Carnegie Mellon University
Research Showcase @ CMU

Department of Philosophy | Dietrich College of Humanities and Social Sciences

2003

Expert Statistical Testimony and Epidemiological Evidence: the Toxic Effects of Lead Exposure on Children

Stephen E. Fienberg
Carnegie Mellon University

Clark Glymour
Carnegie Mellon University

Richard Scheines
Carnegie Mellon University

Follow this and additional works at: http://repository.cmu.edu/philosophy

Part of the Philosophy Commons

Recommended Citation
Fienberg, Stephen E.; Glymour, Clark; and Scheines, Richard, "Expert Statistical Testimony and Epidemiological Evidence: the Toxic Effects of Lead Exposure on Children" (2003). *Department of Philosophy.* Paper 287.
http://repository.cmu.edu/philosophy/287

This Article is brought to you for free and open access by the Dietrich College of Humanities and Social Sciences at Research Showcase @ CMU. It has been accepted for inclusion in Department of Philosophy by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

ELSEVIER

Journal of Econometrics 113 (2003) 33–48

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Expert statistical testimony and epidemiological evidence: the toxic effects of lead exposure on children☆

Stephen E. Fienberg[a,*], Clark Glymour[b], Richard Scheines[b]

[a] *Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA*
[b] *Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA*

Accepted 16 July 2002

Abstract

The past two decades have seen a dramatic growth in the use of statisticians and economists for the presentation of expert testimony in legal proceedings. In this paper, we describe a hypothetical case modeled on real ones and involving statistical testimony regarding the causal effect of lead on lowering the IQs of children who ingest lead paint chips. The data we use come from a well-known pioneering study on the topic and the analyses we describe as the expert testimony are similar to ones that can be found in, major scientific journals. The battle of the experts in this hypothetical case resembles that which many encounter as expert witnesses. The paper concludes with some observations and advice.
© 2002 Published by Elsevier Science B.V.

JEL classification: K40; C1

Keywords: Bayesian methods; Causal modeling; Measurement error; Multiple regression; Omitted-variable bias

1. Introduction

Because statistics is used in so many different fields, statisticians appear as expert witnesses in diverse litigation contexts, ranging from the reliability of forensic evidence, DNA fingerprinting, mass torts, and discrimination cases, to confidentiality of

☆ An earlier version of the material in this paper was developed for use as a case study for the Federal Judicial Center and presented at the Fourth International Conference on Forensic Statistics, Raleigh, NC, December, 1999.

* Corresponding author.
E-mail address: fienberg@owen.vanderbilt.edu (S.E. Fienberg).

0304-4076/03/$ - see front matter © 2002 Published by Elsevier Science B.V.
PII: S0304-4076(02)00165-3

scientific research data, cheating on examinations, drug smuggling, voting fraud, and the adjustment of the US decennial census. For a wide variety of examples of the role of this testimony in specific cases see DeGroot et al. (1986), Evett and Weir (1998), Fienberg (1989, 1997a), Finkelstein and Levin (2001), Gastwirth (1988, 2000), and Zeisel and Kaye (1997).

In many legal contexts the central issue is causal. In some, the causes of a single event are under dispute. Did the leading driver's sudden stop to avoid hitting a dog cause the accident, or was it the failure of the anti-lock braking system in the trailing driver's car? In others, a causal generalization is at issue. Does chronic low-level exposure to asbestos cause lung cancer? Does the level of radiation received by workers at a nuclear power plant cause leukemia? If so, how strong is the effect, i.e., what is the dose–response relationship?

In most cases where a causal generalization lies at the core of a legal action, for practical or ethical reasons, no experimental evidence is available. On occasions where experimental evidence is available, it usually involves laboratory animals and the results are very difficult to extrapolate to humans. As a result, causal generalizations are typically supported (or refuted) by non-experimental statistical evidence. In this article, we lay out a prototypic case in which statistical and epidemiological evidence is brought to bear by opposing experts on a causal claim: does chronic exposure to low levels of lead cause cognitive deficits in children? Using actual data sets, and a fictitious exchange between a plaintiffs' and a defendant's expert, based in large part on published exchanges in the statistical literature, we illustrate several classic argumentative strategies used by econometric, epidemiological and statistical expert witnesses to establish or refute causal generalizations from non-experimental statistical evidence.

After briefly describing the case background, we present the core of the plaintiffs' expert testimony. The expert begins by showing a "raw" association between the putative cause and the effect, articulates possible confounders (covariates or exogenous variables in statistical or econometric language), measures them, and then statistically selects the ones that are worthy of further study. The expert then uses multiple regression to show that the "raw" association between cause and effect does not disappear after controlling for such confounders. The same strategy is used not only in tort cases but also in litigation involving discrimination as well as other areas of the law.

