions correctly. On average, those in the upper half of the PGS EA distribution scores 10% higher than those in the lower half, which shows that genetics-related differences drive differences in cognitive performance. This paper studies whether air pollution mitigates or exacerbates this inequality.

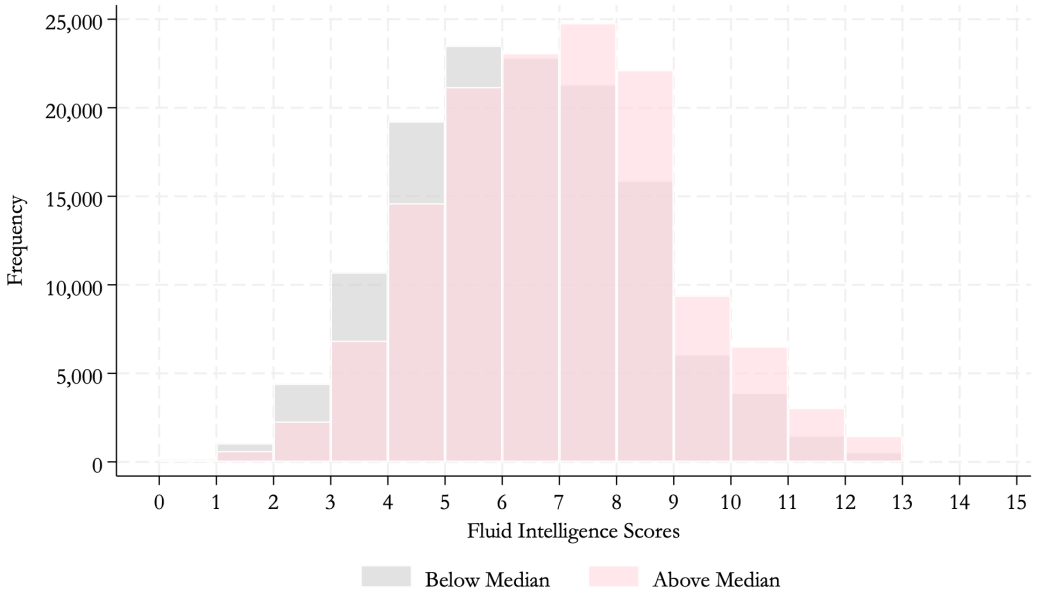


Figure 1: Fluid intelligence distribution by EA polygenic score group.

Because the analysis focuses on the contemporaneous effects of air pollution on cognitive performance, I assign exposure at the location and date of testing. Pollution at the assessment center serves as a proxy for ambient conditions during the assessment visit, which may differ from those at participants' homes. The UK Biobank provides information on participants' residential locations through geocoded home address data collected at the time of the assessment. Since air pollution is defined at the assessment center - date level, I infer the assessment center that each participant

goes to by matching each participant's home address with the nearest assessment center. The vast majority of participants are matched to a center within 40km of their home address, as intended by the UK Biobank. About 8,000 participants reported a home address further than 40km to the nearest assessment center. Since I cannot be sure of the assessment center these participants went to, I exclude them in the main analysis. The results, however, are robust to the inclusion of these participants[1] Furthermore, since the UK Biobank sample exhibits volunteer bias, which can hinder the external validity of the results, I also include inverse probability weights, as computed by van Alten et al. (2024), in all of my analysis to make the sample more representative of the targeted UK Biobank sample.

3.2 Weather data

I collected weather data from the Copernicus Climate Change Service (C3S) (Copernicus Climate Change Service, 2018a, 2018b). C3S contains reanalysis data that combines observational data with physics-based numerical model to produce an estimate of the atmosphere (Auffhammer et al., 2013). The resulting dataset records a wide array of meteorological variables hourly, at both the surface and 37 pressure levels, on a 0.25×0.25-degree grid. The data collected for this study spans the entirety of the United Kingdom and includes temperature (surface and across different pressure levels), as well as surface pressure, dew temperature, precipitation, cloud coverage, wind speed, and wind direction.

I downloaded hourly data at 12AM, 6AM, 12PM, and 6PM for each day from 2006 to 2010. For each hourly observation at each grid point, I created a thermal inversion indicator by determining whether the temperature at the pressure level immediately above the surface exceeds the ground temperature, in which case I set the indicator to one. Once the hourly thermal inversion indicator was created for all grid points, at all hours, I aggregated the indicator to the daily level. The daily thermal inversion indicator was one if there was any thermal inversion during one of the four hours I observed in the day. For all weather variables other than thermal inversions, I aggregated the hourly data to the daily level by calculating the average of the four hourly observations within each day.

I assign daily thermal inversion and weather data to each UK Biobank assessment center by calculating the weighted average of all data points within a 60 km radius of the center. Each data point is weighted by the inverse of its distance to the center, so that closer data points receive higher weights in the calculation. The resulting variable can thus be seen as the weighted spatial coverage of inversion in the area surrounding an assessment center, and can range from zero to one. This measure captures how widespread inversion coverage was within 60 km on a given day, giving greater weight to locations closer to the assessment center. A value of 0.6, for example, means

[1]The results for the non-restricted sample are presented in tables A7 to figure A18 in the Appendix.

that locations experiencing at least one observed inversion account for 60% of the total spatial weight. For robustness, I conduct a robustness check using a narrower 20 km radius. The results are largely consistent across both radii; however, the 20 km radius leads to a significant reduction in the number of observations due to the sparsity of air pollution data points.

3.3 Air pollution data

I use hourly air pollution data from version 4 of the European Air Quality (EAC4) dataset provided by the Copernicus Atmosphere Monitoring Service (CAMS), covering the period from 2006 to 2010 in the United Kingdom (Inness et al., 2019). EAC4 is a gridded dataset that provides global estimates of atmospheric pollutant on a 0.75×0.75 degree grid. The data includes information on key particulate matter pollutants, including PM2.5, PM1, and PM10. The data is derived from a combination of satellite observations and ground-based measurements, processed using atmospheric models, which provide high spatial and temporal resolution.

For each grid point, I compute the daily average using all hourly observations within the day. I assign air pollution levels to each UK Biobank assessment center by computing the average of all grid points within a 60 km radius of the center, weighted by the inverse of each grid point's distance to the center.

Figure A1 in the Appendix illustrates the mean daily PM2.5 concentrations across all UK Biobank centers, averaged over all assessment days available for each center. The map reveals a pronounced spatial gradient, with higher average PM2.5 concentrations concentrated in southern and eastern England, particularly in and around major urban and industrial areas, and lower concentrations observed in northern England, Scotland, Wales. This pattern closely aligns with well-documented sources of air pollution, including population density, traffic intensity, and industrial activity, which are more prevalent in the South East and Midlands.

4 Estimation strategy

4.1 Fixed-effects approach

I first employ a fixed effects technique, controlling for time-invariant characteristics of district, month fixed effects, year fixed effects, and day-of-week fixed effects. This approach captures unobserved heterogeneity across districts, as well as common time-specific heterogeneity that may influence both pollution exposure and cognitive performance. My estimation equation is thus as follows:

$$FI_{ict} = \alpha + \beta PM2.5_{ct} + \gamma_c + \lambda_m + \tau_y + \theta_w + \mu' \mathbf{W}_{ct} + \epsilon_{ict} \tag{1}$$

Where FI_{ict} denotes the fluid intelligence scores of participant i who attended center c on day t.

$PM2.5_{ct}$ is the average daily ambient PM2.5 concentration around assessment center c on day t. γ_c, λ_m, τ_y, θ_w are center, year, month, and day-of-week fixed effects, and $\mathbf{W}$ is a vector of weather controls, which includes ground temperature, dew temperature, cloud coverage, precipitation, wind speed, and wind direction, as well as their squared and cubed terms. While the fixed-effects approach controls for a wide range of confounding factors, it may not fully address the potential endogeneity of pollution exposure. Specifically, β may still be biased if there are unobservable, time-varying differences across participants who attended different assessment centers that also correlate with both cognitive performance and air pollution. Participant may have also selected into different assessment days based on air quality conditions. If such avoidance behaviors lead to systematic differences in who attended assessment centers on days with high air pollution, estimated effects may also capture selective attendance rather than the causal impact of air pollution on cognitive performance. Moreover, since the PM2.5 grid cells are relatively large, the air pollution level assigned to each district may be subject to some measurement error. Given that fixed effects models are sensitive to classical measurement error, my estimates may be biased downward.

4.2 Instrumental variable approach

I thus consider an instrumental variable approach, which will mitigate bias from both endogeneity and classical measurement error. I specifically utilize a meteorological phenomenon known as thermal inversion. Under a normal temperature gradient in which the air temperature falls with increasing altitude, the air vertically circulates as warm, thin air rises and cool, dense air sinks. When pollutant particles are present in the air, this vertical circulation carries some of these pollutants upwards into the upper troposphere. Thermal inversions occur when this normal temperature gradient is inverted, i.e., there is a layer of warm air on top of a layer of cool air, thus preventing the air layers from mixing vertically and trapping pollutants on the ground (Jacobson, 2002). In conjunction with industrial, residential, and vehicle emissions, thermal inversion episodes therefore exacerbate the concentration of air pollutant on the ground (Robinson, 1971).

Thermal inversions arise from meteorological conditions and local geography. Because the weather conditions associated with inversion formation may also directly affect cognitive performance, I control for temperature, humidity, wind speed, cloud coverage, and precipitation using linear, quadratic, and cubic terms. Identification relies on the variation in thermal inversions that remains after conditioning on these controls. The identifying assumption is that the included weather controls adequately capture the direct effects of weather on cognitive performance, and that any remaining weather interactions associated with inversion formation affect cognitive performance only through inversion-induced changes in air pollution. It is unlikely that participants would select into different assessment days based on these unpredictable interactions. Since thermal inversions arguably only affect individuals through their effects on increasing air pollution concen-

tration, they have often been used as instruments for pollution exposure to examine health and non-health outcomes (e.g., Arceo et al., 2016; S. Chen et al., 2022; Jans et al., 2018; Molina, 2021).

Figure A2 in the Appendix displays the weighted spatial coverage of thermal inversion occurrence across UK Biobank assessment centers, averaged over all available assessment days for each center. The map highlights substantial geographic heterogeneity in inversion coverage, with higher coverage in parts of southern and eastern England, and notably lower coverage in northern England and Scotland. This spatial pattern is consistent with meteorological conditions conducive to thermal inversions. Inversions are more frequent in areas characterized by stable atmospheric conditions, such as inland, low-lying regions with weaker wind speeds, which are more prevalent in southern and eastern England. In contrast, areas with stronger winds, such as those near coasts or with more complex topography, like Scotland and parts of Wales, show significantly lower inversion frequencies.

By acting as a lid on near-surface air, thermal inversions can trap multiple pollutants simultaneously. Consequently, the variation in air pollution induced by inversions need not be confined to PM2.5, and a single-pollutant 2SLS specification may capture the effects of other pollutants that also influence cognitive performance, such as NO_2 and SO_2. I therefore treat the reduced form as my preferred specification: under the identifying assumptions discussed above, it captures the cognitive effect of inversion-induced increase in air pollution concentration. Alongside these estimates, I report 2SLS estimates for PM2.5, which express the reduced-form effect relative to the accompanying change in PM2.5 concentration. This provides a scale for comparison with the PM2.5 fixed-effects estimates and previous studies, although one should keep in mind differences in identifying variation and pollution composition. The 2SLS estimates should therefore not be interpreted as isolating the causal effect of PM2.5. The reduced form equation is as follows:

$$FI_{ict} = \alpha_{RF} + \beta_{RF} Inv_{ct} + \gamma_c + \lambda_m + \tau_y + \theta_w + \mu' \mathbf{W}_{ct} + \epsilon_{ict} \tag{2}$$

Inv_{ct} indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center c on day t, and can range from zero to one. All other variables are defined like in equation (1). All other variables are defined like in equation (1). My 1st-stage and 2nd-stage equations for the effects of PM2.5 on fluid intelligence scores are as follows:

$$PM2.5_{ct} = \alpha_{FS} + \beta_{FS} Inv_{ct} + \gamma_c + \lambda_m + \tau_y + \theta_w + \rho' \mathbf{W}_{ct} + \epsilon_{ict} \tag{3}$$

$$FI_{ict} = \alpha_{SS} + \beta_{SS} \widehat{PM2.5}_{ct} + \gamma_c + \lambda_m + \tau_y + \theta_w + \mu' \mathbf{W}_{ct} + \epsilon_{ict} \tag{4}$$

$\widehat{PM2.5}_{ct}$ is the average daily concentration of PM2.5 as instrumented by thermal inversion.

4.3 Heterogeneity by genetic propensity

To examine whether the effect of air pollution on cognitive performance varies by genetic propensity, I extend the specifications in equation (2)[2] and 4 to allow for heterogeneity by polygenic scores (PGS) for educational attainment:

$$\begin{aligned} FI_{ict} =& \alpha_{SS} + \beta_{SS}\widehat{PM2.5}_{ct} + \eta_{SS}PGS_i + \kappa(\widehat{PM2.5}_{ct} \times PGS_i) + \\ & \gamma_c + \lambda_m + \tau_y + \theta_w + \mu'\mathbf{W}_{ct} + \phi\mathbf{PC}_i + \epsilon_{ict} \end{aligned} \tag{5}$$

$$\begin{aligned} FI_{ict} =& \alpha_{RF} + \beta_{RF}Inv_{ct} + \eta_{RF}PGS_i + \kappa_{RF}(Inv_{ct} \times PGS_i) + \\ & \gamma_c + \lambda_m + \tau_y + \theta_w + \mu'\mathbf{W}_{ct} + \phi\mathbf{PC}_i + \epsilon_{ict} \end{aligned} \tag{6}$$

PGS_i is the polygenic scores for educational attainment and is standardized to have mean zero and unit standard deviation. $\mathbf{PC}_i$ is a vector of 40 genetic principal components, and the remaining terms are defined as in the previous specifications. The principal components summarize genome-wide ancestry differences and are standard in analyses using UK Biobank genetic data. Their inclusion serves to control for population stratification, i.e., systematic differences in allele frequencies across ancestry groups that may be correlated with environmental exposures, geographic location, or socio-economic characteristics. Conditioning on a rich set of genetic principal components therefore helps ensure that the estimated associations involving PGS are not driven by ancestry-related confounding (Biroli et al., 2026).

A further concern is that the interaction between genetic propensity and air pollution could partly capture interactions between genetic propensity and the weather. Including weather controls additively does not account for this. I therefore also estimate specifications that interact PGS with all included weather controls, including their linear, quadratic, and cubic terms. The results of these specifications are presented in the Appendix and do not significantly differ from our specifications above.

Importantly, as mentioned in section 2.2, while the main effect of air pollution is identified using exogenous variation from thermal inversions, the interaction term between pollution/inversion and PGS may capture more than just a direct causal gene–environment interaction. Although genetic variation is predetermined and fixed at conception, the interaction coefficient κ may reflect a combination of channels, including direct biological moderation of pollution effects by genetic propensity, but also indirect and non-causal pathways. As such, the interaction term should be

[2]Here I assume that, conditional on the included controls and fixed effects, the effect of thermal inversions on PM2.5 concentrations does not vary with individuals' polygenic scores for educational attainment. Under this first-stage homogeneity assumption, differences in the response of cognitive performance to inversions across PGS levels are not driven by systematic differences in the inversion-induced change in PM2.5

interpreted as describing heterogeneity in the causal pollution effect across individuals with different genetic propensity, rather than as isolating a purely causal biological mechanism.

5 Results

Figure A5 in the Appendix visualizes the relationship between raw fluid intelligence (FI) scores and the average concentration of PM2.5 on the assessment day in a binned scatter plot, without any controls. The figure suggests a positive relationship between PM2.5 and fluid scores, likely reflecting selection into locations, whereby individuals with higher fluid intelligence scores are more likely to reside in more polluted areas (e.g., cities). Indeed, Table A2 in the Appendix shows that exposure to PM2.5 are correlated with multiple socioeconomic variables, and is higher among highly educated individuals. The remainder of this section therefore moves beyond raw correlations by introducing fixed effects and an instrumental-variables design, allowing us to distinguish selection-driven associations from the causal impact of air pollution on fluid intelligence.

5.1 Fixed effects model

Table 1 summarizes the results from estimating equation (1). Column (1) includes center, month, year, and weekday fixed effects, exploiting within-center and within-time variation in pollution exposure. The estimated coefficient on PM2.5 is small and statistically insignificant. Column (2) additionally controls for a rich set of weather variables, including temperature, dew point temperature, wind speed, wind direction, and total cloud cover, along with their squared and cubed terms. The inclusion of these controls does not affect the magnitude or precision of the estimate, which remains close to zero.

The absence of a detectable association in these fixed-effects specifications does not necessarily imply the absence of a causal effect. Since the original grid of air pollution data is quite sparse, inferred air pollution at each assessment center likely contains some measurement error. Moreover, air pollution levels may be correlated with unobserved, time-varying factors even after conditioning on fixed effects and weather controls, which may mask a causal effect of air pollution on cognitive performance. These considerations motivate the use of an instrumental-variables approach in the next section, which leverages exogenous variation in air pollution induced by thermal inversions to isolate the causal effect of air pollution on fluid intelligence.

Table 1: Fixed effects estimates

	Fluid intelligence scores	
	(1)	(2)
PM2.5	-0.0004	-0.0001
	(0.0009)	(0.0009)
R^2	0.01218	0.01275
Observations	125,224	125,224
Weather controls		✓

Notes: The dependent variable is the fluid intelligence score, standardized to have mean zero and standard deviation one. PM2.5 is the average daily concentration of PM2.5 and is in $\mu g/m^3$. Robust standard errors clustered at the assessment center level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

5.2 Instrumental variables

The validity of my instrument relies on whether the instrument sufficiently predicts air pollution concentration (relevance), that the instrument is as good as randomly assigned (exogeneity), and that it only affects the outcome through air pollution (exclusion restriction). Figure 2 plots the relationship between inversion coverage and PM2.5, before including any controls. There is a strong and positive relationship between PM2.5 concentration and whether inversion occurs during the day. To formalize this relationship, Table 2 reports the first-stage estimates of equation (3), where I regress the average concentration of PM2.5 during the assessment day on the occurrence of thermal inversions. The estimated coefficient is positive and statistically significant. A complete inversion coverage in the area around an assessment center raises the average concentration of PM2.5 by 2.5 $\mu g/m^3$, or about a third of its standard deviation (Table A1 in the Appendix). This shows that inversions are a strong predictor of air pollution concentration on the ground and aligns with the theoretical understanding of the pollutant-trapping effects of thermal inversions (Jacobson, 2002).

