

The Lead-Crime Hypothesis: A Meta-Analysis^a

Abstract

Does lead pollution increase crime? We perform the first meta-analysis of the effect of lead on crime, pooling 542 estimates from 24 studies. The effect of lead is overstated in the literature due to publication bias. Our main estimates of the mean effect sizes are a partial correlation of 0.16, and an elasticity of 0.09. Our estimates suggest the abatement of lead pollution may be responsible for 7-28% of the fall in homicide in the US. Given the historically higher urban lead levels, reduced lead pollution accounted for 6-20% of the convergence in US urban and rural crime rates. Lead increases crime, but does not explain the majority of the fall in crime observed in some countries in the 20th century. Additional explanations are needed.

Jel codes: C83, K42, Q53.

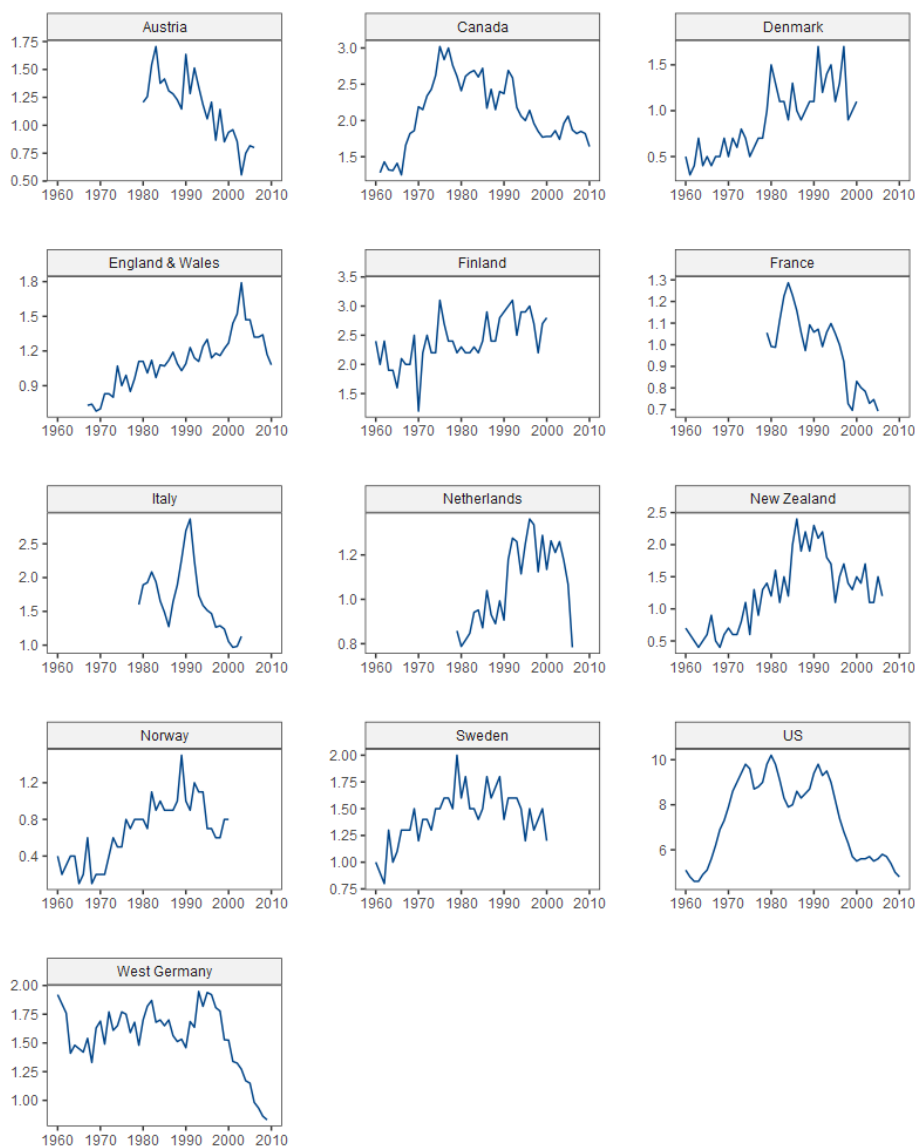
Keywords: Meta-analysis; Publication selection bias; pollution; lead; crime.

^aWe thank Paul Ferraro, Kevin Schnepel, Tom Stanley, Hector Gutierrez Rufrancos, Conny Wollbrant, Jonathan Norris, and two anonymous reviewers for their comments on this paper. We also thank participants at the European Association of Environmental and Resource Economists (EAERE) conference and the Meta-Analysis in Economics Research (MAER) colloquium. Any remaining errors are our own. We are also grateful for the financial support from the UK Economic and Social Research Council.

1. Introduction

Homicide rates spiked and then fell in a consistent pattern across many western countries in the 20th century (figure I). In the US alone the homicide rate has halved since the 1980s, when it was as high as the road fatality rate is today. In other countries the falls are not so great in magnitude, but still amount to many lives saved. If the causes of this fall were known, many more deaths and trauma could be prevented.

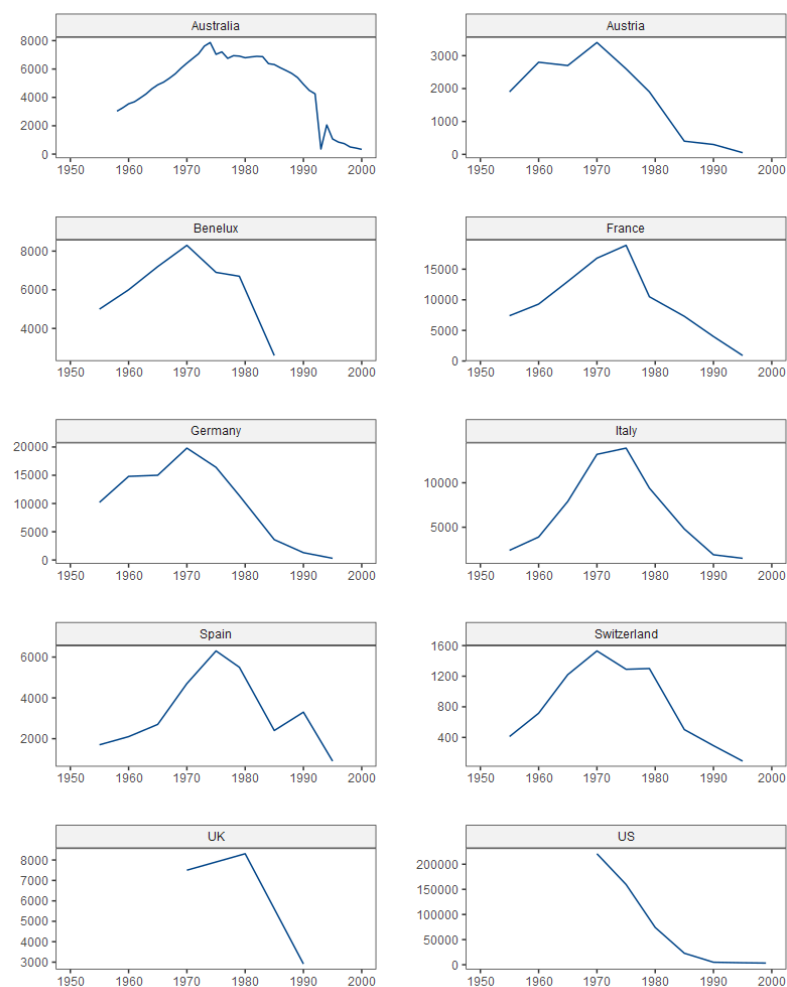
Figure I – Homicide Rate per 100,000 by Country



Sources: New Zealand Police (2018); Buonanno *et al.* (2011), UK Home Office (2019); Uniform Crime Reports for the United States (2019); Falck, Von Hofer & Storgaard (2003); Statistics Canada (2019); Birkel and Dern (2012).

Is lead pollution responsible? Lead is a toxic metal linked to harmful health and behavioural outcomes (see section 2). Studies have pointed to falling lead levels in the environment as a cause of the falls in homicide, and as a factor in reducing crime rates in general. Some have claimed that lead emissions account for as much as 90% of the fall in violent crime (Nevin, 2000, 2007). The reduction in lead pollution over time is largely due to falling emissions from leaded gasoline (figure II), but also due to less lead pollution from water pipes, paint, food, and soil.

Figure II – Lead Emissions by Country (1000 kg Y⁻¹)



Source: Dore *et al.* (2006), Schwikowski *et al.* (2004), Kristensen (2015), Statistical Abstract of the United States (2009).

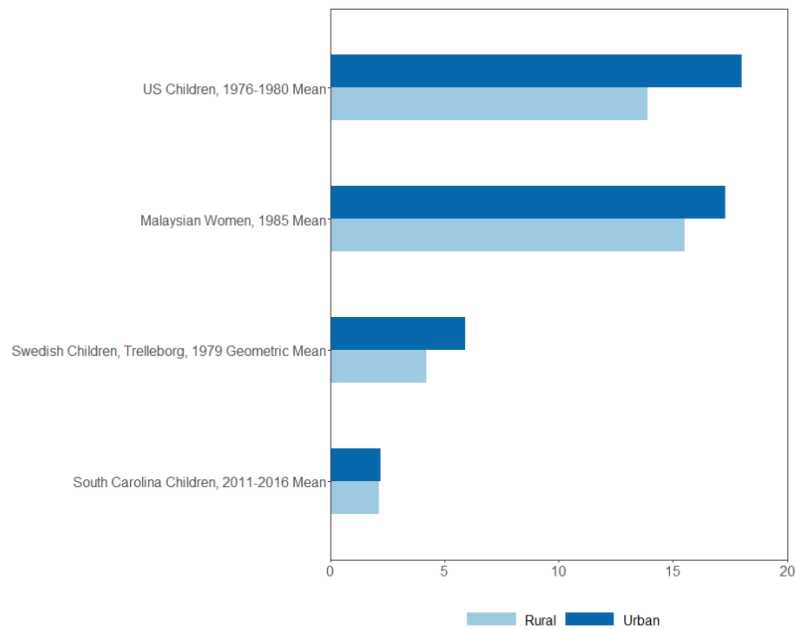
Crucially, this reduction in exposure to lead pollution over time has been spatially uneven. Pollution tends to be more concentrated within urban areas (Carrozzi and Roth,

2020; Borck and Schrauth, 2021) and lead is no exception. The lead burden is likely to be higher in urban areas for several reasons (Levin et al., 2021, O’Flaherty and Sethi, 2015). Urban road traffic is higher, and urban residents often live closer to congested roads, a risk factor before the phase out of leaded gasoline in most countries. Urban dwellers tend to live in closer proximity to lead working sites. Urban areas also have less turnover in soil, which therefore accumulates a larger concentration of lead.

Figure III shows that blood lead levels were generally higher in urban areas than rural areas. Similarly, the left-hand chart in figure IV shows that blood lead levels in urban areas were relatively higher for children under 5 years of age in the US, at least in the period before the phase out of leaded gasoline. In the 1970s and 80s, blood lead levels in Metropolitan Statistical Areas (MSAs) with populations greater than 1 million were 15% higher than levels in other parts of the country. The chart also illustrates the swift convergence in blood lead levels across rural and urban areas in the 1990s. We see a similar pattern in the US crime trends in the right-hand chart of figure IV. The urban crime rate, as measured by the National Crime and Victimization Survey (NCVS) was 70% higher than the rural rate in the early 1990s. There then followed a convergence in crime rates in the 21st century, although the urban rate remains somewhat higher.

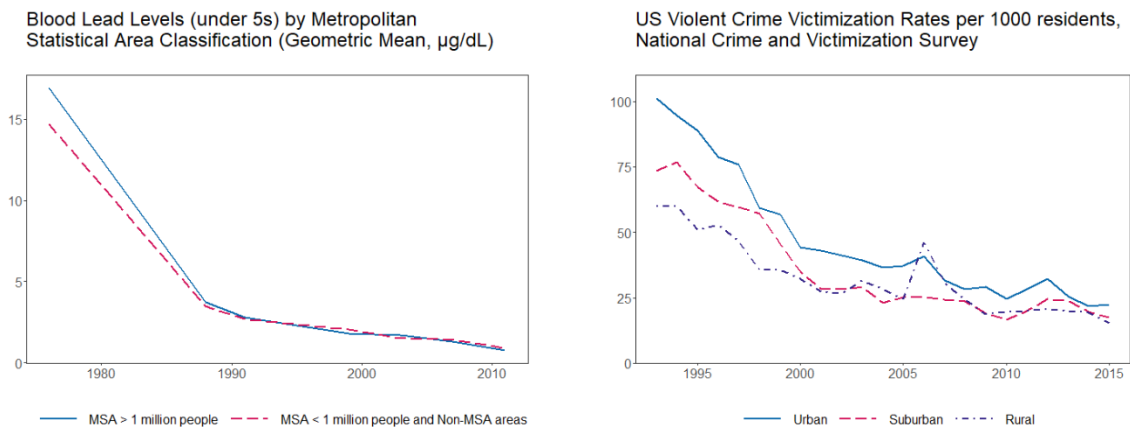
These trends indicate lead *could* explain a large part of the observed variation in crime, both over time and between urban and rural areas. However, the rise and fall pattern in figure I is by no means uniform. Furthermore, Buonanno *et al.* (2011) show that while total crime has behaved similarly to homicide in the US, it has not in Europe (figure V). Similarly, outside the US, population density is associated with lower rather than higher crime rates. Ahlfeldt and Pietrostefani (2019), synthesising the literature on the economic effects of density, estimate that a log-point increase in density is associated with a decline in crime of 0.085 log-points. In the US they find the opposite, density is associated with higher crime.

Figure III – Urban and Rural Blood Lead Levels ($\mu\text{g}/\text{dl}$)



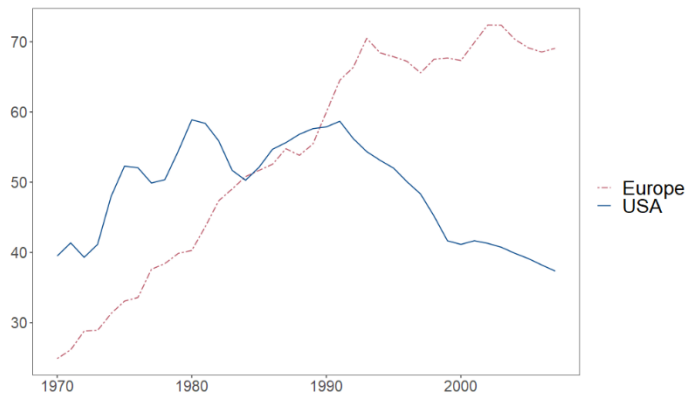
Source: Mahaffey *et al.* (1982), Lim *et al.* (1985), Strömberg Shütz and Skerfving (1995), and Aelion and Davis (2019).

Figure IV – Urban/Rural Blood Lead Levels and Violent Crime in the US



Source: Egan *et al.* (2021) and Bureau of Justice Statistics (2022)

Figure V – Total Recorded Crime Rate per 100,000 in USA and Seven European Countries



Source: Buonanno *et al.* (2011). The countries are: Austria, France, Germany, Italy, The Netherlands, Spain, and the UK.