The defendant's expert responds by challenging the validity of the initial presentation of evidence. The defendant's expert argues that the plaintiffs' expert has omitted important confounders from consideration, and has neglected to take into account measurement error, especially in the confounders. The defendant's expert then presents a sensitivity analysis showing that there is indeed no effect of low levels of lead on IQ for "reasonable" ranges of measurement error.

On rebuttal, the plaintiffs' expert shows that the way in which the defendant's expert has included the supposedly omitted variables is wholly unreliable and conveniently misleading. He demonstrates that measurement error in the confounders can be taken into account, even modeled statistically, and that doing so only strengthens the conclusions he drew.

Our presentation attempts to illustrate the technical aspects of the "battle of experts" that an economist or statistician can expect to experience in a legal context (for some

related discussions on strategic and other non-technical aspects see Fienberg, 1997b; Fisher, 1986; Meier, 1986). It also illustrates the difficulties of making causal inferences from non-experimental data, how such evidence can be legitimately challenged and where it cannot.

In comments on an early draft of this paper, reviewers expressed some surprise at the "unsophisticated" and simplistic nature of some of the "evidence" presented by the experts in this fictitious case. Our response is that this is precisely how evidence is presented in real cases. The expert's job is to educate the trier of the fact, either a judge or a jury, and to do so by building up from easily understood statistical ideas to more complex ones (Fisher, 1986; Fienberg, 1989; Fienberg et al., 1995). Thus, while econometricans, if left to their own devices, might begin with simultaneous equations models involving carefully selected instrumental variables and perhaps a fully Bayesian analysis replete with Markov chain Monte Carlo methods, most expert testimony at trial begins instead with correlation coefficients, 2×2 contingency tables, and perhaps Fisher's exact test (see Good, 2001). The more complete and intellectually justifiable analyses typically come later in the case, if at all.

2. The hypothetical case: lead and IQ

The fictional case that we describe here draws on actual epidemiological data concerning the effects of exposure to lead on children's IQ scores. Empirical studies in this area have had an important influence on national policy and have been the focus of a series of scientific disputes. The case arises under a tort claim raised by the residents of a tenement dwelling and brought against the landlord. The plaintiffs claim that the landlord has failed to meet his responsibility to maintain a safe and healthy dwelling and protect his tenants from undue exposure to lead. We have based the "facts" of the case on a somewhat disguised tort case described by Frank (1992) as "The Case of Sam Brown." The focus of Frank's analysis is the economic calculation of damages on the basis of (1) medical costs of roughly \$40,000, (2) losses in lifetime earnings—estimated to be \$189,000, and intangible factors—\$70,000. These calculations involved the use of statistics through regression analysis, etc. We address none of these issues here.

Rather we cast the individual case of Sam Brown in the context of a hypothetical lawsuit brought by multiple plaintiffs, for which the evidence in the case of Sam Brown is illustrative. Our focus is on the testimony by statistical experts intended to establish or to contradict proof as to the effect of low-level lead exposure on cognitive function. Only if convincing proof of such a causal effect is presented to the court can the issue move to the assessment of damages.

2.1. Basic facts

"Samuel Brown was hospitalized at age 7 for severe lead intoxication (blood level of over 70 mg/dl). Samuel was hospitalized in a tertiary care hospital, where chelation therapy was administered. The duration of the hospital stay was 4 weeks. A psychological

evaluation was performed subsequent to hospital discharge and it concluded that Samuel had suffered an 18 point IQ loss, from 109 Weschler full scale IQ to 91" (see Frank, 1992, pp. 262–263). The other plaintiffs are children who lived in the same building and were exposed to low-level quantities of lead through lead paint that had once been used throughout the building. Information on health and behavioral problems that are seemingly linkable to substantial lead exposure is also available and not in dispute; however, there is a dispute regarding the implications of *low-level* lead exposure. All of the children have current IQs below 100 and evidence available to indicate that, at least in several instances their IQs are lower than when measured several years earlier.

In support of their claims, the plaintiffs present a case based on testimony of a statistical expert with background in epidemiology. In actual cases, it is more likely that there would be multiple experts, e.g., an epidemiologist who might have conducted a study of the effects of low level lead exposure, as well as the statistician who analyzed the data, and perhaps others to testify as to the extent to which the evidence is consistent with other studies. In the case of the effects of lead on IQ, there is an extensive epidemiological and statistical literature on which the experts could opine, e.g., see Needleman and Gatsonis (1990), Pocock et al. (1994), and Waternaux et al. (1998).