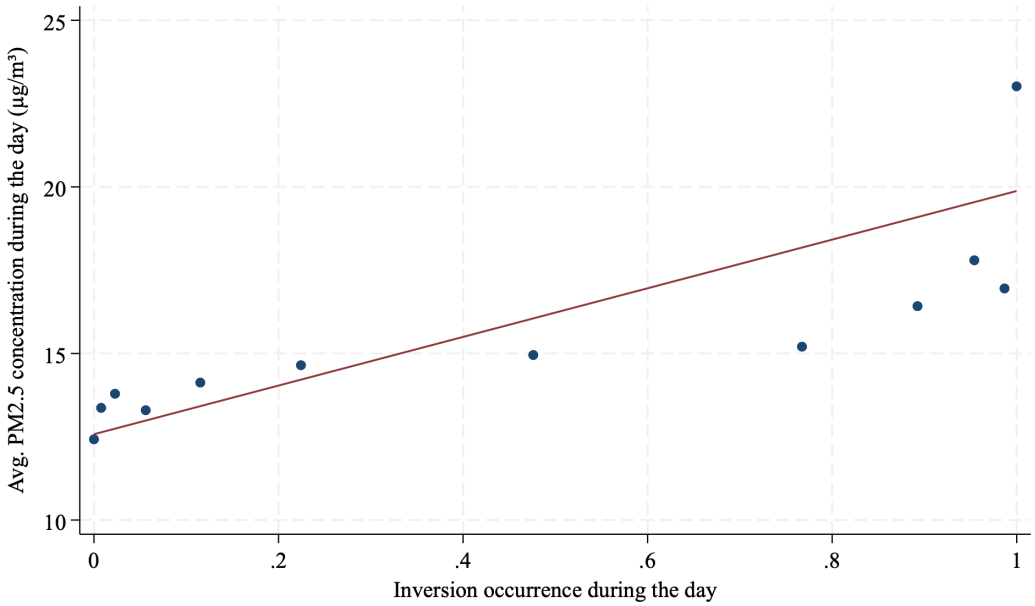


Figure 2: Binned scatter plot between inversion coverage and PM2.5

Table 2: First-stage estimates

	PM2.5 (1)
Inversion coverage	2.532***
	(0.2888)
Partial R^2	0.04571
F-statistic	76.864
Observations	125,224
Weather controls	✓
Fixed effects	✓

Notes: Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center on a given day, and can range from zero to one. PM2.5 is the average daily concentration of PM2.5 and is in $\mu g/m^3$. The first-stage F-statistic is a cluster-robust Wald test of the null hypothesis that the excluded instrument has a zero coefficient. Robust standard errors clustered at the assessment center level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

To gauge the exogeneity of my thermal inversion instrument, I conduct two exercises. First, I regress a host of pre-determined variables, including PGS EA, sex, body mass index (BMI), birth year, household size, and educational attainment on my instrument. Since these variables are pre-determined prior to the assessment and/or should not be affected over the time span of my data, any statistically significant relationship between these variables and thermal inversions would

suggest that the variation in inversions, even after controlling for the weather and my set of fixed effects, are not entirely random. Table A8 show that the estimated effects of inversion coverage on these pre-determined variables are small and close to zero, indicating that there is no relationship between these variables and inversion coverage. In Table A3 of the Appendix, I additionally regress predetermined objective health variables, including prior diagnoses of respiratory and cardiovascular diseases, on thermal inversions and similarly do not find any effects of thermal inversions on these outcomes.

Second, as a falsification check, I regress fluid intelligence scores on the one-day lead of thermal inversion coverage. If my set of weather controls and fixed effects manage to tease out the exogenous variation in contemporaneous inversions, I should not observe any relationship between fluid intelligence scores done on the assessment day and inversion coverage on the day *after*. Table 4 shows that indeed there is no relationship between FI scores and the lead of thermal inversion coverage.

The third assumption refers to the exclusion restriction underlying the thermal inversion instrument. Thermal inversions affect air quality by trapping pollutants close to the ground, but this mechanism is not specific to PM2.5 alone. Inversion events simultaneously increase concentrations of multiple co-emitted pollutants, such as NO_2 and other fine particulates. As a result, the exclusion restriction that thermal inversions affect cognitive performance solely through PM2.5 may be violated if these other pollutants independently influence fluid intelligence. For this reason, while I present two-stage least squares estimates that instrument PM2.5 using thermal inversions to facilitate comparison with the existing literature, I do not interpret these estimates as isolating the effect of PM2.5 alone. Moreover, I also present reduced-form estimates that relate thermal inversion coverage directly to fluid intelligence outcomes. These reduced-form results capture the aggregate cognitive effect of air pollution induced by inversion events without assumptions on the specific contribution of individual pollutants.

Table 3: Exogeneity assessment

	PGS EA	Sex	BMI	Birth year	HH size	Education years
	(1)	(2)	(3)	(4)	(5)	(6)
Inversion coverage	-0.0335	0.0102	0.0377	0.1921	-0.0204	-0.0580
	(0.0236)	(0.0076)	(0.0963)	(0.1095)	(0.0128)	(0.0372)
R^2	0.01485	0.00116	0.01018	0.01946	0.00590	0.02664
Observations	125,224	125,224	124,762	125,224	125,224	125,224
Weather controls	✓	✓	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓	✓	✓

Notes: The set of fixed effects include center, month, year, and day-of-week fixed effects. The set of weather controls include ground temperature, dew temperature, wind speed, wind direction, cloud coverage, and total precipitation, as well as their squared and cubed terms. Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center on a given day, and can range from zero to one. Robust standard errors clustered at the assessment center level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 4: Falsification test

	Fluid intelligence scores
	(1)
Inversion coverage lead	0.0134
	(0.0116)
R^2	0.01277
Observations	125,217
Weather controls	✓
Fixed effects	✓

Notes: The set of fixed effects include center, month, year, and day-of-week fixed effects. The set of weather controls include ground temperature, dew temperature, wind speed, wind direction, cloud coverage, and total precipitation, as well as their squared and cubed terms. The dependent variable is the fluid intelligence score, standardized to have mean zero and standard deviation one. Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center during the day *after the assessment day*, and can range from zero to one. Robust standard errors clustered at the assessment center level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

A remaining concern is that participation in the fluid intelligence module is voluntary, i.e, participants can skip this module and move on to the next section during the UK Biobank baseline assessment. If thermal inversions induce selective participation into this module, the reduced-form estimate may partly reflect changes in the composition of test takers rather than the effect of thermal inversions on cognition. To examine whether this is the case, I regress whether a participant takes

the fluid intelligence test on thermal inversion and present the results in Table 5. The coefficient is positive but not statistically significant at conventional levels. I therefore do not find evidence that thermal inversions induce selective participation into the fluid intelligence test.

Table 5: Effects of thermal inversion on test taking

	Takes FI test (1)
Inversion coverage	0.0275 (0.0177)
R^2	0.63005
Observations	150,787
Weather controls	✓
Fixed effects	✓

Notes: The set of fixed effects include center, month, year, and day-of-week fixed effects. The set of weather controls include ground temperature, dew temperature, wind speed, wind direction, cloud coverage, and total precipitation, as well as their squared and cubed terms. Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center on a given day, and can range from zero to one. The outcome is 1 if participant decided to take the fluid intelligence test, and 0 if they did not. The sample is restricted to only participants who were offered the fluid intelligence test. Robust standard errors clustered at the assessment center level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

5.3 Reduced form and 2SLS

Table 6 presents the estimated effects of air pollution and thermal inversion coverage on fluid intelligence scores. Column (1) reports the reduced-form relationship between thermal inversion coverage and fluid intelligence. The coefficient on inversion coverage is negative and statistically significant, indicating that larger inversion coverage reduces fluid intelligence performance. Specifically, fluid intelligence scores decrease by about 0.02 standard deviations when inversion coverage around the assessment center increases from zero to full coverage[3]. Column (2) reports the 2SLS estimates in which PM2.5 is instrumented using thermal inversion coverage. As discussed previously, 2SLS estimate may incorporate the cognitive effects of other pollutants trapped by inversions and should not be interpreted as isolating the causal effect of PM2.5 alone. My 2SLS estimate shows a decline of approximately 0.10 standard deviations per 10 μg/m^3 increase in PM2.5, similar in magnitude to the 0.08-standard-deviation decline per 10 μg/m^3 increase in PM10 reported by Carneiro et al. (2021) for Brazilian university entrance examinations. It is larger than the corresponding PM2.5 estimates of 0.033 standard deviations for mathematical subjects in Finnish matriculation examina-

[3]Here, zero coverage means that no grid point around the assessment center observed an inversion during the day, and full coverage means that every grid point observed an inversion at some point during the day

tions (Koivuranta et al., 2024) and approximately 0.049 standard deviations for adult brain-training performance (La Nauze and Severnini, 2025).

Table 6: Effects of air pollution/thermal inversion on FI scores

	Fluid intelligence scores	
	(1)	(2)
Inversion coverage	-0.0248**	
	(0.0108)	
PM2.5		-0.0098**
		(0.0040)
R^2	0.01281	0.01152
Observations	125,224	125,224
Weather controls	✓	✓
Fixed effects	✓	✓

Notes: The set of fixed effects include center, month, year, and day-of-week fixed effects. The set of weather controls include ground temperature, dew temperature, wind speed, wind direction, cloud coverage, and total precipitation, as well as their squared and cubed terms. Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center on a given day, and can range from zero to one. PM2.5 is the average daily concentration of PM2.5 and is in $\mu g/m^3$. The dependent variable is the fluid intelligence score, standardized to have mean zero and standard deviation one. Robust standard errors clustered at the assessment center level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 7 shows the 2SLS (column 1) and reduced-form estimates (column 2) for the effects of air pollution on fluid intelligence, including their interactions with PGS for educational attainment. The main effects of PM2.5 and inversion coverage remains negative and statistically significant. These effects, however, do not seem to differ between individuals with low PGS EA versus those with high PGS EA. The interaction terms in both columns are insignificant and close to zero. Previous findings have indicated that air pollution can alter the returns to endowments over time. von Hinke and Sørensen (2023), for example, find evidence that exposure to the London Smog before birth or in early childhood reduced educational attainment more among individuals with higher polygenic scores for education. Our paper shows that, at least in the short term, air pollution affects everyone across different genetic propensity similarly.

Likewise, previous findings have indicated that those with lower cognitive ability, as measured by test scores, tend to suffer more from air pollution (Carneiro et al., 2021; La Nauze and Severnini, 2025). That we found no differential effects across genetic endowments suggest that this additional vulnerability to air pollution is not an inherent feature of their endowments. Instead, it suggests that the heightened sensitivity observed in lower-ability populations may itself be a product of the environment they were exposed to, rather than of a fixed endowment determined at the point of conception.

Table 7: Effects of air pollution/thermal inversions on FI scores
Heterogeneity by PGS EA

	Fluid intelligence scores	
	(1)	(2)
Inversion coverage	-0.0199*	
	(0.0103)	
Inversion coverage × PGS EA	0.0018	
	(0.0063)	
PM2.5		-0.0071*
		(0.0039)
PM2.5 × PGS EA		0.0003
		(0.0009)
R^2	0.05627	0.04982
Observations	125,224	125,224
Weather controls	✓	✓
Fixed effects	✓	✓

Notes: The set of fixed effects include center, month, year, and day-of-week fixed effects. The set of weather controls include ground temperature, dew temperature, wind speed, wind direction, cloud coverage, and total precipitation, as well as their squared and cubed terms. Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center on a given day, and can range from zero to one. PM2.5 is the average daily concentration of PM2.5 and is in $\mu g/m^3$. The dependent variable is the fluid intelligence score, standardized to have mean zero and standard deviation one. PGS EA is standardized at 0 with standard deviation 1. Robust standard errors clustered at the assessment center level are reported in parentheses.

5.4 Heterogeneity of air pollution effects by demographic and health characteristics

The first four rows in figure 3 shows whether the reduced-form effect of thermal inversions on standardized fluid intelligence varies across demographic characteristics. Overall, more educated individuals are more resilient to acute pollution shocks - for each year of additional education, the effects of inversions on fluid intelligence is attenuated by 0.009 standard deviation (see Table A5 in the Appendix). In contrast, the effects of air pollution remains consistent regardless of age, sex, and BMI. To the extent that educational attainment relates to test-taking performance and cognitive ability, this result echoes those of Carneiro et al. (2021) and La Nauze and Severnini (2025), which found that individuals with lower ability saw larger air pollution-induced performance losses in exams and brain training tasks. Combined with the evidence in section 5.3 on the limited moderating effects of PGS EA, my results here suggests that this ability-gradient in vulnerability to air pollution likely operates through acquired human capital rather than predetermined genetic endowment.

The last four rows in figure 3 present the interaction terms between exposure to inversion and health characteristics. Interestingly, the effect of air pollution on cognitive performance seems to be *less* pronounced among those who were hospitalized with a respiratory and/or cardiovascular diagnosis, although we are underpowered to detect these interactions with precision. Similarly, the interaction terms on subjective health and chronic conditions are not significantly different from zero.

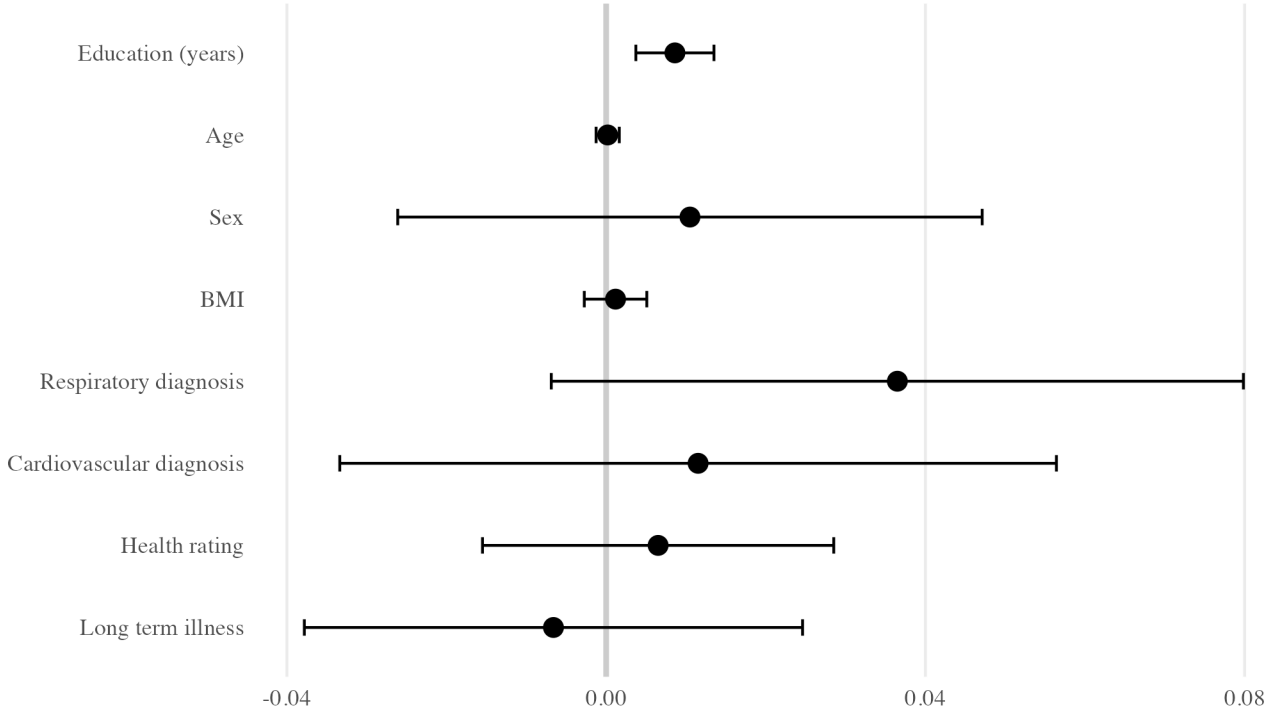


Figure 3: Heterogeneity by demographic and health characteristics

The figure plots the coefficient point estimate and the 95% confidence interval of the interaction term between inversion coverage and each demographic/health variable. All estimations include weather controls and year, month, and day-of-week fixed effects. Respiratory diagnosis is equal to 1 if a participant has ever been hospitalized for a respiratory disease. Cardiovascular diagnosis is equal to 1 if a participant has ever been hospitalized for a cardiovascular disease. Health rating is the subjective health rating participants gave during the baseline, and range from 1 to 4, with 1 being excellent health and 4 being poor health. Long term illness is equal to 1 if a participant has a long term illness. Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center on a given day, and can range from zero to one. The dependent variable is the fluid intelligence score, standardized to have mean zero and standard deviation one. Robust standard errors clustered at the assessment center level.

5.5 Effects on other cognitive domain

Having established that air pollution negatively affects reasoning-related performance, captured by fluid intelligence, I next examine whether this effect extends to other cognitive outcomes measured in the UK Biobank, namely memory and reaction. In the UK Biobank, memory and reaction tests

capture cognitive domains distinct from those in the fluid intelligence module and thus provide complementary evidence on whether air pollution affects cognitive performance more broadly, or whether the effects are concentrated in reasoning-related performance.

Table 8: Effects of thermal inversion on other cognitive domains

	Visual memory (1)	Prospective memory (2)	Numeric memory (3)	Reaction time (4)
Inversion coverage	-0.0038 (0.0032)	0.0025 (0.0036)	-0.0670*** (0.0232)	-0.5379 (0.5867)
R^2	0.00246	0.00639	0.00910	0.00761
Observations	391,064	127,635	40,381	388,795
Weather controls	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓

Notes: The set of fixed effects include center, month, year, and day-of-week fixed effects. The set of weather controls include ground temperature, dew temperature, wind speed, wind direction, cloud coverage, and total precipitation, as well as their squared and cubed terms. Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center on a given day, and can range from zero to one. Note that the outcomes are not yet standardized. Standardization is pending reopening of the UK Biobank Research Analysis Platform. Coefficient magnitudes are therefore not directly comparable across outcomes. Robust standard errors clustered at the location level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 8 reports reduced-form estimates of the effect of thermal inversion coverage on the memory and reaction time tests in the UK Biobank. The estimates provide mixed evidence that inversion-induced pollution affects cognitive domains beyond fluid intelligence. I find no statistically significant effect on visual memory, prospective memory, or reaction time. In contrast, inversion coverage significantly reduces numeric memory performance by 0.067 points. This pattern suggests that the cognitive costs of short-run pollution exposure are not uniform across all measured domains.

The memory findings differ across studies. Whereas I find an adverse effect on numeric memory but no statistically detectable effect on visual memory, La Nauze and Severnini (2025) report adverse effects on a visuospatial working-memory task measuring spatial recall and visual memory span. Bedi et al. (2021), meanwhile, find no statistically detectable effect on working memory, although their estimates cannot rule out modest effects. These mixed findings suggest that further research using comparable measures of different forms of memory is needed to establish more conclusively which aspects of memory are affected by short-run air pollution exposure and whether these effects persist across populations and settings.

6 Sensitivity analyses

Empirical estimates of the causal effects of air pollution are often sensitive to modeling choices, including the definition of exposure, the choice of instrument, the set of control variables, and the spatial and temporal aggregation of data (Bagilet, 2026; Pannullo et al., 2016). Given the high dimensionality of plausible specifications in this setting, reporting results from a single preferred model risks overstating precision and masking the extent to which findings depend on particular researcher decisions. To address this concern, I conduct a comprehensive specification-curve analysis following the approach proposed by Simonsohn et al. (2020). Specifically, I consider possible combinations of reasonable sets of weather controls (5 choices), fixed effects (4 choices), inversion variables (4 choices), air pollution types (3 choices), and spatial aggregation methods (3 choices), resulting in 660 unique specifications for estimating the effects of air pollution on cognitive performance[4]. I briefly explain the possible choices in each set below.

Controls

Weather controls are necessary since the weather, e.g., temperature, may also affect cognitive performance independent of air pollution. Prior studies have incorporated weather controls flexibly, including high-order terms, as well as bins, to capture non-linear effects of the weather. However, since thermal inversion by itself is also a meteorological phenomenon driven by the weather, very flexible weather controls may also lead to useful exogenous variation being absorbed, leaving the thermal inversion variable to mostly capture noise and its standard error highly inflated. Therefore, besides controlling for linear and higher-order terms of the weather variables (the chosen specification), I also estimated regressions in which I only include (1) only temperature controls, with linear, quadratic, and cubic terms, (2) temperature controls with linear and higher-order terms, and dew temperature, cloud coverage, total precipitation, wind speed, wind direction, only in linear terms, (3) temperature with linear and high-order terms and other weather controls, only with linear and quadratic terms (no cubic term). These choices represent less saturated weather specifications. Ex-ante, they may result in more precise estimates, but potentially at a cost of higher bias. The specification curve therefore is useful to decipher the trade-offs associated with the choice of weather controls. Additionally, I included specifications without any weather control to examine how estimates behave without controlling for weather confounders.

