The Lead-Crime Hypothesis: A Meta-Analysis*

Anthony Higney[i†], Nick Hanley[ii], Mirko Moro[iii]

Current Version: November 30th 2021

Abstract

Does lead pollution increase crime? We perform the first meta-analysis of the effect of lead on crime by pooling 542 estimates from 24 studies. We analyse the full sample, converted to partial correlations, and a subsample for which elasticities are available. In both samples we find evidence of publication bias across a range of tests. This publication bias means that the effect of lead is overstated in the literature. We perform over 1 million meta-regression specifications, controlling for this bias, and conditioning on observable between-study heterogeneity. The estimated mean effect size is a partial correlation of 0.16 in the full sample, and an elasticity of 0.11 in the subsample. When we restrict our analysis to only high-quality studies that address endogeneity the estimated mean partial correlation is close to zero. An elasticity of 0.11 suggests the fall in lead over recent decades may be responsible for a 10% fall in homicide in the US. Our results imply that although lead pollution does have a positive effect on crime, it does not explain the majority of the fall in crime observed in many countries in the 20th century, and additional explanations are needed.

Jel codes: C83, K42, Q53.

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

*We thank Paul Ferraro, Kevin Schnepel, Tom Stanley, Hector Gutierrez Rufrancos, Conny Wollbrant and Jonathan Norris, 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.
[i] Division of Economics, University of Stirling, Stirling FK9 4LA.
[ii] Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow.
[iii] Division of Economics, University of Stirling, Stirling FK9 4LA.
†Corresponding author: Email: a.c.higney@stirling.ac.uk

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.

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 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 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. However, the rise and fall pattern in figure I is by no means uniform. Further, Buonanno *et al*. (2011) show that while total crime has behaved similarly to homicide in the US, it has not in Europe (figure III). 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 of estimates. We convert the estimates to comparable effect sizes. For the full sample we use partial correlation coefficients. We also convert estimates to elasticities, where it is possible to do so, and analyse this subsample. 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 0.11 in the subsample. We also find there are substantial differences between the average effect size when we use the full sample, and when we use only the high-quality study designs that address endogeneity. The mean partial correlation coefficient for the "addressing endogeneity" sample is only 0.01, far smaller than the full sample estimate, and within one standard deviation of zero.. The elasticity subsample was too small to separate in this way. The sample of studies that have crime in an area as the focus of analysis have a larger mean effect size compared to that 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 elasticity estimate suggests the fall in blood lead levels is responsible for a 10% fall in homicide from its peak. The homicide rate fell by more than half in this period, leaving around 80% of the fall unaccounted for. 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. 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 artic ice 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[4]. 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 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

[4] They both convert easily to ions with 2+ charge.

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]
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[6] flow diagram (figure IV). A review and description of the studies included is given in appendix A.

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 did not fulfil criteria four and were excluded at the eligibility stage. One other paper (Masters and

[5] By which we mean they must be convertible to a common estimate such as a partial correlation coefficient or an elasticity. See discussion below.
[6] PRISMA is Preferred Reporting Items for Systematic Reviews and Meta-Analyses. It is a standard across sciences for how to report any process where a systematic literature search with filtering is performed.

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 criteria 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 results we use all estimates, 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. We also were able to convert some study estimates into elasticities. We carry out all analysis on the elasticity sample as well as the full sample. However, the elasticity sample is smaller, due to many studies not reporting enough estimation details or descriptive statistics, which does not allow us to split it into further subsamples, as we do with the full sample.

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 quantitively 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.

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. It also cannot be separated into further subsamples, such as only studies which address endogeneity, whereas we can do this with the full sample using PCCs.

Table I presents the mean, median and weighted average PCC for each study (with weights being equal to the precision, 1/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.

4. Methods

4.1 General Approach[7]

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.

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

[7]This sections 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.

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):

(1) $$\theta_j \sim N(\theta, \tau^2)\ \forall\ j$$

(2) $$\hat{\theta}_{ij} \sim N\left(\theta_j, \sigma_{ij}^2\right)\ \forall\ i \text{ and } \forall\ j$$

(3) $$\hat{\theta}_{ij} \mid \theta, \sigma_{ij}^2, \tau^2 \sim N\left(\theta, \sigma_{ij}^2 + \tau^2\right)\ \forall\ i \text{ and } \forall\ j$$

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 $\tau^2$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. $\tau^2 \to \infty$ 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 V and doing the same with the elasticities in figure VI.

We show the common and random effects estimates at the bottom of each figure. 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.

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

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

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 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. $\hat{H}^2$ 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 an estimate of the variance of the effect size distribution in (3) using the DerSimonian-Laird (1986) method and is less sensitive to the number of studies but does not give a sense of how important between-study heterogeneity is compared to within-study sampling variation.

Looking at table III we can see which variables seem important for heterogeneity and the different estimated average effect sizes. The 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.

Studies that look at individual 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 and 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.

The first and most common step is to simply chart the data and visually inspect for bias, using a funnel plot. Figure VII 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 one or more 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 VII 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 precision, as there seems to be in figure VII, would indicate the presence of publication bias (see appendix D). This naturally leads to the estimating equation (5).

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

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. (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 VII. 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).

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

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[8]; 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 use 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 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 for the full sample in PCCs, and table V for the elasticity sample. Linear methods allow for not only an effect beyond bias estimate but

[8] The Funnel Asymmetry test and Precision Effect Estimate with Standard Error (FAT-PEESE). Stanley and Doucouligas (2014) find the FAT-PEESE can sometimes perform better in simulations. They find the FAT-PEESE seems to especially perform better when the true effect is not equal to zero.

an indication of the strength of bias in the FAT coefficient. These are all positive with only the elasticity IV 95% interval not covering zero. The effect beyond bias estimates are all smaller than the random effects estimate for both the full sample PCCs and the elasticity sample. The linear estimates for the full sample are all close to zero, while for the elasticity sample they vary from 0.11-0.13, except for the IV which is close to zero.

The full-sample non-linear point estimates are similar to the linear ones and close to zero, save the Andrews-Kasy estimate which is -0.77. However, the 95% confidence interval covers zero, and this estimate is the outlier. For the elasticity sample, the WAAP and trim and fill methods show similar results to the non-IV linear methods, while the Andrews-Kasy method is close to zero. 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, $\boldsymbol{x_{ij}}$, 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).

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

Where $\boldsymbol{z_{ij}}$ 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 VI. 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 high-quality 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 e 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[9]. 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

[9] Grateful to Paul Ferraro for his comments on this.

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 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. 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. The vast majority of studies 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[10]. However, with this many covariates and modern computational power it is possible to estimate all 2^K specifications[11]. 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.

[10] For an example, see Gechert, Havranek, Irsova, and Kolcunova (2020)

[11] 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 0.12, about the same as we find using this method.

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 VII. 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), but we include the densities of the distributions of the estimated coefficients in figure VIII and the mean, median and standard deviations of the distributions in table VIII in case of interest.

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 using the sample averages or 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 predicated value (i.e. removing the publication bias). 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 with GMM (i.e. base case compared to OLS and 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 IX. 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 the 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 IX. 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.

The elasticity effect sizes are plotted in figure X. They are both unimodal with a mean and median elasticity evaluated at the sample averages of 0.11, and at the "ideal" specification of 0.07 and 0.08.

We next restrict the sample to only high-quality studies that estimate a causal effect rather than an association: our "addressing endogeneity" sub-sample. 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 XI (excluding those variables that cannot be included in the estimation, see table VII). 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 high-quality causal estimating studies and the rest of the sample.

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.

4.5 Explaining the 20th Century Crime Decline

Our calculated elasticities allow us to estimate how much of the fall in crime in the second half of the 20th century was caused by lead. We 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. Given our elasticity estimate of 0.11 (with standard deviation of 0.01), this implies a fall in homicide of 9%-12% with the point estimate at the mean being 10%. The US homicide rate fell 54% from its peak in 1989 to 2014. This would mean that lead accounts for around 20% of the decrease in homicide.

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. 80% of the fall in homicide in the US is unaccounted for.

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, we find an average effect size of 0.16 as a partial correlation, or 0.11 as an elasticity. We find that the average meta-analysis estimate for high-quality studies that address endogeneity is much smaller than for the full sample, or for the correlational sample. Similarly, 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 homicide in the US that is accounted for by reductions in blood lead levels. We estimate reduced blood lead levels reduced homicides by 10%. A substantial decrease. However, this left 80% of the fall in homicide unaccounted for.

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-analysis. 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.

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 IX 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 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

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

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. https://doi.org/10.1016/j.neuro.2009.07.007

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.

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

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.

Doucouliagos, 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

Doucouliagos, 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 2020 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).

Gechert, S., Havranek, T., Irsova, Z., Kolcunova, D., 2020. Measuring Capital-Labor Substitution: The Importance of Method Choices and Publication Bias.

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

Havránek, T., Stanley, T.D., Doucouliagos, H., Bom, P., Geyer-Klingeberg, 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.

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

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

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).

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 neurobehavioural 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 Electroencephlography, 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

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., Gnambs, 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

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

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

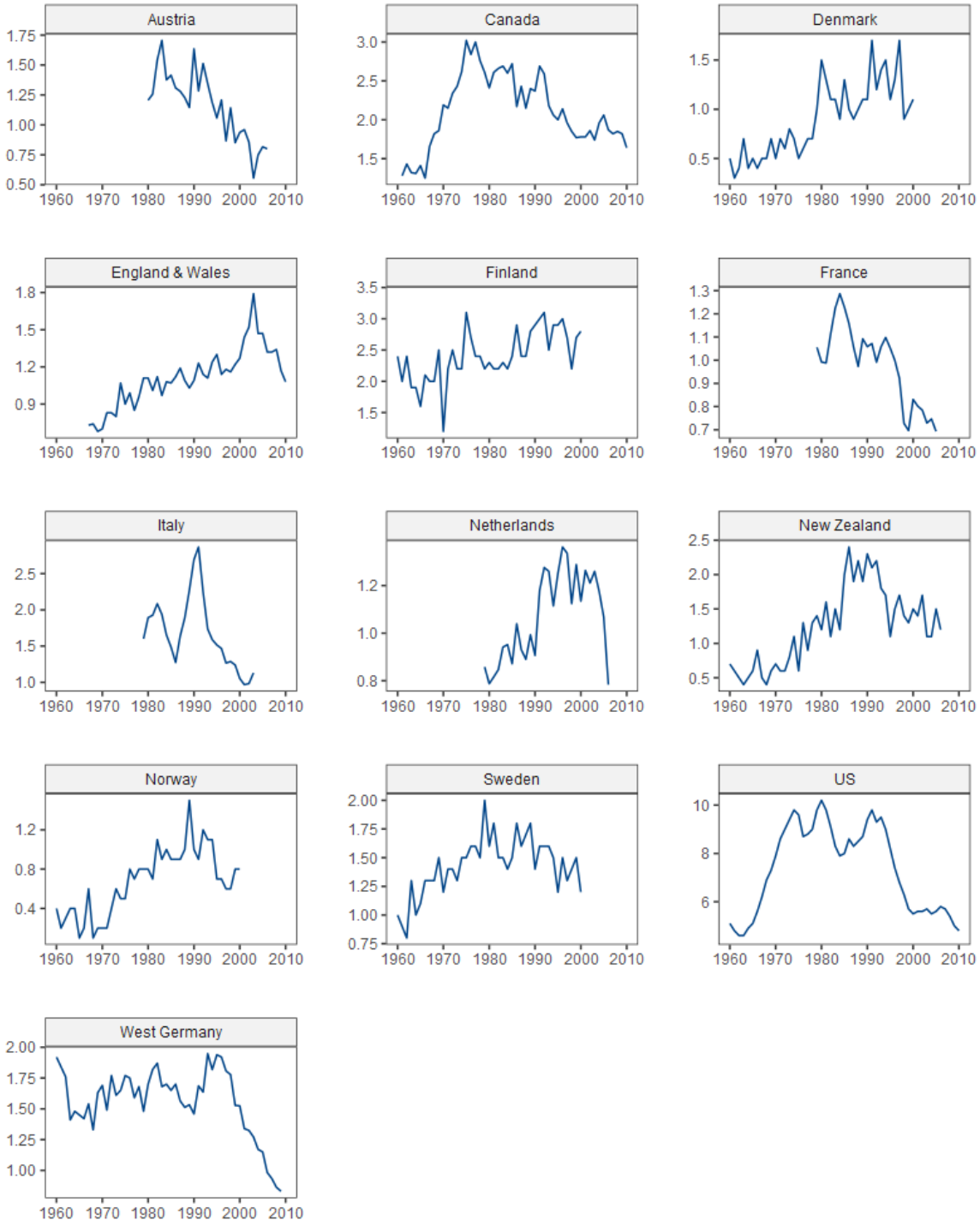
Reyes, J.W., 2015. Lead exposure and behavior: Effects on antisocial and risky behavior among children and adolescents. Econ. Inq. 53, 1580–1605. https://doi.org/10.1111/ecin.12202