2.2. The Massachusetts study

In the hypothetical case we take as the principal data used for statistical analyses to support the plaintiffs' claim come from an observational study of 270 children in Massachusetts (Needleman et al., 1979). Each child in the study had just lost baby teeth which were measured for concentrations of lead. Unlike blood levels of lead, which can fluctuate widely over a 24 h period and thus are poor measures of cumulative exposure, the deposit of lead in baby teeth is a slow and stable response to the lead a child has been exposed to over the course of his or her entire childhood.

Each child's IQ was also measured, along with a variety of physical and social factors, including weight, height, head circumference, sex, etc. Several measures of parental status were also collected, e.g., time from conception to delivery of the child, number of previous live deliveries by the child's mother, the mother's IQ, socioeconomic status and education.[1] Values for some of the variables are missing for some of the children. The data are *non-experimental*, that is, the exposure to which each child was subjected was not under the control of scientific investigators.

[1] We are indebted to Dr. Herbert Needleman of the University of Pittsburgh for providing us with his original data, which we have used in these illustrative analyses. Because of different choices of variables and our use of different statistical analysis packages, the results we obtained from various statistical procedures do not strictly agree with those he has published, and are not intended to do so. See Needleman et al. (1979, 1985).

3. Plaintiffs' statistical case

3.1. The raw association between lead exposure and IQ

By various standard statistical measures, for the children in the study sample, the concentration of lead in teeth is negatively associated with IQ. One measure of dependency is obtained by introducing a cutoff between high and low lead concentrations and a cutoff between high and low IQ. In the absence of a good reason for another value, a natural threshold is the median or middle value. The plaintiffs' statistical expert begins by following the epidemiological approach and presents the numbers of children who are high in both features (lead concentration and IQ), high in one but not the other, or low in the form of a 2×2 contingency table (Table 1).

Note that there are a total of 229 observations in this table, not 270, the number of children in the study. That is because the table is based on all complete cases with no missing observations for a subset of the variables including IQ and lead concentration. Discarding observations with missing data is not usually considered a good statistical practice (e.g., see Little and Rubin, 1987), but it is often done on practical grounds and in the present instance this simplification does not change the inferences reached by the expert (cf., Kadane and Terrin, 1997).

The dependency between binary-valued variables can be measured by the cross-product or odds ratio, which in this case equals 0.58, indicating a negative association between high levels of lead and high IQ, and is significant at a p-value $= 0.042$. The choice of different cutoff or threshold values in this particular data set will produce different cross-product ratio values. But for almost any choice of cutoff the sample data yields a significant cross product ratio. The *sample correlation coefficient* between lead and IQ is -0.253, which is significant at $p < 0.01$. The expert explains that these analyses show that, even at low levels, lead exposure and IQ are negatively associated to a degree that cannot be explained by chance.

3.2. Confounders

The association between lead concentrations in children's teeth and their IQs is clearly not "logical"—one variable is not an aspect of the meaning of the other. There is nothing in the measurement process, or in the way children were selected for the sample, this suggests that the association is an artifact of measurement or sample selection.

Table 1
Contingency table for IQ and lead

	High lead	Low lead	Total
High IQ	51	66	117
Low IQ	64	48	112
Total	115	114	229

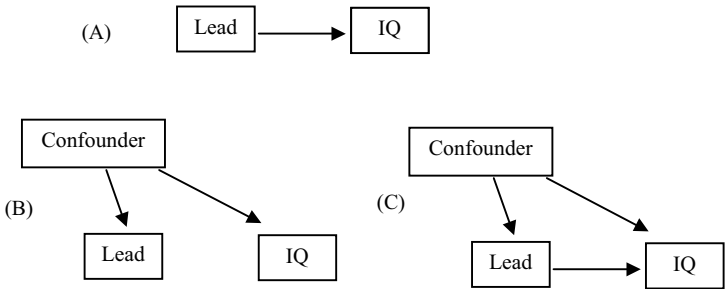

Fig. 1. Alternative explanations of an association between lead and IQ.

Variables that influence both a possible cause and its supposed effect are said in the epidemiological literature to be confounders. In statistics they are often referred to as covariates and in econometrics as exogenous variables. The plausible alternatives therefore are these:

1. The association found in the sample is due to the influence of lead on IQ (Fig. 1a).
2. The association found in the sample is due to the influence of other variables (confounders) on both lead exposure and IQ (Fig. 1b).
3. The association found in the sample is due to the influence of other variables (confounders) on both lead exposure and IQ, and also to the influence of lead on IQ (Fig. 1c).