Fixed Effects

The choice of fixed effects affect the source of identifying variation. My preferred specification consist of year, month, weekday, and center fixed effects. However, stricter fixed effects can be specified, for

[4]There are fewer specifications than 720, which represents the total number of possible combinations, since some combination of these choices result in specifications that cannot be estimated. For example, inverse distance weighting with a maximum radius of 20km results in a substantial sample size reduction, making models with very fine fixed effects difficult to estimate.

example, by interacting month and year fixed effects (year-month), or even day, month, and year (day-month-year). While this eliminates potential remaining biases due to common unobserved shocks across time, it may also dramatically shrink effective variation and attenuate estimates if pollution shocks are highly spatially correlated, since it restricts identification to cross-location differences within the same calendar day. One can also interact center fixed effects with month-year fixed effects to control for local weather patterns, which can substantially decrease potential biases from time-varying heterogeneity across locations. However, identification here would only come from short-run local fluctuations, which may be very noisy. Given these trade-offs, I included four different sets of fixed effects to examine the robustness of the results to these choices.

Thermal inversion operationalization

I include different operationalizations of thermal inversions that have been used so far in the literature. While my preferred specification uses inversion coverage, which is equal to one if inversion occurs during the day, other papers have also used the number of inversions in a day, or inversion strength, which is the temperature difference between ground temperature and temperature. I thus include these operationalizations in my robustness check.

Air pollution measure

I estimate models using alternative particulate measures (PM1, PM2.5, and PM10) to examine whether results depend on the specific operationalization of particulate exposure. Given the high correlation across particle size fractions and the fact that meteorological shocks typically shift them jointly, similar estimates across measures should be interpreted as reflecting the common component of ambient particulate variation rather than distinct size-specific causal effects.

Spatial Aggregation

I consider alternative spatial matching procedures to assign air pollution and weather exposures to assessment centers, comparing inverse distance weighting (IDW) with a 60 km maximum radius, IDW with a 20 km maximum radius, and a nearest-neighbor (NN) approach within a 60 km radius. Using IDW with a 60 km radius smooths local measurement error by averaging across multiple grid points and reduces sensitivity to idiosyncratic grid point noise; however, it may attenuate spatial heterogeneity by incorporating observations from relatively distant locations, potentially introducing exposure misclassification if pollution gradients are steep. Restricting the IDW radius to 20 km increases spatial precision and better captures local variation, but at the cost of greater sensitivity to measurement error and missingness. The nearest-neighbor approach increases local specificity and avoids spatial smoothing, yet it is again highly sensitive to individual grid point noise.

Results

Figure 4 shows the specification curves for 2SLS estimates of air pollution on fluid intelligence scores. Each point in Panel A of figure 4 corresponds to a two-stage least squares estimate obtained from

a distinct combination of reasonable modeling and data choices mentioned above. The estimates are ordered from left to right by magnitude, from the most negative to the most positive. Panel B displays the underlying specification choices associated with each estimate, making explicit how changes in modeling decisions map into variation in the estimated coefficients. The specification curves thus aims to characterize the robustness of the findings across defensible specifications.

Across the vast majority (90%) of specifications , the estimated effect of air pollution is negative. About half of these negative estimates are statistically significant, with point estimates clustered within a relatively narrow range. The negative association is robust to alternative sets of weather controls, including flexible temperature polynomials and comprehensive meteorological adjustments. Results are also stable across alternative fixed-effects structures, although the most restrictive specifications (e.g., day-month-year fixed effects) yield somewhat attenuated and less precisely estimated coefficients, consistent with the reduced identifying variation under those designs. With respect to the instrument definition, estimates remain broadly similar whether thermal inversions are captured through counts, binary indicators, temperature differences, or inversion strength, indicating that findings are not driven by a particular operationalization of the first stage. Likewise, the choice of particulate measure (PM1, PM2.5, PM10) produces highly comparable estimates. Finally, alternative spatial aggregation procedures (nearest neighbor versus inverse distance weighting with different radii) yield qualitatively consistent effects, though the smaller radius produces slightly more dispersed estimates, reflecting the trade-off between spatial precision and measurement noise. Overall, the specification curve indicates that the negative effect of particulate exposure on cognitive performance is robust across a wide range of defensible modeling choices, with variation in magnitude primarily reflecting differences in identifying variation and precision rather than reversals in sign.

In the Appendix, I present specification curves for the reduced form estimates, F-statistics, and estimates for the interaction term. Overall, the qualitative results remain the same to those produced by my preferred specification presented in the main results, across different specification and data choices.

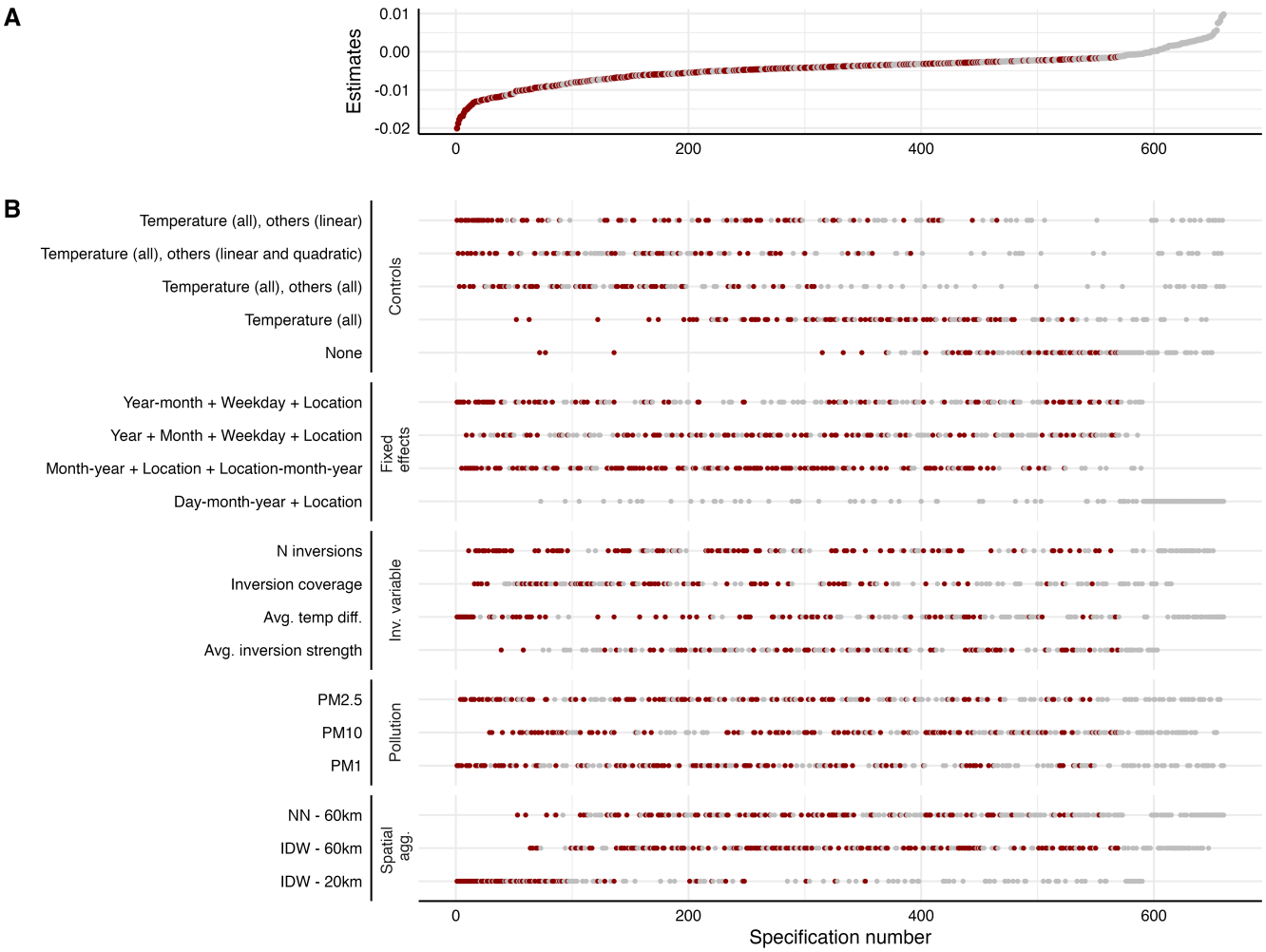


Figure 4: Specification curves for 2SLS estimates of air pollution on fluid intelligence scores

Significant point estimates at the 95% level are colored in red. Regressions in this figure are estimated without polygenic scores or its interactions. In the **Controls** set, *Temperature (all)* denotes linear, quadratic, and cubic terms for ground temperature. *Others (all)* denotes linear, quadratic, and cubic terms for all other weather controls, which are dew temperature, cloud coverage, total precipitation, wind speed, and wind direction. *None* refers to specification without any weather controls, and is included to examine how estimates behave without controlling for weather confounders. In the **Inv. variables** set, *N inversions* denote the spatially weighted daily inversion frequency in the area surrounding an assessment center; *Inversion coverage* indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center on a given day; *Avg. temp diff.* denotes the difference between temperature at the upper layer and the ground temperature (where a positive difference implies a thermal inversion), and *Avg. inversion strength* is the same as *Avg. temp diff.*, except that all negative differences are truncated at 0. In the **Spatial agg.** set, *NN* refers to the nearest neighbor method for assigning air pollution concentration to assessment centers, and *IDW* refers to the inverse distance weighted matching scheme.

7 Conclusion

This paper studies the short-run effects of air pollution on cognitive performance and examines whether these effects differ systematically by genetic endowments. Using high-frequency environmental data from Copernicus merged with the UK Biobank, I find that acute exposure to fine particulate matter ($PM_{2.5}$) significantly impairs fluid intelligence performance on the day of testing. Even in the relatively low-pollution setting of the United Kingdom, these results suggest that air pollution acts as a contemporaneous tax on the cognitive inputs that drive modern economic productivity.

A central challenge in this literature is the potential endogeneity of ambient pollution measures. While standard fixed-effects specifications yield small and statistically insignificant estimates, leveraging exogenous variation in pollution induced by thermal inversions reveals clear negative effects. These findings align with a growing body of causal evidence documenting immediate cognitive impairments from short-term exposure, underscoring that even modest, transient increases in air pollution have measurable consequences cognitive performance.

The analysis of other cognitive domains provides additional evidence of an adverse effect on numeric memory. By contrast, I find no statistically significant effects on visual memory, prospective memory, or reaction time. These results extend the evidence of cognitive impairment beyond fluid intelligence, although they do not establish a broader decline across all measured domains.

To address concerns about specification sensitivity, I implement a comprehensive specification-curve analysis that systematically varies modeling choices, including pollution measures, instruments, fixed effects, weather controls, and spatial aggregation methods. Across this broad set of plausible specifications, the vast majority of estimates indicate a negative relationship between air pollution and cognitive performance. While statistical significance varies with precision, the direction of the effect is remarkably stable, reinforcing confidence in the robustness of the main findings.

A key contribution of this study is the exploration of whether air pollution widens or compresses existing cognitive inequalities, for example, by disproportionately penalizing those with lower endowment. I find no evidence of such heterogeneity; the interaction between pollution shocks and genetic propensity is small and statistically insignificant. This absence of differential effects suggests that the cognitive penalty of $PM_{2.5}$ is a universal burden. Regardless of one's genetic propensity or latent potential for scholastic achievement, the immediate impact on cognition appears remarkably uniform.

Taken together, these results provide a strong argument for air pollution mitigation even in high-income countries with moderate air quality. The welfare benefits of cleaner air extend beyond traditional health metrics and include gains in cognitive performance and thus productivity. A relevant avenue for future research is to investigate whether differential effects across genetic

endowments emerge when looking at longer exposure, and if so, how such heterogeneity arises. This would not only help inform policies aimed at mitigating the long-term consequences of environmental exposure, but also, combined with the null contemporaneous interaction documented in this paper, help clarify the conditions under which gene-environment interactions in cognitive production emerge over time.

References

Almond, D., Currie, J., & Duque, V. (2018). Childhood Circumstances and Adult Outcomes: Act II. *Journal of Economic Literature*, *56*(4), 1360–1446. https://doi.org/10.1257/jel.20171164

Arceo, E., Hanna, R., & Oliva, P. (2016). Does the Effect of Pollution on Infant Mortality Differ Between Developing and Developed Countries? Evidence from Mexico City. *The Economic Journal*, *126*(591), 257–280. https://doi.org/10.1111/ecoj.12273

Archsmith, J., Heyes, A., & Saberian, S. (2018). Air Quality and Error Quantity: Pollution and Performance in a High-Skilled, Quality-Focused Occupation. *Journal of the Association of Environmental and Resource Economists*, *5*(4), 827–863. https://doi.org/10.1086/698728

Auffhammer, M., Hsiang, S. M., Schlenker, W., & Sobel, A. (2013). Using Weather Data and Climate Model Output in Economic Analyses of Climate Change. *Review of Environmental Economics and Policy*, *7*(2), 181–198. https://doi.org/10.1093/reep/ret016

Bagilet, V. (2026, May). Accurate Estimation of Small Effects: Illustration Through Air Pollution and Health. https://doi.org/10.2139/ssrn.6758998

Baker, S., Biroli, P., van Kippersluis, H., & von Hinke, S. (2024). Advantageous early-life environments cushion the genetic risk for ischemic heart disease. *Proceedings of the National Academy of Sciences of the United States of America*, *121*(27), e2314056121. https://doi.org/10.1073/pnas.2314056121

Barbey, A. K. (2018). Network Neuroscience Theory of Human Intelligence. *Trends in Cognitive Sciences*, *22*(1), 8–20. https://doi.org/10.1016/j.tics.2017.10.001

Barcellos, S. H., Carvalho, L., & Turley, P. (2021, May). The Effect of Education on the Relationship between Genetics, Early-Life Disadvantages, and Later-Life SES. https://doi.org/10.3386/w28750

Barcellos, S. H., Carvalho, L. S., & Turley, P. (2018). Education can reduce health differences related to genetic risk of obesity. *Proceedings of the National Academy of Sciences*, *115*(42), E9765–E9772. https://doi.org/10.1073/pnas.1802909115

Bedi, A. S., Nakaguma, M. Y., Restrepo, B. J., & Rieger, M. (2021). Particle Pollution and Cognition: Evidence from Sensitive Cognitive Tests in Brazil. *Journal of the Association of Environmental and Resource Economists*, *8*(3), 443–474. https://doi.org/10.1086/711592

Biroli, P., Galama, T., von Hinke, S., van Kippersluis, H., Rietveld, C. A., & Thom, K. (2026). The Economics and Econometrics of Gene–Environment Interplay. *The Review of Economic Studies*, *93*(1), 144–180. https://doi.org/10.1093/restud/rdaf034

Block, M. L., & Calderón-Garcidueñas, L. (2009). Air pollution: Mechanisms of neuroinflammation and CNS disease. *Trends in Neurosciences*, *32*(9), 506–516. https://doi.org/10.1016/j.tins.2009.05.009

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., & Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203–209. https://doi.org/10.1038/s41586-018-0579-z

Carneiro, J., Cole, M. A., & Strobl, E. (2021). The Effects of Air Pollution on Students' Cognitive Performance: Evidence from Brazilian University Entrance Tests. *Journal of the Association of Environmental and Resource Economists*, *8*(6), 1051–1077. https://doi.org/10.1086/714671

Chang, T. Y., Graff Zivin, J., Gross, T., & Neidell, M. (2019). The Effect of Pollution on Worker Productivity: Evidence from Call Center Workers in China. *American Economic Journal: Applied Economics*, *11*(1), 151–172. https://doi.org/10.1257/app.20160436

Chen, S., Oliva, P., & Zhang, P. (2022). The effect of air pollution on migration: Evidence from China. *Journal of Development Economics*, *156*, 102833. https://doi.org/10.1016/j.jdeveco.2022.102833

Chen, X., Chen, L., Xie, W., Mueller, N. D., & Davis, S. J. (2023). Flight delays due to air pollution in China. *Journal of Environmental Economics and Management*, *119*, 102810. https://doi.org/10.1016/j.jeem.2023.102810

Clay, K., Lewis, J., & Severnini, E. (2018). Pollution, Infectious Disease, and Mortality: Evidence from the 1918 Spanish Influenza Pandemic. *The Journal of Economic History*, *78*(4), 1179–1209. https://doi.org/10.1017/S002205071800058X

Cook, N., Heyes, A., & Rivers, N. (2023). Clean Air and Cognitive Productivity: Effect and Adaptation. *Journal of the Association of Environmental and Resource Economists*, *10*(5), 1265–1308. https://doi.org/10.1086/724951

Copernicus Climate Change Service. (2018a). ERA5 hourly data on pressure levels from 1940 to present. https://doi.org/10.24381/CDS.BD0915C6

Copernicus Climate Change Service. (2018b). ERA5 hourly data on single levels from 1940 to present. https://doi.org/10.24381/CDS.ADBB2D47

Cunha, F., & Heckman, J. (2007). The Technology of Skill Formation. *American Economic Review*, *97*(2), 31–47. https://doi.org/10.1257/aer.97.2.31

de Souza, E. A., Silva, S. A., Vieira, B. H., & Salmon, C. E. G. (2023). fMRI functional connectivity is a better predictor of general intelligence than cortical morphometric features and ICA parcellation order affects predictive performance. *Intelligence*, *97*, 101727. https://doi.org/10.1016/j.intell.2023.101727

Deryugina, T., Heutel, G., Miller, N. H., Molitor, D., & Reif, J. (2019). The Mortality and Medical Costs of Air Pollution: Evidence from Changes in Wind Direction. *American Economic Review*, *109*(12), 4178–4219. https://doi.org/10.1257/aer.20180279

Ebenstein, A., Lavy, V., & Roth, S. (2016). The Long-Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution. *American Economic Journal: Applied Economics*, *8*(4), 36–65. https://doi.org/10.1257/app.20150213

Elder, A., Gelein, R., Silva, V., Feikert, T., Opanashuk, L., Carter, J., Potter, R., Maynard, A., Ito, Y., Finkelstein, J., & Oberdörster, G. (2006). Translocation of Inhaled Ultrafine Manganese Oxide Particles to the Central Nervous System. *Environmental Health Perspectives*, *114*(8), 1172–1178. https://doi.org/10.1289/ehp.9030

Faherty, T., Raymond, J. E., McFiggans, G., & Pope, F. D. (2025). Acute particulate matter exposure diminishes executive cognitive functioning after four hours regardless of inhalation pathway. *Nature Communications*, *16*(1), 1339. https://doi.org/10.1038/s41467-025-56508-3

Fawns-Ritchie, C., & Deary, I. J. (2020). Reliability and validity of the UK Biobank cognitive tests. *PLoS ONE*, *15*(4), e0231627. https://doi.org/10.1371/journal.pone.0231627

Frampton, M. W., Pietropaoli, A., Dentler, M., Chalupa, D., Little, E. L., Stewart, J., Frasier, L., Oakes, D., Wiltshire, J., Vora, R., & Utell, M. J. (2015). Cardiovascular effects of ozone in healthy subjects with and without deletion of glutathione-S-transferase M1. *Inhalation Toxicology*, *27*(2), 113–119. https://doi.org/10.3109/08958378.2014.996272