Alternative hypotheses for the observed fall in crime in some countries range from falling poverty levels (Rosenfeld and Fornango, 2007, and Messner, Raffalovich, and Mcmillan, 2001), to demographic transition, where an ageing population is less likely to be victimised by or engage in crime (Fox, 2005, chap. 9; Baumer, Rosenfeld, and Wolff, 2012), increased/better policing or incarceration (Levitt, 1996, 1997, 2004; Marvell and Moody, 1996; and Corman and Mocan, 2000), to more controversial hypothesis such as legalized abortion reducing the number of children born into “adverse home environments” (Donohue and Levitt, 2001, 2019; Buonanno *et al.*, 2011). Tcherni-Buzzeo (2019) provides a recent summary of potential causes.

Against this background, our paper conducts the first meta-analysis of the effect of lead on crime. We systematically review the literature and construct a dataset containing 542 estimates from a total of 24 studies. We convert these estimates to comparable effect sizes. For this full sample we use partial correlation coefficients. We also convert estimates to elasticities, where it is possible to do so, and analyse this subsample of 312 estimates from 11 papers. Throughout the paper we account for the importance of the research design in identifying credible treatment effects by running separate analyses on the subsample of papers that address sorting or endogeneity bias of lead and crime explicitly. They do this by examining natural experiments where there is plausibly exogenous variation in lead exposure. These studies estimate effects using research designs such as difference-in-difference or instrumental variables. For simplicity, we

label this subsample as the “addressing endogeneity” sample. This subsample consists of 7 studies and 220 estimates when using partial correlation coefficients. The sample declines to 5 studies and 211 estimates when using elasticities.

We perform tests for publication bias and find that the effect of lead on crime is overstated in the literature due to this bias. Furthermore, we find substantial between-study heterogeneity in our sample. We therefore use meta-regression to estimate an average effect size accounting for both publication bias and the observable between-study heterogeneity. We take into account model uncertainty by estimating over 1 million meta-regression specifications, using every combination of our covariates on both the full sample, the elasticity subsample, and several subsamples which exhibit less between-study heterogeneity. We plot the distributions of the estimated average effect size of lead on crime and calculate its mean.

Our main finding is that the estimated mean effect size, evaluated at sample averages, is a partial correlation of 0.16 in the full sample, and an elasticity of 0.09 in the subsample. We also find there are differences between the average effect size when we use the full sample, and when we use only study designs that address endogeneity with quasi-experimental methods. The mean partial correlation coefficient for the “addressing endogeneity” sample is only 0.01, far smaller than the full sample estimate. However, when we use the smaller sample of studies which address endogeneity and have elasticity estimates the mean elasticity range is 0.05-0.17.

We also distinguish between studies in which estimates are based on regional or area-level (e.g., US states) from individual-level data. The sample of studies that use crime in an area as the focus of analysis have a larger mean effect size compared to those of studies which focus on individual behaviour. Conversely, we do not find evidence of differences for the effect of lead on different types of crime when we use homicide, violent, and non-violent crime samples.

Finally, we examine the share of the fall in crime in the late 20th century that lead pollution accounts for. Using the example of homicide in the US, our range of elasticity estimates suggests the fall in blood lead levels is responsible for 4-15 percentage points of the 54% fall in homicide from its peak, with our main estimate being 8. This would mean lead explains around 7-28% of the fall in crime, leaving 93-72% unaccounted for. When we

estimate the share of US urban/rural violent crime convergence explained by falling lead levels, we obtain a figure of 6-20%, with our main estimate being 11%. Our findings suggest that, while the effect of lead pollution on crime is positive, it is not responsible for the majority of the fall in crime observed in some countries in the 20th century, or the majority of the urban/rural crime convergence. Therefore, other explanations require further investigation.

2. Lead and Crime

Lead has long been part of the human environment. It was used in cosmetics, paint, and as coinage in ancient China (Schafer, 1956). Similar uses were recorded in ancient Egypt, India, and across the Bronze Age world (Needleman, 1992). The sweet taste of lead acetate meant that the Roman Empire, and later medieval Europe, used lead to sweeten wine, cider, and food (Lessler, 1988). The Romans had many other uses for lead, using it for cooking utensils, pottery, and water pipes (Hernberg, 2000). Indeed, Roman use of lead was prodigious, with estimates from Greenland arctic ice cores putting the increase in atmospheric lead pollution at around 4000 metric tons a year at its peak 2000 years ago (Hong *et al.* 1994). This is equivalent to the UK's lead pollution emissions in the mid-1980s, when leaded gasoline had not yet been phased out.

Lead is a useful but toxic metal. At high levels of exposure even adults will experience lead poisoning. Acute lead poisoning is rare but can kill quickly. Chronic poisoning can still kill and is associated with abdominal pain, organ failure, tumours, and exhaustion, amongst other symptoms (WHO, 2010a). Although chronic lead poisoning in adults still happens, and appears to affect behaviour, it is primarily the long-term lead exposure of children that is thought to influence crime rates.

Children are especially vulnerable to lead pollution. Children not only absorb more lead per unit body weight than adults, but, as the brain and nervous system are still developing, lead has more harmful long-term effects even at low levels (WHO, 2010b). Lead is chemically similar to calcium. Calcium is important for cell growth, and synaptic functioning, as well as a myriad of other body processes (Sanders *et al.*, 2010). Therefore, lead is particularly harmful to the developing brain and nervous system, and thus in the womb and early infancy are the worst time to be exposed to lead (WHO, 2010b).

The extent to which children have been exposed to lead pollution has varied substantially, both over time and spatially. As detailed in the introduction, for many OECD countries lead air pollution rose sharply in the mid-20th century before peaking in the 70s and 80s (figure II). Children in urban areas tended to have higher blood lead levels during this period (figures III and IV). The highest average blood lead levels for children today are in low and middle income countries, with one estimate putting the share of children with elevated blood levels (above 5µg/dL) at one third (GBD, 2019).

Yet even today, in countries that have reduced blood lead levels, there remain pockets with higher pollution. Cities with low pH water supplies tend to have higher lead levels if they also have lead pipes, because the water reacts more strongly upon the lead piping. Feigenbaum and Muller (2016), using distance to a lead refinery as an instrument, find these cities to have higher homicide rates in the early 20th century. Aizer and Currie (2019) find that blood lead levels are higher for those living near a road, but this only applies in the period before the phasing out of leaded gasoline. Tanaka *et al.*, (2022) show that pollution around lead-acid battery recycling plants in the US sharply reduced after an air-quality law was introduced in 2009, but this led to offshoring of lead battery recycling to Mexico. Infants living near the Mexican plants began to experience worse health outcomes as a result. Highly concentrated lead pollution, and higher blood lead levels, have also been found near airports (Zahran *et al.*, 2017), lead smelters (Stromberg *et al.*, 1995) and NASCAR racetracks (Hollingsworth and Rudik, 2021). This inequality in lead exposure means that any effect of lead on crime will also be spatially uneven.

The causal chain of lead to crime starts with the biological changes it induces at this young age. The mechanism for these changes is laid out in Sanders *et al.* (2010), and there is an array of evidence for lead's negative effects. These include impaired nerve conduction (Sindhu and Sutherling, 2015), damaged myelination in the nerve system (Brubaker *et al.*, 2009), impeded brain development (Lanphear, 2015), and reduced brain matter (Cecil *et al.*, 2008).

The next link in the chain is from biological change to behavioural change in later life. Meta-analyses have found that lead exposure is associated with aggressiveness and other conduct problems (Marcus, Fulton, & Clarke, 2010), lower IQ (Schwartz, 1994), and impaired cognitive functioning (Vlasak *et al.*, 2019, and Seeber *et al.*, 2002).

The final link is from behavioural changes to an increased propensity to commit crime. There are several possible mechanisms. Needleman pioneered research on lead exposure and aggressiveness (1996), suggesting it is linked to violent crime in particular. In contrast, Denno (1990) and Fergusson, Boden and Horwood (2008) argue that the link is through lower education outcomes, leading to worse life outcomes, which causes increased criminality. This mechanism is consistent with Becker's (1968) economic theory of crime, where lower opportunity cost makes crime relatively more attractive, and suggests lead would show a stronger link to property crime than violent crime. A third mechanism was proposed by Gottfredson and Hirschi (1990), where lack of self-control, combined with opportunity, causes higher crime rates. Lead has been associated with increases in impulsivity (Winter and Sampson, 2017), and so may cause an increase in crime through this process. If this mechanism were true we might expect increases in violent crime, non-violent crime, or both. Separating the different types of crime may help identify which, if any, mechanism lead acts through. However, whilst a range of mechanisms have been laid out linking lead in the environment to the propensity to commit crime, the strength of this link is a matter of empirical enquiry. The main objective of this paper is to quantify the strength of this link from the range of empirical work reported to date. To do this, we use meta-analysis.

3. Data

Meta-analysis data collection begins by specifying the criteria which studies must fulfil to be accepted into the analysis.

The criteria we chose were:

1. The explanatory variable must be some quantitative measure of lead exposure.
2. Outcome variable must measure crime in some way (i.e. not other types of behaviour such as aggressiveness or depression).
3. Must have original estimates, i.e. no review papers.
4. Must have estimates that can be combined into a meta-analysis.
5. Be published before December 2019.
6. Study must be available in English.

We then undertook a systematic literature review for papers on Web of Science, PubMed, and Google Scholar in 2019. We also searched on NBER and REPEC for working papers to include as much “grey” literature as possible. The keyword combinations used were:

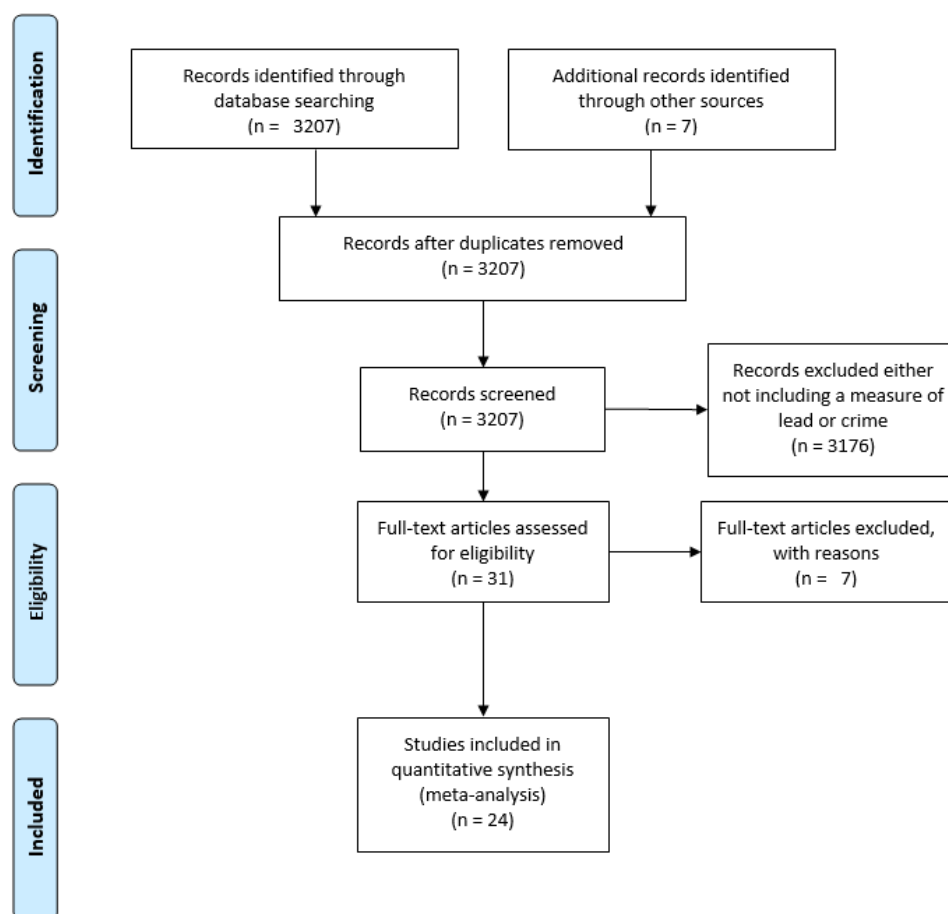
“lead”, or “lead” AND “pollution”, or “lead” AND “poisoning”, or “lead” AND “exposure”, or “lead” AND “blood”, or “lead” AND “air”, or “lead” AND “paint”, or “lead” AND “water”

Combined with:

“crime” or “conviction” or “arrest” or “jail” or “prison”

After searching, papers were screened to see if they fulfilled the criteria, as laid out in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram (figure VI). A review and description of the studies included is given in appendix A.

Figure VI – PRISMA Flow Diagram of Studies Selection Process



The vast majority of the studies identified in the literature review did not fulfil criteria one or two and therefore did not estimate the lead-crime relationship. These were then filtered out at the screening stage. 31 papers did estimate the lead-crime relationship, but 7 of these could not be converted into comparable effect sizes, failing criterion four. Criterion four is needed because estimates must be combined in a meta-analysis. Estimates are made comparable by converting into a common metric, such as the partial correlation coefficient (PCC), or an elasticity. Most regression coefficients and simple correlations can be converted into PCCs easily. Odds ratios and standardised mean differences can also be converted into PCCs. However, five papers used risk ratios (Boutwell *et al.*, 2016; Boutwell *et al.*, 2016; Haynes *et al.*, 2011; Stretesky and Lynch, 2001; and Write *et al.*, 2008). Risk ratios can be converted into odds ratios, which can then be converted to PCCs, but need a base rate risk to do so. It was not possible to infer a base rate risk from the data available in the papers. Therefore, these papers were excluded at the eligibility stage. One other paper (Masters and Coplan, 1999) contained charts but not enough information to make PCCs and was excluded. Similarly, Denno (1990) did not have enough information to use the estimates. No papers were excluded based on criterion six, but search terms were only in English. This left 24 papers in the final meta-analysis dataset.

We organised accepted papers into a dataset following the guidelines for meta-analysis in economics in Havránek *et al.* (2020). Every paper gave multiple estimates for the effect of lead on crime. Meta-analyses tend to either select one estimate from each study as a “representative” estimate; or take all estimates and account for the potential clustering of estimates from the same study. Both are defensible. Taking all estimates means more information available for the meta-analysis. Representative estimates, on the other hand, may be less biased. For example, a researcher may show a simple OLS estimate before giving reasons for why it will be biased. They then go on to use their preferred method of estimation, which attenuates this bias. In most of our analysis we use all estimates from the studies, but as a robustness check we also test our results by using one representative estimate from each study in appendix E. The results are similar.

In the full sample, there are 542 estimates from the 24 studies. The dataset forms an unbalanced panel, with each estimate being an observation and observations grouped by

study. The studies included span across a variety of disciplines including economics, sociology, medicine, epidemiology, and criminology.

Study effect sizes were then converted to the common effect size. Conversion is necessary because both lead and crime are measured in different ways in each paper, and therefore must be converted to be comparable. All studies in the full sample could be converted to PCCs. See appendix B for more details of how PCCs and the PCC standard errors are calculated.