If either alternatives 2 or 3 hold, then the raw association between lead and IQ is not a reliable guide to the strength of the causal effect of lead on IQ. The expert explains to the court that there are two basic strategies for dealing with this problem: randomly assign treatment (in this case lead exposure), or measure the confounders and statistically control for them. By randomly assigning treatment, we eliminate the influence of any possible confounder on the putative cause, and thus create a situation in which the resulting raw association is the appropriate measure of the causal effect. In the present instance, and many others, however, randomly assigning the level of lead exposure would have been neither ethical nor practical.

The expert explains that this leaves us with the second strategy to pursue.

3.3. Confounder selection

The expert explains that it is well understood in the statistical literature that confounders that can substantially bias an estimate of a causal effect, if they are not measured and statistically controlled for, are those variables that are highly associated with the putative cause and effect. Thus the Massachusetts study research team focused on measuring variables that are plausibly common causes of the putative cause (lead) and effect (IQ), and that are highly associated with both—they identified and measured over 40 potential confounders. At this point, the expert presents to the court a brief synopsis of multiple regression and controlling for confounders (which the readers of this journal need not receive).

Table 2
Regression results of lead regression

Predictor	Coef	Stdev	T	p-Value
Constant	−33.10	45.54	−0.73	0.468
lead	−0.15169	0.07879	−1.93	0.056
ht	0.2937	0.4872	0.60	0.547
wt	−0.0348	0.1016	−0.34	0.733
headc	1.6797	0.6826	2.46	0.015
sex	−1.605	1.941	−0.83	0.409
birwt	−0.05342	0.05522	−0.97	0.335
pregl	0.1838	0.5184	0.35	0.723
med	−0.1166	0.8101	−0.14	0.886
faed	0.9190	0.4906	1.87	0.063
sesmo	0.488	1.844	0.26	0.792
sesda	−1.015	1.213	−0.84	0.404
numpr	−2.031	1.031	−1.97	0.051
nlb	1.158	1.220	0.95	0.344
schol	−0.016330	0.009905	−1.65	0.101
teacg	0.00584	0.01304	0.45	0.655
piq	0.35529	0.07831	4.54	0.000
mab	0.5287	0.3616	1.46	0.145
fab	−0.4898	0.2512	−1.95	0.053
$S = 12.32$	R-Sq=35.4%		R-Sq(adj)=28.6%	

Simply measuring every possible confounder one can think of and sticking it into a multiple regression is by no means optimal, however. For one, if two variables are redundant measures of the same basic confounder, and are themselves highly correlated, then the statistical estimation can become unstable and hence unreliable. For another, one pays a price in statistical power with every extra variable included in a regression—thus the apparent relationship between cause and effect might come out numerically the same in two regressions, but it might appear significant relative to one set of confounders and insignificant relative to another. (This is a lesson often lost in the presentation of regression results in some economics journals.)

Having explained the statistical methodology, with data from the lead and IQ study, the expert then presents the result of a multiple linear regression of IQ on 18 of the 40 confounder variables (Table 2).

The coefficient for lead in this regression equation is −0.15169, indicating that IQ falls approximately 0.15 points for every extra unit change in lead concentration found in baby teeth, even controlling for all the other confounders. All of the regressors together account for 35.4% of the variance of IQ in the sample. The coefficient for lead is significant at the 0.06 level but not at the 0.05 level ($p = 0.056$).

Several of the confounders in this regression are far from being significant predictors of IQ, however, and thus cost us in the power for detecting whether the estimated effect of lead on IQ is truly significant. Further, several are themselves highly correlated, making the regression results unstable. A more reasonable strategy is to conceive of and

Table 3
Correlations and p-values ($n = 221$)

	lead	fab	nlb	med	mab	piq	ciq
lead	1.00						
fab	0.08	1.00					
	(0.23)						
nlb	0.11	0.39	1.00				
	(0.10)	(0.00)					
med	−0.14	0.02	−0.18	1.00			
	(0.04)	(0.78)	(0.01)				
mab	−0.15	0.85	0.47	0.003	1.00		
	(0.02)	(0.00)	(0.00)	(0.96)			
piq	−0.06	0.17	0.03	0.53	0.16	1.00	
	(0.39)	(0.01)	(0.70)	(0.00)	(0.02)		
IQ	−0.23	−0.0003	−0.17	0.41	0.05	0.40	1.00
	(0.00)	(0.99)	(0.01)	(0.00)	(0.43)	(0.00)	

Table 4

Predictor	Coef	SE Coef	T	p
Constant	56.091	7.635	7.35	0.000
lead	−0.16267	0.07028	−2.31	0.022
med	1.3621	0.5158	2.64	0.009
nlb	−1.3977	0.6277	−2.23	0.027
piq	0.27862	0.06636	4.20	0.000
mab	0.5948	0.2957	2.01	0.045
fab	−0.3629	0.2116	−1.71	0.088
$S = 12.66$	R-Sq=26.0%		R-Sq(adj)=24.0%	

measure as many confounders as is practical, but then to select those that are important with one of the many variable selection procedures available, and then estimate the effect using the set of variables selected.