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

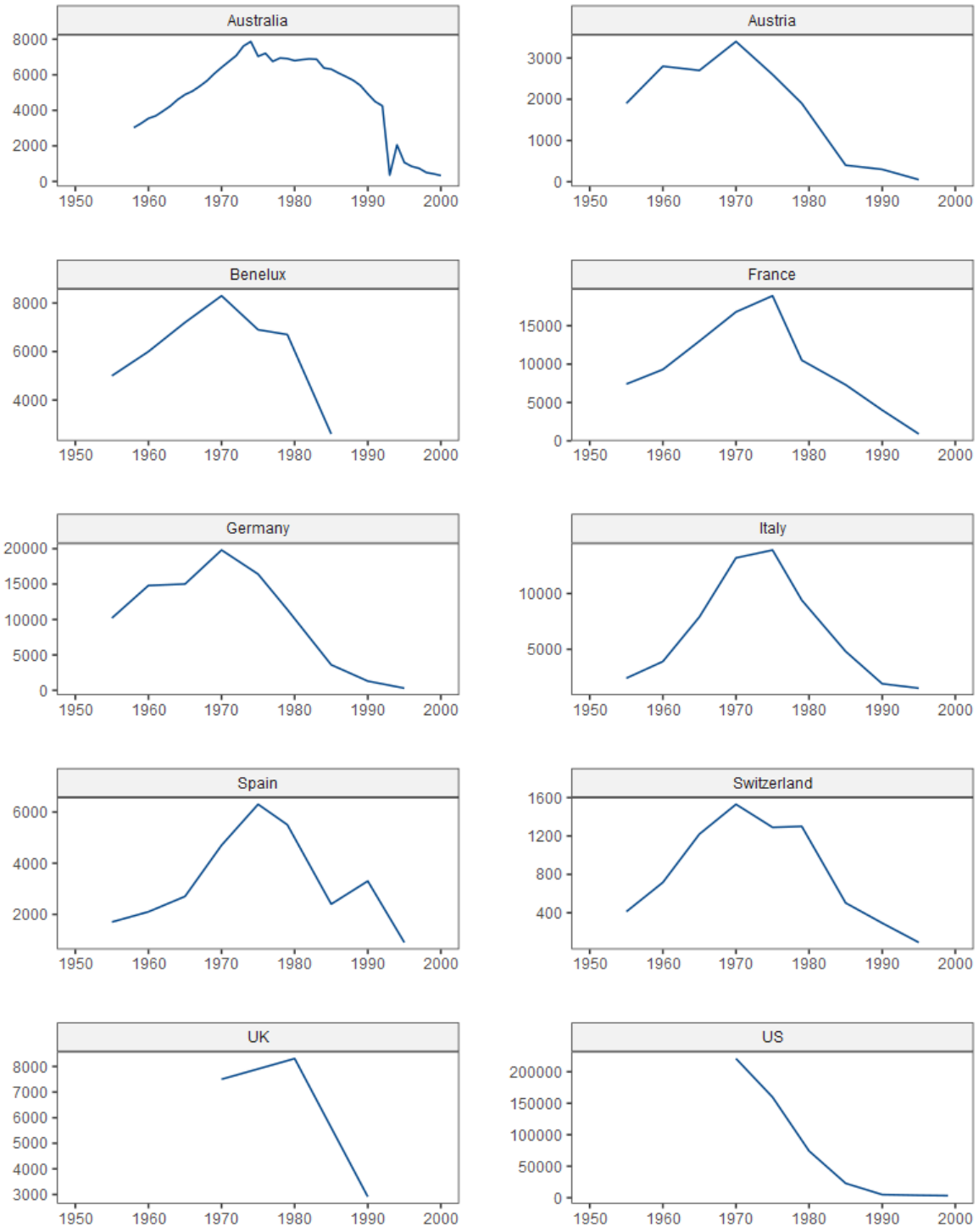
Taylor, M.P., Forbes, M.K., Opeskin, B., Parr, N., Lanphear, B.P., 2018. Further analysis of the relationship between atmospheric lead emissions and aggressive crime: an ecological study. Environ. Heal. 17. https://doi.org/10.1186/s12940-018-0354-5

Figure I – Homicide Rate per 100,000 by Country



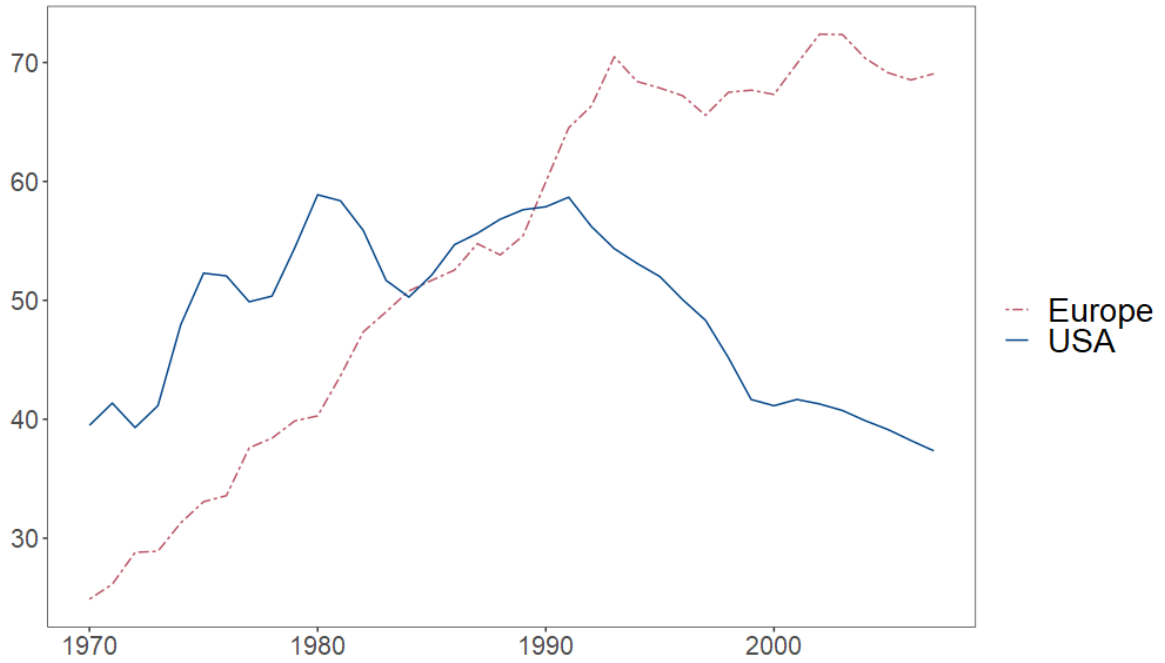
Sources: New Zealand Police (2018); Buonanno *et al*. (2011), UK Home Office (2019); Uniform Crime Reports for the United States (2019); Falck, Von Hofer & Storgaard (2003); Statistics Canada (2019); Birkel and Dern (2012). Back to place in paper.

Figure II – Lead Emissions by Country (1000 kg Y^{-1})



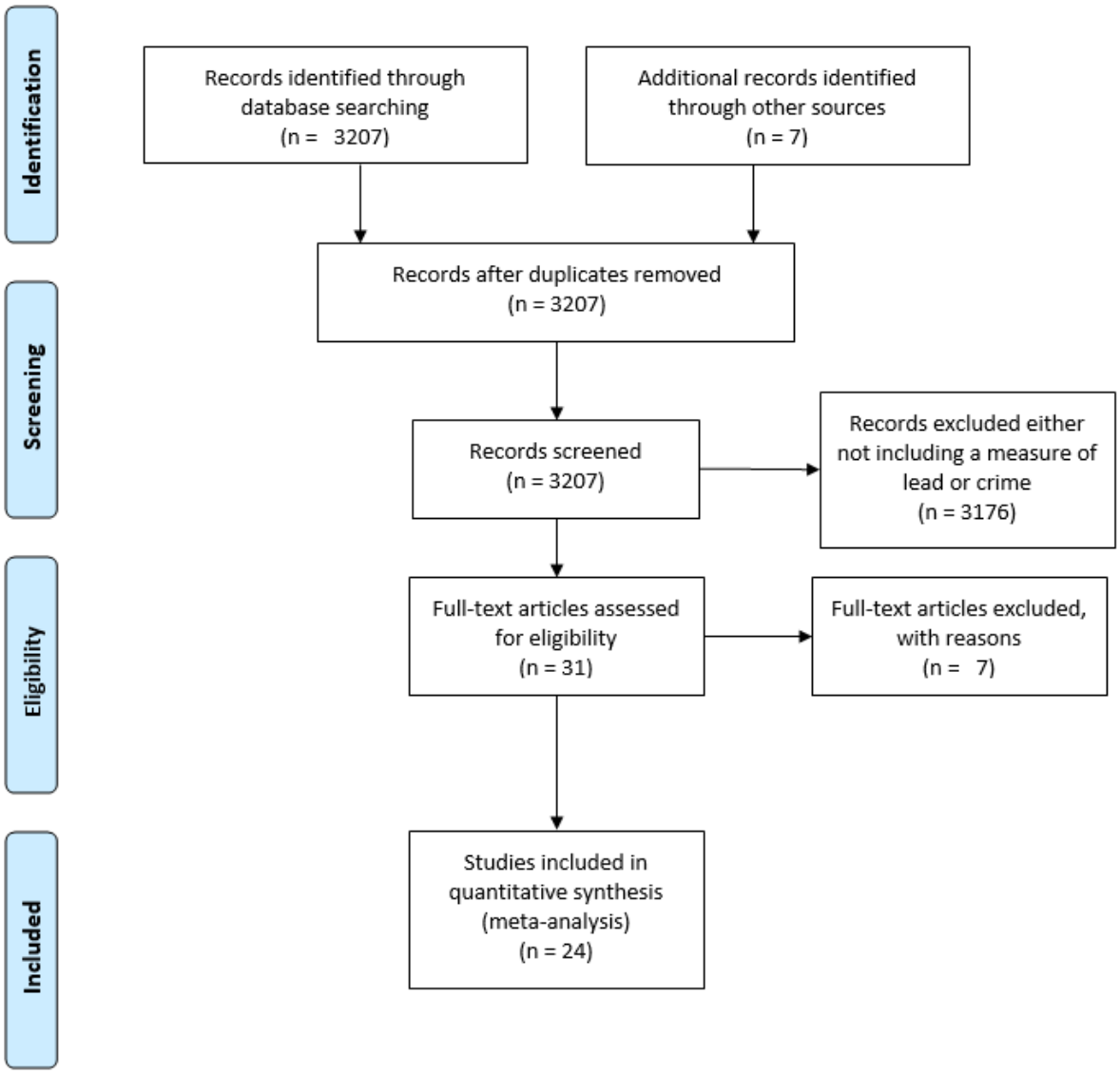
Source: Dore *et al*. (2006), Schwikowski *et al*. (2004), Kristensen (2015), Statistical Abstract of the United States (2009). Back to place in paper.