PCCs measure the correlation between two variables holding other variables in the model constant. Their sizes are not intuitive. They have no unit and cannot be interpreted quantitatively in a meta-analysis with varied measurements of outcome (Doucouliagos, 2011). However, as they are bounded from -1 to 1, they do offer a sense of the magnitude and direction of an effect. In a survey of economic effect sizes Doucouliagos (2011) offers the following rough guidelines: 0.07-0.17 is a small effect size, 0.18-0.33 is a moderate one, and above 0.33 a large one. For most of the paper, we follow this taxonomy, but a small effect combined with a large absolute change in a variable can still mean it is significant for welfare.

We were also able to convert some study estimates into elasticities. The elasticities measure the percent change in some measure of crime, given a percent change in some measure of lead pollution. They provide a better measure of the real effect rather than the measure of statistical strength the PCCs provide. The trade-off is that the sample is smaller and therefore may be less representative of the literature. There are 11 studies and 312 estimations in what we label for simplicity the “elasticity subsample”.

Table I presents the mean, median and weighted average PCC for each study (with weights being equal to the precision, $1/\text{standard error of the PCC}$). It also includes some information on the characteristics of each study. We do the same for the elasticity sample in table II.

Table I – Partial Correlation Coefficients from the Studies Used in Full Sample Meta-analysis

Study & Year	Median	Mean	Weighted Average	Type of Crime	Individual or Area-level	Addresses Endogeneity
Aizer & Currie (2019)	0.027	0.019	0.019	Violent and non-violent	Individual	Yes
Barrett (2017)	0.556	0.556	0.589	Violent	Area	No
Beckley <i>et al.</i> (2018)	0.065	0.061	0.063	Violent and non-violent	Individual	No
Billings & Schnepel (2018)	0.122	0.113	0.103	Violent and non-violent	Individual	Yes
Curci & Masera (2018)	0.027	0.043	0.029	Violent	Area	Yes
Dills, Miron & Summers (2008)	0.022	0.021	0.021	Violent and non-violent	Area	No
Feigenbaum & Muller (2016)	0.054	0.056	0.053	Only Homicide	Area	Yes
Fergusson <i>et al.</i> (2008)	0.080	0.079	0.080	Violent and non-violent	Individual	No
Grönqvist, Nilsson and Robling (2019)	0.002	0.003	0.003	Violent and non-violent	Individual	Yes
Lauritsen <i>et al.</i> (2016)	0.740	0.495	0.742	Violent and non-violent	Area	No
Lersch & Hart (2014)	0.043	0.043	0.043	Violent and non-violent	Area	No
Manduca & Sampson (2019)	0.087	0.087	0.087	Violent and non-violent	Individual	No
Masters <i>et al.</i> (1998)	0.051	0.061	0.061	Violent and non-violent	Area	No
McCall & Land (2004)	-0.017	-0.017	-0.017	Only Homicide	Individual	No
Mielke & Zahran (2012)	0.526	0.497	0.515	Violent	Area	No
Needleman <i>et al.</i> (2002)	0.336	0.307	0.324	Non-violent	Individual	No
Nevin (2000)	0.914	0.912	0.937	Violent	Area	No
Nevin (2007)	0.808	0.710	0.874	Violent and non-violent	Area	No
Nkomo <i>et al.</i> (2017)	0.004	0.052	0.088	Violent	Individual	No
Reyes (2007)	0.059	0.053	0.053	Violent and non-violent	Area	Yes
Reyes (2015)	0.026	0.036	0.029	Violent and non-violent	Individual	Yes
Sampson and Winter (2018)	-0.065	-0.046	-0.046	Violent and non-violent	Individual	No
Stretesky & Lynch (2004)	0.396	0.352	0.331	Violent and non-violent	Area	No
Taylor <i>et al.</i> (2018)	0.371	0.377	0.429	Violent	Area	No

Notes. Table shows median and mean partial correlation coefficient (PCC) estimates from each study of the effect of lead on crime. These averages are computed from 542 estimates from 24 studies used for the full sample meta-analysis. It also shows an average where estimates are combined in a weighted average with the weights equal to one divided by the standard error. Table also shows what type of crime was used as dependent variable in each study, whether the study unit of interest was an individual or a geographic area, and whether any estimates in the study used a design that attempted to account for endogeneity. All coding is done at an estimate level, so a study may include both “addresses endogeneity” and “correlational” estimates, violent and non-violent estimates etc.

Table II – Estimated Elasticities in Studies Used in Elasticity Subsample Meta-analysis

Study & Year	Median	Mean	Weighted Average	Type of Crime	Individual or Area-level	Addresses Endogeneity
Barrett (2017)	0.68	0.68	0.61	Violent	Area	No
Curci & Masera (2018)	0.20	0.22	0.12	Violent	Area	Yes
Feigenbaum & Muller (2016)	0.72	0.73	0.32	Only Homicide	Area	Yes
Fergusson et al. (2008)	2.45	2.14	0.94	Violent and non-violent	Individual	No
Grönqvist, Nilsson and Robling (2019)	0.04	0.06	0.07	Violent and non-violent	Individual	Yes
Mielke & Zahran (2012)	0.53	0.53	0.48	Violent	Area	No
Reyes (2007)	0.74	0.61	0.29	Violent and non-violent	Area	Yes
Reyes (2015)	0.50	0.64	0.40	Violent and non-violent	Individual	Yes
Sampson and Winter (2018)	-0.22	-0.29	-0.12	Violent and non-violent	Individual	No
Stretesky & Lynch (2004)	0.15	0.15	0.15	Violent and non-violent	Area	No
Taylor et al. (2018)	0.24	0.25	0.26	Violent	Area	No

Notes. Table shows median and mean elasticity estimates from each study of the effect of lead on crime. These averages are computed from 312 estimates from 11 studies used for the “elasticity” subsample. It also shows an average where estimates are combined in a weighted average with the weights equal to one divided by the standard error. Table also shows what type of crime was used as dependent variable in each study, whether the study unit of interest was an individual or a geographic area, and whether any estimates in the study used a design that attempted to account for endogeneity. All coding is done at an estimate level, so a study may include both “addresses endogeneity” and “correlational” estimates, violent and non-violent estimates etc.

4. Methods and Results

4.1 General Approach²

Let θ_j be an effect size of interest in study j . Study j uses some method to estimate θ_j and these we denote as $\hat{\theta}_{ij}$, for estimate i of study j . Researchers are often interested in both how close $\hat{\theta}_{ij}$ is to θ_j (internal validity), and in how useful θ_j would be in predicting results from a similar event or study. This can be interpreted as the degree of external validity of a study.

²This section owes much to the excellent expositions in Meager (2019), Rubin (1981), and Röver (2018). Much of their explanation deals with Bayesian methods but works equally well for non-Bayesian methods up to the point we arrive at.

If θ_j is a draw from some distribution with a likelihood function $\psi(\cdot | \Theta)$ such that $\theta_j \sim \psi(\cdot | \Theta) \forall j$, then there exists some parameter(s) Θ which can give information about a new draw θ_{j+1} from that distribution. It is the parameters contained in Θ that are estimated in a meta-analysis. There may be several parameters of interest, but in practice meta-analyses usually estimate two: θ , the mean of the distribution, and the variance τ^2 . This is because meta-analyses tend to impose the assumption $\theta_j \sim N(\theta, \tau^2) \forall j$ in the interests of efficient estimation. Even if this is not the true shape of the distribution McCulloch and Neuhaus (2011) show, both in theory and simulation, that maximum likelihood estimates are robust to different distributions of θ_j around θ . If we also assume, as the individual studies themselves usually do, that $\hat{\theta}_{ij}$ follows a normal distribution with mean θ_j and variance σ_{ij}^2 , then this leads to the normal-normal hierarchical model of Rubin (1981):

$$\begin{aligned}
 (1) \quad & \theta_j \sim N(\theta, \tau^2) \forall j \\
 (2) \quad & \hat{\theta}_{ij} \sim N(\theta_j, \sigma_{ij}^2) \forall i \text{ and } \forall j \\
 (3) \quad & \hat{\theta}_{ij} | \theta, \sigma_{ij}^2, \tau^2 \sim N(\theta, \sigma_{ij}^2 + \tau^2) \forall i \text{ and } \forall j
 \end{aligned}$$

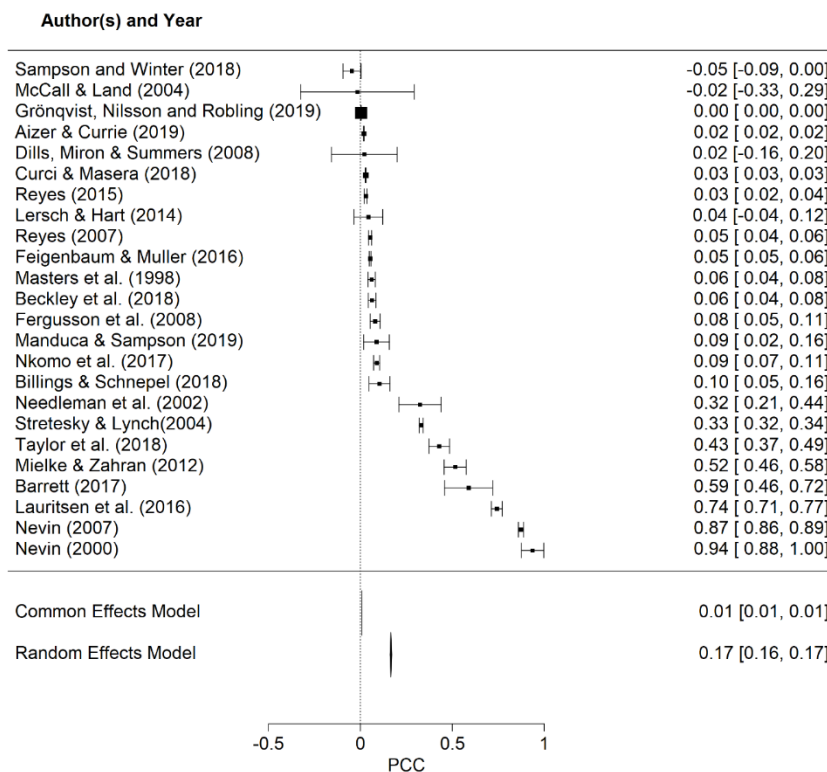
where the last expression follows from the previous two but is expressed in marginal form, as in Röver (2018). This marginal form can be further extended to be conditional on observable variables, common across the $\hat{\theta}_{ij}$'s, as we do in our meta-regression analysis.

The variance of the effect size distribution τ^2 is a crucial measure of how useful aggregation of estimates will be. If τ^2 is zero, then all studies are estimating the exact same effect and it is only the study variances that affect how well they can predict θ_{j+1} . This we call the common effect model following the Rice, Higgins, and Lumley (2018) terminology. As τ^2 grows larger, aggregation becomes less useful. $\tau^2 \rightarrow \infty$ represents an “apples and oranges” comparison where meta-analysis should never be undertaken.

4.2. Between-Study Heterogeneity

We begin investigating between-study heterogeneity in effect sizes by plotting each study's weighted average PCC along with their 95% confidence intervals in figure VII and doing the same with the elasticities in figure VIII.

Figure VII – Forest Plot, Partial Correlations

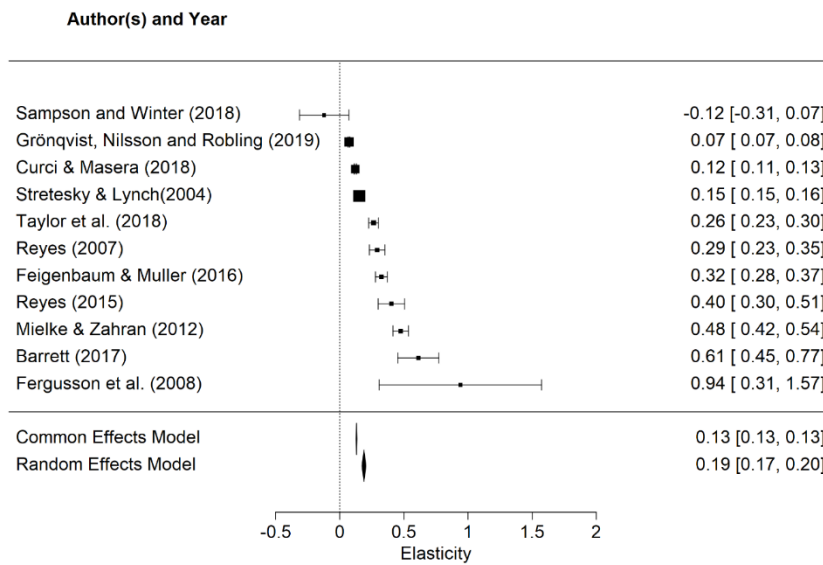


Notes. Chart shows weighted average partial correlation coefficients (PCCs) of each study's effect size along with corresponding 95% confidence intervals. The weighted averages are calculated by first normalizing the PCCs so that confidence intervals can be constructed, then the fixed effects average is calculated, finally the estimates are converted back to PCCs (see appendix B for details). Bottom of table shows common effects and random effects estimates for all studies combined (see appendix C for details). Numbers on right are the point estimates and the 95% confidence intervals.

We show the common and random effects estimates at the bottom of each figure. Both estimates are weighted averages, where more precise estimates get more weight. However, the random effects estimate will give more equal weight to each study the larger the estimate of between-study heterogeneity. See appendix C for more details on

the calculations. The PCC common effects point estimate is 0.01 and the random effects 0.17, while the elasticity common effects estimate is 0.13, and the random effects is 0.19. The difference between the common and random effects estimates indicates that between-study heterogeneity is important, as the lower the estimated heterogeneity between studies, the closer the random effects estimate will be to the common effects.

Figure VIII – Forest Plot, Elasticities



Notes. Chart shows weighted average of each study's effect sizes converted to elasticities along with corresponding 95% confidence intervals. Bottom of table shows common effects and random effects estimates for all studies combined (see appendix C for details). Numbers on right are the point estimates and the 95% confidence intervals.

It is unlikely that the only source of this heterogeneity is the random, unobservable variances σ_{ij}^2 and τ^2 . Distribution (3) can be extended to be conditional on a $1 \times K$ vector of variables \mathbf{x}_{ij} . In this case the study specific estimates θ_j are a function of this variation in \mathbf{x} and we have the conditional distribution:

$$(4) \quad \hat{\theta}_{ij} \mid \sigma_{ij}^2, \tau^2, \mathbf{x}_{ij}, \boldsymbol{\beta} \sim N(\mathbf{x}_{ij}'\boldsymbol{\beta}, \sigma_{ij}^2 + \tau^2) \forall i \text{ and } \forall j$$

If these variables are observable, we can include them in our estimation. To investigate sources of observable between-study heterogeneity, table III splits the data into further sub-samples, based on common characteristics. These characteristics are also used as covariates in the meta-regression analysis and described fully in section 4.4. We then

compare three measures of between-study heterogeneity for each sample, $\hat{\tau}^2$, \hat{I}^2 , and \hat{H}^2 . For each of these measures, the higher they are, the higher the estimated between-study heterogeneity.