Franz, C. E., Gustavson, D. E., Elman, J. A., Fennema-Notestine, C., Hagler, D. J., Baraff, A., Tu, X. M., Wu, T.-C., De Anda, J., Beck, A., Kaufman, J. D., Whitsel, N., Finch, C. E., Chen, J.-C., Lyons, M. J., & Kremen, W. S. (2023). Associations Between Ambient Air Pollution and Cognitive Abilities from Midlife to Early Old Age: Modification by APOE Genotype. *Journal of Alzheimer's disease: JAD*, *93*(1), 193–209. https://doi.org/10.3233/JAD-221054

Fukushima, N., Hinke, S. v., & Sørensen, E. N. (2024, September). The long-term human capital and health impacts of a pollution reduction programme [arXiv:2409.11839 [econ.GN]]. https://doi.org/10.48550/arXiv.2409.11839

Gawryluk, J. R., Palombo, D. J., Curran, J., Parker, A., & Carlsten, C. (2023). Brief diesel exhaust exposure acutely impairs functional brain connectivity in humans: A randomized controlled crossover study. *Environmental Health*, *22*(1), 7. https://doi.org/10.1186/s12940-023-00961-4

Heckman, J. J. (2007). The economics, technology, and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences*, *104*(33), 13250–13255. https://doi.org/10.1073/pnas.0701362104

Hoek, G., Krishnan, R. M., Beelen, R., Peters, A., Ostro, B., Brunekreef, B., & Kaufman, J. D. (2013). Long-term air pollution exposure and cardio- respiratory mortality: A review. *Environmental Health*, *12*(1), 43. https://doi.org/10.1186/1476-069X-12-43

Holub, F., & Thies, B. (2023, March). Air Quality, High-Skilled Worker Productivity and Adaptation: Evidence from Github. https://doi.org/10.2139/ssrn.4398088

Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V.-H., Razinger, M., Remy, S., Schulz, M., & Suttie, M. (2019). The CAMS reanalysis of atmospheric composition. *Atmospheric Chemistry and Physics*, *19*(6), 3515–3556. https://doi.org/10.5194/acp-19-3515-2019

Jacobson, M. Z. (2002). *Atmospheric Pollution: History, Science, and Regulation*. Cambridge University Press. https://doi.org/10.1017/CBO9780511802287

Jans, J., Johansson, P., & Nilsson, J. P. (2018). Economic status, air quality, and child health: Evidence from inversion episodes. *Journal of Health Economics*, *61*, 220–232. https://doi.org/10.1016/j.jhealeco.2018.08.002

Josephson, A., Michler, J. D., Kilic, T., & Murray, S. (2026). The mismeasure of weather: Using earth observation data for estimation of socioeconomic outcomes. *Journal of Development Economics*, *178*, 103553. https://doi.org/10.1016/j.jdeveco.2025.103553

Karimi, B., & Samadi, S. (2024). Long-term exposure to air pollution on cardio-respiratory, and lung cancer mortality: A systematic review and meta-analysis. *Journal of Environmental Health Science and Engineering*, *22*(1), 75–95. https://doi.org/10.1007/s40201-024-00900-6

Koivuranta, M., Korhonen, M., & Lehto, J. (2024). The impact of ambient PM2.5 air pollution on student performance: Evidence from finnish matriculation examinations. *Economics Letters*, *235*, 111542. https://doi.org/10.1016/j.econlet.2024.111542

Künn, S., Palacios, J., & Pestel, N. (2023). Indoor air quality and strategic decision making. *Management Science*, *69*(9), 5354–5377. https://doi.org/10.1287/mnsc.2022.4643

La Nauze, A., & Severnini, E. (2025). Air Pollution and Adult Cognition: Evidence from Brain Training. *Journal of the Association of Environmental and Resource Economists*, *12*(1), 221–255. https://doi.org/10.1086/730390

Laumbach, R. J., Kipen, H. M., Kelly-McNeil, K., Zhang, J., Zhang, L., Lioy, P. J., Ohman-Strickland, P., Gong, J., Kusnecov, A., & Fiedler, N. (2011). Sickness Response Symptoms among Healthy Volunteers after Controlled Exposures to Diesel Exhaust and Psychological Stress. *Environmental Health Perspectives*, *119*(7), 945–950. https://doi.org/10.1289/ehp.1002631

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T. A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., Fontana, M. A., Kundu, T., Lee, C., Li, H., Li, R., Royer, R., Timshel, P. N., Walters, R. K., Willoughby, E. A., … Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics*, *50*(8), 1112–1121. https://doi.org/10.1038/s41588-018-0147-3

Lombardo, D., & Kaufmann, T. (2023). Different patterns of intrinsic functional connectivity at the default mode and attentional networks predict crystalized and fluid abilities in childhood. *Cerebral Cortex Communications*, *4*(3), tgad015. https://doi.org/10.1093/texcom/tgad015

Molina, T. (2021). Pollution, Ability, and Gender-Specific Investment Responses to Shocks. *Journal of the European Economic Association*, *19*(1), 580–619. https://doi.org/10.1093/jeea/jvaa005

Muslimova, D., Dias Pereira, R., von Hinke, S., van Kippersluis, H., Rietveld, C. A., & Meddens, S. F. W. (2023). Rank concordance of polygenic indices [Number: 5]. *Nature Human Behaviour*, *7*(5), 802–811. https://doi.org/10.1038/s41562-023-01544-6

Muslimova, D., van, K. H., Rietveld, C. A., von, H. S., & Meddens, S. F. W. (2026). Gene-Environment Complementarity in Educational Attainment. *Journal of Labor Economics*, *0*(0), 000–000. https://doi.org/10.1086/734087

Muslimova, D., van Kippersluis, H., Rietveld, C. A., von Hinke, S., & Meddens, S. F. W. (2023, July). Nature-nurture interplay in educational attainment [arXiv:2012.05021 [econ, q-fin]]. https://doi.org/10.48550/arXiv.2012.05021

Office for National Statistics. (2024, May). Local Authority Districts Boundaries UK. Retrieved February 3, 2026, from https://geoportal.statistics.gov.uk/datasets/ons::local-authority-districts-may-2024-boundaries-uk-bgc-2/about

Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S. M., Sidorenko, J., Kweon, H., Goldman, G., Gjorgjieva, T., Jiang, Y., Hicks, B., Tian, C., Hinds, D. A., Ahlskog, R., Magnusson, P. K. E., Oskarsson, S., Hayward, C., Campbell, A., ... Young, A. I. (2022). Polygenic prediction of educational attainment within and between families from genome-wide association analyses in 3 million individuals [Number: 4]. *Nature Genetics*, *54*(4), 437–449. https://doi.org/10.1038/s41588-022-01016-z

Pannullo, F., Lee, D., Waclawski, E., & Leyland, A. H. (2016). How robust are the estimated effects of air pollution on health? Accounting for model uncertainty using Bayesian model averaging. *Spatial and Spatio-temporal Epidemiology*, *18*, 53–62. https://doi.org/10.1016/j.sste.2016.04.001

Pereira, R. D., Rietveld, C. A., & van Kippersluis, H. (2020). The Interplay between Maternal Smoking and Genes in Offspring Birth Weight. *medRxiv*, 2020.10.30.20222844. https://doi.org/10.1101/2020.10.30.20222844

Qi, Y., Wei, S., Xin, T., Huang, C., Pu, Y., Ma, J., Zhang, C., Liu, Y., Lynch, I., & Liu, S. (2022). Passage of exogeneous fine particles from the lung into the brain in humans and animals. *Proceedings of the National Academy of Sciences*, *119*(26), e2117083119. https://doi.org/10.1073/pnas.2117083119

Robinson, A. J. (1971). Ii. Air Pollution. *Journal of the Royal Society of Arts*, *119*(5180), 505–519. Retrieved February 20, 2024, from https://www.jstor.org/stable/41370757

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2020). Specification curve analysis. *Nature Human Behaviour*, *4*(11), 1208–1214. https://doi.org/10.1038/s41562-020-0912-z

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

Sunyer, J., Esnaola, M., Alvarez-Pedrerol, M., Forns, J., Rivas, I., López-Vicente, M., Suades-González, E., Foraster, M., Garcia-Esteban, R., Basagaña, X., Viana, M., Cirach, M., Moreno, T., Alastuey, A., Sebastian-Galles, N., Nieuwenhuijsen, M., & Querol, X. (2015). Association between Traffic-Related Air Pollution in Schools and Cognitive Development in Primary School Children: A Prospective Cohort Study. *PLOS Medicine*, *12*(3), e1001792. https://doi.org/10.1371/journal.pmed.1001792

van Alten, S., Domingue, B. W., Faul, J., Galama, T., & Marees, A. T. (2024). Reweighting UK Biobank corrects for pervasive selection bias due to volunteering. *International Journal of Epidemiology*, *53*(3), dyae054. https://doi.org/10.1093/ije/dyae054

Virolainen, S. J., VonHandorf, A., Viel, K. C. M. F., Weirauch, M. T., & Kottyan, L. C. (2023). Gene–environment interactions and their impact on human health. *Genes & Immunity*, *24*(1), 1–11. https://doi.org/10.1038/s41435-022-00192-6

von Hinke, S., & Sørensen, E. N. (2023). The long-term effects of early-life pollution exposure: Evidence from the London smog. *Journal of Health Economics*, *92*, 102827. https://doi.org/10.1016/j.jhealeco.2023.102827

Weuve, J., Puett, R. C., Schwartz, J., Yanosky, J. D., Laden, F., & Grodstein, F. (2012). Exposure to Particulate Air Pollution and Cognitive Decline in Older Women. *Archives of Internal Medicine*, *172*(3), 219–227. https://doi.org/10.1001/archinternmed.2011.683

Woolgar, A., Parr, A., Cusack, R., Thompson, R., Nimmo-Smith, I., Torralva, T., Roca, M., Antoun, N., Manes, F., & Duncan, J. (2010). Fluid intelligence loss linked to restricted regions of damage within frontal and parietal cortex. *Proceedings of the National Academy of Sciences*, *107*(33), 14899–14902. https://doi.org/10.1073/pnas.1007928107

Xu, X., Ha, S. U., & Basnet, R. (2016). A Review of Epidemiological Research on Adverse Neurological Effects of Exposure to Ambient Air Pollution. *Frontiers in Public Health*, *4*. https://doi.org/10.3389/fpubh.2016.00157

Yamada, H. (2026, August). When the Air Doesn't Move the Data: The Thermal Inversion Instrument in Developing Countries. Retrieved October 18, 2026, from https://ies.keio.ac.jp/en/publications/27955/

Yao, Z., Zhang, W., Ji, X., & Weng, W. (2023). Short-Term Exposure to Air Pollution and Cognitive Performance: New Evidence from China's College English Test. *Environmental and Resource Economics*, *85*(1), 211–237. https://doi.org/10.1007/s10640-023-00765-7

Appendix

A Tables and figures

Table A1: Summary statistics

	Mean	SD	Min	Max	N
Fluid intelligence scores	6.18	2.10	0.00	13.00	125224.00
Avg. PM2.5 concentration during the day	15.31	6.30	3.92	44.88	125224.00
Avg. PM1 concentration during the day	11.43	6.11	1.80	38.63	125224.00
Avg. PM10 concentration during the day	22.60	8.61	5.57	63.33	125224.00
Inversion coverage	0.40	0.43	0.00	1.00	125224.00
Avg. ground temperature	10.76	5.20	-5.21	22.02	125224.00
Avg. dew temperature	280.48	4.71	266.68	290.99	125224.00
Avg. cloud coverage	0.72	0.23	0.00	1.00	125224.00
Avg. wind speed	4.14	1.73	0.77	10.99	125224.00
Avg. wind direction	3.36	1.31	0.30	5.88	125224.00
Avg. total precipitation	0.00	0.00	0.00	0.00	125224.00
PGS EA	0.01	1.00	-4.12	4.36	125224.00

Notes: The table presents summary statistics for fluid intelligence scores, air pollution, inversion occurrence, weather controls, and polygenic scores for educational attainment. The sample has been restricted to that used in the analysis. Average total precipitation has a mean of 0.00008 and a standard deviation of .0002094, with the minimum being 0 and maximum being .003731.

Table A2: Bivariate regression between PM2.5 and socioeconomic variables

	Education years (1)	Sex (2)	BMI (3)	Birth year (4)	HH size (5)
PM2.5	0.0014***	0.0002*	-0.0063***	0.0157***	0.0005*
	(0.0005)	(0.0001)	(0.0011)	(0.0019)	(0.0003)
Observations	391,068	391,068	389,931	391,068	391,068
Weather controls	×	×	×	×	×
Fixed effects	×	×	×	×	×

Notes: The table shows the coefficients of PM2.5 obtained from regressing each of the socioeconomic variables on PM2.5. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

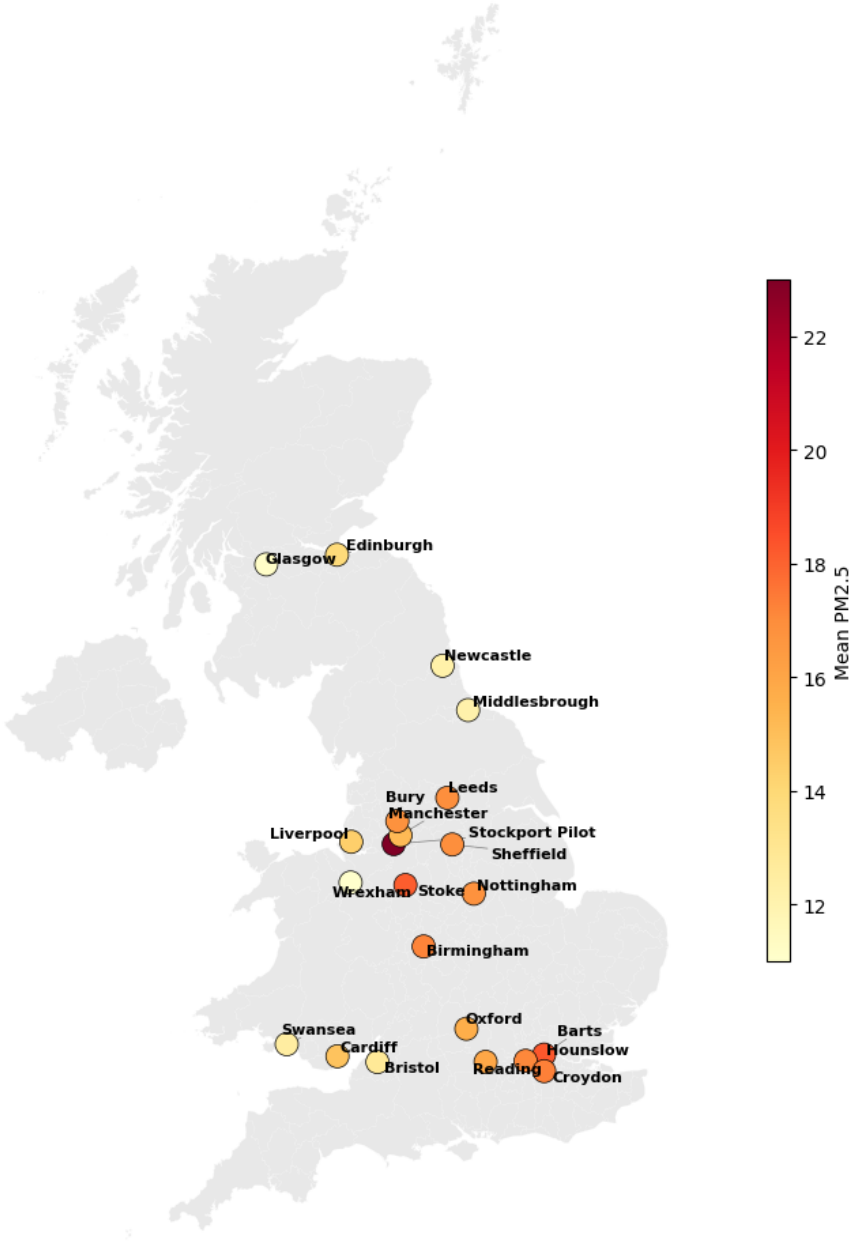


Figure A1: Mean PM2.5 across UK Biobank assessment centers, averaged over all assessment days.

Notes: The map shows the average daily PM2.5 concentration (in $\mu g/m^3$) across 22 UK Biobank assessment centers. For each assessment center, the average is computed based on all available dates *in the dataset*. That is, dates in which at least one participant attended that assessment center. Thus, the dates used to calculate the average for one assessment center are not necessarily the same dates used to calculate the average of another assessment center. The background map of the United Kingdom is taken from the Office for National Statistics (2024).

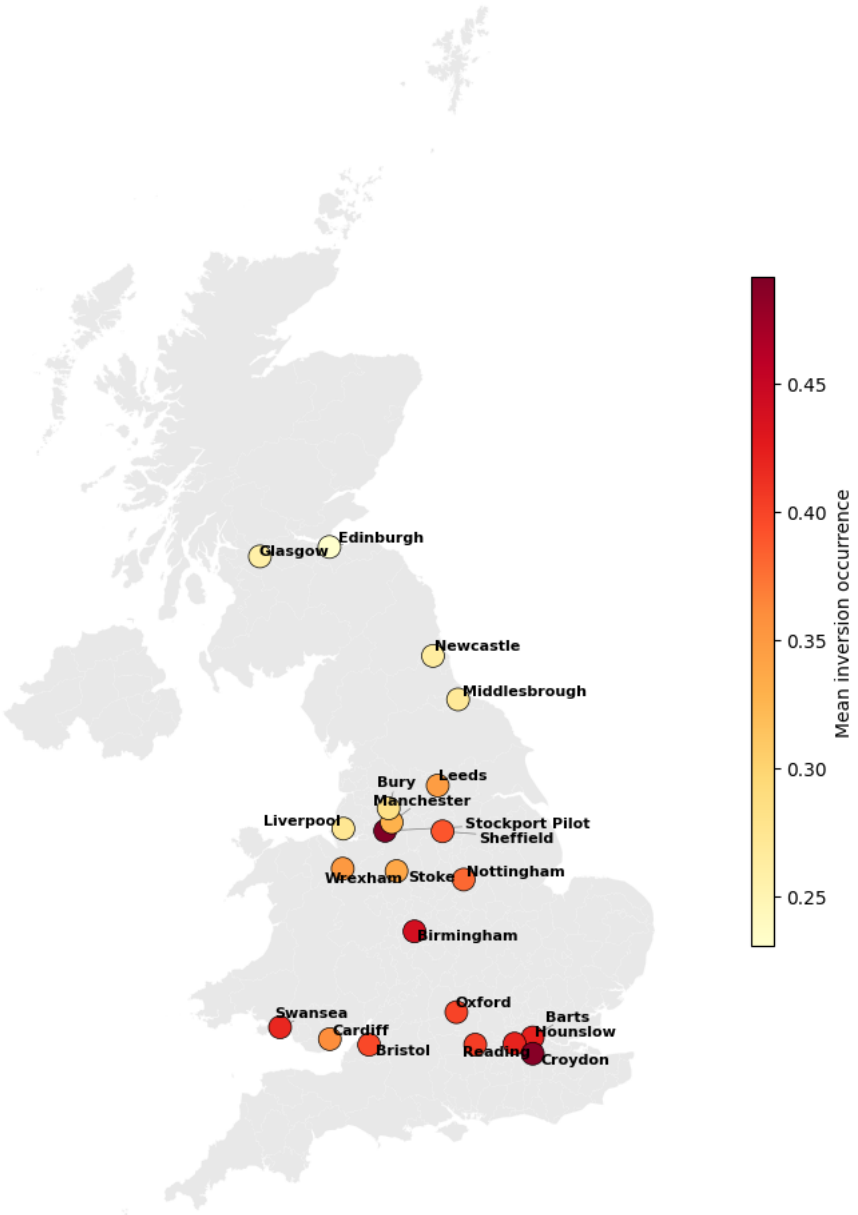


Figure A2: Weighted spatial coverage of inversion across UK Biobank assessment centers, averaged over all assessment days.