Using backwards-stepwise regression, Needleman et al. (1985) found that the following confounders should be controlled for when estimating the effects of lead on IQ: mab—mother's age at birth, fab—father's age at birth, med—mother's level of education in years, nlb—number of live births previous to the sampled child, piq—parent's IQ scores.

The correlations among these variables and lead and IQ, as well as the p-values of these correlations (in parentheses) are given in Table 3.

Performing the multiple regression using lead and only these confounders, we get the results given in Table 4.

In this case, the coefficient for the effect of lead on IQ is quite similar, −0.163, but now statistically significant ($p = 0.022$).

Variations on the set of confounders controlled for in fact make little difference in the estimate of the effect of lead on IQ, an important piece of evidence in this case.

Using TETRAD's PC algorithm to select the confounders (see Spirtes et al., 2000, or Scheines et al., 1994) the expert testifies that he included only a subset of the five found by Needleman et al. (1985): med—mother's level of education in years, piq—parent's IQ scores.

The resulting regression, given in standardized form, with t-statistics in parentheses, is as follows:

$$IQ = -0.177 \text{ lead} + 0.251 \text{ med} + 0.253 \text{ piq}$$
$$(2.89) \qquad (3.5) \qquad (3.59).$$

All coefficients are significant ($p < 0.05$). The expert summarizes his testimony by reminding the court that, in almost every case described by these regression results, a child can expect to lose approximately 1 and $\frac{1}{2}$ IQ points for every extra unit of lead exposure. This, he claims, is strong evidence that lead has a deleterious effect on the cognitive capacities of children even at low levels, and those who expose children to it should be held responsible.

4. The defendant's expert response

The defendant's expert begins by explaining that the scientific case presented by the plaintiffs' experts, which is thin statistically even assuming their models are correct, rests on two assumptions that are clearly false. First, although they have considered several possible confounders, they have omitted a few that have been shown in previous studies to be crucial. Second, although they are using multiple regression to control for confounders, they neglect to model the measurement error for the confounders. When one corrects for these assumptions, first by imputing the missing confounder and second by doing a sensitivity analysis on the levels of potential measurement error, the expert claims that the effect of low-level lead on IQ vanishes (cf., Marais and Wecker, 1998). Here are some of the details.

4.1. Omitted variables

Although the plaintiffs' experts measured several properties of the mother and the child, they omitted the father's IQ, a crucial confounder that will surely bias the results if not controlled for statistically. As a consequence of his review the literature on lead exposure and IQ, the expert claims that it is clear that the following factors are correlated with lead exposure and also play a significant role in determining IQ: mother's IQ, father's IQ, and quality of the home environment.

The data used by the plaintiffs' experts do not include the father's IQ, and uses the mother's educational level as a proxy for the Home environment, which is dubious at best. By omitting the father's IQ, the plaintiffs' experts have performed an analysis in which the effect of lead on IQ will be estimated with *omitted variable bias*. Since the correlation between father's IQ and lead is negative, and since the correlation between father's IQ and the child's IQ is positive, omitting father's IQ will most likely result in a negative bias. That is, the estimate of the effect of lead on IQ in a model that

Table 5
Partially completed correlation matrix

	iq	momiq	dadiq	home	lead
iq	1				
momiq	0.477	1			
dadiq			1		
home				1	
lead	−0.253	−0.124			1

Table 6
Completed correlation matrix

	iq	momiq	dadiq	home	lead
iq	1				
momiq	0.477	1			
dadiq	0.5110	0.2550	1		
home	0.4500	0.2300	0.1800	1	
lead	−0.253	−0.124	−0.1370	−0.4700	1

omits father's IQ will be more negative than it should. The same is true for the quality of the home environment (home). In the plaintiffs' expert analysis, it might well be that the bias from omitting the father's IQ and home is big enough to account for the entire estimated negative effect.