$\hat{\tau}^2$ is an estimate of the variance of the effect size distribution in (3) using the DerSimonian-Laird (1986) method. It is measured in the same units as the effect sizes, which is either PCCs or elasticities in our analysis. The larger is $\hat{\tau}^2$ then the greater the dispersion of the “true” effect sizes each study is attempting to estimate.

\hat{I}^2 is an estimate of the proportion of observed variance between effect sizes that is due to effect size heterogeneity, as opposed to sampling variation. It is a figure between 0% and 100%. If 100%, it means all the observed variation is due to between-study effect size heterogeneity. If 0% it means the effect being estimated is homogeneous between studies, and all observed variation is due to sampling error.

\hat{H}^2 is more complicated to interpret. It is the residual standard deviation from regressing the t-statistic of each effect size on its precision. \hat{H}^2 of 1 means that all studies are estimating the exact same effect. The larger \hat{H}^2 is, the greater the between-study effect size variation.

\hat{H}^2 and \hat{I}^2 are sensitive to the number of estimates and the variation in the standard error of those estimates. \hat{I}^2 tends to 100 as the number of estimates included increases. $\hat{\tau}^2$ is less sensitive to the number of studies used in the analysis compared to \hat{I}^2 and \hat{H}^2 , but it does not give a sense of how important between-study heterogeneity is compared to within-study sampling variation.

Table III – Random Effects and Heterogeneity Estimates by Subsample

Sample	RE Estimate	SE	$\hat{\tau}^2$	\hat{I}^2	\hat{H}^2	Studies	Estimates (N)
Full Sample	0.166	0.002	0.002	99	108	24	542
Addressing Endogeneity	0.014	0.001	0.000	90	10	7	220
Correlational	0.505	0.014	0.059	99	159	20	322
Individual-level	0.008	0.001	0.000	95	20	11	125
Area-level	0.388	0.010	0.033	99	123	13	417
Homicide	0.172	0.012	0.010	94	18	8	103
Violent Crime	0.261	0.008	0.016	99	72	18	339
Non-Violent Crime	0.492	0.040	0.120	99	145	15	82
Total Crime	0.077	0.003	0.001	99	152	11	119
North America	0.217	0.006	0.011	98	58	19	386
Europe	0.069	0.003	0.001	100	201	2	85
Direct Lead Measure = TRUE	0.092	0.026	0.031	95	19	9	54
Direct Lead Measure = FALSE	0.171	0.002	0.002	99	118	15	488
Representative Estimate = TRUE	0.186	0.020	0.006	98	54	24	24
Representative Estimate = FALSE	0.167	0.002	0.002	99	111	24	518
Control Gender = TRUE	0.007	0.001	0.000	95	20	8	103
Control Gender = FALSE	0.355	0.007	0.017	99	123	18	439
Control Race = TRUE	0.084	0.008	0.005	97	29	13	114
Control Race = FALSE	0.190	0.003	0.002	99	128	14	428
Control Income = TRUE	0.028	0.002	0.000	97	31	13	174
Control Income = FALSE	0.399	0.008	0.016	99	139	16	368
Control Education = TRUE	0.006	0.001	0.000	95	19	11	106
Control Education = FALSE	0.345	0.007	0.015	99	124	17	436
Elasticity Sample*	0.189	0.008	0.010	91	12	11	312
Elasticity Sample (Addressing Endogeneity)*	0.198	0.012	0.016	88	8	5	211

Notes. *Indicates values are in elasticities, not PCCs. RE Estimate is a random effects, meta-analysis estimate computed using DerSimonian-Laird (1986) method. SE is the standard error of the RE estimate. τ^2 , I^2 , and H^2 are estimates of between-study heterogeneity. See section 4.2 for more details.

Looking at table III we can see which variables seem important for heterogeneity and the different estimated average effect sizes. The PCC subsample of studies which control for endogeneity has a lower estimated heterogeneity and a smaller effect size compared to the correlational sample. Endogeneity can arise from unobserved variables correlated with both crime and lead. These could bias upwards the estimate of the effect of lead on crime. We cannot rule out that these variables may cause individuals both to commit more crime and be more exposed to lead, rather than lead being the cause. Therefore, the difference between the “addressing endogeneity” sample and the full

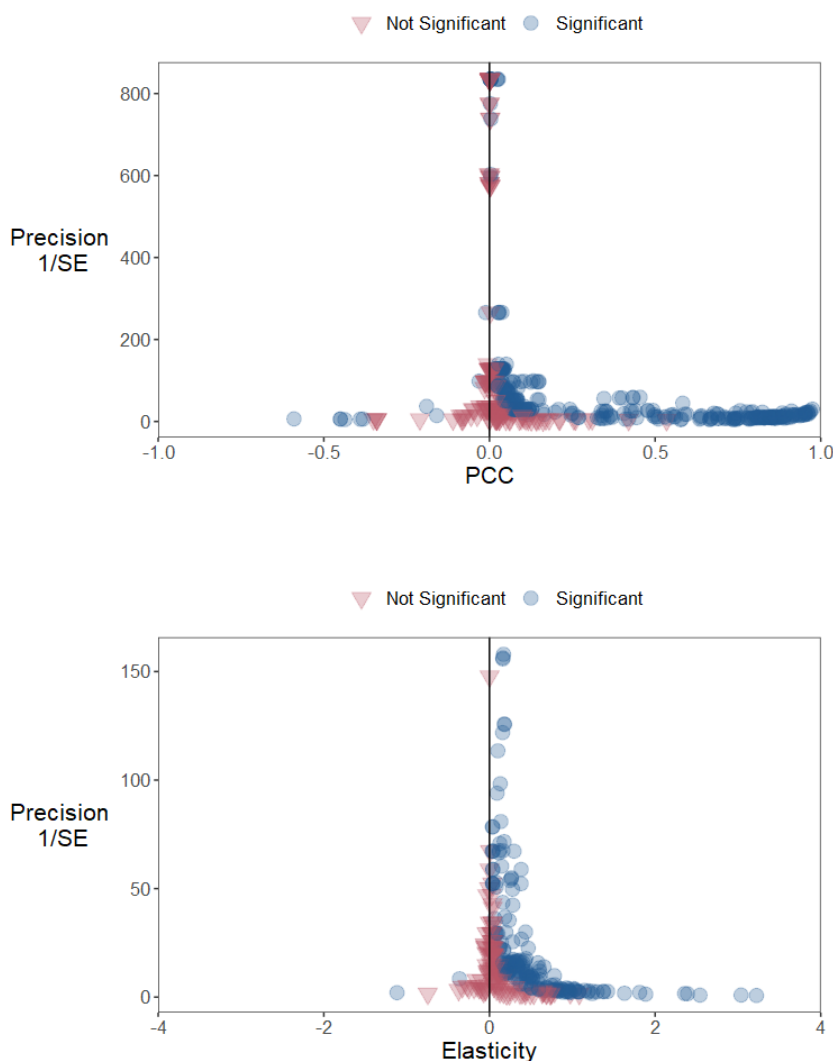
sample could be related to these factors. The elasticity subsample also shows lower heterogeneity than the full sample, but is similar to the “elasticity and addressing endogeneity” sample.

Studies that look at individual-level data on the propensity to commit crime have lower estimated heterogeneity and estimated effect size compared to studies that look at crime committed within a geographic area. Studies which use homicide as the dependent variable appear to have less heterogeneity and find a smaller effect size. This reduction in heterogeneity may be due to lower measurement error in homicide data compared to other types of crime, combined with more similar classification of this crime across countries, and therefore less noise in the data. Finally, when race, gender, education, and income covariates are included in an estimation, these tend to lower the effect size. These subsamples also show less between-study heterogeneity than those which do not include these covariates. The estimated differences in effect size and heterogeneity between subsamples indicates observable variation is important and must be considered when we estimate an average effect. We incorporate the observable variation indicated in table III into our meta-regression analysis in section 4.4. A further, and common source of heterogeneity in effect sizes in meta regression analysis comes from publication bias. We investigate this in the next section.

4.3. Publication bias

Publication bias is a well-known problem across disciplines (see for example: DeLong and Lang, 1992; Ioannidis, 2005; Ioannidis, Stanley and Doucouliagos, 2014; and Ferraro and Shukla, 2020). Papers which contain statistically significant effect sizes are more likely to be published than those which show no effects, or those which contain counter-intuitive results (also known as the bottom-drawer problem). It is standard practice to test for the presence of publication bias in meta-analysis.

Figure IX – Funnel Charts



Notes. PCC = Partial Correlation Coefficient. Precision is one divided by the standard error. “Significant” means statistically significant at the 95% confidence level using two-sided critical values of a normal distribution.

The first and most common step is to simply chart the data and visually inspect for bias, using a funnel plot. Figure IX plots effect sizes against their precision. The upper funnel shows the PCCs for the full sample, and the lower the elasticities for that subsample. A funnel with no bias should be symmetrical around central tendencies. The estimates will tend to spread out as the precision decreases, but they should do so symmetrically if this is only due to sampling noise. Figure IX shows a pronounced asymmetry in the estimates, suggesting there may be a positive bias. There appears to be less asymmetry in the elasticity panel. This suggests these studies may be more similar, and/or have less bias. Some of the studies with the largest effect sizes did not report enough information

for elasticities to be calculated, which may be the reason for this. Although there is asymmetry in both panels, suggesting publication bias, it is also possible this is due to heterogeneity within the sample. We explore this possibility in section 4.4.

More formal testing of publication bias is also possible. There are many tests for publication bias. We use seven methods, which we split into linear and non-linear methods. Linear tests involve regressions of a measure of sampling uncertainty on the estimated effect. A linear relationship between the estimate and its standard error, as figure IX implies, would indicate the presence of publication bias (see appendix D). This naturally leads to the estimating equation (5).

$$(5) \quad \hat{\theta}_{ij} = \theta + \beta_F \hat{\sigma}_{ij} + u_j + \epsilon_{ij}; \text{ where } \epsilon_{ij} \sim N(0, \sigma_{ij}^2) \text{ and } u_j \sim N(0, \tau^2)$$

This is the combined Funnel Asymmetry Test (FAT) and Precision Effect Test (PET). Here the FAT is β_F , and is an estimate of the size and sign of publication bias. It is a function of the inverse Mills' ratio. If positive then estimates that are positive are more likely to be published than negative ones. This test also gives an estimate of θ that takes into account this bias, called the PET. Equation (5) nests the common effects model where τ^2 is zero.

The test in (5) would be subject to heteroskedasticity, as can be observed from figure IX. We have estimates of the heteroskedasticity in $\hat{\sigma}_{ij}$. These can therefore be used to weight the regression and we estimate the test with weighted least squares following Stanley (2008).

$$(6) \quad \hat{t}_{ij} = \theta \frac{1}{\hat{\sigma}_{ij}} + \beta_F + v_j + e_{ij}$$

Here the dependent variable \hat{t}_{ij} is now the t-ratio, rather than the estimate alone. The intercept of the regression is the FAT and the coefficient on $\frac{1}{\hat{\sigma}_{ij}}$ is the PET.

We estimate four variations of linear publication bias tests. First with OLS and clustered standard errors by study, but no study fixed effects; second, a variation of this where we regress on the variance rather than the standard error (Stanley and Doucouligas, 2014); third a full hierarchical FAT-PET with study fixed effects. We estimate this with

restricted maximum likelihood (REML), as Monte Carlo simulations suggest REML performs well for unbalanced panels (Baltagi, Song and Jung, 2000). Finally, we use the square root of the sample size as an instrumental variable for the precision. This last method allows for the fact some estimation techniques may be less efficient but lead to unbiased estimates.

We also run three non-linear methods. The Weighted Average of Adequately Powered Estimates (WAAP) of Stanley, Doucouliagos, and Ioannidis (2017) estimates which studies are post-hoc “adequately powered” and only uses these to calculate an average effect size. The Trim and Fill (TF) method (Duval & Tweedie, 2000) adds imputed studies on the sparse side of the funnel before calculating an average effect. The Andrews and Kasy (2019) method reweights all observations by estimated relative publication probabilities and calculates an average effect size after reweighting. See appendix D for a full discussion of all methods.

Table IV shows the results of all tests. We estimate the tests with four different samples. Panel A is the full sample using PCCs, panel B is all studies which address endogeneity (PCCs), Panel C is only studies with elasticity estimates available, and panel D is studies which both address endogeneity and have elasticities. Linear methods allow for not only an effect beyond bias estimate but an indication of the strength of bias in the FAT coefficient. In all four panels every estimate of publication bias is positive, indicating positive estimates are more likely to be published. Only the FAT-PEESE estimate in panel B, and the IV estimate in panel C, have 95% intervals that cover zero. In every panel, the effect beyond bias estimates are all smaller than the random effects estimate of table III, indicating the effect size is overstated due to publication bias.

Table IV – Effect Beyond Bias and Publication Bias Estimates

	FAT-PET	FAT- PEESE	Multi- level FP	IV	WAAP	TF	AK
Panel A – Full Sample, PCCs							
<i>Effect Beyond Bias</i>	-0.003 (0.002)	0.005 (0.002)	0.006 (0.004)	-0.004 (0.002)	0.005 (0.002)	0.008 (0.018)	-0.773 (0.438)
<i>Publication bias</i>	5.026 (1.283)	32.227 (8.638)	3.502 (0.885)	5.062 (1.297)	.	.	.
<i>Groups</i>	24	24	24	24	.	.	24
<i>Observations</i>	542	542	542	542	362	542	542
Panel B – Only Addressing Endogeneity Sample, PCCs							
<i>Effect Beyond Bias</i>	0.001 (0.001)	0.004 (0.001)	0.001 (0.001)	0.001 (0.001)	0.003 (0.000)	0.007 (0.002)	0.001 (0.002)
<i>Publication bias</i>	2.159 (0.431)	11.305 (10.186)	1.982 (0.434)	2.159 (0.430)	.	.	.
<i>Groups</i>	7	7	7	7	.	.	7
<i>Observations</i>	220	220	220	220	55	220	220
Panel C – Only Elasticity Sample*							
<i>Effect Beyond Bias</i>	0.110 (0.029)	0.128 (0.021)	0.107 (0.010)	-0.056 (0.087)	0.126 (0.022)	0.145 (0.018)	0.025 (0.069)
<i>Publication bias</i>	1.202 (0.545)	3.355 (0.805)	1.966 (0.681)	4.579 (2.935)	.	.	.
<i>Groups</i>	11	11	11	11	.	.	11
<i>Observations</i>	312	312	312	312	122	312	312
Panel D –Only Elasticity and Addressing Endogeneity Sample*							
<i>Effect Beyond Bias</i>	0.040 (0.007)	0.084 (0.019)	0.084 (0.016)	0.013 (0.014)	0.116 (0.028)	0.081 (0.015)	0.018 (0.021)
<i>Publication bias</i>	1.801 (0.440)	4.371 (1.127)	1.392 (0.619)	2.186 (0.514)	.	.	.
<i>Groups</i>	5	5	5	5	.	.	5
<i>Observations</i>	211	211	211	211	70	211	211

Notes. *Indicates effects are elasticities rather than PCCs. Estimates are presented with their standard errors in brackets. FAT-PET is Funnel Asymmetry test and Precision Effect Test (Stanley and Doucouliagos, 2014). FAT-PEESE is Funnel Asymmetry Test and Precision Effect Estimate with Standard Error. The multi-level FP is a FAT-PET multi-level model with fixed effects for each study. IV is a FAT-PET regression with square root of sample size used as an instrumental variable for the precision using two stage least squares. WAAP (Stanley, Doucouliagos, & Ioannidis, 2017) is the Weighted Average of Adequately Powered Estimates, where studies below a certain estimated power are removed before calculating the effect. TF is Trim and fill (Duval & Tweedie, 2000), which removes outlier studies and then adds imputed studies before calculation an average effect. AK is the Andrews-Kasy method (Andrews & Kasy,

2019), which is a step function selection model which reweights the observed sample with estimated publication probabilities. See Appendix D for full explanation of each method.