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



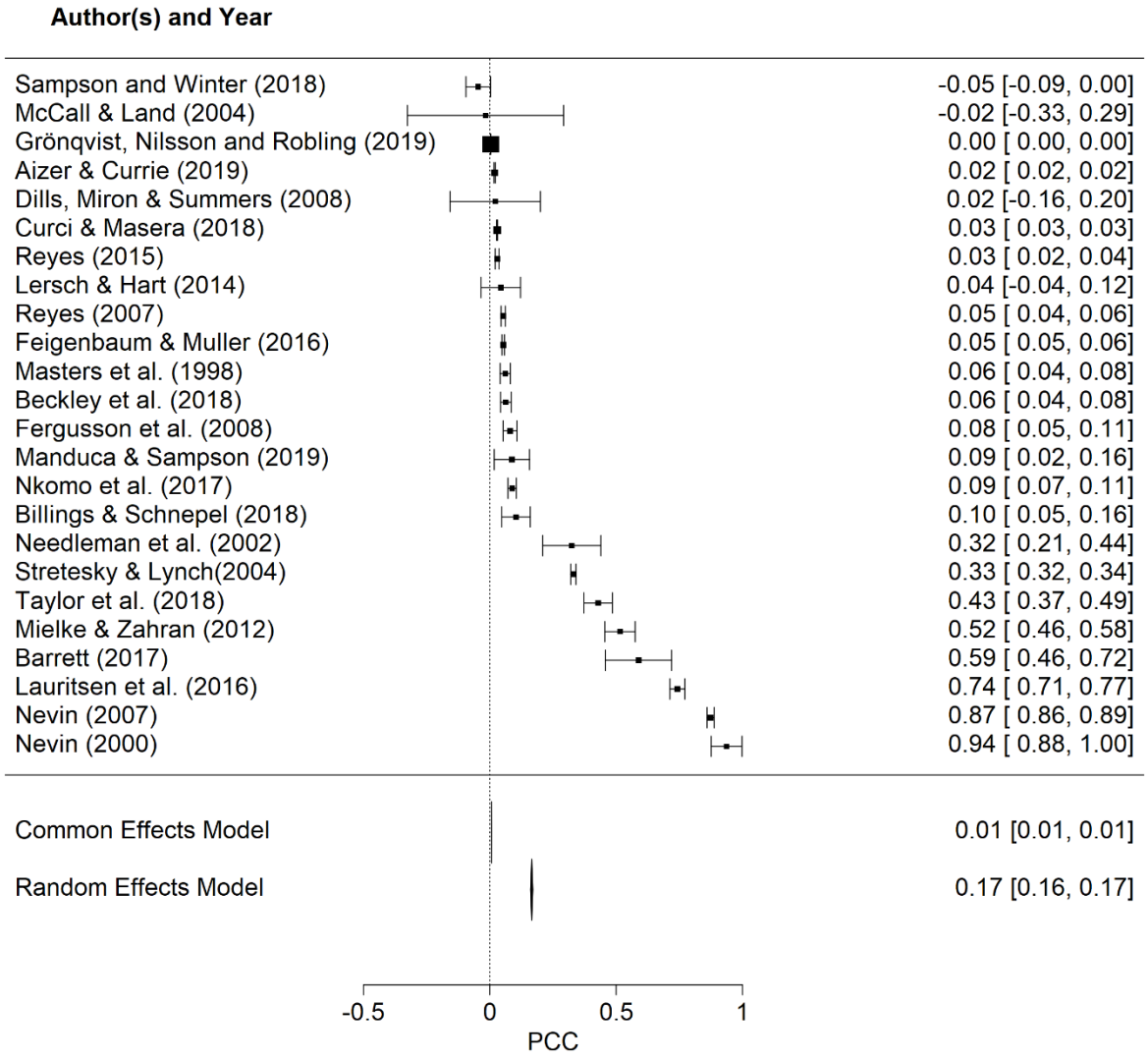
Source: Buonanno *et al*. (2011). The countries are: Austria, France, Germany, Italy, The Netherlands, Spain, and the UK.
Back to place in paper.

Figure IV – PRISMA Flow Diagram of Selection Process



Back to place in paper.

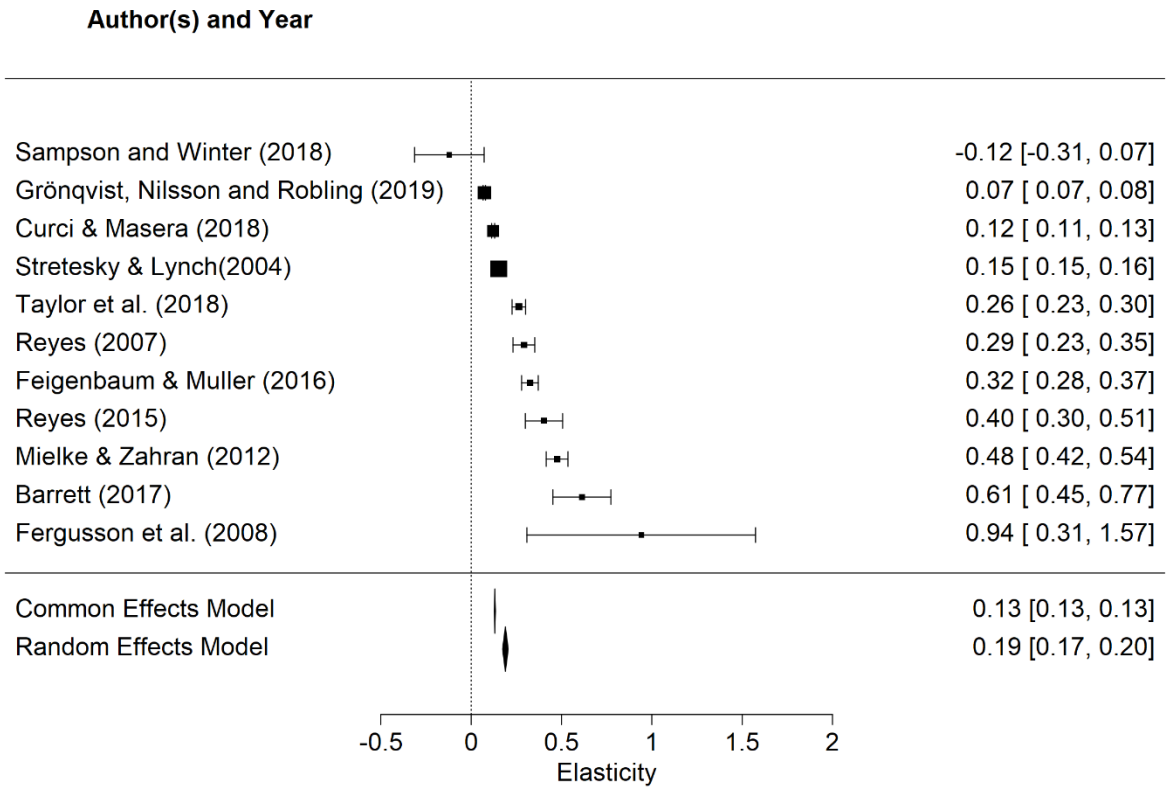
Figure V – Forest Plot, Partial Correlations



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 for details).

Bottom of table shows common effects and random effects estimates for all studies combined (see appendix for details). Numbers on right are the point estimates and the 95% confidence intervals. Back to place in paper.

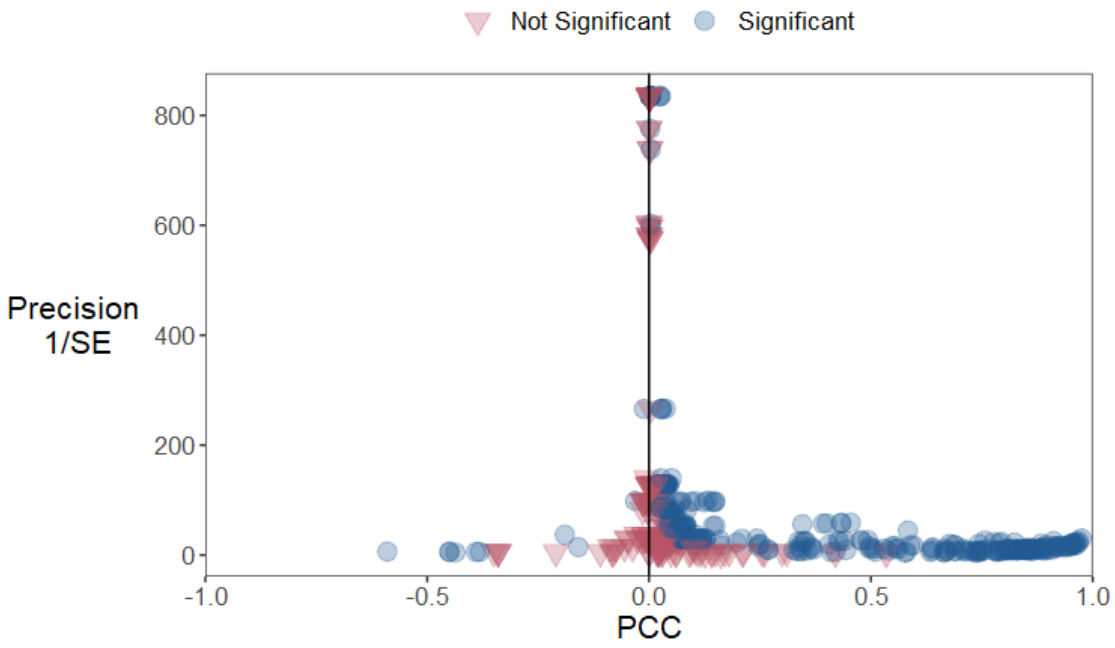
Figure VI – Forest Plot, Elasticities

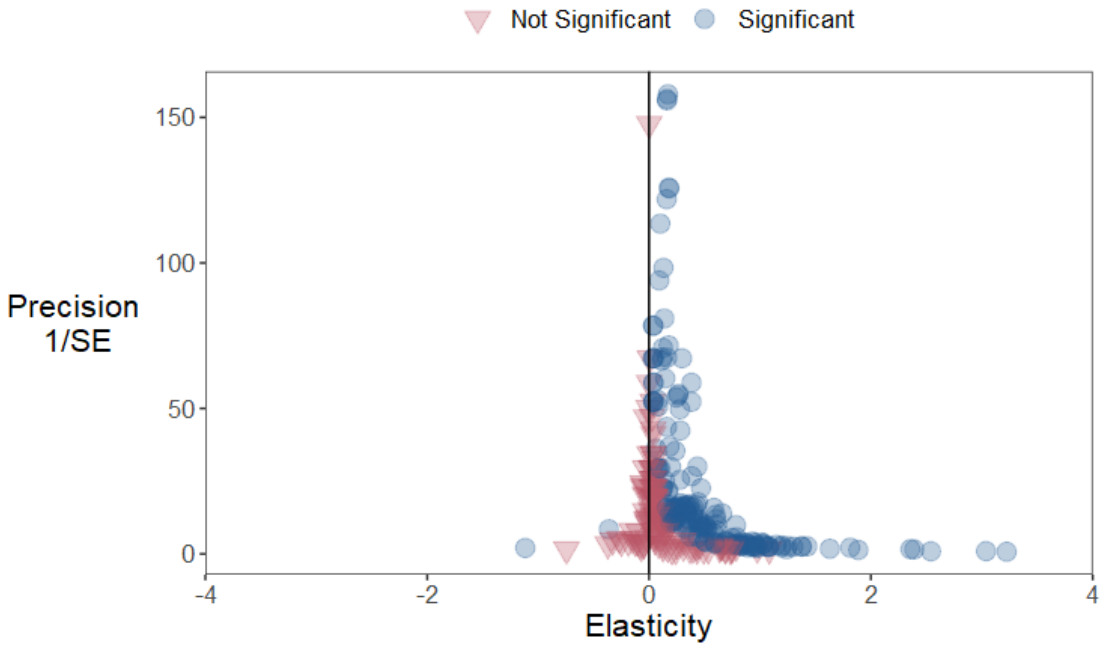


Notes. Chart shows weighted average of each study's effect sizes converted to elasticities along with corresponding 95% confidence intervals.

Bottom of table shows common effects and random effects estimates for all studies combined (see appendix for details). Numbers on right are the point estimates and the 95% confidence intervals. Back to place in paper.