In order to explore these issues, the expert has used other studies in the literature to impute, or fill in, the correlations between father's IQ, home, and the variables used by the plaintiffs in order to statistically examine the nature of the plaintiffs' omitted variable bias. He began with the correlation matrix shown in Table 5, where the only values we have are from the study used by the plaintiffs.

For the remaining values, he used the mean value of other published studies in which these variables are measured, or we use reasonable scientific background knowledge. For example, in order to impute the correlation between home and lead, we found seven published studies that include this correlation. The median of these published correlations is −0.470. In order to impute the correlation between the father's and child's IQ, he claims to use the theory of genetics, which tells us that the contribution of the father and mother to the child's genome is on average equal. He thus imputes a value close to 0.477. In similar fashion, he arrives at the correlation matrix in Table 6.

4.2. Measurement error

In the statistical model described by the plaintiffs' expert, the estimates of the effect of lead on IQ were obtained with multiple linear regression. In a regression analysis, the variables are usually separated into the *outcome* (in this case the child's IQ), and

the *regressors* (lead, momiq, faiq, home, etc.). As is well known in the statistical and economics community, the estimates from such an analysis are biased if the regressors are *measured with error*. That is, if the absolute value of the correlation between the real confounder and the variable used to measure it is less than 1, then there is a measurement error. The lower this correlation, the more the measurement error. Since the test–retest correlation of IQ scores is in between 0.9 and 0.95, we know that the IQ regressors are measured with an error. We also have data on the lead measurements, as well as the home score. The defendant's expert explains that we can correct the results from multiple linear regression to take the measurement error into account, but the plaintiffs' expert neglected to do so.

4.3. Sensitivity analysis

Rather than to commit ourselves to a single value for the imputed correlations or the exact amount of measurement error in each regressor, the defendant's expert performed 10,000 analyses, each of which slightly perturbed the original values that we took as our best guess. He sampled values for each correlation that we had to impute from a low of $\frac{1}{2}$ of our best guess to twice our best guess. He also did similar perturbations for a wide range of measurement errors. For each of the 10,000 analyses performed, he stored an estimate of the effect of lead on IQ and of the significance level of the estimates. In *none* of the 10,000 did he find a significant effect of lead on IQ. Therefore, when omitted variables and measurement error are taken into account, he claims that there is no scientific evidence to support the claim that low levels of lead has a deleterious effect on IQ.

5. The plaintiffs' expert rebuttal

5.1. Omitted variables

Although the defense has applied a lot of energy in trying to negate two decades of evidence on the effect of lead on IQ, there are several problems with their analysis (Waternaux et al., 1998). On rebuttal, the plaintiffs' expert describes these problems, the most important of which is the flawed reasoning with respect to imputing "missing" correlations. The defendant's expert made it appear that his analysis was robust to small errors of imputation by doing a sensitivity analysis in which he varied the imputed correlation values by a factor of two both up and down. The plaintiffs' expert then claims that he will show that this range is much to narrow and that in the most important cases the results were way off in a direction that, unsurprisingly, benefited instead of weakened the defendant's case.

The expert points out that, whereas it seems almost negligent for his study and previous ones to have omitted the father's IQ, there is a good reason that they did so: namely, that in the population of children who are typically exposed to significant lead, the fathers are not around to take an IQ test. In fact, in many instances the father is unknown! The defendant's expert imputed correlations between the father's

Table 7

Correlation	Defense's imputed values	More reasonable estimates
Home–momiq	0.230	0.410
Home–lead	−0.470	−0.100

IQ and the other variables in the study by (1) using other studies in which father's IQ and any other variable in their analysis occurs, or (2) by reasoning from background theory. The problem with the first strategy is that *correlations are population specific*. That is, the correlation between the father's IQ and the amount of lead his child was exposed to might be very different in a middle class suburban population than it is in a poor inner city. By using studies from different domains to impute correlations into the Massachusetts study, the expert neglected this basic fact of sampling. In several of the correlations that he imputed, the other studies cited vary considerably depending on the population studied, and in fact varied over a larger range than he used in his sensitivity analysis.

After scrutinizing these external sources more carefully, the plaintiffs' expert made more reasonable guesses as to the correlations to be imputed, and then repeated the analyses. His results (Table 7) were completely at odds with the ones presented by the defendant's expert, in particular, linked to two imputed correlations: home and momiq, and home and lead.

By using the defendant's expert values, the home variable's correlation with lead and its correlation with IQ make it an omitted variable that, when included, accounts for most of the association between lead and IQ, rendering the estimate of the direct effect of lead on IQ insignificant. By using reassessed values, which he claims are much more plausible for the population under study, the expert argues that the reverse is true: the home variable does not account for enough of the association to make the effect of lead on IQ fall below standard levels of significance.