The estimates for the full sample and addressing endogeneity sample (panels A and B) are all close to zero, save the full sample Andrews-Kasy estimate which is -0.77. However, the 95% confidence interval covers zero, and this estimate is the outlier. For the elasticity sample (panel C) they vary from 0.15 to -0.06, but most estimates are around 0.11. For elasticity estimates that address endogeneity, the estimates range from 0.01-0.08. As a robustness check, we also estimate all methods using only representative estimates in appendix E and the results are similar.

All tests suggest publication bias is present in the sample. This should not be a surprise as Stanley and Doucouliagos (2013) show that bodies of literature with theoretically implausible signs or sizes tend to exhibit more publication bias. It is, of course, theoretically implausible that an increase in lead pollution would cause a decrease in crime, and therefore it may be researchers do not write up papers showing such findings. Nevertheless, we should expect negative estimates due to sampling noise. This may explain the finding of publication bias in all tests and the asymmetry in the funnel plots.

The tests also suggest the true mean effect size of lead on crime may be close to zero, but this could be due to the relatively small sample, or to characteristics of the studies. These characteristics can be investigated more thoroughly with meta-regression analysis.

4.4. Meta-Regression Analysis

Meta-regression analysis (MRA) follows from (4) where we include common observable variation in our estimation. Given all tests suggest the presence of publication bias we include the FAT in all regressions. We also weight all regression covariates by the standard errors as in (6). Therefore, the specification is the same as in (6) except we now also regress on a vector of observable covariates, x_{ij} , weighted by the standard errors of the estimate. This includes the precision, and the coefficient on the precision is now only

an estimate of the average effect size when all other covariates are set to zero. The meta-regression is shown in (7).

$$(7) \quad \hat{t}_{ij} = \beta_F + \mathbf{z}'_{ij}\boldsymbol{\beta} + v_j + e_{ij}$$

Where \mathbf{z}_{ij} is a $1 \times K$ vector of weighted observable covariates.

The covariates included are based on common characteristics of the studies that are suggested by the literature. Their descriptive statistics are included in table V. The majority are dummy variables indicating whether that characteristic is present for that estimate. All variables are coded at estimate level, not at study level. That is, different estimates from the same study may have different characteristics, and therefore have different values for the covariates. There is a dummy variable that equals one when an estimate comes from a quasi-experimental study design that attempts to deal with endogeneity concerns. There is a dummy variable which is one when an estimate is of crime in an area, and zero when it is at the individual level. There are four dummy variables which indicate whether specific controls were included in the estimation. Lead exposure is correlated with poverty (Baghurst et al. 1999) and race (Sampson and Winter, 2016), may have different effects on men and women (Denno, 1990), and may have a relationship with educational outcomes (Fergusson, Boden and Horwood, 2008). Therefore, when an estimation includes these variables we might expect it to influence the estimate. The interpretation of the effect of these variables depends on where they are in the causal chain. If these variables are confounders, causing changes in lead and changes in crime, then omitting them will tend to overstate the effect of lead on crime (given they change both in same direction). If they are mediators, changed by lead and then changing crime, then conditioning on them can lead to understating the effect of lead on crime. This is especially important when study designs do not use some method to deal with endogeneity issues. Of course, there are other variables that may be important controls, but these were not found to be common enough across studies to include.

Next there are three dummy variable that describe what type of crime was used as the dependent variable (homicide, violent, and non-violent), with a reference group of total crime. This allows us to test whether the different mechanisms proposed in section 2 matter. The violent crime category nests homicide within it. They are separate categories because homicide data is thought to be the best quality crime data, and thus less likely to

suffer from bias (Fox and Zatz, 2000). We next have two dummy variables representing possible estimation effects. One for if simple OLS was used, another for if maximum likelihood was used. The reference group is any other estimation such as GMM or mean differences. We have two dummy variables for further estimation effects. One for if panel data were used, and another for if the results are reported as odds ratios.

Table V – Descriptive Statistics of Covariates used in the Meta-Regression Analysis

Variable	Mean	Median	Standard Deviation
Control_gender	0.19	0	0.39
Control_race	0.21	0	0.41
Control_income	0.32	0	0.47
Control_education	0.20	0	0.40
Homicide	0.19	0	0.39
Violent	0.63	1	0.48
Non_Violent	0.15	0	0.36
Both	0.22	0	0.41
Area	0.77	1	0.42
OLS	0.39	0	0.49
ML	0.13	0	0.34
Odds_Ratio	0.03	0	0.17
Panel	0.67	1	0.47
Addressing Endogeneity	0.41	0	0.49
North America	0.71	1	0.45
Europe	0.16	0	0.36
Direct Lead Measure	0.10	0	0.30
Publication Year*	2013	2015	6
Number of Covariates*†	445	13	802
Sample Size*	64478	901	186,709

Notes:*Indicates variables have been standardised.

†Includes fixed effects for degrees of freedom adjustment

A further two dummy variables are geographic dummies that equal one when an estimate come from a either North America or Europe, with the rest of the world as the reference group. 70% of estimates use data from North America. The final dummy variable equals one when a direct measure of lead, from either blood, bone, or dentine samples, is used in the estimation and zero when a proxy measure or estimate, such as leaded gasoline use in an area, is used. This allows us to test whether there is a systematic difference in effect

sizes found when lead levels are taken directly from subjects, which we might expect to give a more accurate measure of the true effect, rather than proxied. The final three covariates are the publication year, sample size, and the number of covariates included in the estimation. These variables have been standardised to aid the restricted maximum likelihood convergence.

We estimate many specifications due to model uncertainty. Our sample is relatively small and coefficient estimation varies significantly in alternative specifications. The number of different covariate combinations is 2^K where K is the total number of covariates. It is common in the meta-analysis literature to employ some method of model averaging or shrinkage to deal with model uncertainty. However, with this many covariates and modern computational power it is possible to estimate all 2^K specifications³. In addition, table III showed that some subsamples have substantially less heterogeneity than the full sample. It may be that these sub-samples suit aggregation better than the full sample. For example, we might expect studies with individuals as the unit of analysis to share much more common information than those that have a geographic area as the unit of interest. We therefore also estimate all covariate specifications for these subsamples. It is not possible to estimate every combination as some dummy variables no longer have any variation in the subsamples, leading to collinearity. This can also lead to other variables being excluded as they become the new base case (for example if there are no studies from outside Europe or North America in a subsample, then Europe becomes the base case). A full list of the covariates included for each subsample is in table VI. We estimate every possible combination of covariates for the full sample and the subsamples. We include the FAT, the estimate of publication bias. We estimate with REML and include study fixed effects.

We do not interpret the coefficients on the covariates following best practice (see Westreich and Greenland, 2013 and Stevenson and Elwert, 2020), as they are not identified. Instead, we use the information from each meta-regression specification to construct a distribution of estimates of the average effect of lead on crime. We are now estimating an average effect conditional on the observable heterogeneity in our specifications. In practice, meta-analysis tends to do this in two ways, either by using the

³ As a robustness check we perform Bayesian Model Averaging in appendix F. The posterior mean PCC using the full sample and evaluated at the sample averages is 0.09, lower than the method we use here. The elasticity posterior mean is also lower at 0.07.

sample averages or by taking some “ideal” specification. We do both. That is, for each specification we generate a predicted estimate of the effect of lead on crime, using both the sample averages, or by using an ideal specification, and not including the FAT in the predicted value (i.e. removing the publication bias).

Table VI – Variables Used in Combinations For Each Sample Estimation

Sample	Variables Used
Full Sample	Control gender, Control race, Control income, Control education, Homicide, Violent, Non-Violent, Area dummy, OLS, ML, Odds Ratio, Panel dummy, Addressing Endogeneity, North America, Europe, Direct Lead Measure, Publication Year, Covariates, Sample Size
Addressing Endogeneity Sample	Control gender, Control race, Control income, Homicide, Violent, Non-Violent, Area dummy, OLS, Panel dummy, Publication Year, Covariates, Sample Size
Correlational Sample	Control gender, Control race, Control income, Control education, Homicide, Violent, Non-Violent, Area dummy, OLS, ML, Odds Ratio, Panel dummy, North America, Direct Lead Measure, Publication Year, Covariates, Sample Size
Area-level Sample	Control race, Control income, Control education, Homicide, Violent, Non-Violent, OLS, ML, Panel dummy, Addressing Endogeneity, Direct Lead Measure, Publication Year, Covariates, Sample Size
Individual-level Sample	Control gender, Control race, Control income, Control education, Violent, Non-Violent, OLS, ML, Odds Ratio, Panel dummy, Addressing Endogeneity, Direct Lead Measure, Publication Year, Covariates, Sample Size
Homicide Sample	Control race, Control income, OLS, Panel dummy, Addressing Endogeneity, Publication Year, Covariates, Sample Size
Violent Crime Sample	Control gender, Control race, Control income, Control education, Area dummy, OLS, ML, Panel dummy, Addressing Endogeneity, North America, Direct Lead Measure, Publication Year, Covariates, Sample Size
Non-Violent Crime Sample	Control gender, Control race, Control income, Control education, Area dummy, OLS, ML, Odds ratio, Panel dummy, Addressing Endogeneity, North America, Direct Lead Measure, Publication Year, Covariates, Sample Size
Elasticity Sample	Control gender, Control race, Control income, Control education, Homicide, Violent, Non-Violent, Area dummy, OLS, ML, Panel dummy, Addressing Endogeneity, North America, Direct Lead Measure, Publication Year, Covariates, Sample Size
Elasticity and Addressing Endogeneity Sample	Control gender, Control race, Control income, Homicide, Violent, Non-Violent, Area dummy, Publication Year, Covariates, Sample Size

Notes. Table shows which covariates were included for each sub-sample estimation. Inclusion depended on whether there was variation in the covariate for that subsample.

The ideal specification we use is one that includes controls for race, education, income and gender, that uses individual data, directly measured lead levels, controls for endogeneity, uses panel data, is estimated without just using simple OLS or ML, uses total crime as the dependent variable, uses North American data (as most of our sample is from there), and uses the sample averages for the publication year, sample size, and number of covariates. This ideal specification is chosen to represent a robust and high-quality

estimation, and as such we would expect it be generally lower than the sample averages estimates.

The means, medians, and standard deviations of the full sample and subsample estimates are presented in table VII. The top panel shows the estimates effect sizes evaluated at the sample averages, while the bottom shows effect sizes evaluated at the “ideal” specification. The table also shows the number of specifications for each sample. The final column shows how many of the estimates fell outside of the feasible interval of the PCC [-1,1]. This indicates whether there may be a misspecification issue with that particular sample estimation.

The distribution of coefficient sizes for the full sample estimation is in plotted in figure X, panel A. The left figure shows effect sizes evaluated at the sample averages, while the right shows effect sizes evaluated at the “ideal” specification. In each there is a distribution of 524,288 estimated effect sizes. The mean and median PCC for the sample averages distribution are 0.16 and 0.18 respectively, which is “moderately positive” according to the Doucouliagos (2011) taxonomy. The distribution appears to be bimodal with one peak close to zero and the other around 0.2. The distribution of the ideal specification is not bimodal and is roughly symmetrical. The mean and median are 0.13 and 0.09 respectively. As expected the ideal specification is lower than the sample averages.

We next restrict the sample to only the studies that estimate a causal effect with quasi-experimental methods rather than an association: our “addressing endogeneity” subsample. This consists of seven studies and 220 estimates. It is common in meta-analysis to exclude correlational studies altogether (e.g., Kraft, Blazer and Hogan, 2018). Although we have not excluded those studies in this meta-analysis, we now examine what a meta-analysis estimate with only causal studies would be. We saw in table III that the addressing endogeneity subsample has lower between-study heterogeneity than the full sample, so aggregation may yield comparatively more information.

We plot the sub-sample average specification and ideal specification in figure X, panel B (excluding those variables that cannot be included in the estimation, see table VI). The distribution of the sample average predicated values is tight around zero with a mean and median of 0.01, and a sample standard deviation of 0.01. The “ideal” specification also has

a mean and median of 0.01. The results suggest there is a systematic difference between the “addressing endogeneity” studies and the rest of the sample.

Table VII - Meta-Analysis Average Estimates for The Full Sample and Each Subsample

Sample averages

Sample	Mean	Median	SD	N	% < -1 or > 1
Full Sample	0.16	0.18	0.07	524288	0%
Addressing Endogeneity Sample	0.01	0.01	0.01	4096	0%
Correlational Sample	0.29	0.29	0.10	131072	0%
Area-level Sample	0.25	0.26	0.06	16384	0%
Individual-level Sample	0.03	0.03	0.01	65536	0%
Homicide Sample	0.58	0.54	0.22	256	0%
Violent Crime Sample	0.39	0.39	0.22	16384	0%
Non-violent Crime Sample	0.75	0.71	0.24	32768	14%
Elasticity Sample*	0.09	0.09	0.03	131072	.
Elasticity and Addressing Endogeneity Sample*	0.10	0.09	0.04	1024	.

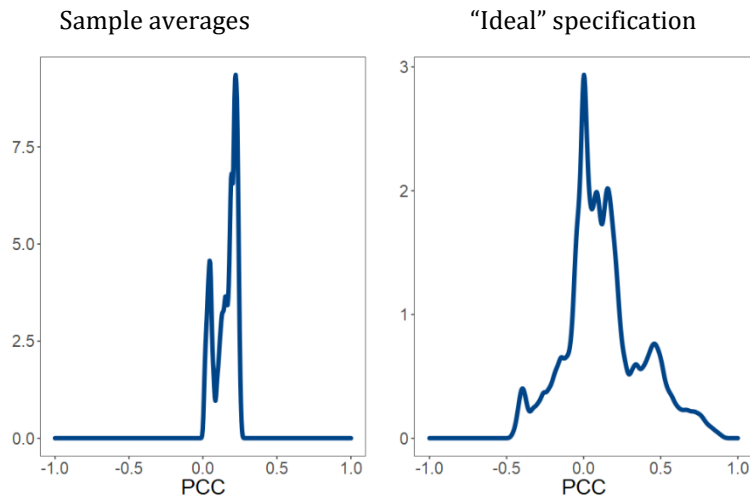
“Ideal” specification

Sample	Mean	Median	SD	N	% < -1 or > 1
Full Sample	0.13	0.09	0.25	524288	0%
Addressing Endogeneity Sample	0.01	0.01	0.02	4096	0%
Correlational Sample	0.49	0.37	0.6	131072	15%
Area-level Sample	0.23	0.20	0.22	16384	0%
Individual-level Sample	0.02	0.02	0.04	65536	0%
Homicide Sample	0.28	0.27	0.17	256	0%
Violent Crime Sample	0.57	0.14	1.29	16384	36%
Non-violent Crime Sample	1.26	0.58	3.50	32768	64%
Elasticity Sample*	0.09	0.05	0.20	131072	.
Elasticity and Addressing Endogeneity Sample*	0.17	0.16	0.16	1024	.