Notes: The map shows the weighted spatial coverage of inversion within 60km of an assessment center, for all 22 UK Biobank assessment centers. For each assessment center, the average is computed based on all available dates *in the dataset*. That is, dates in which at least one participant attended that assessment center. Thus, the dates used to calculate the average for one assessment center are not necessarily the same dates used to calculate the average of another assessment center. The background map of the United Kingdom is taken from the Office for National Statistics (2024).

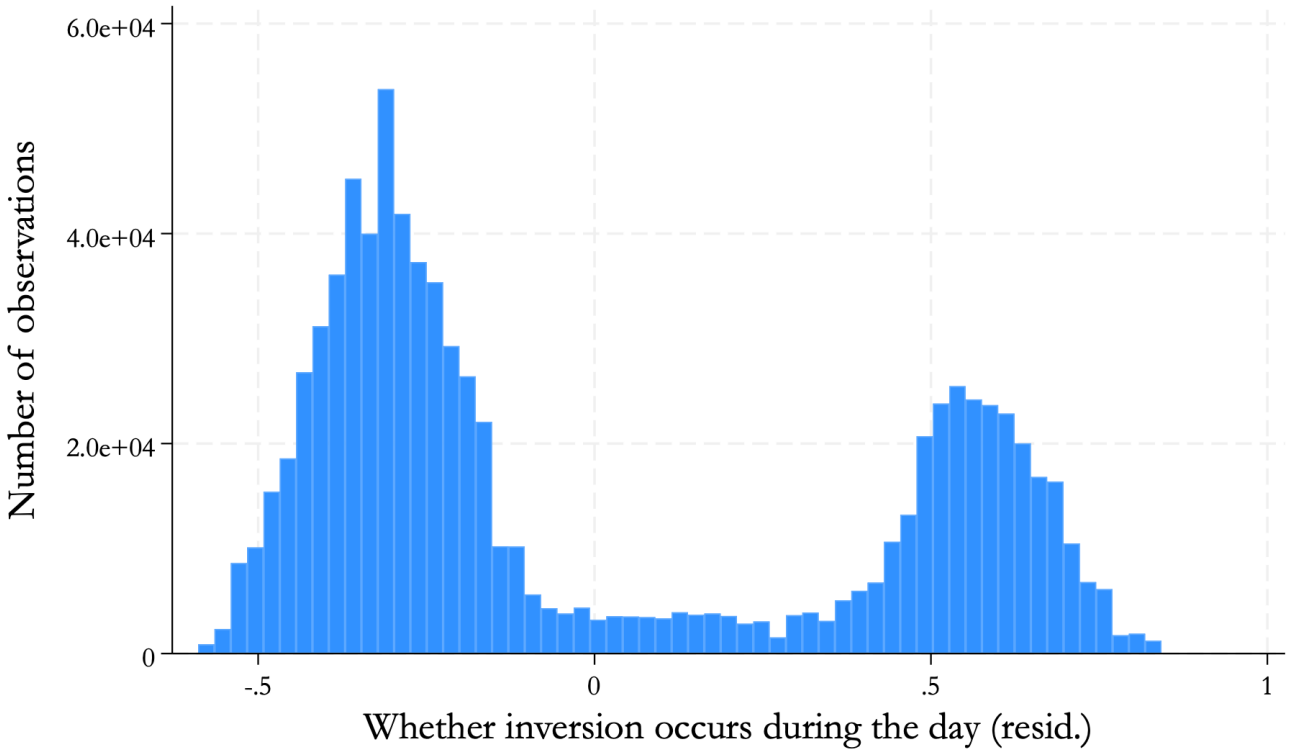


Figure A3: Histogram for residualized inversion coverage

inversion coverage are residualized from a regression of inversion coverage on center, month, year, and day-of-week fixed effects.

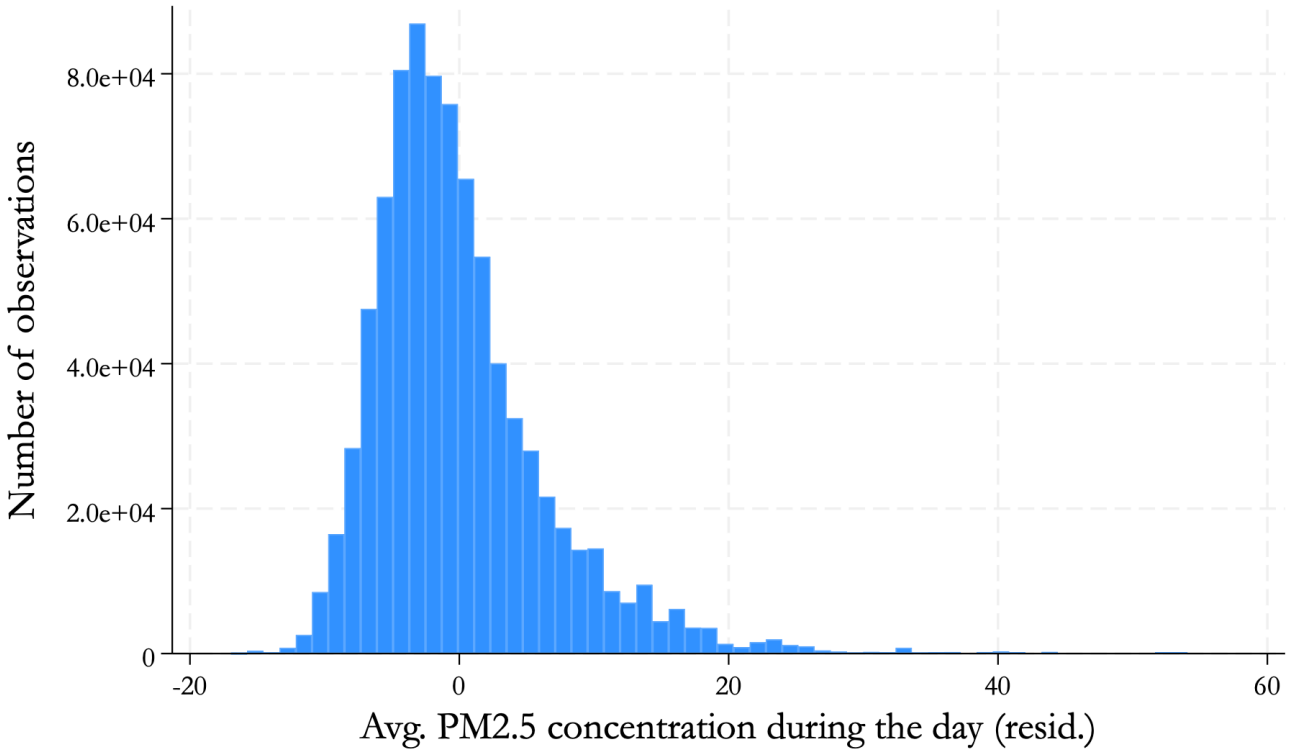


Figure A4: Histogram for residualized PM2.5

PM2.5 are residualized from a regression of inversion coverage on center, month, year, and day-of-week fixed effects.

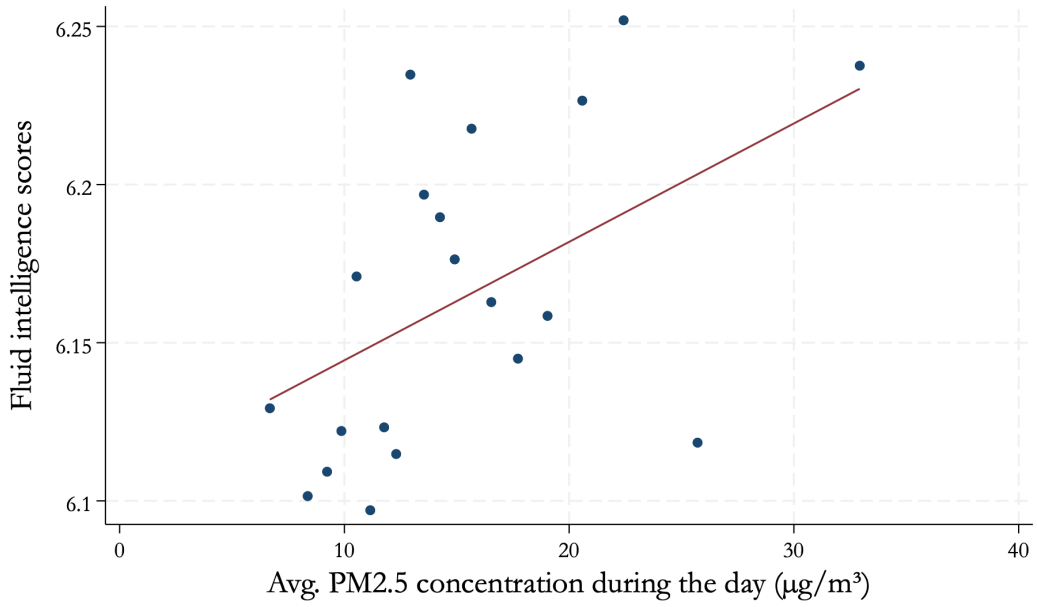


Figure A5: Binned scatter plot for fluid intelligence (FI) scores and average PM2.5 concentration during the assessment day.

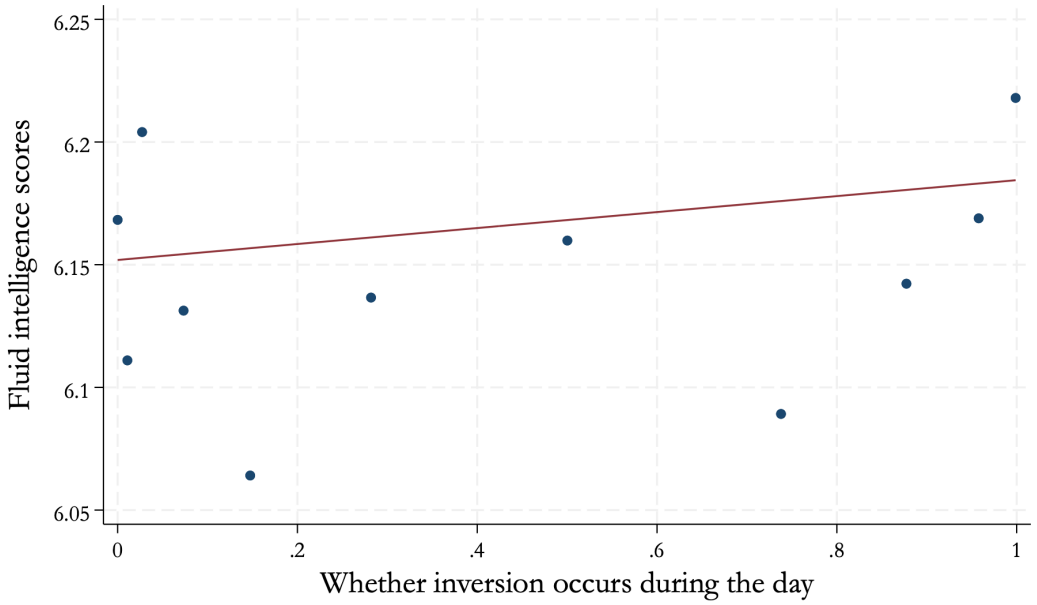


Figure A6: Binned scatter plot for fluid intelligence (FI) scores and raw thermal inversion coverage

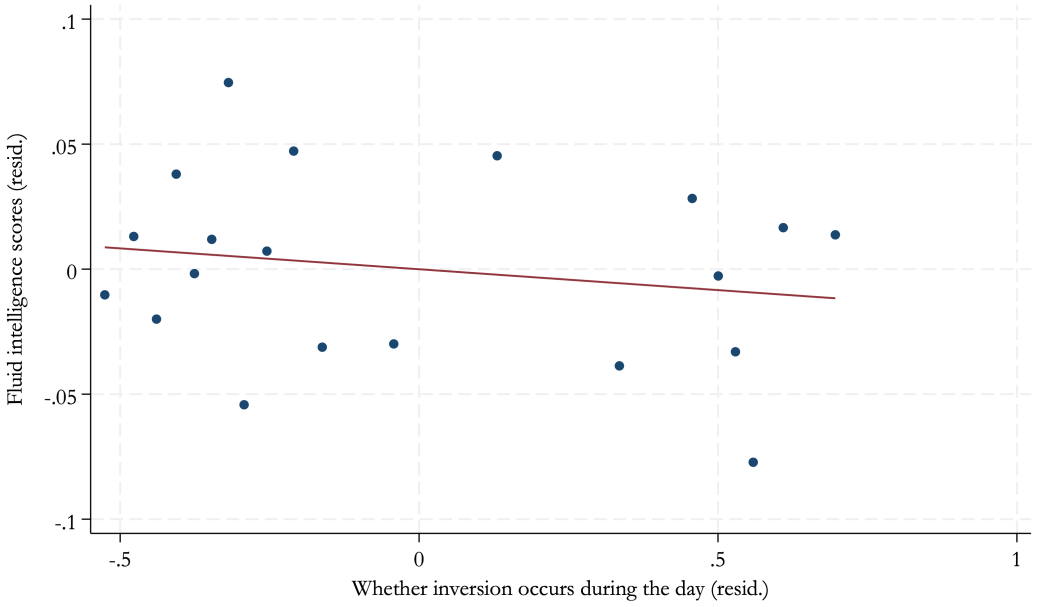


Figure A7: Binned scatter plot for fluid intelligence (FI) scores and residualized thermal inversion coverage

inversion coverages are residualized from a regression of inversion coverage on center, month, year, and day-of-week fixed effects.

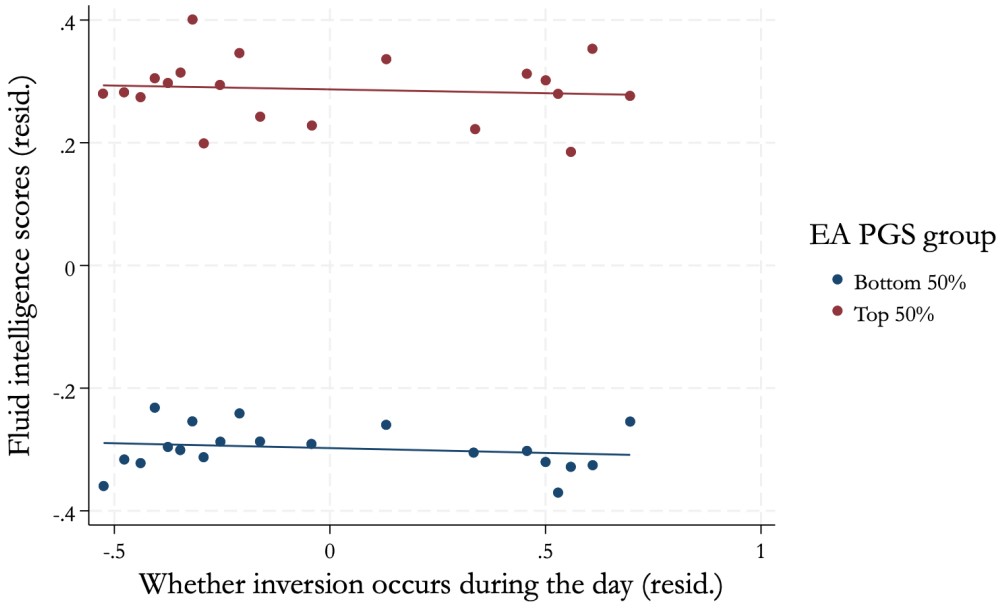


Figure A8: Binned scatter plot for fluid intelligence (FI) scores and residualized thermal inversion coverage, by PGS EA

inversion coverages are residualized from a regression of inversion coverage on center, month, year, and day-of-week fixed effects. PGS EA are divided at the median.

Table A3: Exogeneity assessment - health outcomes

	Respiratory (1)	Cardiovascular (2)	Long-term illness (3)
Inversion coverage	0.0037	-0.0015	0.0008
	(0.0043)	(0.0033)	(0.0070)
R^2	0.00344	0.00595	0.00970
Observations	125,032	125,048	122,549
Weather controls	✓	✓	✓
Fixed effects	✓	✓	✓

Notes: Respiratory (column 1) is equal to 1 if a participant has ever been hospitalized for a respiratory disease. *Cardiovascular* (column 2) is equal to 1 if a participant has ever been hospitalized for a cardiovascular disease. *Long − term illness* (column 3) is equal to 1 if a participant has a long term illness. Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center on a given day, and can range from zero to one. Robust standard errors clustered at the location level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A4: Effects of air pollution/thermal inversions on FI scores
Heterogeneity by PGS EA

	Fluid intelligence scores	
	(1)	(2)
Inversion coverage	-0.0207*	
	(0.0103)	
Inversion coverage × PGS EA	-0.0057	
	(0.0066)	
PM2.5		-0.0073*
		(0.0039)
PM2.5 × PGS EA		-0.0017
		(0.0021)
R^2	0.05674	0.05014
Observations	125,224	125,224
Weather controls	✓	✓
Fixed effects	✓	✓
Interaction between weather and PGS EA	✓	✓

Notes: The set of fixed effects include center, month, year, and day-of-week fixed effects. The set of weather controls include ground temperature, dew temperature, wind speed, wind direction, cloud coverage, and total precipitation, as well as their squared and cubed terms. Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center on a given day, and can range from zero to one. PM2.5 is the average daily concentration of PM2.5 and is in $\mu g/m^3$. The dependent variable is the fluid intelligence score, standardized to have mean zero and standard deviation one. PGS EA is standardized at 0 with standard deviation 1. Robust standard errors clustered at the assessment center level are reported in parentheses.