5.2. Measurement error

The expert goes on to agree with his counterpart that the regressors are probably measured with error. What he then argues is that the defendant's expert neglected to mention that measurement error can often attenuate the apparent effect estimated by multiple regression. In this case, measurement error might make the effect of lead on IQ appear closer to zero than it is in reality. He concludes his testimony by describing an attempt to include measurement error in his model. He does this by imposing particular values on the amount of measurement error present in each regressor, a technique which does not allow for any uncertainty over the amount of measurement error. By varying how much measurement error is assumed to exist in their sensitivity analysis, they include uncertainty, but only qualitatively. The analysis by the defendant's expert simply imposed a range on the amount of measurement error on each regressor, but no information about what the *likely* value is, i.e., they put no probability distribution over how much measurement error exists for each regressor.

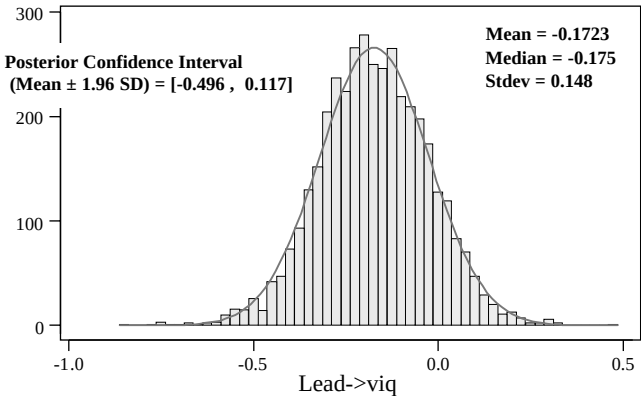

Fig. 2. Posterior distribution of the effect of lead on IQ.

The alternative he explains is to do a fully Bayesian analysis to incorporate the entire distribution rather than simply consider the endpoints of an interval. He goes on to describe such an analysis in which the regressors included lead, mother's IQ, head circumference, number of prior pregnancies, and father's education (e.g., see Scheines et al., 1999, for a more complete description of the model and its specification). The histogram in (Fig. 2) shows the posterior distribution over the causal strength of the effect from lead to IQ. The median in the posterior is -0.175, the mean -0.1723, and the standard deviation in the posterior is 0.148. Although there are some values in the posterior over this effect that are positive, the bulk of the mass is negative, indicating that lead has a deleterious influence on IQ. Almost all of the posterior probability is assigned to values that imply lead exposure reduces IQ.

5.3. *Meta-analyses*

In science, replication is crucial to establishing hypotheses, and it is extremely rare that a single study establishes any causal mechanism firmly, especially in epidemiology. In the case of lead and IQ, there have been literally dozens of studies examining the effect of low levels of lead exposure on a child's IQ. For example, see Pocock et al. (1994).

A few authors have gone through the trouble of systematically collecting the results of previous studies via a *meta-analysis*. In particular, Pocock et al. investigated 14 studies in which lead was measured by taking blood samples, and 7 in which it was measured by Needleman's (1979) technique of assessing its concentration in baby teeth. In all the studies catalogued, the authors tried to control parental IQ, the quality of the home, socioeconomic factors, and other potential confounders. They summarize the results as follows:

> Overall synthesis of this evidence, including a meta-analysis, indicates that a typical doubling of body lead burden (from 10 to 20 μg/dl) (blood lead or from 5

to 10 μg/g tooth lead) is associated with a mean deficit in full scale IQ of 1–2 IQ points (p. 1189).

Testimony about this and other meta-analyses would likely be introduced at trial by other expert witnesses and the defense would surely have witnesses to dispute the findings!

6. Summary

We have described a hypothetical case where the expert testimony began with a variety of fairly standard statistical methods and progressed through multiple regression, sensitivity analyses, and a Bayesian approach to measurement error. These methods cannot be relied upon to demonstrate causation in many instances, and so we borrowed analysis from the recent statistical literature to demonstrate how, in the "battle of the experts," one can expect to see an escalation of the methodology and efforts to use "new results and approaches," ones that may not have been fully vetted in the professional literature. In such circumstances, judges may now exclude parts of statistical and other expert testimony under the Daubert standard (for a description of *Daubert v. Merill Dow Pharmaceuticals*, 113 S. Ct. 2783 (1993), and its implications for statistical testimony, see Fienberg et al., 1995). Our description has focused on the technical details of the expert testimony, but one must always keep in mind that this testimony is part of an adversarial process not controlled by the experts (e.g., Fisher, 1986; Meier, 1986; Fienberg, 1997b).