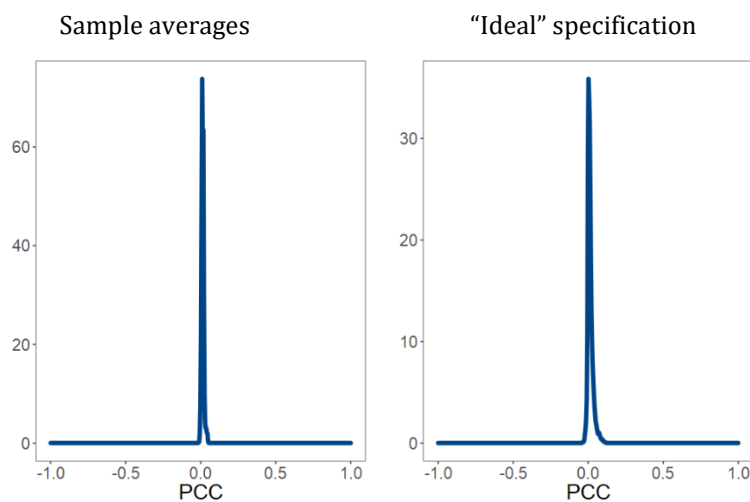
Notes. *Indicates values are elasticities rather than PCCs. Table shows results from combining multiple meta-regression estimates, each using different specifications. All regressions carried out by restricted maximum likelihood. This is done for the full sample and subsamples. N is the number of regressions carried out, each a different specification. The mean and median are the summary statistics of the average effect size from these regressions, given in Partial Correlation Coefficients (PCCs) or elasticities. PCCs are bounded between -1 and 1. The last column gives the percent of effects which fall outside this range.

Figure X – Density of Meta-Analysis Average Effect Size Estimates from Full Sample

Panel A – Full Sample (PCCs)



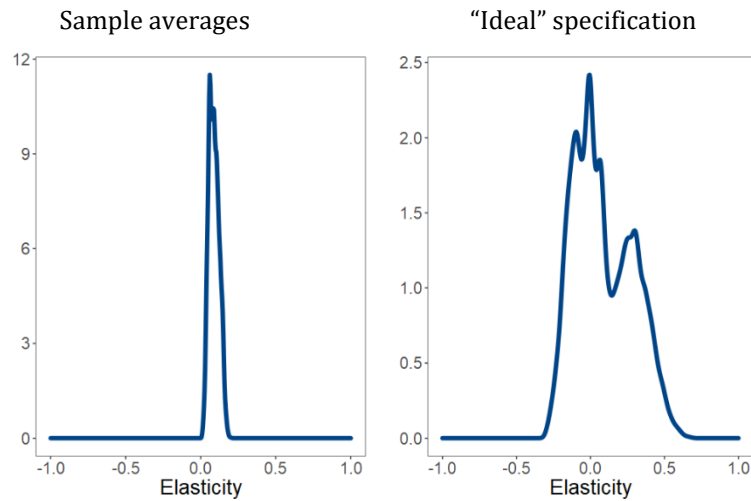
Panel B – Addressing Endogeneity Sample (PCCs)



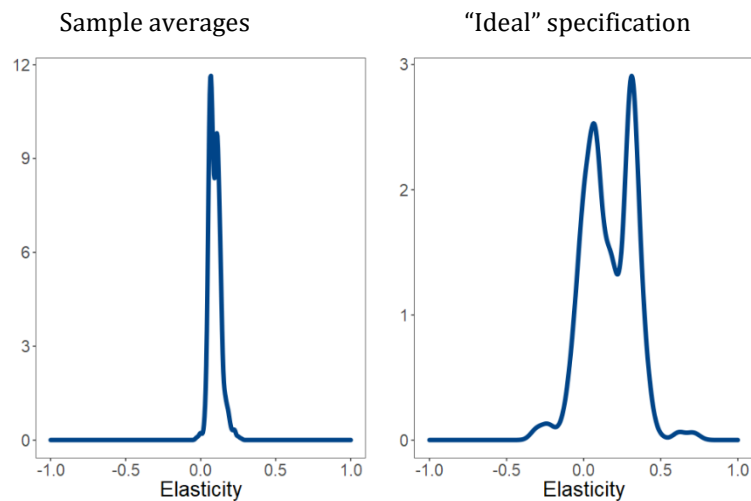
Notes. Chart shows densities for the distribution of meta-regression estimated average effect sizes. Chart on left shows estimated average effect for each specification evaluated at the sample averages. Chart on right shows estimated average effect for each specification evaluated-at an “ideal” specification. X axis truncated at feasible interval of a PCC, [-1,1].

Figure XI – Density of Meta-Analysis Average Effect Estimates for Elasticity Subsample

Panel A – Full Elasticity Sample



Panel B –Elasticity and Addressing Endogeneity Sample



Notes. Chart shows densities for the distribution of meta-regression estimated average effect sizes for the addressing elasticity sub-sample.. Chart on left shows estimated average effect for each specification evaluated at the sample averages. Chart on right shows estimated average effect for each specification evaluated-at an "ideal" specification.

In figure XI, we carry out the same exercise except only for those studies that have elasticity estimates available. The elasticity effect sizes in panel A, figure XI are for the full elasticity sample. The mean and median effect size, evaluated at the sample averages, are both an elasticity of 0.09. Evaluated at the “ideal” specification they are 0.09 and 0.05 respectively. The “ideal” distribution shows more heterogeneity and is bimodal. The standard deviation is 0.2, much higher than the 0.03 evaluated at the sample averages.

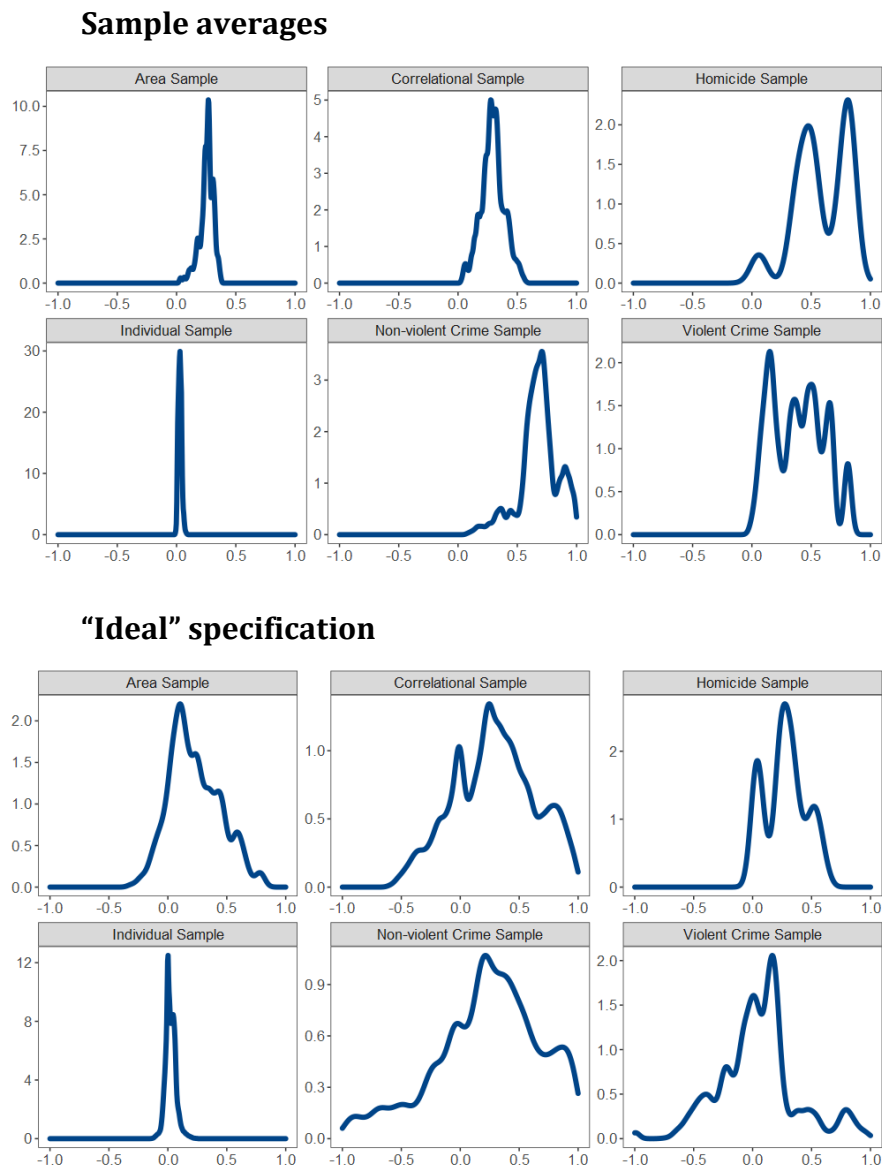
Panel B of figure XI, the addressing endogeneity and elasticity subsample, is very similar to panel A. The mean is 0.10 and the median elasticity is 0.09 when evaluated at the sample averages. When evaluated at the “ideal” specification the addressing endogeneity, elasticity sample mean is 0.17 and the median is 0.16. These are considerably larger than when evaluated at the sample averages, or when looking at the “ideal” specification for the full elasticity sample. In both panel A and B the “ideal” specification distributions have a larger variance than the sample average distributions. 40% of the “ideal” specifications yield a negative elasticity when using the full elasticity sample, and 15% are negative when using the addressing endogeneity, elasticity sample. In contrast almost no estimates are negative when evaluated at the sample averages in panel A or B. This suggests the “ideal” specification is much more sensitive to model changes than when we evaluate at the sample averages.

We next plot several other subsample distributions of interest in figure XII. The difference between the area and individual sample is striking. The area sample means and medians are much larger than the individual sample for both the sample average specification and the ideal specification. The individual sample mean and median PCCs are small and the distributions are tight around the means compared to the area sample. This suggests that covariates matter less for the individual sample effect sizes compared to the area sample. Similar to the area-individual comparison, the correlational sample has much higher means and medians than the addressing endogeneity sample.

Comparing homicide, violent, and non-violent crime samples we can see they all have large mean and median PCCs, but the non-violent and violent subsamples have a portion of the distribution outside $[-1,1]$, suggesting misspecification and that the results may not be reliable. The standard deviations for these tend to be much larger as well. Furthermore, due to the lack of homicide estimates, only 256 specifications could be run without convergence issues. Overall, the results suggest that lead affects all types of

crime, but we cannot say if it has a bigger effect on some types than others. We cautiously suggest that if lead does have an effect on crime it is across all categories of crime.

Figure XII – Densities of Meta-Analysis Average Effect Estimates From Subsamples



Notes. Chart shows densities for the meta-regression estimated average effect sizes for a number of subsamples. Top chart shows estimated average effect for each specification evaluated at the sample average for each subsample. Bottom chart shows estimated average effect for each specification evaluated at an “ideal” specification. X axes truncated at feasible interval of a PCC, [-1,1]

4.5 Explaining the 20th Century Crime Decline

Our calculated elasticities allow us to estimate how much of the fall in crime observed in the second half of the 20th century was caused by lead. We first use the dramatic fall in homicide in the US as an example. The median blood lead level in children in the US fell 88% from 1976-2009. The US homicide rate fell 54% from its peak in 1989 to 2014. Given our main elasticity estimate of 0.09 for the full elasticity sample evaluated at the sample averages (with standard deviation of 0.03), this 5-11 percentage points of the 54% fall are due to lead, with the point estimate being 8. This would mean that lead accounts for 15% of the decrease in homicide. If we use the full range of elasticity mean estimates in table VII we have values from 0.05-0.17. These would imply 4-15 percentage points of the 54% fall were accounted for by lead. This would mean 7-28% of the fall in homicide was due to falling lead levels.

Our estimates imply lead pollution is an important factor in reducing homicides, and lead abatement has saved lives, but it does not account for the majority of the fall. Depending on the specification, we conclude that 93%-73% of the fall in homicide in the US is unaccounted for.

We next carry out estimates of how much of the urban/rural violent crime convergence in figure IV can be explained by the relatively higher blood lead levels in urban areas in the 1970s. Average blood lead levels in under 5s (geometric mean) declined by 15 μ g/dL in large population MSAs from 1976 to 2011, and by 12.7 in smaller MSAs and rural areas over the same period. The difference in the violent victimization rate per 1,000 people between urban and rural areas, as measured by the NCVS, was 41 in 1993, and 7 in 2015. Given the lag in time between childhood lead exposure to adult criminal acts, we believe these differing periods give enough time for the lead change to take effect. The gap in victimizations declined by 34 per 1,000 people in this period. Using an elasticity of 0.09, we estimate the relative change in blood lead levels would account for 3.6 of these victimizations. That is, the difference in lead levels accounts for 11% of the convergence in the victimization rate. Using the 0.05-0.17 range of elasticity estimates in table VII, means that lead accounts for between 2 and 7 of the victimisation gap difference. This would explain 6%-20% of the convergence in violent victimisation rates. While not negligible, this leaves a large part of the convergence in urban/rural crime rates unexplained as well.

5. Discussion and Conclusion

Changes to the amount of lead in the environment have been put forward as one of the main causes of the decrease in crime, especially homicide, in many western countries. We performed the first meta-analysis of the effect of lead on crime. We find there is publication bias in the lead-crime literature, and that meta-analysis estimates that do not control for this will overstate the effect of lead on crime. Using meta-regression, taking into account publication bias and between-study heterogeneity, our main estimates are an average effect size of 0.16 as a partial correlation, or 0.09 as an elasticity. When using the larger PCC sample, we find that the average meta-analysis estimate for studies that address endogeneity is much smaller than for the full sample, or for the correlational sample. However, when using the elasticity sample, the average meta-analysis estimate for studies that address endogeneity tended to be similar to the full elasticity sample, except when evaluated at the “ideal” specification, in which case it was larger. The average effect size estimate for studies that have individuals as the unit of interest is much smaller than for the sample of studies that have a geographic area as the unit of interest. When we examined the differences between lead’s effect on homicide, violent and non-violent crime, we could not confidently state there was any difference between them.

Finally, we performed calculations to estimate the share of the decline in crime in the US that is accounted for by reductions in blood lead levels. We estimate that of the total 54% fall in homicides observed in the US in 1976-2009, reduced blood lead levels accounted for 4-15 percentage points. A substantial decrease. However, this was only a 7-28% share of the total fall, leaving 93-72% unaccounted for. Similarly, we find that the relative changes in blood lead levels account for 6-20% of the convergence in urban and rural violent crime rates observed in the US, leaving much of this convergence unexplained.