Figure VII – Funnel Charts



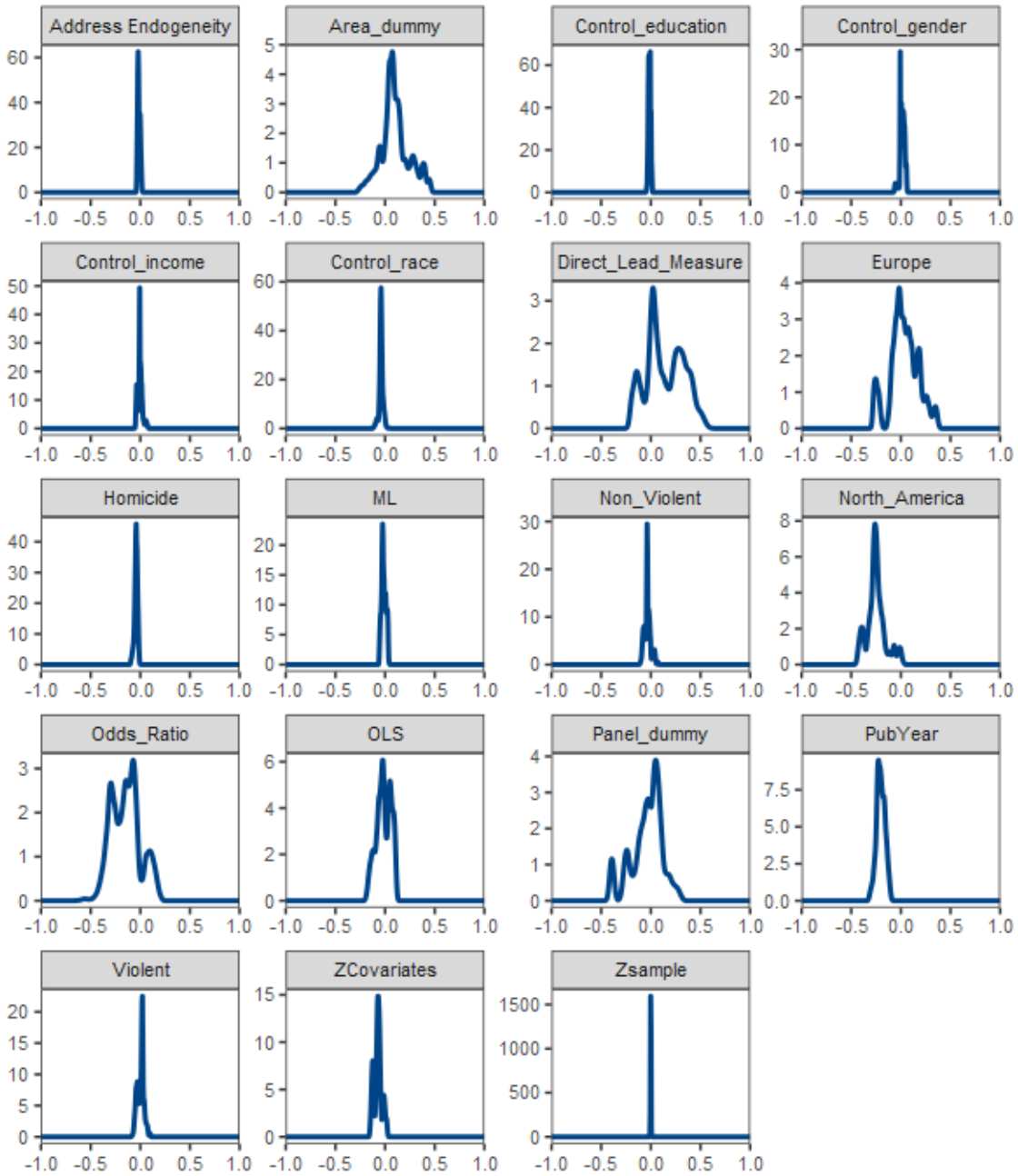


Notes. PCC = Partial Correlation Coefficient

Precision is one divided by the standard error.

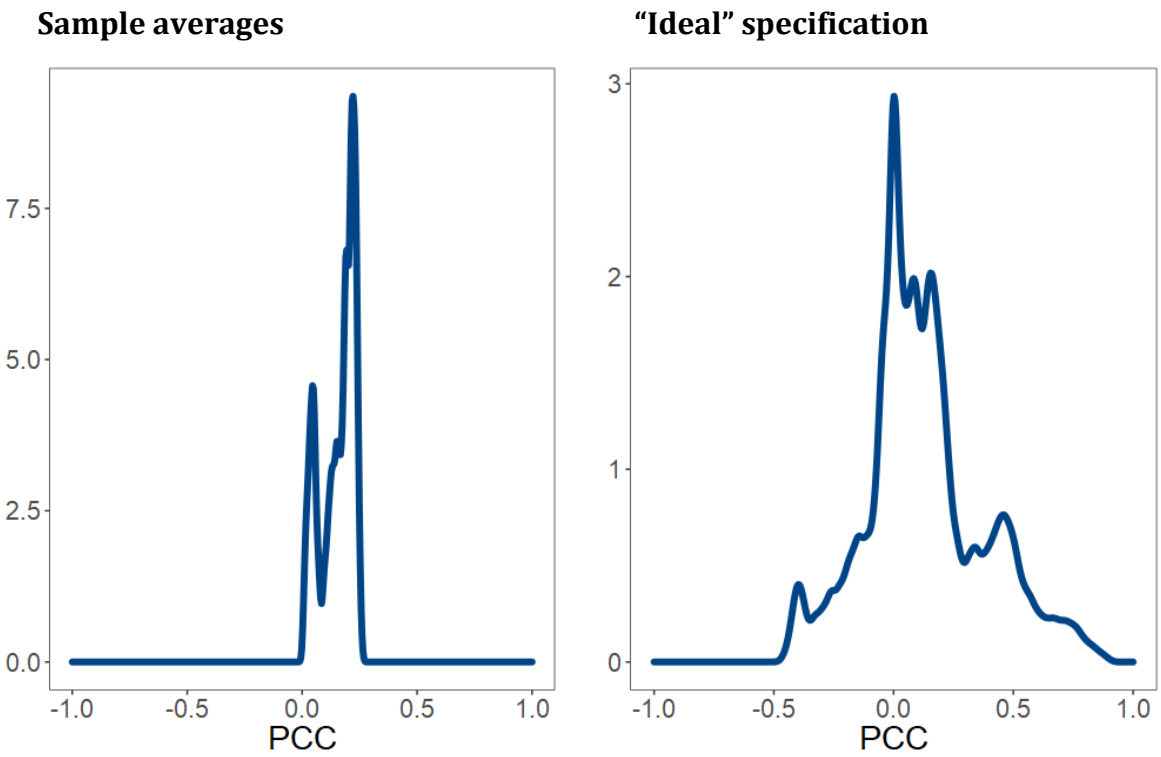
"Significant" means statistically significant at the 95% confidence level using two-sided critical values of a normal distribution. Back to place in paper.

Figure VIII – Density of Meta-Regression Coefficients on Covariates, PCCs



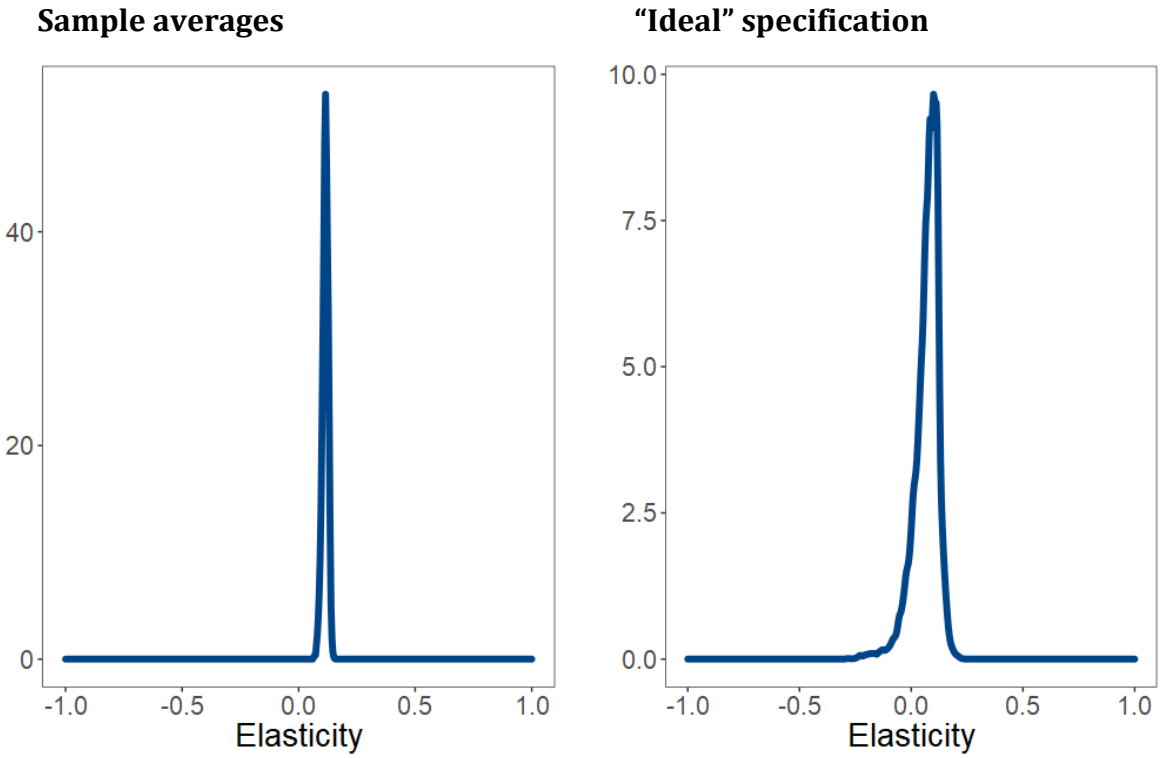
Notes. Chart shows densities for the distribution of meta-regression estimated coefficients on covariates for the full sample. Each chart shows the distribution of over 500,000 estimates of the coefficient. X axis truncated at feasible interval of a PCC, [-1,1]. Back to place in paper.

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



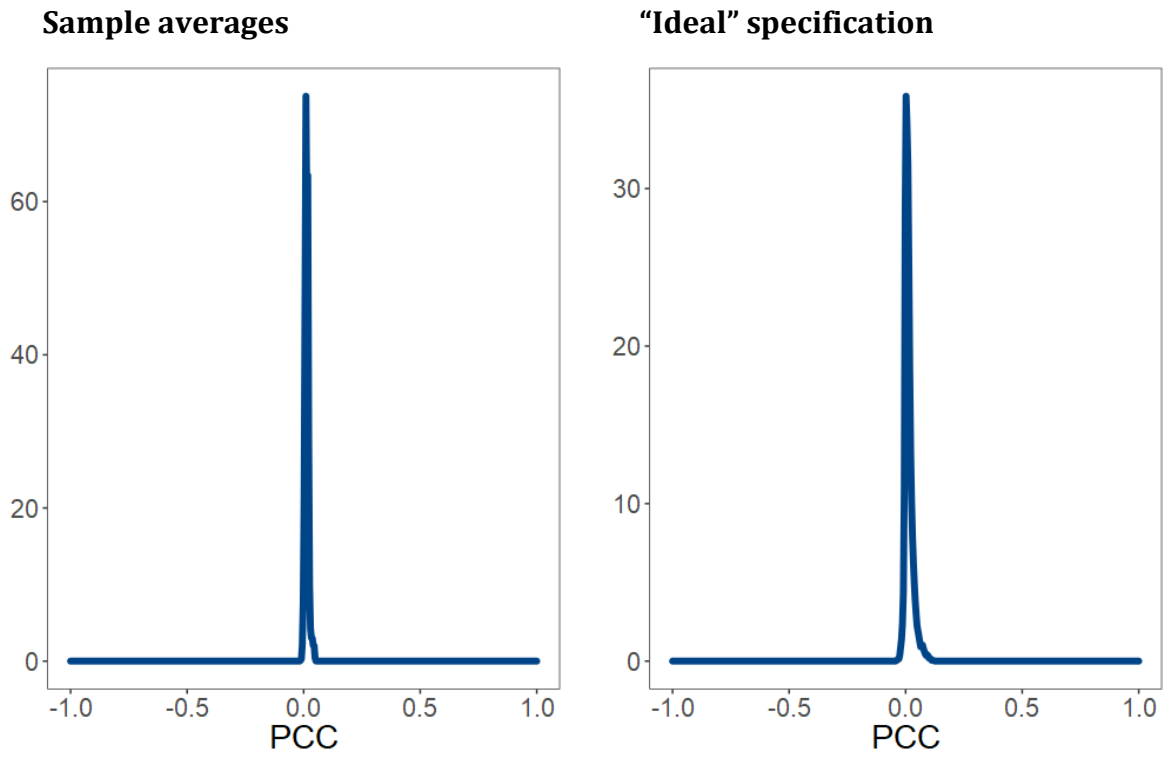
Notes. Chart shows densities for the distribution of meta-regression estimated average effect sizes. Chart on left shows estimated average effect for each specification evaluated at the sample averages. Chart on right shows estimated average effect for each specification evaluated-at an "ideal" specification.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

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



Notes. Chart shows kernel 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. Back to place in paper.