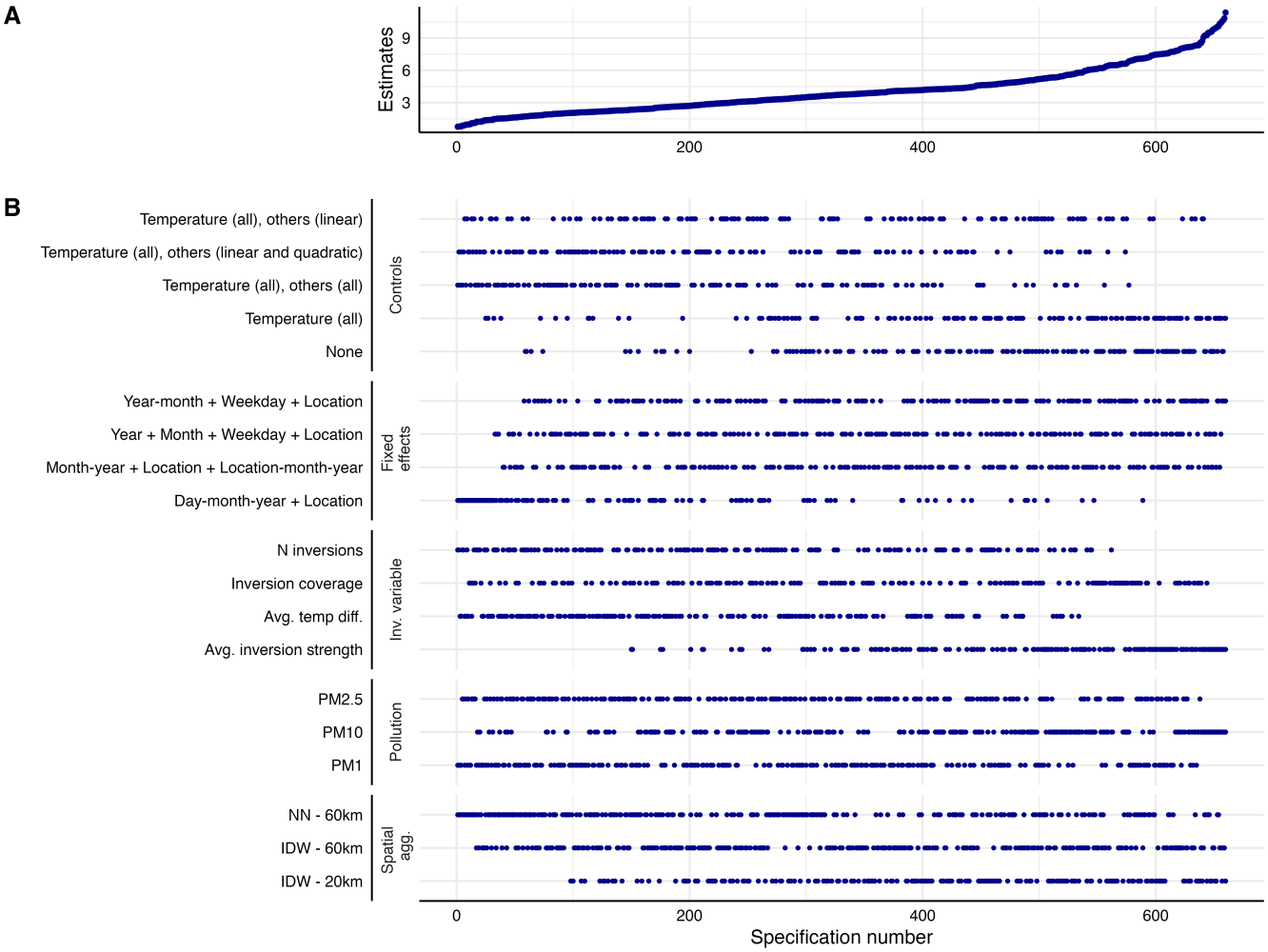


Figure A9: Specification curves for 1st stage estimates - effects of inversion coverage on air pollution

Significant and positive point estimates at the 95% level are colored in blue. Regressions in this figure are estimated without polygenic scores or its interactions.

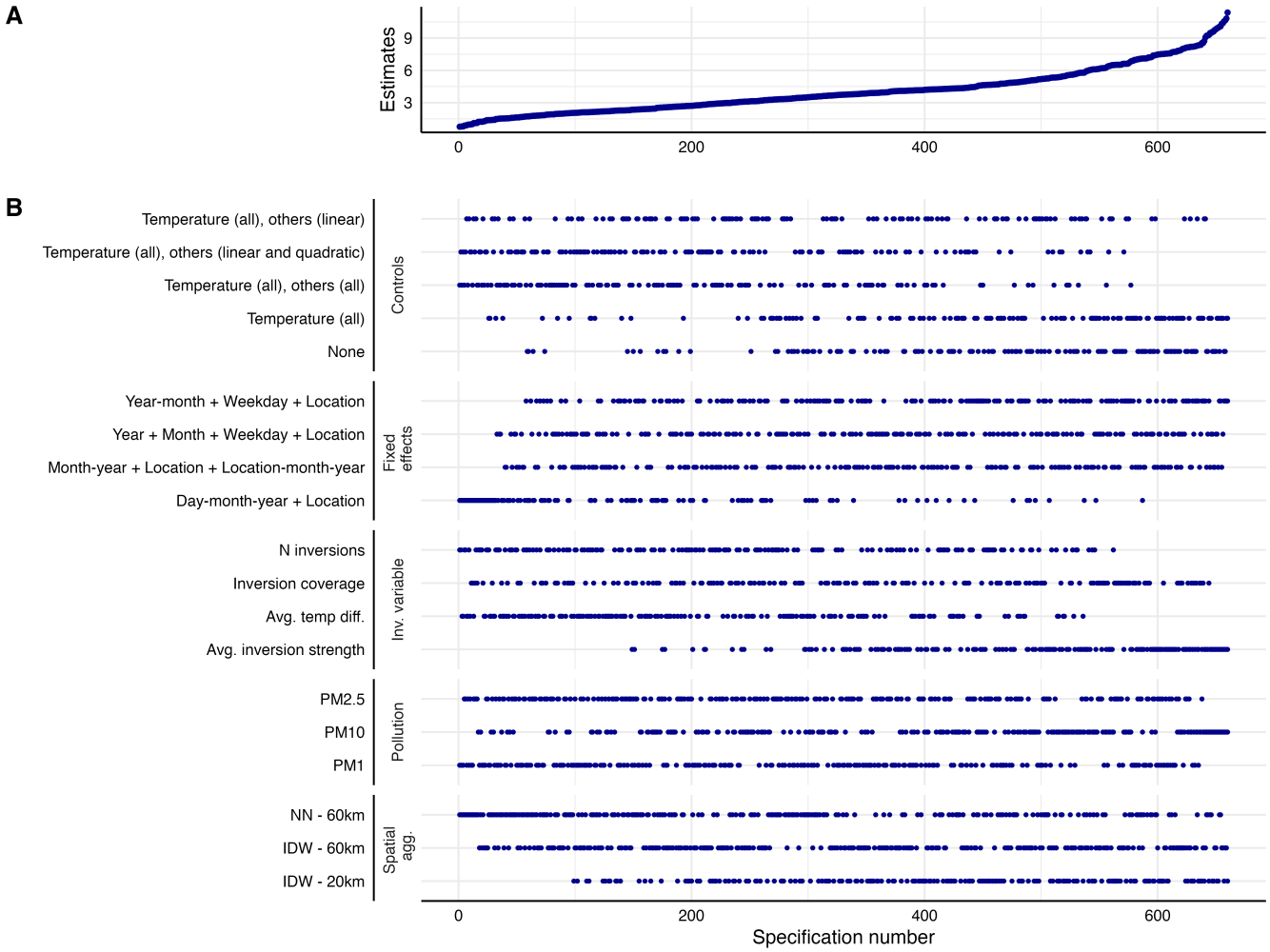


Figure A10: Specification curves for 1st stage estimates - effects of inversion coverage on air pollution (specifications with PGS EA)

Significant and positive point estimates at the 95% level are colored in blue. Regressions in this figure are estimated with polygenic scores or its interactions.

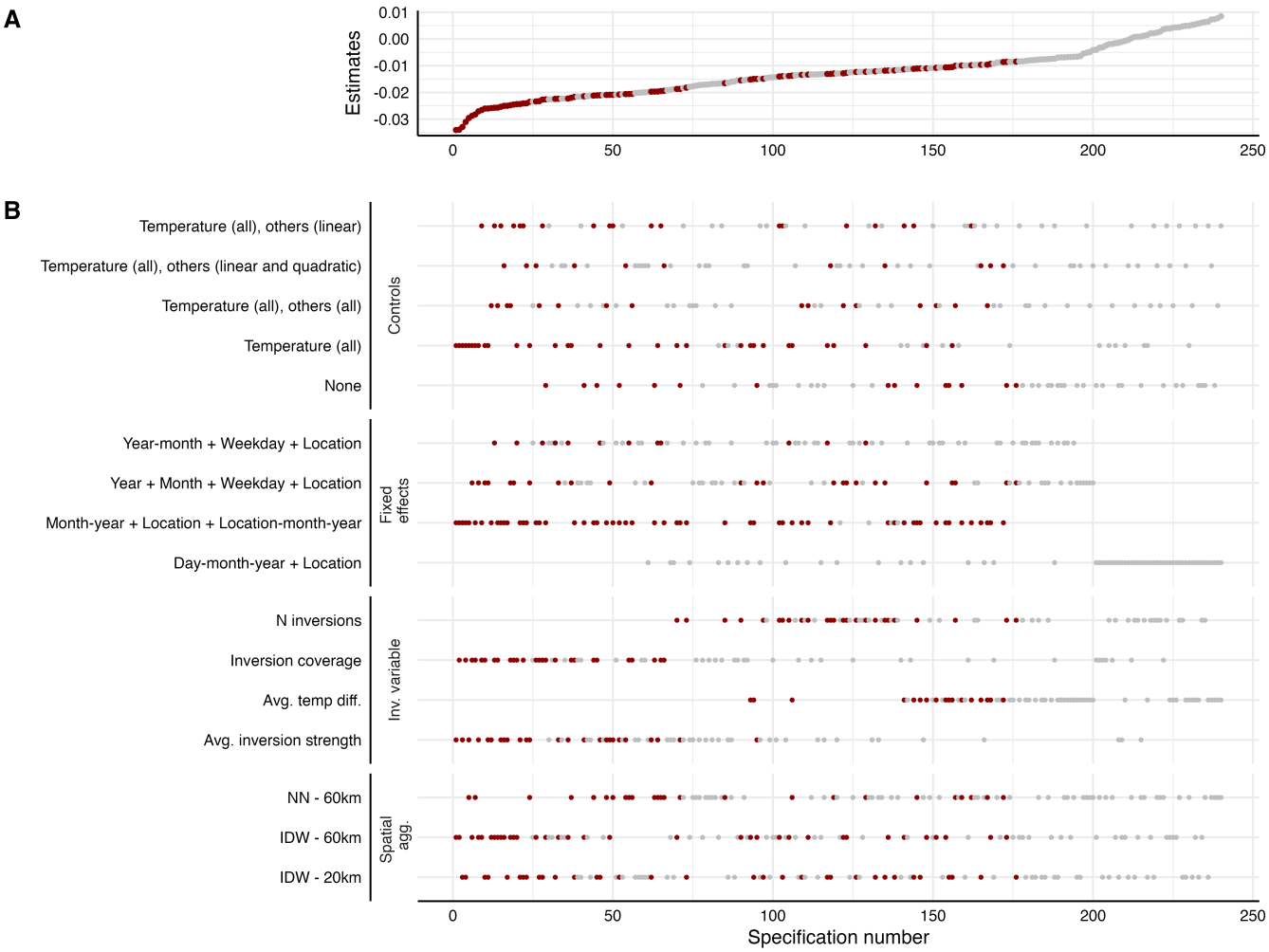


Figure A11: Specification curves for reduced-form estimates of air pollution on fluid intelligence scores

Significant point estimates at the 95% level are colored in red. Regressions in this figure are estimated without polygenic scores or its interactions.

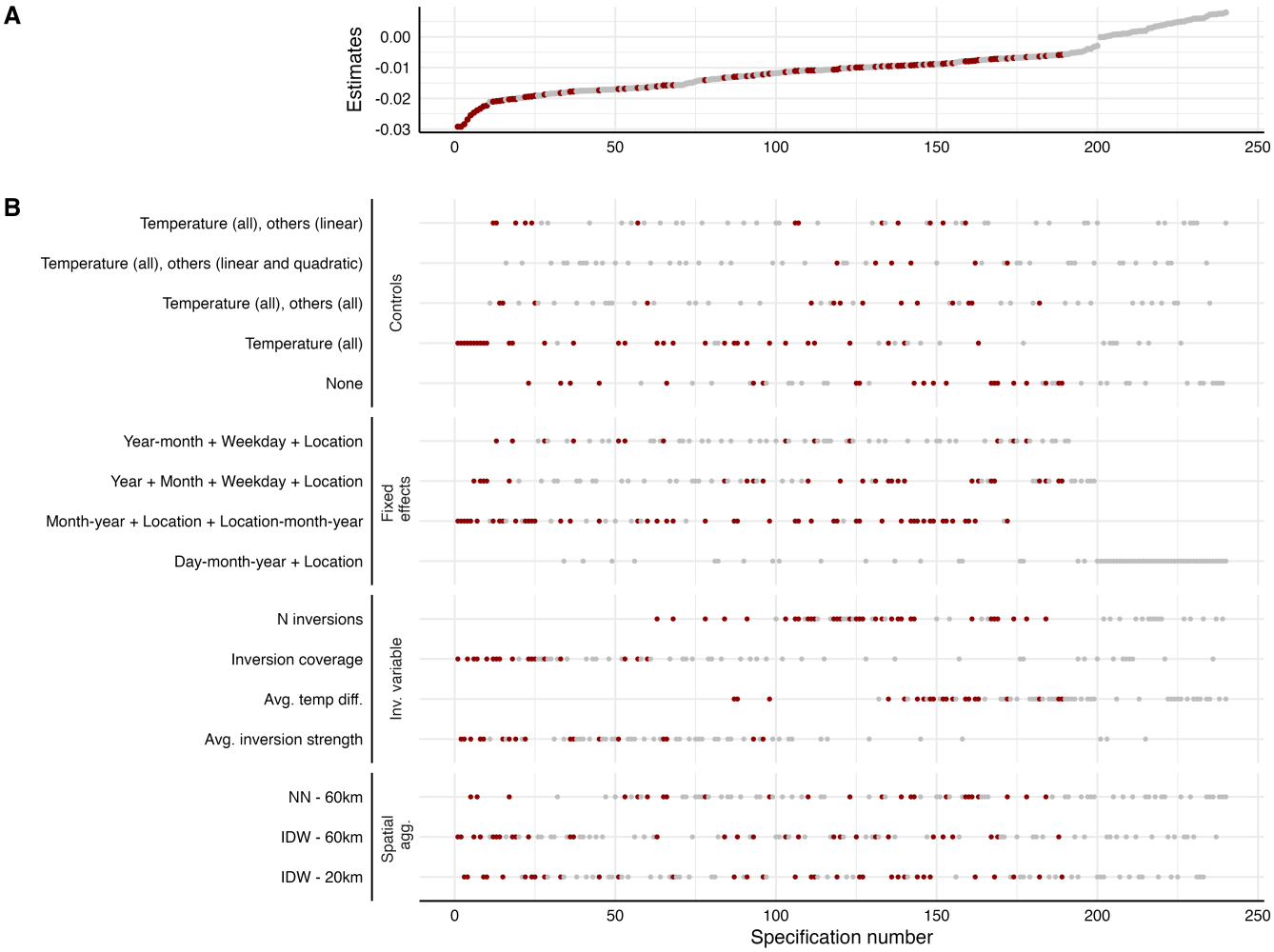


Figure A12: Specification curves for reduced-form estimates of air pollution on fluid intelligence scores (specifications with polygenic scores)

Significant point estimates at the 95% level are colored in red. Regressions in this figure are estimated wit polygenic scores or its interactions.

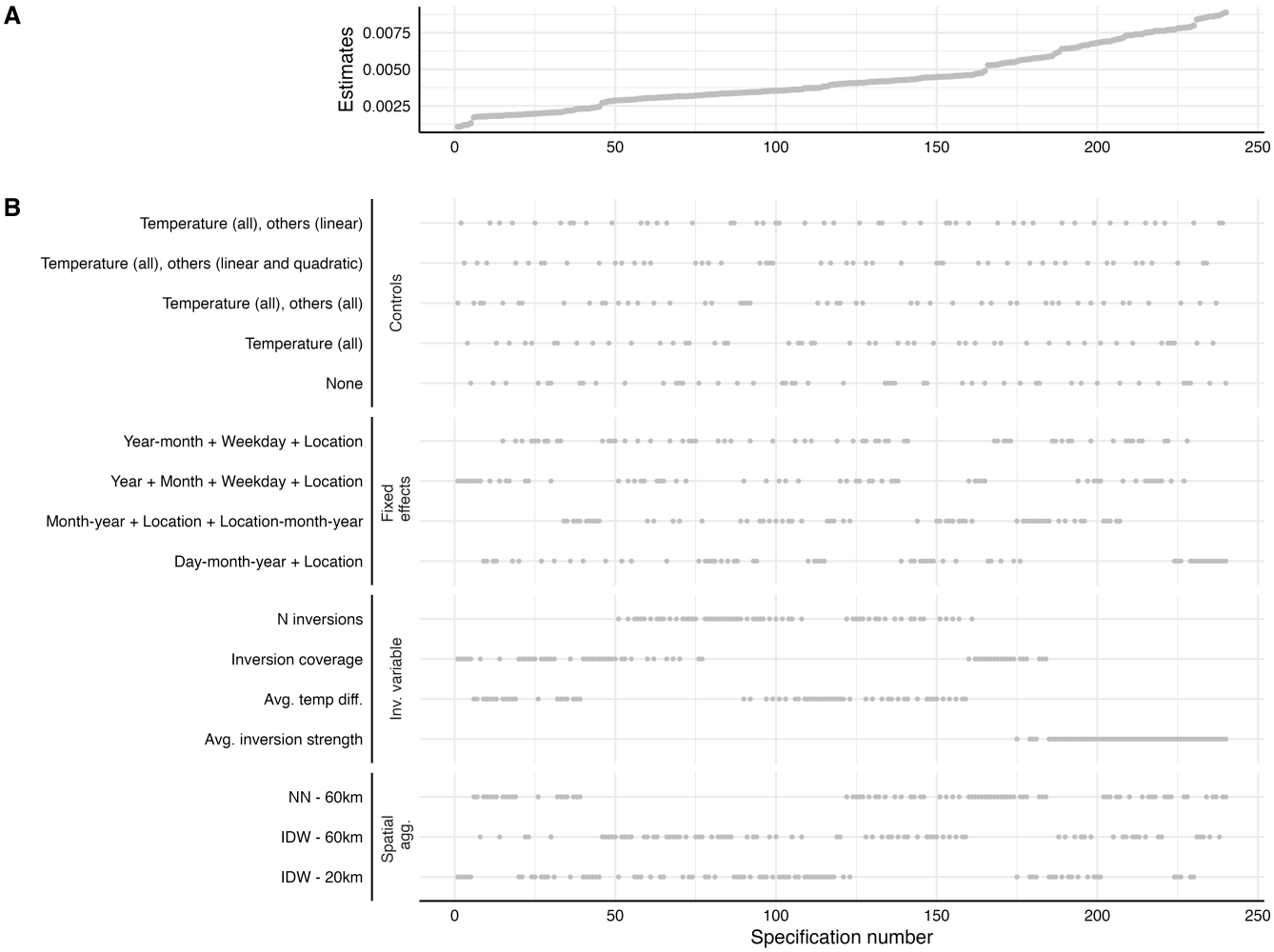


Figure A13: Specification curves for reduced-form estimates of the interaction between air pollution and PGS EA

Significant point estimates at the 95% level are colored in red.