Overall, the results suggest that declines in lead pollution reduce crime but are not the cause of the majority of the fall in crime observed in many western countries. We are unable to provide estimates on the size of other causes here but hope our results can provide a rough benchmark for relative importance in future meta-analyses. It is possible that the large differences in our samples can be reconciled. For example, the large

difference between the individual and area samples may be because crime has fallen at the extensive margin rather than the intensive margin. Tcherni-Buzzeo (2019) observe that around 5% of the population are responsible for 50% of crime, and that the fall in crime in the US is likely due to falls in this high-crime population, rather than less crimes per individual in that population. If less lead pollution only meant less probability of committing crime for this small slice in the population, it might nevertheless lead to a large fall in crime at the area level. A second possibility is that relatively small effects of lead at the individual level can be exacerbated by peer effects from other lead affected individuals⁴. Recent work has found these peer effects can even affect those without elevated blood lead levels (Gazze, *et al.*, 2021). In areas with high levels of lead, the individual effects of lead may be compounded by peers also having high levels of lead, leading to a much larger impact at the area level.

There are several limitations to our analysis. Most importantly, the sample size is not large. We have 24 studies and 542 estimates, this is not unusual for a meta-analysis but, particularly for our subsample estimates, this could play a part in the differences. It may explain why so much of the distribution for the different types of crime in table VII were outside the feasible PCC interval of $[-1,1]$. We attempt to mitigate this by using various tests for publication bias, and estimating many different specifications, but we cannot rule out that the results are due to small sample effects. Secondly, the between-study heterogeneity is large in our sample. This calls into question how comparable the studies are. This is to be expected as studies use different concepts and measures of crime and lead, different units of interest, and different estimation techniques. We try to mitigate this by converting to PCCs or elasticities, using different sub-samples that have lower between-study heterogeneity, and using meta-regression with covariates. However, even with these mitigations, it may be that the literature is not comparable and therefore meta-analysis estimates will be noise. In this case it casts doubt on the external validity of the studies examining the lead-crime hypothesis. The solution would be far more studies that estimate elasticities using comparable measures of lead and crime.

For policymakers, our results are a warning against assuming the large crime levels in past decades cannot return now that lead pollution is much lower. The results are not a signal that lead abatement is fruitless. As outlined in section 2, the evidence of harmful

⁴ We thank an anonymous reviewer for this suggestion.

biological and health changes due to lead is overwhelming. There is no known safe level of lead. Even if outcomes higher up the causal chain, such as crime, are not as affected by lead, the evidence still shows lead abatement will increase health outcomes, especially for the very young.

For future research, we have two main suggestions. The first is that there are enough low sample size, correlational studies in the lead-crime literature. What is needed now is high power, high-quality causal estimates of the effect of lead on crime. The value added of such studies would be increased by testing the effect on different types of crime, and the possible interaction of lead with other potential causes. The second is that more high-quality causal estimates of the elasticity of other causes of crime are needed. Our results suggest lead is not responsible for the majority of the fall in crime since the 80s and therefore leaves open room for other explanations. These explanations must account for the fact homicide has fallen across many (but not all!) western countries at roughly the same time. They must also account for the fact that total crime has risen in Europe and fallen in the US, while the homicide rate has fallen in both. Further comparison of the relative shares of responsibility for the fall in crime, as well as the interaction between causes, may also be fruitful and we suggest further meta-analyses, using modern methods, would be helpful in this area.

References

Aelion, C.M., Davis, H.T., 2019. Blood lead levels in children in urban and rural areas: Using multilevel modeling to investigate impacts of gender, race, poverty, and the environment. *Sci. Total Environ.* 694, 133783.

<https://doi.org/10.1016/j.scitotenv.2019.133783>

Andrews, I., Kasy, M., 2019. Identification of and Correction for Publication Bias. *Am. Econ. Rev.* 109, 2766–2794. <https://doi.org/10.1257/aer.20180310>

Baghurst, P.A., Tong, S., Sawyer, M.G., Burns, J., McMichael, A.J., 1999. Sociodemographic and behavioural determinants of blood lead

concentrations in children aged 11–13 years. The Port Pirie Cohort Study. *Med. J. Aust.* 170, 63–7.

Baltagi, B.H., Song, S.H., Jung, B.C., 2002. A comparative study of alternative estimators for the unbalanced two-way error component regression model, *Econometrics Journal*.

Baumer, E.P., Rosenfeld, R., Wolff, K.T., 2012. Expanding the Scope of Research on Recent Crime Trends.

Becker, G.S., 1968. Crime and Punishment: an Economic Approach, in: *The Economic Dimensions of*

Crime. Palgrave Macmillan UK, pp. 13–68.

https://doi.org/10.1007/978-1-349-62853-7_2

Birkel, C., Dern, H., 2012. Homicide in Germany, in: Handbook of European Homicide Research: Patterns, Explanations, and Country Studies. Springer New York, pp. 313–328. https://doi.org/10.1007/978-1-4614-0466-8_20

Boutwell, B.B., Nelson, E.J., Emo, B., Vaughn, M.G., Schootman, M., Rosenfeld, R., Lewis, R., 2016. The intersection of aggregate-level lead exposure and crime. Environ. Res. 148, 79–85. <https://doi.org/10.1016/j.envres.2016.03.023>

Boutwell, B.B., Nelson, E.J., Qian, Z., Vaughn, M.G., Wright, J.P., Beaver, K.M., Barnes, J.C., Petkovsek, M., Lewis, R., Schootman, M., Rosenfeld, R., 2017. Aggregate-level lead exposure, gun violence, homicide, and rape. PLoS One 12. <https://doi.org/10.1371/journal.pone.0187953>

Brody, D.J., Kramer, R.A., Flegal, K.M., Pirkle, J.L., Matte, T.D., Gunter, E.W., Paschal, D.C., 1994. Blood Lead Levels in the US Population: Phase 1 of the Third National Health and Nutrition Examination Survey (NHANES III, 1988 to 1991). JAMA 272, 277–283. <https://doi.org/10.1001/IAMA.1994.03520040039038>

Brubaker, C.J., Schmithorst, V.J., Haynes, E.N., Dietrich, K.N., Egelhoff, J.C., Lindquist, D.M., Lanphear, B.P., Cecil, K.M., 2009. Altered myelination and axonal integrity in adults with childhood lead exposure: A diffusion tensor imaging study. Neurotoxicology.

Buonanno, P., Drago, F., Galbiati, R., Zanella, G., 2011. Crime in Europe and the United States: dissecting the ‘reversal of misfortunes.’ Econ. Policy 26, 347–385. <https://doi.org/10.1111/j.1468-0327.2011.00267.x>

Bureau, U.C., 2009. Statistical Abstract of the United States: 2009.

Bureau of Justice Statistics, 2020. *National Crime Victimization Survey*. United States.

Carrozzi, F., Roth, S., 2020. Dirty Density: Air Quality and the Density of American Cities. IZA DP No. 13191. <https://ftp.iza.org/dp13191.pdf>

Cecil, K.M., Brubaker, C.J., Adler, C.M., Dietrich, K.N., Altaye, M., Egelhoff, J.C., Wessel, S., Elangovan, I., Hornung, R., Jarvis, K., Lanphear, B.P., 2008. Decreased brain volume in adults with childhood lead exposure. PLoS Med. <https://doi.org/10.1371/journal.pmed.0050112>

Corman, H., Mocan, H.N., 2000. A time-series analysis of crime, deterrence, and drug abuse in New York City. Am. Econ. Rev. 90, 584–604. <https://doi.org/10.1257/aer.90.3.584>

De Long, J.B., Lang, K., 1992. Are all Economic Hypotheses False? J. Polit. Econ. 100, 1257–1272. <https://doi.org/10.1086/261860>

Denno, D.W., 1990. Biology and Violence: From Birth to Adulthood. Cambridge University Press. <https://doi.org/10.1017/CBO9780511752803>

DerSimonian, R., Laird, N., 1986. Meta-analysis in clinical trials. Control. Clin. Trials. [https://doi.org/10.1016/0197-2456\(86\)90046-2](https://doi.org/10.1016/0197-2456(86)90046-2)

Donohue, J.J., Levitt, S.D., 2001. The impact of legalized abortion on crime. Q. J. Econ. <https://doi.org/10.1162/00335530151144050>

Donohue, J., Levitt, S., 2019. The Impact of Legalized Abortion on Crime over the Last Two Decades. Cambridge, MA. <https://doi.org/10.3386/w25863>

Dore, C.J., Watterson, J.D., Murrells, T.P., Passant, N.R., Hobson, M.M., Baggott, S.L., Thistlethwaite, G., Goodwin, J.W.L., King, K.R., Adams, M., Walker, C., Downes, M.K., Coleman, J., Stewart, R.A., Wagner, A., Sturman, J., Conolly, C., Lawrence, H., Li, Y., Jackson, J., Bush, T., Grice, S., Brophy, N., Misselbrook, T., Brown,

L., Dragosits, U., Sutton, M., Mile, R., Mobbs, D., 2006. UK Emissions of Air Pollutants 1970 to 2004.

Doucouliaos, C. and Stanley, T., 2013. Are all Economic Facts Greatly Exaggerated? Theory Competition and Selectivity. *Journal of Economic Surveys*, 27: 316-339. <https://doi.org/10.1111/j.1467-6419.2011.00706.x>

Doucouliaos, H., 2011. "How large is large? Preliminary and relative guidelines for interpreting partial correlations in economics," Working Papers eco_2011_5, Deakin University, Department of Economics.

Duval, S., & Tweedie, R., 2000. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455-463. <https://doi.org/10.1111/j.0006-341x.2000.00455.x>

Falck, S., von Hofer, H., Storgaard, A., 2003. Nordic Criminal Statistics 1950-2000. Department of Criminology Stockholm University.

Ferraro, P., and Shukla, P., 2020. Feature—Is a Replicability Crisis on the Horizon for Environmental and Resource Economics? *Review of Environmental Economics and Policy* 14:2, 339-351

Fox, J.A., 2005. Demographics and U.S. Homicide, in: *The Crime Drop in America*, Revised Edition. Cambridge University Press, pp. 288-317. <https://doi.org/10.1017/CBO9780511616167.010>

Fox, J.A., Marianne Zawitz, 2000. Homicide Trends in the United States: 2000 Update [WWW Document]. Bur. Justice Stat. URL https://www.researchgate.net/publication/246383500_Homicide_Trends_in_the_United_States_2000_Update (accessed 1.14.21).

Gazze, L., Persico, C., Spirovska, S., 2021. The Long-Run Spillover Effects of Pollution: How Exposure to

Lead Affects Everyone in the Classroom. SSRN Electron. J. <https://doi.org/10.2139/SSRN.3842757>

GBD, Global Burden of Disease, 2019 | Institute for Health Metrics and Evaluation [WWW Document]. Glob. Burd. Dis. - IMHE. URL <https://www.healthdata.org/gbd/2019> (accessed 1.26.22).

Gottfredson, M.R., Hirschi, T., 1990. *A General Theory of Crime*. Stanford University Press, Stanford.

Egan, K.B., Cornwell, C.R., Courtney, J.G., Ettinger, A.S., 2021. Blood Lead Levels in U.S. Children Ages 1-11 Years, 1976-2016. *Environ. Health Perspect.* 129. <https://doi.org/10.1289/EHP7932>

Havráněk, T., Stanley, T.D., Doucouliagos, H., Bom, P., Geyer-Klingeborg, J., Iwasaki, I., Reed, W.R., Rost, K., Aert, R.C.M., 2020. Reporting Guidelines For Meta-Analysis In Economics. *J. Econ. Surv.* 34, 469-475. <https://doi.org/10.1111/joes.12363>

Haynes, E.N., Chen, A., Ryan, P., Succop, P., Wright, J., Dietrich, K.N., 2011. Exposure to airborne metals and particulate matter and risk for youth adjudicated for criminal activity. *Environ. Res.* 111, 1243-1248. <https://doi.org/10.1016/j.envres.2011.08.008>

Hernberg, S., 2000. Lead Poisoning in a Historical Perspective, *American Journal Of Industrial Medicine*.

Higgins, J.P.T., Thompson, S.G., 2002. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* 21, 1539-1558. <https://doi.org/10.1002/SIM.1186>

Hollingsworth, A., Rudik, I., 2021. The Effect of Leaded Gasoline on Elderly Mortality: Evidence from Regulatory Exemptions. *Am. Econ. J. Econ. Policy* 13, 345-73. <https://doi.org/10.1257/POL.20190654>

Hong, S., Candelone, J.P., Patterson, C.C., Boutron, C.F., 1994. Greenland ice evidence of hemispheric lead pollution two millennia ago by Greek and Roman

civilizations. *Science* (80-.). 265, 1841–1843.

<https://doi.org/10.1126/science.265.5180.1841>

Ioannidis, J.P.A., 2005. Why most published research findings are false, in: *Getting to Good: Research Integrity in the Biomedical Sciences*.

<https://doi.org/10.1371/journal.pmed.0020124>

Ioannidis, J.P.A., Stanley, T.D., Doucouliagos, H., 2017. The Power of Bias in Economics Research. *Econ. J. h*

Keele, L., Stevenson, R., & Elwert, F., 2020. The causal interpretation of estimated associations in regression models. *Political Science Research and Methods*, 8(1), 1-13.

Kristensen, L.J., 2015. Quantification of atmospheric lead emissions from 70 years of leaded petrol consumption in Australia. *Atmos. Environ.* 111, 195–201.

<https://doi.org/10.1016/j.atmosenv.2015.04.012>

Lanphear, B.P., 2015. The Impact of Toxins on the Developing Brain. *Annu. Rev. Public Health* 36, 211–230. <https://doi.org/10.1146/annurev-publhealth-031912-114413>

Lessler, M.A., 1988. Lead and Lead Poisoning from Antiquity to Modern Times. *Ohio Journal of Science: Volume 88, Issue 3*.

<https://kb.osu.edu/handle/1811/23252>

Levin, R., Zilli Vieira, C.L., Rosenbaum, M.H., Bischoff, K., Mordarski, D.C., Brown, M.J., 2021. The Urban Lead (Pb) Burden in Humans, Animals and the Natural Environment. *Environ. Res.* 193, 110377.

<https://doi.org/10.1016/j.ENVRES.2020.110377>

Levitt, S.D., 1996. The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation. *Q. J. Econ.* 111, 319–351.

<https://doi.org/10.2307/2946681>

Levitt, S.D., 2004. Understanding why crime fell in the 1990s: Four factors that explain the decline and six

that do not. *J. Econ. Perspect.*

<https://doi.org/10.1257/089533004773563485>

Levitt, S.D., 1997. Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime. *Am. Econ. Rev.* <https://doi.org/10.2307/2951346>

Lim, H.H., Ong, C.N., Domala, Z., Phoon, W.O., Khoo, H.E., 1985. Blood lead levels in Malaysian urban and rural pregnant women. *Public Health* 99, 23–29.

[https://doi.org/10.1016/S0033-3506\(85\)80123-2](https://doi.org/10.1016/S0033-3506(85)80123-2)

O' Flaherty, Brendan, Sethi, Rajiv, 2015. Urban Crime, Handbook of Regional and Urban Economics, in: Gilles Duranton & J. V. Henderson & William C. Strange (ed.), Handbook of Regional and Urban Economics, edition 1, volume 5, chapter 0, pages 1519-1621, Elsevier.