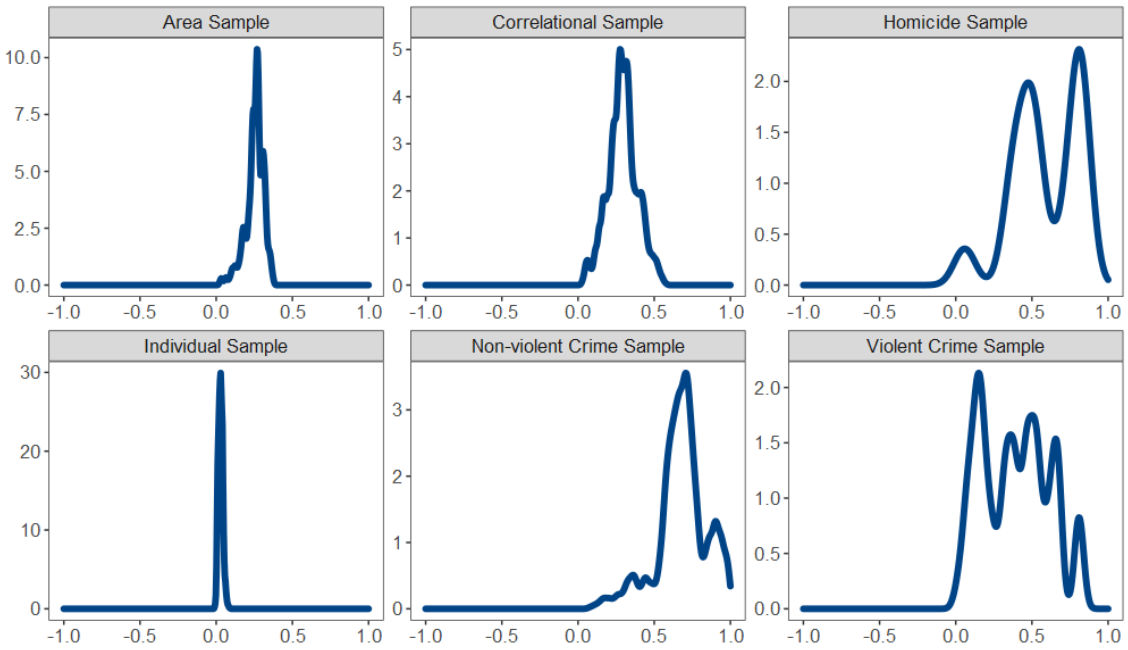
Figure XI – Density of Meta-Analysis Average Effect Estimates for "Addressing Endogeneity" Subsample



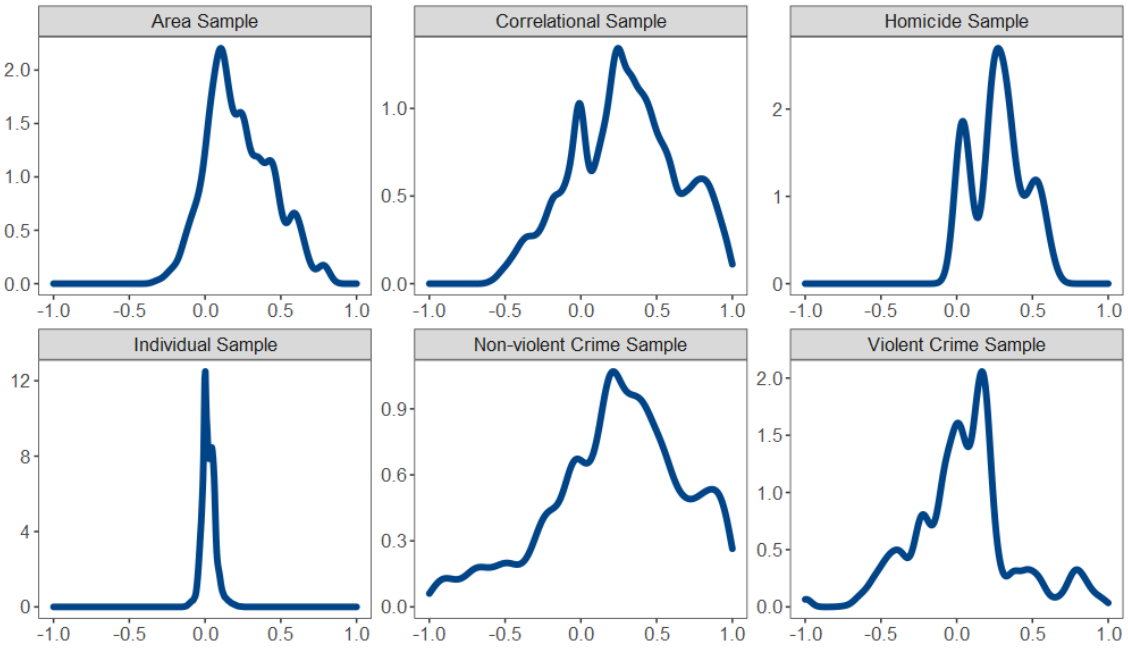
Notes. Chart shows kernel densities for the distribution of meta-regression estimated average effect sizes for the addressing endogeneity sub-sample. Chart on left shows estimated average effect for each specification evaluated at the sample averages. Chart on right shows estimated average effect for each specification evaluated-at an "ideal" specification. X axis truncated at feasible interval of a PCC, [-1,1]. Back to place in paper.

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

Sample averages



"Ideal" specification



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]. Back to place in paper.

Table I – Studies Used in Full Sample Meta-analysis

Study & Year	Median	Mean	Weighted Average	Type of Crime	Individual or Area	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 *et al.* (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 *et al.* (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 *et al.* (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 *et al.* (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 *et al.* (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 *et al.* (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 *et al.* (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. 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. Back to place in paper.

Table II – Studies Used in Elasticity Subsample Meta-analysis

Study & Year	Median	Mean	Weighted Average	Type of Crime	Individual or Area	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. 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. Back to place in paper.

Table III – Random Effects and Heterogeneity Estimates by Subsample

Sample	RE Estimate	SE	$\hat{\tau}^2$	$\hat{I}^2$	$\hat{H}^2$	Studies	Estimates
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	0.008	0.001	0.000	95	20	11	125
Area	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

Notes. RE Estimate is a random effects, meta-analysis estimate computed using DerSimonian-Laird (1986) method. All in PCCs except for elasticity sample. SE is the standard error of the RE estimate. τ^2, $\hat{I}^2$, and $\hat{H}^2$ are estimates of between-study heterogeneity. See section 4.2 for more details. Back to place in paper.
*Values in elasticities, not PCCs

Table IV – Effect Beyond Bias and Publication Bias Estimates, Partial Correlations

Panel A – Linear Models

	FAT-PET	FAT-PEESE	Multi-level FAT-PET	IV
Effect Beyond Bias	-0.003	0.005	0.006	-0.004
	(0.002)	(0.002)	(0.004)	(0.002)
Publication bias	5.026	32.227	3.502	5.062
	(1.283)	(8.638)	(0.885)	(1.297)
Groups	24	24	24	24
Observations	542	542	542	542

Panel B – Non-Linear Models

	WAAP	Trim and Fill	Andrews-Kasy
Effect Beyond Bias	0.005	0.008	-0.773
	(0.002)	(0.018)	(0.438)
Groups	N/A	N/A	24
Observations	542	542	542

Notes. Estimates are PCCs 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 FAT-PET is a mixed effects-multi-level model with a different slope coefficient 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. Trim and fill (Duval & Tweedie, 2000), removes outlier studies and then adds imputed studies before calculation an average effect. The Andrews-Kasy (Andrews & Kasy,2019) method is a step function selection model which reweights the observed sample with estimated publication probabilities. See Appendix D for full explanation of each method. Back to place in paper.

Table V – Effect Beyond Bias and Publication Bias Estimates, Elasticities

Panel A – Linear Models

	FAT-PET	FAT-PEESE	Multi-level FAT-PET	IV
Effect Beyond Bias	0.110	0.128	0.107	-0.056
	(0.029)	(0.021)	(0.010)	(0.087)
Publication bias	1.202	3.355	1.966	4.579
	(0.545)	(0.805)	(0.681)	(2.935)
Groups	11	11	11	11
Observations	312	312	312	312

Panel B – Non-Linear Models

	WAAP	Trim and Fill	Andrews-Kasy
Effect Beyond Bias	0.126	0.145	0.025
	(0.022)	(0.018)	(0.069)
Groups	N/A	N/A	11
Observations	312	312	312

Notes. Estimates are elasticities 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 FAT-PET is a mixed effects-multi-level model with a different slope coefficient 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. Trim and fill (Duval & Tweedie, 2000), removes outlier studies and then adds imputed studies before calculation an average effect. The Andrews-Kasy (Andrews & Kasy,2019) method is a step function selection model which reweights the observed sample with estimated publication probabilities. See Appendix D for full explanation of each method. Back to place in paper.

Table VI – Descriptive Statistics of MRA Covariates

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

Notes. See section 4.3 for description of variables. Back to place in paper.

*Indicates variables have been standardised.

Table VII – 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 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 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

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. Back to place in paper.

Table VIII – Densities of Coefficients On Covariates Used In Meta-Regression

	Mean	Median	Standard Deviation
Control_gender	0.01	0.02	0.02
Control_race	-0.04	-0.04	0.02
Control_income	0.00	-0.01	0.02
Control_education	-0.01	-0.01	0.01
Homicide	-0.04	-0.04	0.02
Violent	0.00	0.01	0.03
Non_Violent	-0.03	-0.04	0.03
Area_dummy	0.10	0.08	0.14
OLS	-0.01	-0.02	0.07
ML	-0.01	-0.02	0.02
Odds_Ratio	-0.15	-0.15	0.15
Panel_dummy	-0.04	-0.01	0.15
Addressing Endogeneity	-0.01	-0.02	0.01
North_America	-0.25	-0.26	0.09
Europe	0.04	0.03	0.14
Direct_Lead_Measure	0.14	0.12	0.19
Publication Year	-0.20	-0.20	0.04
Number of Covariates	-0.07	-0.07	0.04
Sample Size	0.00	0.00	0.00

Notes. Table shows mean, median and standard deviation of the distributions of the coefficients on covariates used in the full sample, PCC meta-regressions. Back to place in paper.

Table IX - 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 Sample	0.25	0.26	0.06	16384	0%
Individual 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*	0.11	0.11	0.01	131072	

"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 Sample	0.23	0.20	0.22	16384	0%
Individual 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*	0.07	0.08	0.06	131072	

Notes. 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. Back to place in paper.

Appendices

The Lead-Crime Hypothesis: A Meta-Analysis

A. Review of literature used in meta-analysis

There are 24 total studies included in this meta-analysis. The studies use different methods to examine the lead-crime relationship. Longitudinal studies, which track the same people over time, are common. Fergusson, Boden and Horwood (2008) use a longitudinal sample and find a positive association between dentine lead levels at 6-9 years of age and later offending while including race and family socioeconomic status covariates. However, the effect was smaller once variation in education grades was added. They reasoned that the effect of lead was in reducing education outcomes, leading to more crime. Overall, they find that lead only explains 1% of the variation in crime. Nkomo *et al*. (2017) used a longitudinal sample in South Africa and found a positive association between blood lead levels at age 13 and violent crime in later life. Beckley *et al*. (2018) find only a small positive effect of childhood lead levels and both violent and non-violent crime in their longitudinal sample of New Zealand residents. They conclude other factors are much more important for determining crime rates. Finally, Sampson and Winter (2018) follow a longitudinal sample in Chicago and find school age lead levels are not associated with an increase in arrests in later life. Overall, longitudinal studies show a mixed picture, both on whether there is an effect and whether it is a strong one.