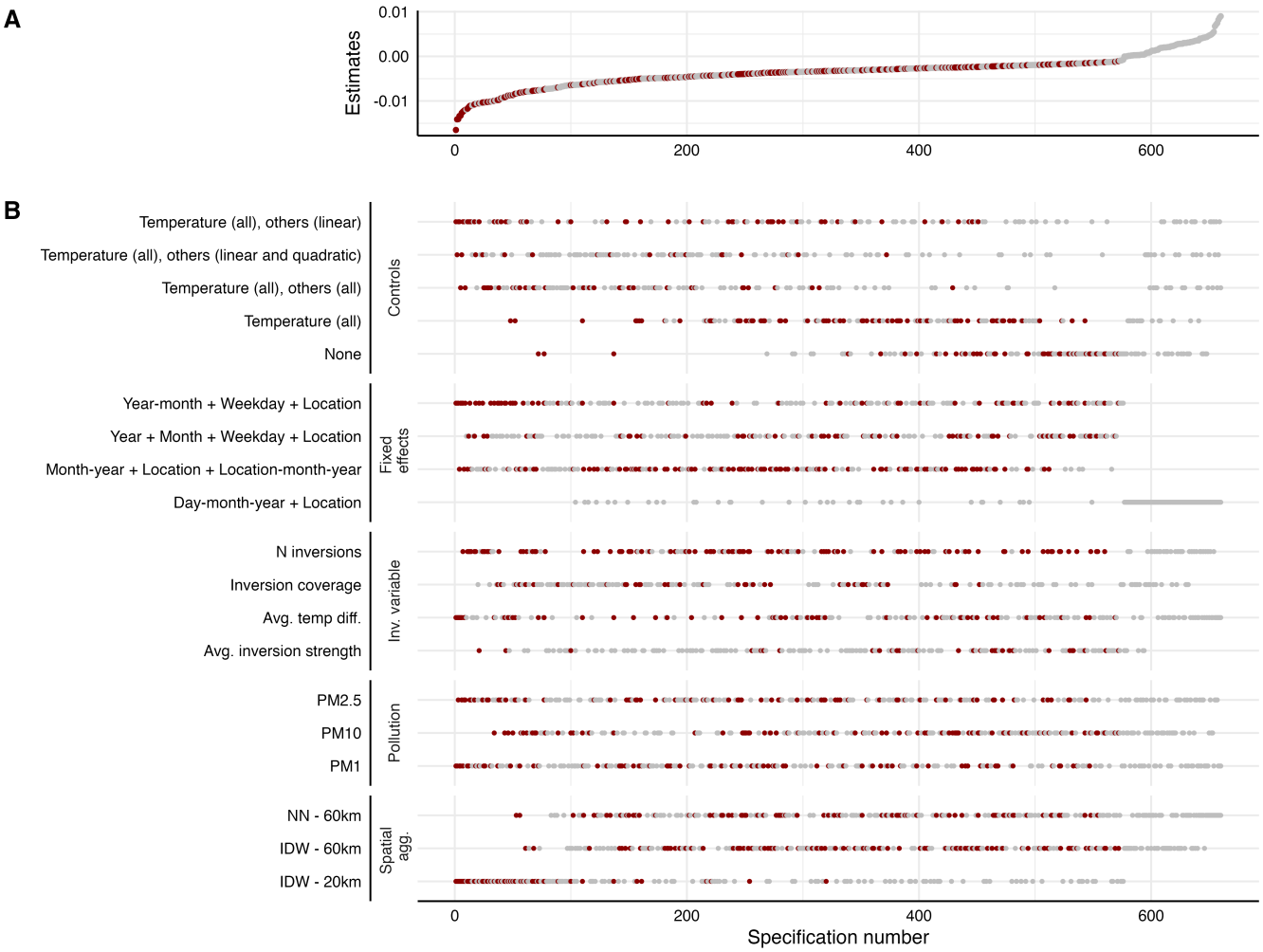


Figure A14: Specification curves for 2SLS estimates of air pollution on fluid intelligence scores (specifications with polygenic scores)

Significant point estimates at the 95% level are colored in red. Results presented in this specification curves are filtered for F-statistics larger or equal to 10. Regressions in this figure are estimated with polygenic scores or its interactions.

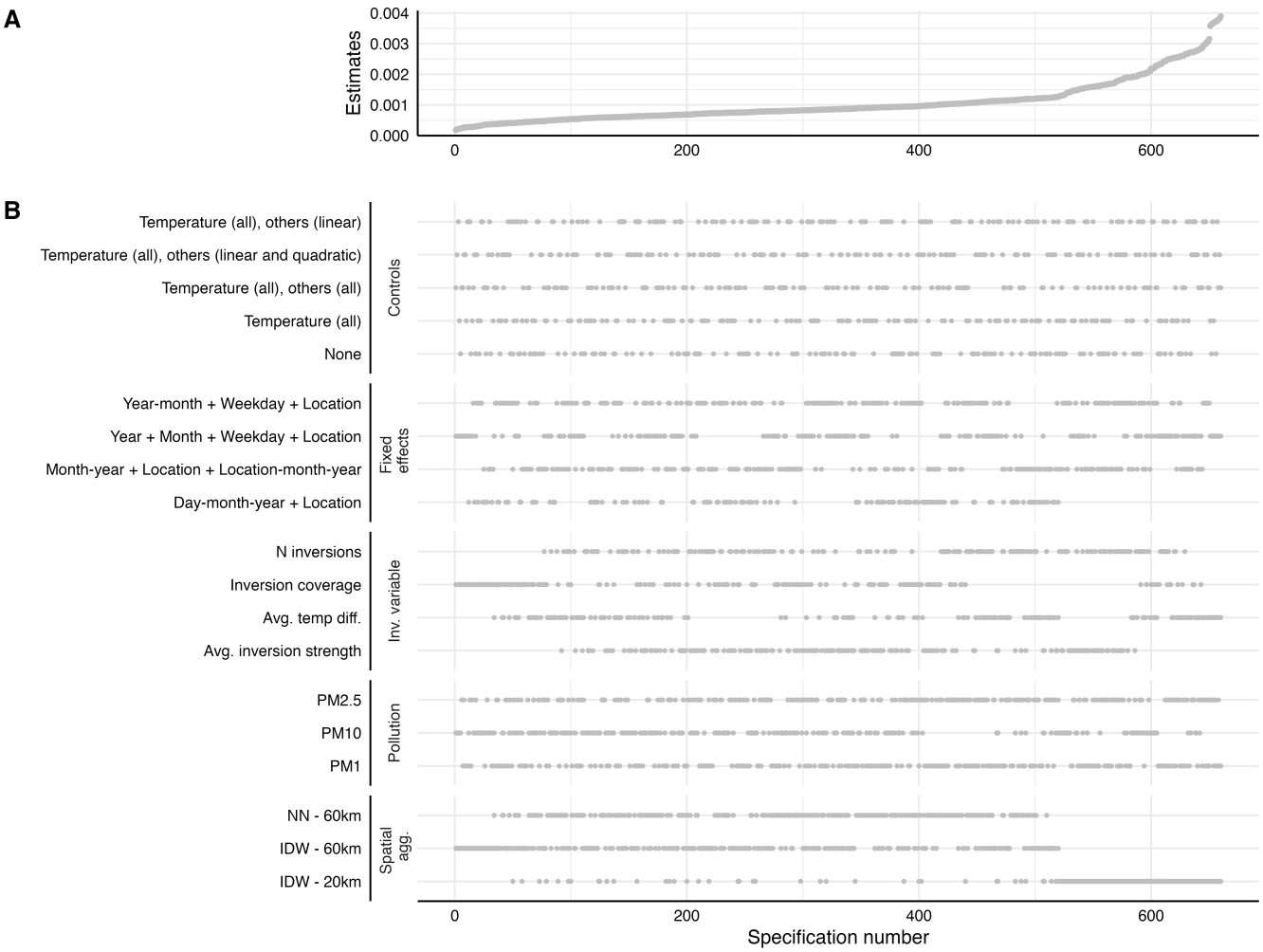


Figure A15: Specification curves for 2SLS estimates of the interaction between air pollution and PGS EA

Significant point estimates at the 95% level are colored in red. Results presented in this specification curves are filtered for F-statistics larger or equal to 10.

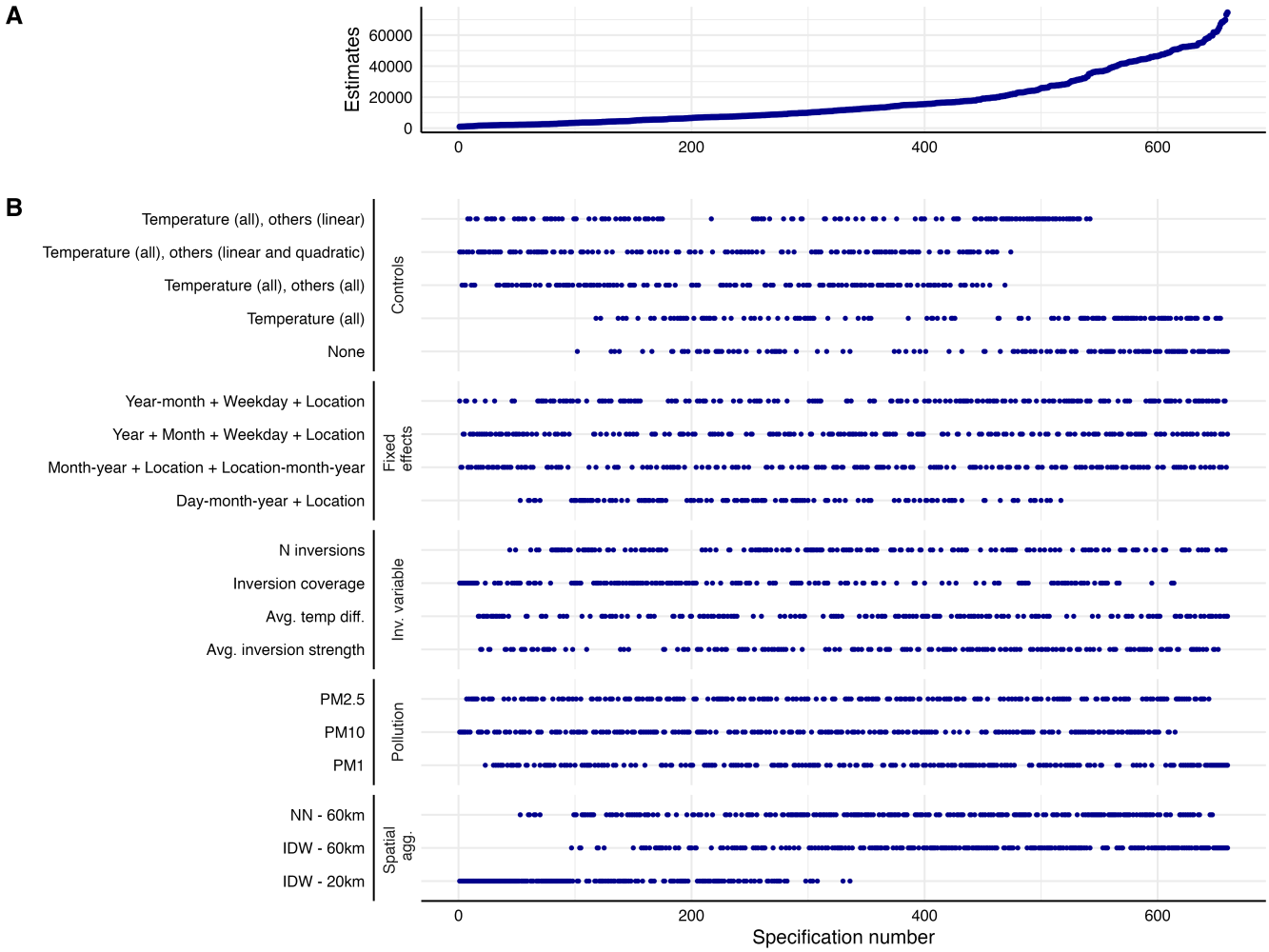


Figure A16: Specification curves for F-stats in 2SLS estimates of air pollution on fluid intelligence scores (specifications without polygenic scores)

F-statistics above 10 are presented in blue.

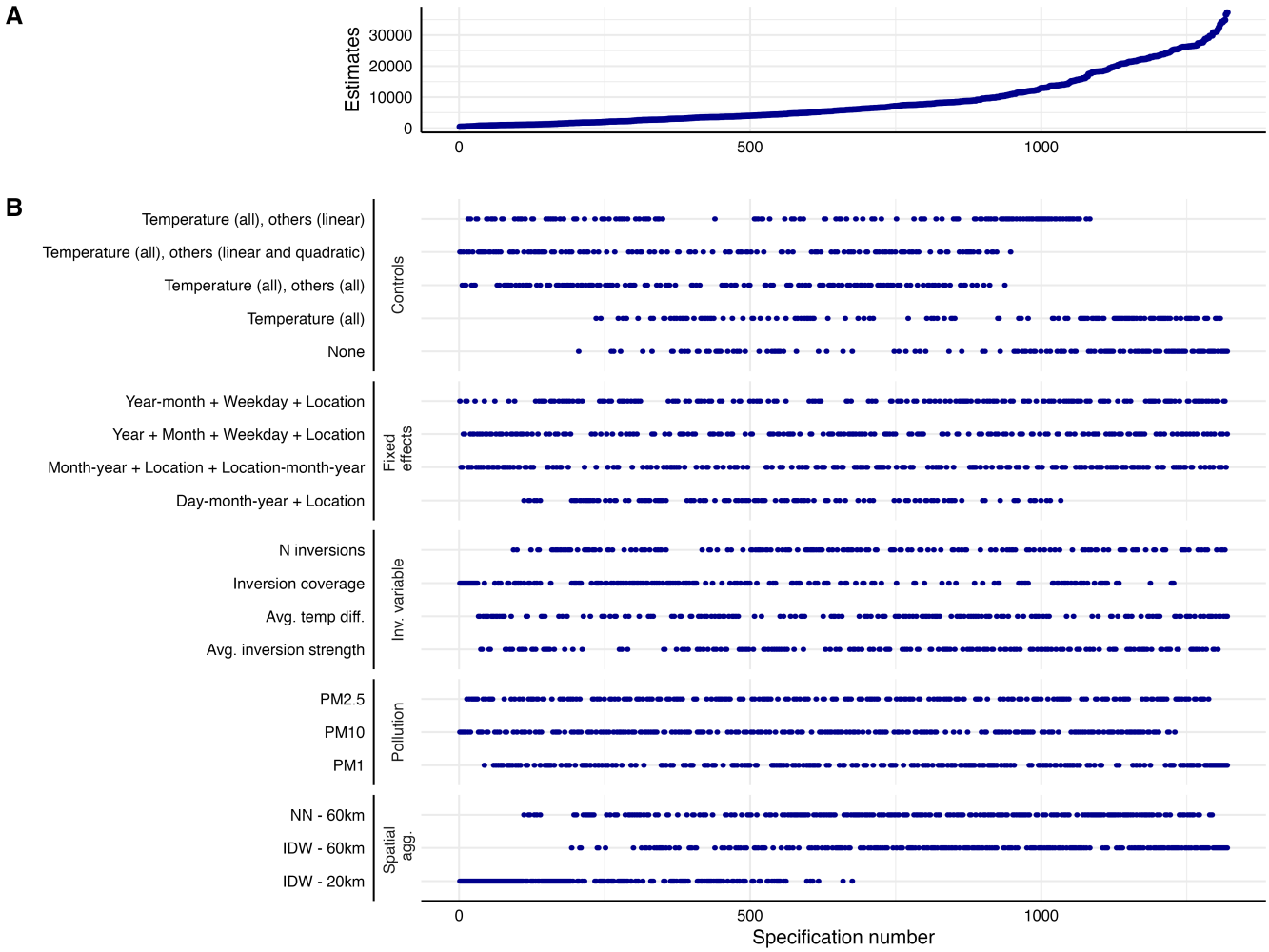


Figure A17: Specification curves for F-stats in 2SLS estimates of air pollution on fluid intelligence scores (specifications with polygenic scores)

F-statistics above 10 are presented in blue.

Table A5: Heterogeneity by demographic characteristics

	Fluid intelligence scores			
	(1)	(2)	(3)	(4)
Inversion coverage	-0.1237***	-0.0371	-0.0315**	-0.0580
	(0.0287)	(0.0377)	(0.0106)	(0.0499)
Inversion × Educ. years	0.0086***			
	(0.0022)			
Inversion × Age		0.0002		
		(0.0007)		
Inversion × Sex			0.0105	
			(0.0166)	
Inversion × BMI				0.0012
				(0.0018)
R^2	0.20825	0.02006	0.01873	0.01634
Observations	125,224	125,224	125,224	124,762
Weather controls	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓

Notes: The set of fixed effects include center, month, year, and day-of-week fixed effects. The set of weather controls include ground temperature, dew temperature, wind speed, wind direction, cloud coverage, and total precipitation, as well as their squared and cubed terms. Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center on a given day, and can range from zero to one. The dependent variable is the fluid intelligence score, standardized to have mean zero and standard deviation one. Robust standard errors clustered at the assessment center level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A6: Heterogeneity by health characteristics

	Fluid intelligence scores			
	(1)	(2)	(3)	(4)
Inversion coverage	-0.0272**	-0.0270**	-0.0417	-0.0251*
	(0.0119)	(0.0114)	(0.0285)	(0.0116)
Inversion × Respiratory	0.0365*			
	(0.0197)			
Inversion × Cardiovascular		0.0115		
		(0.0204)		
Inversion × Health rating			0.0065	
			(0.0100)	
Inversion × Long-term illness				-0.0066
				(0.0142)
R^2	0.01884	0.01898	0.03782	0.01916
Observations	125,032	125,048	125,224	122,549
Weather controls	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓

Notes: The set of fixed effects include center, month, year, and day-of-week fixed effects. The set of weather controls include ground temperature, dew temperature, wind speed, wind direction, cloud coverage, and total precipitation, as well as their squared and cubed terms. Respiratory is equal to 1 if a participant has ever been hospitalized for a respiratory disease. Cardiovascular is equal to 1 if a participant has ever been hospitalized for a cardiovascular disease. Health rating is the subjective health rating participants gave during the baseline, and range from 1 to 4. Long-term illness is equal to 1 if a participant has a long term illness. Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center on a given day, and can range from zero to one. The dependent variable is the fluid intelligence score, standardized to have mean zero and standard deviation one. Robust standard errors clustered at the assessment center level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A7: First-stage estimates - no home distance restriction

	PM2.5
	(1)
Inversion coverage	2.531***
	(0.2884)
R^2	0.65758
Observations	126,135
Weather controls	✓
Fixed effects	✓

Notes: The sample additionally includes respondents living more than 40km from a UK Biobank assessment center. Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center on a given day, and can range from zero to one. PM2.5 is the average daily concentration of PM2.5 and is in $\mu g/m^3$. Robust standard errors clustered at the assessment center level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A8: Exogeneity assessment for demographic outcomes - no home distance restriction

	PGS EA	Sex	BMI	Birth year	HH size	Education years
	(1)	(2)	(3)	(4)	(5)	(6)
Inversion coverage	-0.0304	0.0093	0.0389	0.2011*	-0.0212	-0.0561
	(0.0229)	(0.0072)	(0.0962)	(0.1067)	(0.0122)	(0.0361)
R^2	0.01459	0.00117	0.01001	0.01913	0.00584	0.02600
Observations	126,135	126,135	125,668	126,135	126,135	126,135
Weather controls	✓	✓	✓	✓	✓	✓
Fixed effects	✓	✓	✓	✓	✓	✓

Notes: The sample additionally includes respondents living more than 40km from a UK Biobank assessment center. The set of fixed effects include center, month, year, and day-of-week fixed effects. The set of weather controls include ground temperature, dew temperature, wind speed, wind direction, cloud coverage, and total precipitation, as well as their squared and cubed terms. Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center on a given day, and can range from zero to one. Robust standard errors clustered at the assessment center level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A9: Falsification test - no home distance restriction

	Fluid intelligence scores (1)
Inversion coverage lead	0.0118
	(0.0111)
R^2	0.01240
Observations	126,128
Weather controls	✓
Fixed effects	✓