<https://doi.org/10.1016/B978-0-444-59531-7.00023-5>

Mahaffey, K.R., Annett, J.L., Roberts, J., Murphy, R.S., 2010. National Estimates of Blood Lead Levels: United States, 1976–1980.

<http://dx.doi.org/10.1056/NEJM198209023071001> 307, 573–579.

<https://doi.org/10.1056/NEJM198209023071001>

Marcus, D.K., Fulton, J.J., Clarke, E.J., 2010. Lead and conduct problems: A meta-analysis. *J. Clin. Child Adolesc. Psychol.*

<https://doi.org/10.1080/15374411003591455>

Marvell, T.B., Moody, C.E., 1996. Specification problems, police levels, and crime rates. *Criminology* 34, 609–646. <https://doi.org/10.1111/j.1745-9125.1996.tb01221.x>

Masters, R.D., Coplan, M.J., 1999. Water treatment with silicofluorides and lead toxicity. *Int. J. Environ. Stud.* 56, 435–449.

<https://doi.org/10.1080/00207239908711215>

McCulloch, C. E., & Neuhaus, J. M., 2011. Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter. *Statistical Science*,

26(3), 388–402.

<http://www.jstor.org/stable/23059138>

Meager, R., 2019. Understanding the average impact of microcredit expansions: A Bayesian hierarchical analysis of seven randomized experiments. *Am. Econ. J. Appl. Econ.* 11, 57–91.

<https://doi.org/10.1257/app.20170299>

Messner, S.F., Raffalovich, L.E., Mcmillan, R., 2001. Economic deprivation and changes in homicide arrest rates for white and black youths, 1967-1998: a national time-series analysis. *Criminology* 39, 591–614. <https://doi.org/10.1111/j.1745-9125.2001.tb00934.x>

Needleman, H.L., 1991. Human Lead Exposure - Herbert L. Needleman - Google Books [WWW Document]. URL <https://books.google.co.uk/books?hl=en&lr=&id=e9fel0gM3j0C&oi=fnd&pg=PA23&dq=#v=onepage&q&f=false> (accessed 12.14.19).

Needleman, H.L., 1996. Bone Lead Levels and Delinquent Behavior. *JAMA J. Am. Med. Assoc.* 275, 363. <https://doi.org/10.1001/jama.1996.03530290033034>

New Zealand Police. (2018). Historic murder offences in New Zealand 1926–2017. Retrieved from <https://www.police.govt.nz/sites/default/files/publications/historic-new-zealand-murder-rates.pdf> (accessed November 14, 2018).

Pope, D.G., Pope, J.C., 2012. Crime and property values: Evidence from the 1990s crime drop. *Reg. Sci. Urban Econ.* 42, 177–188. <https://doi.org/10.1016/j.REGSCIURBECO.2011.08.008>

Rainald Borck, Philipp Schrauth, Population density and urban air quality, *Regional Science and Urban Economics*, Volume 86, 2021, 103596, ISSN 0166-

0462,

<https://doi.org/10.1016/j.regsciurbeco.2020.103596>

Rice, K., Higgins, J.P.T., Lumley, T., 2018. A re-evaluation of fixed effect(s) meta-analysis. *J. R. Stat. Soc. Ser. A (Statistics Soc.* 181, 205–227. <https://doi.org/10.1111/rssa.12275>

Rosenfeld, R., Fornango, R., 2007. The impact of economic conditions on robbery and property crime: The role of consumer sentiment. *Criminology* 45, 735–769. <https://doi.org/10.1111/j.1745-9125.2007.00096.x>

Röver, C., 2020. Bayesian random-effects meta-analysis using the bayesmeta r package. *J. Stat. Softw.* 93. <https://doi.org/10.18637/jss.v093.i06>

Rubin, D.B., 1981. Estimation in Parallel Randomized Experiments. *J. Educ. Stat.* 6, 377–401. <https://doi.org/10.3102/10769986006004377>

Sanders, T., Liu, Y., Buchner, V., Tchounwou, P.B., 2009. Neurotoxic effects and biomarkers of lead exposure: A review. *Rev. Environ. Health.* <https://doi.org/10.1515/REVEH.2009.24.1.15>

Schafer, E.H., 1956. The Early History of Lead Pigments and Cosmetics in China. *T'oung Pao* 44, 413–438. <https://doi.org/10.1163/156853256X00135>

Schwartz, J., 1994. Low-level lead exposure and children's IQ: a meta-analysis and search for a threshold. *Environ. Res.* 65, 42–55. <https://doi.org/10.1006/enrs.1994.1020>

Schwikowski, M., Barbante, C., Doering, T., Gaeggeler, H.W., Boutron, C., Schotterer, U., Tobler, L., Van De Velde, K., Ferrari, C., Cozzi, G., Rosman, K., Cescon, P., 2004. Post-17th-Century Changes of European Lead Emissions Recorded in High-Altitude Alpine Snow and Ice. *Environ. Sci. Technol.* 38, 957–964. <https://doi.org/10.1021/es034715o>

- Seeber, A., Meyer-Baron, M., Schäper, M., 2002. A summary of two meta-analyses on neurobehavioral effects due to occupational lead exposure. *Arch. Toxicol.* <https://doi.org/10.1007/s00204-001-0315-5>
- Sindhu, K.K., Sutherling, W.W., 2015. Role of Lead in the Central Nervous System: Effect on Electroencephalography, Evoked Potentials, Electroretinography, and Nerve Conduction. *Neurodiagn. J.* <https://doi.org/10.1080/21646821.2015.1043222>
- Stanley, T.D., 2008. Meta-Regression Methods for Detecting and Estimating Empirical Effects in the Presence of Publication Selection*. *Oxf. Bull. Econ. Stat.* 70, 305–9049. <https://doi.org/10.1111/j.1468-0084.2007.00487.x>
- Stanley, T.D., Doucouliagos, H., 2014. Meta-regression approximations to reduce publication selection bias. *Res. Synth. Methods* 5, 60–78. <https://doi.org/10.1002/jrsm.1095>
- Stanley TD, Doucouliagos H, Ioannidis JP, 2017. Finding the power to reduce publication bias. *Stat Med.* 10;36(10):1580-1598.
- Statistics Canada, 2019. Number, rate and percentage changes in rates of homicide victims [WWW Document]. <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3510006801>
- Stretesky, P.B., Lynch, M.J., 2001. The Relationship Between Lead Exposure and Homicide. *JAMA Pediatr.* 155, 579–582. <https://doi.org/10.1001/archpedi.155.5.579>
- Stromberg, U., Schutz, A., Skerfving, S., 1995. Substantial decrease of blood lead in Swedish children, 1978-94, associated with petrol lead. *Occup. Environ. Med.* 52, 764–769. <https://doi.org/10.1136/OEM.52.11.764>
- Tanaka, S., Teshima, K., Verhoogen, E., n.d. North-South Displacement Effects of Environmental Regulation: The Case of Battery Recycling. *Am. Econ. Rev. Insights.* <https://doi.org/10.1257/AERI.20210201>
- Tcherni-Buzzeo, M., 2019. The “Great American Crime Decline”: Possible Explanations. Springer, Cham, pp. 309–335. https://doi.org/10.1007/978-3-030-20779-3_16
- Ucrdatatool.gov, 2019. Uniform Crime Reporting Statistics. <http://www.ucrdatatool.gov/abouttheucr.cfm> (accessed November 17, 2019).
- UK Home Office, 2012. Historical crime data - GOV.UK. <https://www.gov.uk/government/statistics/historical-crime-data> (accessed November 14, 2019).
- Vlasak, T., Jordakieva, G., Gnams, T., Augner, C., Crevenna, R., Winker, R., Barth, A., 2019. Blood lead levels and cognitive functioning: A meta-analysis. *Sci. Total Environ.* <https://doi.org/10.1016/j.scitotenv.2019.03.052>
- Westreich, D., and Greenland, S., 2013. The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients, *American Journal of Epidemiology*, Volume 177, Issue 4, Pages 292–298, <https://doi.org/10.1093/aje/kws412>
- WHO, 2010a. WHO | Preventing disease through healthy environments: a global assessment of the burden of disease from environmental risks. WHO.
- WHO, 2010b. Childhood Lead Poisoning. WHO
- Winter, A.S., Sampson, R.J., 2017. From lead exposure in early childhood to adolescent health: A Chicago birth cohort. *Am. J. Public Health* 107, 1496–1501. <https://doi.org/10.2105/AJPH.2017.303903>
- Wright, J.P., Dietrich, K.N., Ris, M.D., Hornung, R.W., Wessel, S.D., Lanphear, B.P., Ho, M., Rae, M.N., 2008.

Association of prenatal and childhood blood lead concentrations with criminal arrests in early adulthood. PLOS Med. 5, 732–740. <https://doi.org/10.1371/journal.pmed.0050101>

Zahran, S., Iverson, T., McElmurry, S.P., Weiler, S., 2017. The Effect of Leaded Aviation Gasoline on Blood Lead in Children. <https://doi.org/10.1086/691686> 4, 575–610. <https://doi.org/10.1086/691686>

Studies Included in Meta-Analysis

Aizer, A., Currie, J., 2019. Lead and Juvenile Delinquency: New Evidence from Linked Birth, School, and Juvenile Detention Records. *Rev. Econ. Stat.* 101, 575–587.

https://doi.org/10.1162/rest_a_00814

Barrett, K.L., 2017. Exploring Community Levels of Lead (Pb) and Youth Violence. *Sociol. Spectr.*

<https://doi.org/10.1080/02732173.2017.1319307>

Beckley, A.L., Caspi, A., Broadbent, J., Harrington, H., Houts, R.M., Poulton, R., Ramrakha, S., Reuben, A., Moffitt, T.E., 2018. Association of Childhood Blood Lead Levels With Criminal Offending. *JAMA Pediatr.* 172, 166–173.

<https://doi.org/10.1001/jamapediatrics.2017.4005>

Billings, S.B., Schnepel, K.T., 2018. Life after lead: Effects of early interventions for children exposed to lead. *Am. Econ. J. Appl. Econ.*

<https://doi.org/10.1257/app.20160056>

Curci, F., Masera, F., 2018. Flight From Urban Blight: Lead Poisoning, Crime and Suburbanization. *SSRN Electron. J.* <https://doi.org/10.2139/ssrn.3245090>

Dills, A.K., Miron, J.A., Summers, G., 2008. What do economists know about crime? NBER working paper series. <https://www.nber.org/papers/w13759>

Feigenbaum, J.J., Muller, C., 2016. Lead exposure and violent crime in the early twentieth century. *Explor. Econ. Hist.* 62, 51–86.

<https://doi.org/10.1016/j.eeh.2016.03.002>

Fergusson, D.M., Boden, J.M., Horwood, L.J., 2008. Dentine lead levels in childhood and criminal behaviour in late adolescence and early adulthood. *J. Epidemiol. Community Health* 62, 1045–1050.

<https://doi.org/10.1136/jech.2007.072827>

Grönqvist, H., Nilsson, J.P., Robling, P.O., 2020. Understanding how low levels of early lead exposure affect children's life trajectories. *J. Polit. Econ.*

<https://doi.org/10.1086/708725>

Lauritsen, J.L., Rezey, M.L., Heimer, K., 2016. When Choice of Data Matters: Analyses of U.S. Crime Trends, 1973–2012. *J. Quant. Criminol.* 32, 335–355.

<https://doi.org/10.1007/s10940-015-9277-2>

Lersch, K.M., Hart, T.C., 2014. Environmental Justice, Lead, and Crime: Exploring the Spatial Distribution and Impact of Industrial Facilities in Hillsborough County, Florida. *Sociol. Spectr.*

<https://doi.org/10.1080/02732173.2014.857184>

Manduca, R., Sampson, R.J., 2019. Punishing and toxic neighborhood environments independently predict the intergenerational social mobility of black and white children. *Proc. Natl. Acad. Sci. U. S. A.* 116, 7772–7777.

<https://doi.org/10.1073/pnas.1820464116>

Masters, R.D., Hone, B., Doshi, A., Hone, B., Doshi, A., 1998. ENVIRONMENTAL POLLUTION, NEUROTOXICITY, AND CRIMINAL VIOLENCE 23–58.

<https://doi.org/10.1201/9781482287400-8>

McCall, P.L., Land, K.C., 2004. Trends in environmental lead exposure and troubled youth, 1960–1995: an age-period-cohort-characteristic analysis. *Soc. Sci. Res.* 33, 339–359.

<https://doi.org/10.1016/j.ssresearch.2003.06.002>

Mielke, H.W., Zahran, S., 2012. The urban rise and fall of air lead (Pb) and the latent surge and retreat of societal violence. *Environ. Int.* 43, 48–55.

<https://doi.org/10.1016/j.envint.2012.03.005>

Needleman, H.L., McFarland, C., Ness, R.B., Fienberg, S.E., Tobin, M.J., 2002. Bone lead levels in adjudicated delinquents - A case control study. *Neurotoxicol.*

Teratol. 24, 711–717.

[https://doi.org/10.1016/S0892-0362\(02\)00269-6](https://doi.org/10.1016/S0892-0362(02)00269-6)

Nevin, R., 2000. How lead exposure relates to temporal changes in IQ, violent crime, and unwed pregnancy. *Environ. Res.* 83, 1–22.

<https://doi.org/10.1006/enrs.1999.4045>

Nevin, R., 2007. Understanding international crime trends: The legacy of preschool lead exposure.

Environ. Res. 104, 315–336.

<https://doi.org/10.1016/j.envres.2007.02.008>

Nkomo, P., Mathee, A., Naicker, N., Galpin, J., Richter, L.M., Norris, S.A., 2017. The association between elevated blood lead levels and violent behavior during late adolescence: The South African Birth to Twenty Plus cohort. *Environ. Int.* 109, 136–145.

<https://doi.org/10.1016/j.envint.2017.09.004>

Reyes, J., 2007. Environmental policy as social policy? The impact of childhood lead exposure on crime. *B E J. Econ. Anal. POLICY* 7, 51.

<https://doi.org/10.2202/1935-1682.1796>

Reyes, J.W., 2015. Lead exposure and behavior: Effects on antisocial and risky behavior among children and adolescents. *Econ. Inq.* 53, 1580–1605.

<https://doi.org/10.1111/ecin.12202>

Sampson, R.J., Winter, A.S., 2018. Poisoned development: assessing childhood lead exposure as a cause of crime in a birth cohort followed through adolescence. *Criminology* 56, 269–301.

<https://doi.org/10.1111/1745-9125.12171>

Stretesky, P.B., Lynch, M.J., 2004. The relationship between lead and crime. *J. Health Soc. Behav.*

<https://doi.org/10.1177/002214650404500207>

Taylor, M.P., Forbes, M.K., Opeskin, B., Parr, N., Lanphear, B.P., 2018. Further analysis of the relationship between atmospheric lead emissions and aggressive crime: an ecological study. *Environ. Heal.*

17. <https://doi.org/10.1186/s12940-018-0354-5>