A different strand of research looks at the correlation of lead levels and crime across time and areas, rather than at an individual level. Three studies look at time series of lagged lead levels and crime for the US. Nevin (2000) finds a positive effect, but McCall and Land (2004) find no effect on the age cohorts most affected in youth by the increase in leaded gasoline. They reason that increased lead levels at one time should only affect the crime rates of that cohort, not earlier cohorts, and so only look at crime rates for those certain age ranges. Lauritsen, Rezey, and Heimer (2016) look at two different data series of crime: the National Crime Victimization Survey (NCVS) and the Uniform Crime Reports (UCR). They find that lead is positively correlated with violent crime in the UCR but not the NCVS, which they consider a better measure of violent crime. However, they consider both data sources equally valid for property crime. Stretesky and Lynch (2004) find a strong effect when looking across US countries for both property and violent crime using the UCR. Mielke and Zahran (2012) find a strong effect across six US cities, Lersch and Hart (2014) find the same looking at Florida census tracts. Both Barrett (2017) and

Manduca and Sampson (2019) find a strong positive relationship in census tracts in Chicago using different methods. Looking outside the US, Taylor *et al*. (2018) find positive results for violent crime in Australia, and across six suburbs in New South Wales. Nevin (2007) estimates the relationship for many OECD countries and finds pre-school blood levels are strongly associated with a whole range of violent and non-violent crime. On the whole, studies which look at geographic areas as the unit of interest tend to find the strongest positive associations between lead and crime.

The final strand of the literature are those studies that attempt to identify a casual effect while accounting for endogeneity from unobserved variables correlated with both crime and lead. These could bias the estimate of the effect of lead on crime. Lead exposure is correlated with poverty (Baghurst *e al*. 1999) and race (Sampson and Winter, 2016) and likely with other, unobservable, variables. We cannot rule out that these variables may cause individuals to commit more crime and be more exposed to lead, rather than lead being the cause. Even panel data designs with controls may not account for this endogeneity. The endogeneity threat has led to some, more recent, studies using quasi-experimental methods. Needleman (2002) carried out a "case control" study where young offenders were matched to a "control" group chosen for similar observable characteristics. The offender group was found to have higher bone lead levels. Although this this is an improvement beyond looking at correlation alone, the likelihood of unobservable group differences means that the problem of endogeneity was not adequately resolved.

Reyes (2007) is the first study to use quasi-experimental methods to derive a causal estimate. She uses the different grades and concentration of lead in gasoline in US states as an instrumental variable for lead levels. She finds an effect of lead on violent crime but not property crime. In a later paper (2015) she uses a similar identification strategy with individual-level data. Here she finds a positive effect on both property and violent crime. Feigenbaum and Muller (2016) also use an instrumental variable strategy. They instrument for the presence of lead water pipes in US cities using the distance to the nearest lead refinery in 1899, a period in which thousands of US cities built their water supplies. They find a positive causal effect on homicides in 1921-1936. Aizer and Currie (2018) use nearby traffic volume interacted with year of birth as an instrument for lead and include sibling fixed effects. They find a positive relationship between lead and

incarceration. Curci and Masera (2018) also find a positive association when they look across 300 US cities. Most of the estimates from this paper do not fall under the "addressing endogeneity" category, but in one chart of estimates they use soil quality as an instrument for lead. Grönqvist *et al*. (2019) use a sample of 800,000 Swedish children grouped by neighbourhoods and cohorts. They instrument for blood lead levels by the lead measured in moss in the areas. The estimates are mixed but tend to show a small positive effect on crime. Finally, Billings and Schnepel (2018) match a treatment group of children who had blood lead levels above a 10μg/dL threshold in two tests, with a control group of children who were above the threshold in the first test and just below in the second test, thus failing to qualify for treatment. This, close to randomised control trial, study finds a positive effect of lead on crime, with a stronger effect on property crime than violent crime. Overall, the few studies that use quasi-experimental methods all find a positive effect on crime, but they tend to find a smaller effect than the studies that look at correlations across geographic areas.

B. Converting to common estimates

To conduct a meta-analysis all estimates must be converted to a common metric. We use both elasticities and partial correlation coefficients (PCCs). We calculate the PCC as shown in equation (I):

(I) $$PCC_{ij} = \frac{t_{ij}}{\sqrt{t_{ij+df_{ij}}^2}}$$

Where t_{ij} is the t-ratio for estimate *i* of study *j*, and df_{ij} is the degrees of freedom. The standard error of each PCC is calculated according to equation (II):

(II) $$SE_{ij} = \frac{PCC_{ij}}{t_{ij}}$$

Some papers reported odds ratios rather than correlation coefficients. Following Polanin and Snilstveit (2016), we converted these to PCCs.

(III) $$PCC_{ij} = \frac{ln(OR_{ij})\times\left(\frac{\sqrt{3}}{\pi}\right)}{\sqrt{\left(\ln(OR_{ij})\times\left(\frac{\sqrt{3}}{\pi}\right)\right)^2 + a_{ij}}}$$

Where OR_{ij} is the odds ratio *i* for study *j* and $a_{ij} = \frac{(n_{ij1}+n_{ij2})^2}{n_{ij1}n_{ij2}}$. Here a_{ij} is a correction factor which depends on the sample size in the control and treatment groups (n_{ij1} and n_{ij2}). If the sample sizes are unknown, or there are no treatment and control groups, we follow Borenstein *et al*. (2009) and set them to be equal, which gives $a = 4$.

In a similar way we calculate standard error equivalents for odds ratio estimates. Following the Cochrane Handbook (Higgins and Green, 2011), first we convert the 95% confidence intervals to odds ratio standard errors (ORSE).

(IV) $$ORSE_{ij} = \frac{(\ln(\overline{CI}) - \ln(\underline{CI}))}{3.92}$$

Where $\overline{CI}$ is the upper confidence interval limit and $\underline{CI}$ is the lower confidence interval limit. I then convert this into partial correlation coefficient standard errors.

(V) $$SE_{ij} = \sqrt{\frac{(a^2\times ORSE_{ij}^2\times\left(\frac{3}{\pi^2}\right)}{\left(\left(\log(OR_{ij})\times\left(\frac{\sqrt{3}}{\pi}\right)\right)^2 + a\right)^3}}$$

Only one study (Billings and Schnepel, 2018) has estimates which are similar to randomised control trial estimates, with a mean difference shown between control and treatment groups. These can also be converted to PCCs. For these we follow Borenstein *et al*. (2009) and first compute the within-groups standard deviation $SD_{ij}$for estimate *i* of study *j,* as shown in (VI).

(VI) $$SD_{ij} = \sqrt{\frac{(n_{ij1}-1)\times S_{ij1}^2 + (n_{ij2}-1)\times S_{ij2}^2}{n_{ij1}+n_{ij2}-2}}$$

Here, n_{ij1} is the sample size for the control group for *i* of study *j*, S_{ij1} is the standard deviation for the control group, while n_{ij2} and S_{ij2} are the same from the treatment group.

We use this to calculate Cohen's D:

(VII) $D_{ij} = \frac{\bar{X}_{ij1} - \bar{X}_{ij2}}{SD_{ij}}$

Where $\bar{X}_{ij1}$ is the sample mean for the control group and $\bar{X}_{ij2}$ for the treatment group. Finally, we convert Cohen's D to a PCC by equation (VIII).

(VIII) $PCC_{ij} = \frac{D_{ij}}{\sqrt{D_{ij}^2 + a_{ij}}}$

Here a_{ij} is the same as that for equation (III) except we have the sample sizes for each group so we do not set it to equal 4. The variance for Cohen's D is calculated as in (IX).

(IX) $DVar_{ij} = \frac{n_{ij1} + n_{ij2}}{n_{ij1} \times n_{ij2}} + \frac{D_{ij}^2}{2(n_{ij1} + n_{ij2})}$

This is then used to calculate the standard error of the PCC.

(X) $SE_{ij} = \sqrt{\frac{a_{ij}^2 \times DVar_{ij}}{\left(D_{ij}^2 + a_{ij}\right)^3}}$

One further study only uses simple correlations (Lauritsen *et al*., 2016). The standard errors for these must be approximated. We use the approximation of one divided by n-3 for the correlation standard errors, as n is the same for all estimates, the standard errors are the same for all these estimates.

C. Common effects and random effects meta-analysis

This section explains how common and random effects meta-analysis estimates are calculated.

Before calculating fixed or random effects meta-averages, first we convert all PCCs to normalised estimates with equation (XI), so that correct confidence intervals can be calculated.

(XI) $Z_{ij} = 0.5\ ln\left(\frac{1+PCC_{ij}}{1-PCC_{ij}}\right)$

Where Z_{ij} is the normalised effect size of a PCC. The process is that first PCCs are converted to normalised estimates, we estimate using either common effects or random effects, then the estimates are converted back to a PCC with equation (XII).

(XII) $PCC = \frac{e^{2z}-1}{e^{2z}+1}$

Where in this case the PCC is the meta-analysis estimate as a correlation coefficient, and Z is the estimate obtained from the normalised PCCs.

To calculate the common effects averages we weight each estimate by the inverse of the variance, and then divide the sum of these weighted estimates by the sum of the weights as shown in following two equations:

(XIII) $W_{ij} = \frac{1}{V_{ij}}$

(XIV) $FE = \frac{\sum_{i=1}^{N} W_{ij} Z_{ij}}{\sum_{i=1}^{N} W_{ij}}$

Where V_{ij} is the variance of estimate i of study j , FE is the fixed effects average, and Z_{ij} is normalised PCC. This average is converted back into a PCC by equation (XII). Along with the averages I calculate 95% confidence intervals, first by obtaining the standard errors of FE.

(XV) $SE_{FE} = \sqrt{\frac{1}{\sum_{i=1}^{k} w_{ij}}}$

Then obtaining lower and upper limits in the normal fashion. The fixed effect averages and standard error can be used to calculate Z-scores for hypothesis testing as normal.

Random effects meta-averages are estimated in the same way as fixed effects, except we replace V_{ij} in equation (XIII) with V_{ij}^{*}. Where $V_{ij}^{*} = V_{ij} + T^2$, and T^2 is an estimate of the between-study variation. There are different methods of estimating T^2, we use the DerSimonian-Laird (1986) method.

D. Publication bias adjustment

We use seven methods to obtain an estimate of the average effect after adjusting for publication bias. This section describes those methods in more detail.

All publication bias methods either test or assume that the observed sample distribution is a truncated version of the underling population distribution. We have no details about the missing values (i.e. this is not a censored distribution). Therefore, selection models using observations (such as in Heckman, 1976) are not possible.

The publication bias methods rely on assumptions about the truncation process that generates the selection bias which causes the observed distribution to differ from the population distribution. The observed sample and the selection bias assumptions are combined in some estimation procedure, and this produces an estimate which is adjusted for the publication bias, if it is found to be present. In some cases when tests reject publication bias there is no adjustment, and the estimate collapses into either the common or random effects estimate.

Linear Methods

The first four methods are all linear regressions based on the PET-PEESE method. The PET-PEESE is itself an extension of the Egger (1997) test. The Egger test is a simple regression of the effect size on the standard error. A t-test on the standard error coefficient is a test of publication bias where ***H0*** = no publication bias, and ***H1*** = there is publication bias.

Stanley and Doucouliagos (2014) note the heteroskedasticity in the Egger test, as more precise effect sizes (assuming a shared effect size distribution and that estimates also have sampling error) will tend to be closer together. Therefore, they extend the Egger test by using weighted least squares, with the weights being the inverse of the standard errors themselves, which are an estimate of this heteroskedasticity. The coefficient on the precision (1/SE) is the Funnel Asymmetry Test (FAT). The intercept in this model becomes the Precision Effect Test (PET). The FAT is an estimate of the bias, the sign of

which indicates the direction of the bias. The PET is an estimate of the average effect size when publication bias is zero, i.e., the effect size population mean.

The coefficient on the FAT approximates the inverse Mills' ratio. However, this is not a constant, it varies with the standard error. Therefore, Stanley and Doucouliagos (2014) propose using a Taylor expansion around the standard error to better approximate the inverse Mills' ratio. In theory, any number of additional polynomials could be included in the regression, but sample size restrictions in meta-analysis, and the decreasing returns on including more polynomials, mean that few meta-analyses go beyond a cubic term. Stanley and Doucouliagos (2014) propose constraining the linear term on the standard error to be zero and using a squared term. This is the Precision Effect Estimate with Standard Error (PEESE) test. They find in simulations that this performs better than the FAT-PET when the "true" mean of the population of estimates is not equal to zero. This is the second method we use.

The third method is simply the FAT-PET but including study fixed effects. This is more efficient than the standard the FAT-PET, assuming the common effects model is not true for the population. This is estimated with restricted maximum likelihood, which adjusts the degrees of freedom downward for each study fixed effect, without which the variance of the error is biased downwards.