Notes: The sample additionally includes respondents living more than 40km from a UK Biobank assessment center. Fluid intelligence scores is a raw score that goes from 0 to 13. Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center during the day *after the assessment day*, and can range from zero to one. Robust standard errors clustered at the location level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A10: Exogeneity assessment for health outcomes - no home distance restriction

	Respiratory (1)	Cardiovascular (2)	Long-term illness (3)
Inversion coverage	0.0029	-0.0013	0.0003
	(0.0046)	(0.0032)	(0.0071)
R^2	0.00347	0.00585	0.00962
Observations	125,941	125,957	123,441
Weather controls	✓	✓	✓
Fixed effects	✓	✓	✓

Notes: The sample additionally includes respondents living more than 40km from a UK Biobank assessment center. *Respiratory* (column 1) is equal to 1 if a participant has ever been hospitalized for a respiratory disease. *Cardiovascular* (column 2) is equal to 1 if a participant has ever been hospitalized for a cardiovascular disease. $Long-term\ illness$ (column 3) is equal to 1 if a participant has a long term illness. Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center on a given day, and can range from zero to one. Robust standard errors clustered at the location level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A11: Effects of thermal inversion on test taking
no home distance restriction

	Takes FI test
	(1)
Inversion	0.0271
coverage	(0.0176)
R^2	0.62881
Observations	151,831
Weather controls	✓
Fixed effects	✓

Notes: The sample additionally includes respondents living more than 40km from a UK Biobank assessment center. The set of fixed effects include center, month, year, and day-of-week fixed effects. The set of weather controls include ground temperature, dew temperature, wind speed, wind direction, cloud coverage, and total precipitation, as well as their squared and cubed terms. Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center on a given day, and can range from zero to one. The outcome is 1 if participant decided to take the fluid intelligence test, and 0 if they did not. The sample is restricted to only participants who were offered the fluid intelligence test. Robust standard errors clustered at the assessment center level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A12: Effects of air pollution/thermal inversion on FI scores
no home distance restriction

	Fluid intelligence scores	
	(1)	(2)
Inversion coverage	-0.0249**	
	(0.0103)	
PM2.5		-0.0098**
		(0.0039)
R^2	0.01244	0.01114
Observations	126,135	126,135
Weather controls	✓	✓
Fixed effects	✓	✓

Notes: The sample additionally includes respondents living more than 40km from a UK Biobank assessment center. The set of fixed effects include center, month, year, and day-of-week fixed effects. The set of weather controls include ground temperature, dew temperature, wind speed, wind direction, cloud coverage, and total precipitation, as well as their squared and cubed terms. Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center on a given day, and can range from zero to one. PM2.5 is the average daily concentration of PM2.5 and is in $\mu g/m^3$. The dependent variable is the fluid intelligence score, standardized to have mean zero and standard deviation one. Robust standard errors clustered at the assessment center level are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

Table A13: Effects on FI scores - no home distance restriction
Heterogeneity by PGS EA

	Fluid intelligence scores	
	(1)	(2)
Inversion coverage	-0.0207*	
	(0.0104)	
Inversion coverage × PGS EA	0.0015	
	(0.0068)	
PM2.5		-0.0074*
		(0.0039)
PM2.5 × PGS EA		0.0002
		(0.0009)
R^2	0.05562	0.04914
Observations	126,135	126,135
Weather controls	✓	✓
Fixed effects	✓	✓

Notes: The sample additionally includes respondents living more than 40km from a UK Biobank assessment center. The set of fixed effects include center, month, year, and day-of-week fixed effects. The set of weather controls include ground temperature, dew temperature, wind speed, wind direction, cloud coverage, and total precipitation, as well as their squared and cubed terms. Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center on a given day, and can range from zero to one. PM2.5 is the average daily concentration of PM2.5 and is in $\mu g/m^3$. The dependent variable is the fluid intelligence score, standardized to have mean zero and standard deviation one. PGS EA is standardized at 0 with standard deviation 1. Robust standard errors clustered at the assessment center level are reported in parentheses.

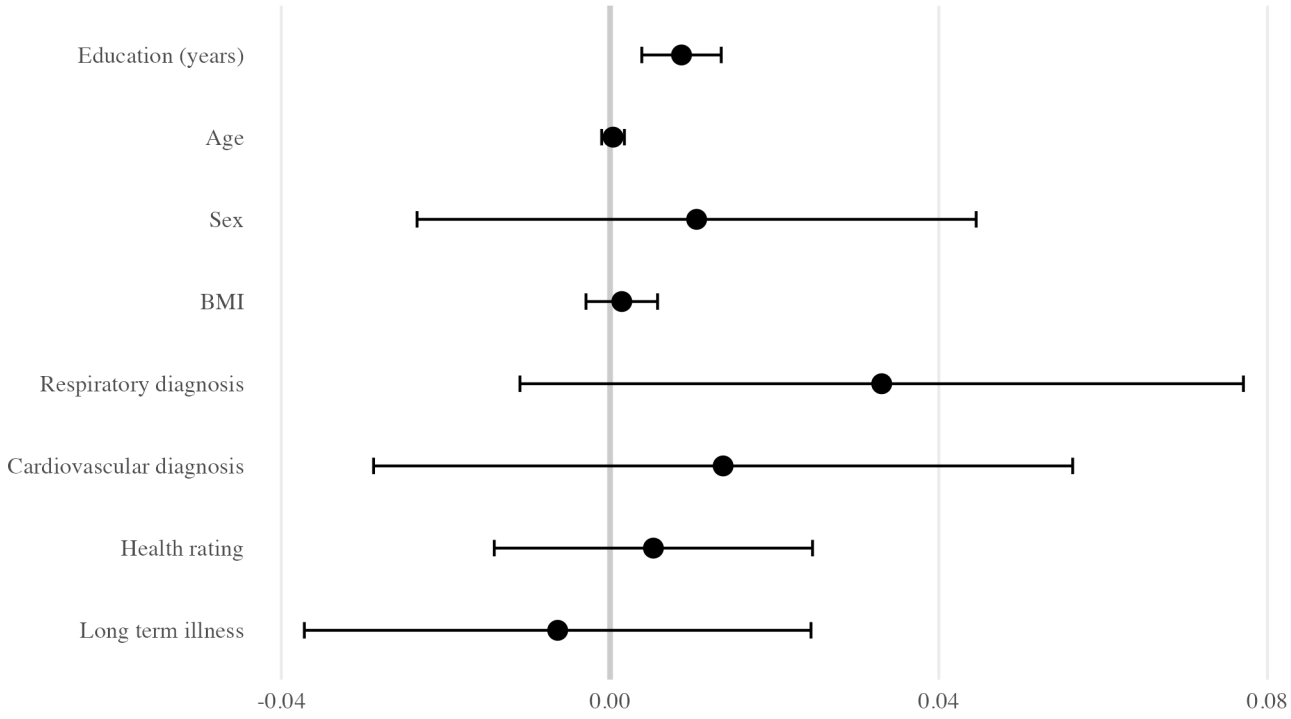


Figure A18: Heterogeneity by demographic and health characteristics - no home distance restriction

The figure plots the coefficient point estimate and the 95% confidence interval of the interaction term between inversion coverage and each demographic/health variable. The sample additionally includes respondents living more than 40km from a UK Biobank assessment center. All estimations include weather controls and year, month, and day-of-week fixed effects. Respiratory diagnosis is equal to 1 if a participant has ever been hospitalized for a respiratory disease. Cardiovascular diagnosis is equal to 1 if a participant has ever been hospitalized for a cardiovascular disease. Health rating is the subjective health rating participants gave during the baseline, and range from 1 to 4, with 1 being excellent health and 4 being poor health. Long term illness is equal to 1 if a participant has a long term illness. Inversion coverage indicates the (weighted) spatial coverage of inversion in the area surrounding an assessment center on a given day, and can range from zero to one. The dependent variable is the fluid intelligence score, standardized to have mean zero and standard deviation one. Robust standard errors clustered at the assessment center level.

B Cognitive tests

B.1 Questions in the fluid intelligence test

The UK Biobank fluid intelligence test measures participants' capacity to solve problems requiring logic and reasoning, independent of acquired knowledge. Participants were given a maximum of two minutes to answer as many questions as possible. The test consisted of 13 questions, covering numerical reasoning, verbal reasoning, logic, and pattern completion. Participants could skip questions if they wished. The fluid intelligence score is constructed as the number of questions answered correctly.

Table B14: UK Biobank fluid intelligence test questions and answer choices

Question	Answer choices
1. Add the following numbers together: 1, 2, 3, 4, 5. Is the answer?	1. 13 2. 14 3. 15 4. 16 5. 17 6. Do not know 7. Prefer not to answer
2. Which number is the largest?	1. 642 2. 308 3. 987 4. 714 5. 253 6. Do not know 7. Prefer not to answer
3. Bud is to flower as child is to?	1. Grow 2. Develop 3. Improve 4. Adult 5. Old 6. Do not know 7. Prefer not to answer
4. 11, 12, 13, 14, 15, 16, 17, 18. Divide the sixth number to the right of twelve by three. Is the answer?	1. 5 2. 6 3. 7 4. 8 5. Do not know 6. Prefer not to answer

Continued on next page

Question	Answer choices
5. If Truda's mother's brother is Tim's sister's father, what relation is Truda to Tim?	1. Aunt 2. Sister 3. Niece 4. Cousin 5. No relation 6. Do not know 7. Prefer not to answer
6. If sixty is more than half of seventy-five, multiply twenty-three by three. If not, subtract 15 from eighty-five. Is the answer?	1. 68 2. 69 3. 70 4. 71 5. 72 6. Do not know 7. Prefer not to answer
7. Stop means the same as?	1. Pause 2. Close 3. Cease 4. Break 5. Rest 6. Do not know 7. Prefer not to answer
8. If David is twenty-one and Owen is nineteen and Daniel is nine years younger than David, what is half their combined age?	1. 25 2. 26 3. 27 4. 28 5. 29 6. Do not know 7. Prefer not to answer

Continued on next page

Question	Answer choices
9. Age is to years as height is to?	1. Long 2. Deep 3. Top 4. Metres 5. Tall 6. Do not know 7. Prefer not to answer
10. 150, 137, 125, 114, 104. What comes next?	1. 96 2. 95 3. 94 4. 93 5. 92 6. Do not know 7. Prefer not to answer
11. Relaxed means the opposite of?	1. Calm 2. Anxious 3. Cool 4. Worried 5. Tense 6. Do not know 7. Prefer not to answer
12. 100, 99, 95, 86, 70. What comes next?	1. 50 2. 49 3. 48 4. 47 5. 46 6. 45 7. Do not know 8. Prefer not to answer

Continued on next page

Question	Answer choices
13. If some flinks are plinks and some plinks are stinks, then some flinks are definitely stinks.	1. False 2. True 3. Neither true nor false 4. Not sure 5. Do not know 6. Prefer not to answer

B.2 Description of secondary cognitive tests

Visual Memory Test

Participants were told that they would play two games of "Pairs," in which they would first see a set of picture cards and should try to remember as many as possible. In the first round, participants were shown three pairs of matching symbol cards arranged in a random pattern for three seconds, as illustrated in B19a. The cards were then turned face down, as shown in B19b, and participants were asked to identify the matching pairs by touching the cards on the screen. The goal was to find all pairs in as few attempts as possible.

After the three-pair round, the task was repeated with six pairs of cards. In this second round, participants were shown twelve cards for five seconds before the cards were turned over.

In each round, the layout of cards was randomly generated in each round, and there was no time limit: participants could continue matching or mismatching cards until all pairs had been found, unless they chose to abandon the task. UK Biobank records the number of rows and columns used in each round, the number of correct matches, the number of incorrect matches, and the time taken to complete each round.

Prospective Memory Test

The UK Biobank prospective memory test is designed to measure whether participants remember to carry out a delayed intention. At the beginning of the cognitive testing section, participants were told that at the end of the games they would be shown four colored shapes and asked to touch the blue square, but that they should instead remember to touch the orange circle, as shown in B20a. After the participant presses "Next," a hidden timer starts, and participants proceed through the Pairs matching test, the fluid intelligence test, and the reaction-time test before the prospective memory task is resumed.

At the end of the cognitive testing sequence, participants are shown a screen with four colored shapes as shown in B20b. The on-screen instruction asks them to touch the blue square and then

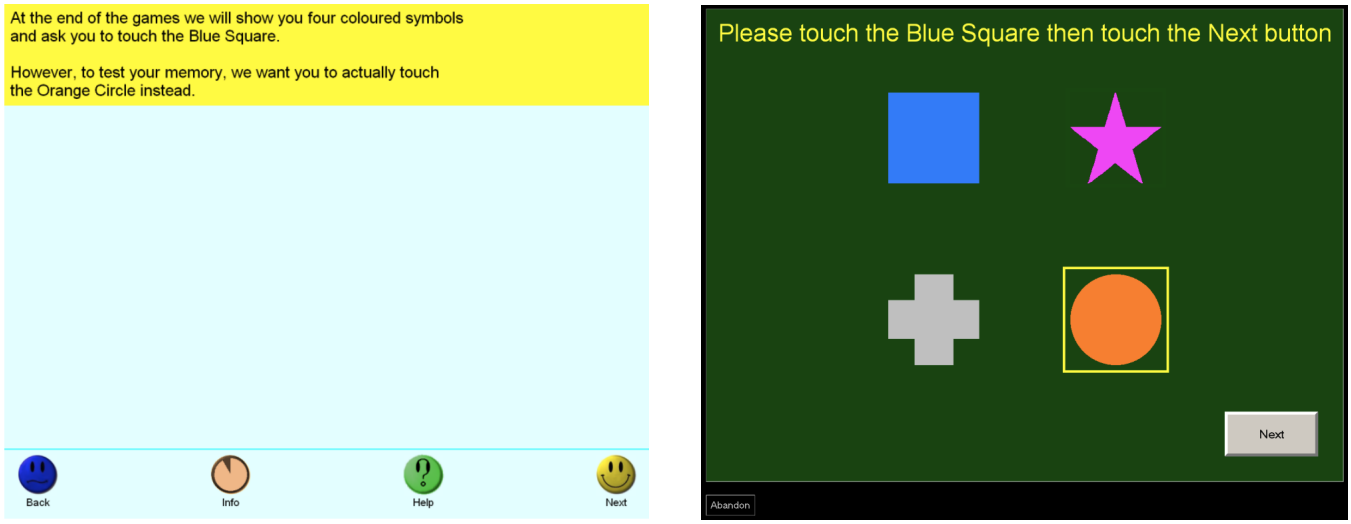


(a) Pairs of matching cards are displayed

(b) Cards turned upside down

Figure B19: Example display of the visual memory test

press "Next." Correct performance therefore requires participants to inhibit the immediate on-screen instruction and instead remember the earlier delayed instruction to select the orange circle. If the participant selects the blue square, they are prompted to remember the earlier instruction; otherwise, the program accepts the selected answer and the test ends. The UK Biobank stores the participant's initial and final answer, whether the final answer was correct, the history and number of attempts, whether the test was completed or abandoned, and timing information.



(a) Display showed prior to the battery of cognitive tests

(b) Display showed at the end of the battery of cognitive tests

Figure B20: Prospective memory test

Numeric Memory Test

The UK Biobank numeric memory test is a touch-screen measure of short-term memory and digit span. Unlike the main cognitive tests used in the baseline assessment, this task was administered only during the pilot phase of recruitment and is therefore available for a substantially smaller subset of participants. The task asked participants to remember increasingly long numbers. Participants were first told that they would be shown a number, that the number would disappear after a short delay, and that they would then have to reproduce the number *in reverse order* using an on-screen number pad. The first number contained two digits. The number was displayed for 2,000 milliseconds plus 500 milliseconds for each digit in the sequence; for example, a two-digit number was shown for 3,000 milliseconds. After the number disappeared, participants waited 3,000 milliseconds before entering the remembered number on the activated screen.

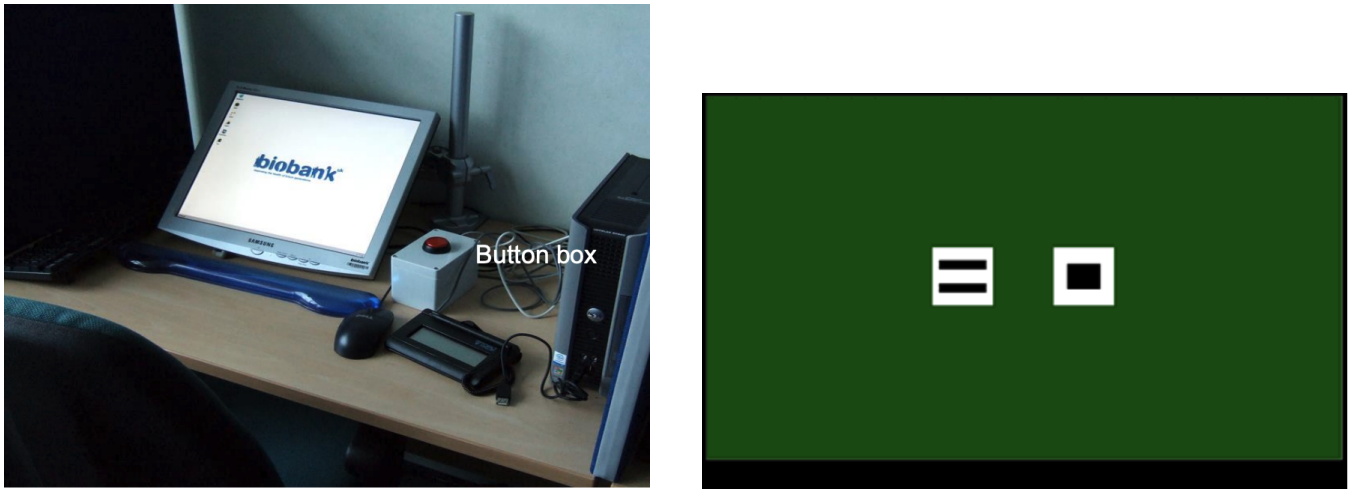
The test became progressively more difficult. It started with a two-digit sequence and increased by one digit each time the participant answered correctly, up to a maximum sequence length of 12 digits. To avoid especially memorable or repetitive sequences, no digit was allowed to be the same as either the immediately preceding digit or the digit before that. The test ended after five successive incorrect answers for two-digit numbers, or after two successive incorrect answers for numbers of three or more digits. UK Biobank records the number of digits shown in each round, the number displayed, the participant's entered response, whether each round was answered correctly, the number of rounds completed, the maximum number of digits remembered, response timing, total completion time, and whether the test was completed, stopped, or abandoned.

Reaction Time Test

The UK Biobank reaction time test is designed to measure simple processing speed. Participants completed the test using their dominant hand, which they held ready over a button box placed at the touch-screen station. Figure B21a shows the button box used in the task. The task presented participants with two cards at a time on the screen. Participants were instructed to press the button as quickly as possible whenever the two cards displayed matching symbols. The task therefore captures the speed with which participants detect a simple visual match and translate that detection into a motor response.

The main version of the test consisted of 12 pairs of cards and did not include an abandon button. UK Biobank records the identity of the two cards shown in each round, the number of times the response button was pressed, the time taken to first press the button in each round, and the derived mean time taken to correctly identify matching pairs. Following the pilot phase of recruitment in 2006, the number of rounds was reduced from 20 to 12 to remove redundancy and shorten the task while retaining the key information from the test. UK Biobank therefore presents pilot and main baseline data separately. The first five rounds are treated as training rounds and

are not included in the derived mean reaction-time measure.



(a) Participant tap on the button upon seeing matching cards

(b) Participants compare two cards shown on the screen

Figure B21: Example set-up and display of the reaction test