The fourth method we use is the FAT-PET with an instrumental variable. There are other reasons beyond publication bias why the effect size might be correlated with the standard error. For example, regression discontinuity designs (although there are none in our sample) converge at a rate at least as slow as the cubed root of the sample size. Whereas OLS converges at a rate of the root of the sample size. A regression discontinuity with the same sample size will tend to have larger standard errors than the simple OLS regression. The effect size will also be different, perhaps because they estimate different estimands, or perhaps because the bias is larger in the OLS sample. Similarly, two stage least squares will tend to have larger errors even if it is estimating the same estimand as OLS. Therefore, the coefficient on the standard error may not be a good approximation of the inverse Mills' ratio.

An alternative strategy is to use the inverse of the square root of the sample size as an instrumental variable for the standard error. The sample size is correlated with the standard error. Assuming no relationship between sample size and the effect size beyond its relationship to the standard error (the exclusion restriction), then it will give a better estimate of publication bias and therefore a better PET estimate.

Non-linear methods

The weighted average of adequately powered estimates (WAAP) developed by Stanley, Doucouliagos, and Ioannidis (2016) estimates a common effects weighted average using only high-powered studies. Studies are discarded if they do not meet some power threshold given by:

(D.1) $$\frac{\hat{\mu}_w}{2.8}$$

Where $\hat{\mu}_w$ is some estimate of the average effect, and the 2.8 denominator comes from the sum of two t-distributed test standard deviations, $t_{1-\frac{\alpha}{2}} + t_{(1-\beta)}$. Following convention, the critical value of the test of the null is set as $\alpha = 0.05$, and the power of the test is set as 80%, so that $\beta = 20\%$. This gives a sum of 1.96 + 0.84 = 2.8. Stanley, Doucouliagos, and Ioannidis (2016) suggest using the common effects estimate as the value $\hat{\mu}_w$. Given the very small common effects estimate in our sample this would only leave only one study, that of Grönqvist, Nilsson and Robling (2019). This would mean the WAAP collapses into the weighted average estimate in table 1. To be more generous to the Lead-Crime hypothesis, we instead use the larger random effects estimate as $\hat{\mu}_w$. The studies and number of estimates from each considered to be adequately powered under this method is given in table D.1.

Table D.1 – Studies and estimates used in WAAP

Study	Estimates
Aizer & Currie (2019)	6
Beckley et al. (2018)	10
Billings & Schnepel (2018)	3
Curci & Masera (2018)	97
Feigenbaum & Muller (2016)	43
Fergusson et al. (2008)	6
Grönqvist, Nilsson and Robling (2019)	54
Lersch & Hart (2014)	2
Manduca & Sampson (2019)	2
Masters et al. (1998)	3
Mielke & Zahran (2012)	1
Nevin (2000)	1
Nevin (2007)	26
Nkomo et al. (2017)	10
Reyes (2007)	65
Reyes (2015)	13
Stretesky & Lynch (2004)	20

Trim and Fill first ranks studies by the absolute value of their effect sizes, then estimates how many effect sizes are missing from either the positive or negative side of the distribution (the negative side in our case). Importantly, these studies are assumed to be not observed with probability one. This contrasts with other methods which estimate the publication probabilities over certain intervals (such as Andrews-Kasy). The trim-and-fill method then uses an iterative algorithm to obtain an average effect estimate.

1. First obtain the random effects estimate from the full sample, use this to estimate the number of missing studies (they propose three different estimators for this).
2. Using the estimate for number of missing studies on the negative side, an equal number of studies are "trimmed" from the sample on the positive side, starting with the largest and moving down.

3. Now obtain another random effects estimate from the trimmed sample and use this to again estimate a number of missing studies.
4. Continue until the random effects estimate of iteration *j* is equal to the estimate of iteration *j* – 1.
5. Now add the "fill", where imputed values are added to the negative side of the distribution, using the estimates obtained in the last iteration and the most positive values in the sample left after "trimming" (see section 5 in their paper).
6. Finally, obtain a new random effects estimate using the full initial sample, plus the imputed "filled" values.

This method adds 226 estimates to the full sample trim and fill, 82 to the elasticity sample, and 11 to the representative estimates sample.

In the Andrews and Kasy (2019) method, they use a step function to estimate the probability of observing an effect over various intervals of the distribution. This contrasts with the trim and fill, where some observations are assumed missing with probability one, and the FAT-PET, which uses an approximation of the inverse Mills' ratio to deal with the truncation.

They observe, however, that the publication probabilities can only be identified up to scale. That is, we cannot know that absolute probability of publication over any one interval. Therefore, we must estimate relative publication probabilities. We do this by setting one publication probability as the reference probability, and then identifying the others up to scale, i.e., relative to this one. In our case the reference probability is the probability of observing a positive effect size that is significant at the 5% level. This probability is set at some arbitrary value (one in our case) and the other probabilities estimated relative to this. If the estimated probabilities are less than one, then they are less likely to be observed than positive values significant at the 5% level, and vice versa.

With relative probabilities estimated, the distribution is reweighted using the relative probabilities to reconstruct the true untruncated distribution. We can use this to get an estimate of the population mean, adjusting for the publication bias. We use the maximum likelihood approach and algorithm in Hedges (1992) as recommended by

Andrews and Kasy (2019) to do this. In the case of only using representative estimates, we did not achieve convergence.

The estimates publication probabilities over different z-score intervals are shown below for the full sample and the elasticity sample.

Figure D.1 – Estimated relative publication probabilities, partial correlations

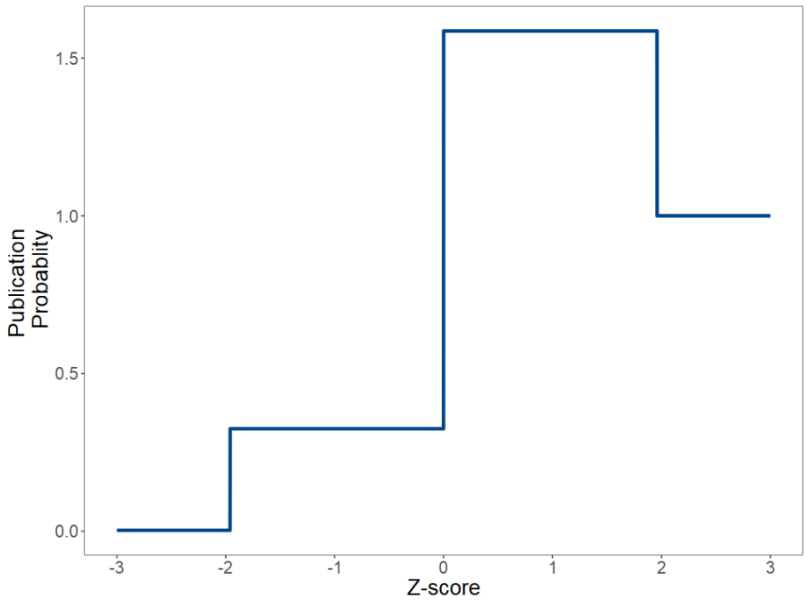
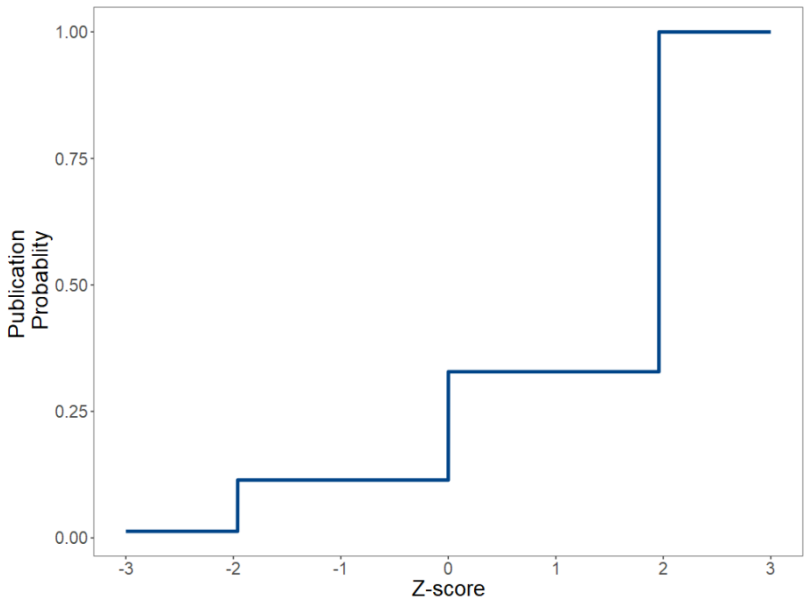


Figure D.2 – Estimated relative publication probabilities, elasticities



E. Analysis using only representative estimates

In most of our analysis we use all estimates. As a robustness check, here we use only one representative estimate from each paper. There was not always a clear representative estimate from each study. Therefore, choosing the estimates involves some subjective judgement. We tried to choose results mentioned in the abstract or as the main result. In general, we chose representative estimates which were less specific (i.e., totals preferred to subsample male/female, white/black results etc.), and estimates obtained using more covariates for correlational results.

In section 4.3 we test for publication bias using all estimates. In table E.1 we repeat the exercise using only the representative estimates. However, we cannot estimate the hierarchical model, or cluster errors as we only have one estimate per study. Furthermore the Andrews-Kasy method, using maximum likelihood, did not converge.

Table E.1 – Effect beyond bias and publication bias estimates using representative estimates, partial correlations

Panel A – Linear Models

	FAT-PET	**FAT-PEESE**	**IV**
Effect Beyond Bias	-0.001	0.007	-0.001
	(0.002)	(0.004)	(0.002)
Publication bias	3.717	12.152	3.733
	(0.894)	(6.998)	(0.880)
Observations	24	24	24

Panel B – Non-Linear Models

	WAAP	**Trim and Fill**
Effect Beyond Bias	0.007	0.015
	(0.004)	(0.059)
Observations	24	24

Notes. Estimates are PCCs 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 FAT-PET is a mixed effects-multi-level model with a different slope coefficient 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. Trim and fill (Duval & Tweedie, 2000), removes outlier studies and then adds imputed studies before calculation an average effect. The Andrews-Kasy (Andrews & Kasy,2019) method is a step function selection model which reweights the observed sample with estimated publication probabilities. See Appendix D for full explanation of each method.

F. Bayesian Model Averaging (BMA)

We carry out two forms of Bayesian model averaging: 1) we obtain an ensemble estimate of the effect beyond bias, using both linear and non-linear publication bias correction models, 2) we take model averages over all covariates used in the meta-regressions.

Table F.1 presents Bayesian model averages of publication bias correction models. We use the RoBMA R package of Bartoš *et al*. (2021).

Table F.1 – Effect beyond bias, Bayesian model averages

	Full Sample, PCCS	**Representative Estimates, PCCs**	**Elasticities**
Effect Beyond Bias	-0.17	-0.09	0.09
Observations	542	24	312

We also carry out Bayesian model averaging with all variables used in our meta-regression analysis. We estimate a normal-gamma conjugate model with a uniform model prior and unit information g-prior. These are the same as in Bajzik *et al*. (2019), see there for more information. The results are given below in table F.2.

Table F.2 – Posterior results from Bayesian model averaging, PCC

Variable	Posterior Mean	Posterior Standard Deviation	Posterior Inclusion Probability
Precision	0.35	0.03	1.00
Control gender	0.04	0.03	0.71
Control race	0.00	0.01	0.09
Control income	-0.04	0.03	0.70
Control education	0.00	0.00	0.05
Homicide	-0.03	0.03	0.54
Violent	0.00	0.02	0.14
Non_Violent	-0.01	0.03	0.34
Area	0.24	0.03	1.00
OLS	0.03	0.03	0.55
ML	0.04	0.03	0.81
Odds_Ratio	-0.04	0.07	0.35
Panel dummy	-0.17	0.02	1.00
Addressing Endogeneity	0.00	0.00	0.05
North_America	-0.41	0.03	1.00
Europe	0.00	0.02	0.07
Direct Lead Measure	-0.39	0.04	1.00
Publication Year	0.00	0.01	0.08
Covariates	-0.07	0.01	1.00
Sample size	0.00	0.00	0.09
FAT	3.40	NA	1.00
Observations	542		

We evaluate the posterior means at the sample averages for each variable (excluding the FAT as normal). This gives a point estimate PCC of 0.09.

We do the same for the elasticity sample in table F.3.

Table F.3 – Posterior results from Bayesian model averaging, elasticity

Variable	Posterior Mean	Posterior Standard Deviation	Posterior Inclusion Probability
Precision	0.13	0.01	1.00
Control gender	-0.01	0.03	0.01
Control race	0.00	0.00	0.98
Control income	0.00	0.00	0.13
Control education	0.00	0.03	0.10
Homicide	0.00	0.00	0.75
Violent	0.00	0.01	0.88
Non_Violent	-0.06	0.02	0.00
Area	0.00	0.01	0.65
OLS	0.00	0.00	0.16
ML	0.00	0.01	1.00
Panel dummy	-0.01	0.01	0.00
Addressing Endogeneity	0.00	0.00	0.70
North_America	0.00	0.01	0.41
Direct Lead Measure	0.00	0.02	0.01
Publication Year	0.00	0.00	0.01
Covariates	0.00	0.00	0.29
Sample size	0.00	0.00	1.00
FAT	1.95	NA	NA
Observations	312		

Again, we evaluate the posterior means at the sample averages for each variable (excluding the FAT as normal). This gives a point estimate elasticity of 0.12.

G. Alternative elasticity estimates

Our full sample includes studies that we could not obtain elasticity estimates from. However, it is a larger and possibly more representative sample of the literature. In this section we therefore convert the PCC estimates from the full sample into plausible

elasticities. The PCC and the elasticity are related, but not in a straightforward manner. This forces us to make some strong assumptions in the interests of welfare analysis.

Given a PCC and the change in a given measure of crime for a given measure of lead, $\frac{\delta Crime}{\delta Lead}$, then the relationship between the two is given in (7).

$$(8) \qquad PCC = \frac{\delta Crime}{\delta Lead}\frac{sd(Lead)}{sd(Crime)}\frac{sd(\widetilde{Lead} - \tilde{\mathbf{z}}'\gamma_1)}{sd(\widetilde{Crime} - \tilde{\mathbf{z}}'\gamma_2)}$$

Where $sd(.)$ means the standard deviation. $\widetilde{Lead} - \tilde{\mathbf{z}}\gamma_1$ are the residuals from a regression of *Lead* on $\boldsymbol{z}$, a vector of variables related to lead and crime, where both lead and $\boldsymbol{z}$ have been standardised. Similarly, $\widetilde{Crime} - \tilde{\mathbf{z}}\gamma_2$ are the residuals from a regression of *Crime* on $\boldsymbol{z}$**,** where both have been standardised. If we wish to attach a causal interpretation to the elasticity, we can think of $\boldsymbol{z}$, following Peters, Bühlmann, and Meinshausen (2016), as the minimum set of variables under which the distribution of *Crime* is invariant when conditioned on both $\boldsymbol{z}$ and *Lead*.

It can be seen that a PCC will always share the same sign as $\frac{\delta Crime}{\delta Lead}$ but will be inflated or deflated according to the relative size of the standard deviations in (7). $\frac{\delta Crime}{\delta Lead}\frac{sd(Lead)}{sd(Crime)}$ is equivalent to a standardised coefficient. The intuition for the last ratio is as follows: the greater the variation in *Lead* that is not explained by $\boldsymbol{z}$, the larger the PCC, because the overlapping variation between the independent effect of *Lead* and *Crime* is relatively greater. The PCC is also greater the larger the amount of variation in *Crime* explained by $\boldsymbol{z}$**.** This is because the share of unexplained variation in *Crime* becomes smaller, so the share of variation jointly explained by *Lead* and $\boldsymbol{z}$ increases. As more of the variation in *Crime* is explained by both *Lead* and $\boldsymbol{z}$**,** their PCCs will tend to 1 or -1.

To evaluate an elasticity at the sample means we multiply both sides by $\frac{\overline{Lead}}{\overline{Crime}}$, where the bar indicates the mean. We can then rearrange (7) to put it in terms of the elasticity η.

$$(9) \qquad \eta = \frac{\overline{Lead}}{\overline{Crime}}\frac{sd(Crime)}{sd(Lead)}\frac{sd(\widetilde{Crime} - \tilde{\mathbf{z}}'\gamma_2)}{sd(\widetilde{Lead} - \tilde{\mathbf{z}}'\gamma_1)}PCC$$

We can see that the size of the PCC relative to the elasticity depends on three ratios. The first two, the relative means and standard deviations, depend on the measures of crime and lead. We use homicide and blood lead data from the US as an illustrative example to examine plausible elasticities, given the fall in both violent and non-violent crime was

particularly pronounced there. The means, standard deviations, and sources are given in table IX. Given these, the relative size of the PCC to the elasticity depends upon the third ratio of residual standard deviations. This ratio could theoretically take any value between zero and infinity, and therefore so could the elasticity (assuming the PCC is positive). We therefore look at what are plausible values for this ratio and what is the range of the elasticity given these values.

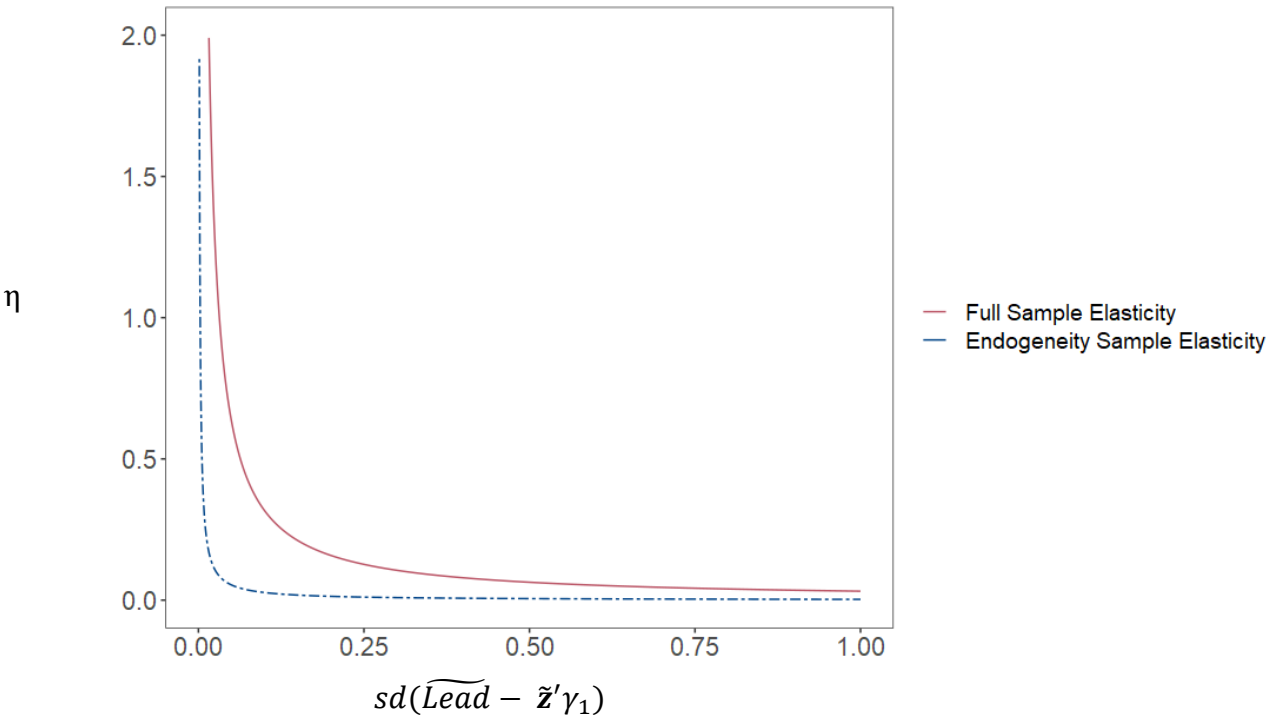
The maximum value the numerator $sd\left(\widetilde{Crime} - \tilde{\boldsymbol{z}}'\gamma_2\right)$ can take is one, representing no common variation between $\boldsymbol{z}$ and *Crime*. We hold it at one, to inflate the PCC as much as possible. The final element of the equation is $sd(\widetilde{Lead} - \tilde{\boldsymbol{z}}'\gamma_1)$. This is the residual variation in *Lead* not explained by $\boldsymbol{z}$**.** The lower this is, the more the PCC will be inflated, and therefore the greater the elasticity. The elasticity is convex in $sd(\widetilde{Lead} - \tilde{\boldsymbol{z}}'\gamma_1)$, decreasing at a decreasing rate.

Figure F.1 plots the relationship between the elasticity and $sd(\widetilde{Lead} - \tilde{\boldsymbol{z}}'\gamma_1)$, given the estimated mean PCCs, the values in table IX, and holding $sd\left(\widetilde{Crime} - \tilde{\boldsymbol{z}}'\gamma_2\right)$ constant at the maximum value of one. The elasticities drop sharply with an increase in the denominator $sd\left(\widetilde{Crime} - \tilde{\boldsymbol{z}}'\gamma_2\right)$, with the elasticity for the addressing endogeneity sample approaching close to zero almost immediately. The elasticity for the full sample slopes down more gently but even so does not suggest a large elasticity except at extremely small values of $sd\left(\widetilde{Crime} - \tilde{\boldsymbol{z}}'\gamma_2\right)$.

We can now propose a range of plausible values for the elasticity. Given the uncertainties around the ratio of unexplained variations in (9), this is somewhat arbitrary, but we hope, given the discussion above, not unreasonably so. There is no compelling reason to suppose $\boldsymbol{z}$ would explain more of the variation in *Lead* than in *Crime*. Nevertheless, if we take as a lower bound that $sd(\widetilde{Lead} - \tilde{\boldsymbol{z}}'\gamma_1)$ is ten times as large as $sd\left(\widetilde{Crime} - \tilde{\boldsymbol{z}}'\gamma_2\right)$, and as a conservative upper bound that they are equal, then we can give a range of values based on our estimated PCCs. For the full sample PCC, this gives an elasticity of 0.32-0.03. For the addressing endogeneity sample PCC, the range is 0.03-0.00, to two decimal places. The median blood lead level in children fell 88% from 1976-2009. The full sample elasticity estimates therefore would suggest the fall in lead has decreased homicide in the US by between 28% and 3%. The equivalent decrease for the addressing endogeneity sample is between 3% and 0%. The US homicide rate fell 54% from its peak in 1989 to

2014. This would mean that lead accounts for between 52% and 6% of the decrease in homicide using the full sample elasticity, and 5%-0% using the addressing endogeneity elasticity. Our generous assumptions of the lower bound on the ratio of residual variation in (8) imply that lead may be the most important factor in the fall in homicide. Our upper bound on that same ratio implies lead accounts for very little of the fall in crime.

Figure G.1 – Estimated Elasticity of on lead on crime



Notes. Chart shows how η, the calculated elasticity of lead on crime, varies with changes in $sd(\widetilde{Lead} - \tilde{\mathbf{z}}'\gamma_1)$, the standard deviation of the residual in a regression of a set of standardised variables $\tilde{\mathbf{z}}$, and the standardised measure of lead $\widetilde{Lead}$.

Table G.1 – Descriptive statistics of data used for elasticity estimation

Variable	**Mean**	**Standard Deviation**
Median blood lead level for children ages 1-5 in US	3.39	4.42
US Homicide rate	6.98	1.81

Sources. NHANES data for blood lead and FBI uniform crime reports for the homicide data.

H. Appendix References

Polanin and Snilstveit, 2016. Converting between effect sizes: Campbell policy note 3. https://doi.org/10.4073/cmpn.2016.3

Bajzik, J., Havranek, T., Irsova, Z., Schwarz, J., 2019. Estimating the Armington Elasticity: The Importance of Data Choice and Publication Bias.

Borenstein, M., Hedges, L. V., Higgins, J.P.T., Rothstein, H.R., 2009. Introduction to Meta-Analysis. John Wiley & Sons, Ltd, Chichester, UK. https://doi.org/10.1002/9780470743386

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

Higgins JPT, G.S. (editors), 2011. Cochrane Handbook for Systematic Reviews of Interventions, Version 5. ed. The Cochrane Collaboration.

Peters, J., Bühlmann, P., Meinshausen, N., 2016. Causal inference by using invariant prediction: identification and confidence intervals. J. R. Stat. Soc. Ser. B (Statistical Methodol. 78, 947–1012. https://doi.org/10.1111/rssb.12167