The case study we have outlined has some lessons for statisticians and econometricians, whether or not they are going to be testifying experts. In assessing arguments about the strengths of causes that rely on regression analysis and related statistical procedures, one must bear in mind that:

- The results typically depend on the variables selected.
- The "statistical significance" slogan is often arbitrary and depends on details of distribution and sample size that are easily hidden in the argument.
- The reliabilities of various statistical procedures are often unstated, and often unknown.
- "Non-standard" procedures, such as alternatives to standard stepwise regression methods and the Bayesian techniques, are often better warranted than more conventional methods.

It is worth noting that the Bayesian methods in expert testimony are gaining increasing currency in the courts although the progress is slower than in the statistical and econometric literatures (e.g., see Evett and Weir, 1998; Fienberg, 1997b; Kadane and Terrin, 1997).

For judges and juries who might be faced by expert testimony of the sort outlined here we note that there is simply no substitute for thinking through the statistical claims and trying to understand the rationale behind competing claims. This is what the *Daubert* decision by the supreme Court now demands.

We have saved our conclusions about the substance of the analyses we have outlined here on lead and IQ. Most of the analyses we describe in this paper are ours, and we find those presented as the plaintiffs' rebuttal case quite compelling. Moreover, the meta-analyses referred above are perfectly consistent with the analyses that "the plaintiffs' expert reported here." In virtue of such a consistent preponderance of replicable evidence, we believe there is little doubt as to the fact that even low levels of lead have a deleterious effect on a child's IQ. The implication of such a finding for the awarding of monetary damages in "The Case of Sam Brown" are another matter and involve economic, statistical, and legal analysis.

7. For further reading

The following references may also be of interest to the reader: Fienberg and Finkelstein, 1996; Fienberg and Kaye, 1991; Fienberg and Straf, 1991.

References

DeGroot, M.H., Fienberg, S.E., Kadane, J.B., (Eds.), 1986. Statistics and the Law. Wiley, New York.

Evett, I.W., Weir, B.S., 1998. Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists. Sinauer, Sunderand, MA.

Fienberg, S.E. (Ed.), 1989. The Evolving Role of Statistical Assessments as Evidence in the Courts. Springer, New York.

Fienberg, S.E., 1997a. Statistics In: Frieman, M.J., Berenblut, M.L. (Eds.), The Litigator's Guide to Expert Witnesses. Canada Law Book, Aurora, Ont., pp. 101–123 (Chapter 7).

Fienberg, S.E., 1997b. Ethics and the expert witness: statistics on trial. Journal of the Royal Statistical Society, Series A 160, 321–333.

Fienberg, S.E., Finkelstein, M.O., 1996. Bayesian statistics and the law. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, Vol. 5. Oxford University Press, Oxford, pp. 129–146.

Fienberg, S.E., Kaye, D.H., 1991. Legal and statistical aspects of some mysterious clusters. Journal of the Royal Statistical Society, Series A 154, 61–74.

Fienberg, S.E., Straf, M.L., 1991. Statistical evidence in the US courts: an appraisal. Journal of the Royal Statistical Society, Series A 154, 49–59.

Fienberg, S.E., Krislov, S., Straf, M.L., 1995. Understanding and evaluating statistical evidence in litigation. Jurimetrics 36, 1–32.

Finkelstein, M.O., Levin, B.A., 2001. Statistics for Lawyers, 2nd Edition. Springer, New York.

Fisher, F.M., 1986. Statisticians, econometricians, and adversary proceedings. Journal of the American Statistical Association 81, 277–286.

Frank, R.G., 1992. Economic aspects of the litigation for harm due to lead poisoning. In: Needleman, H.L. (Ed.), Human Lead Exposure. CRC Press, Boca Raton, FL, pp. 259–266 (Chapter 16).

Gastwirth, J.L., 1988. Statistical Reasoning in Law and Public Policy: Tort Law, Evidence, and Health. Academic Press, New York.

Gastwirth, J.L. (Ed.), 2000. Statistical Science in the Courtroom, Springer, New York.

Good, P.I., 2001. Applying Statistics in the Courtroom: A New Approach for Attorneys and Expert Witnesses. CRC Press, Boca Raton, FL.

Kadane, J.B., Terrin, N., 1997. Missing data in the forensic context. Journal of the Royal Statistical Society, Series A 160, 351–357.

Little, R.J.A., Rubin, D.B., 1987. Statistical Analysis with Missing Data. Wiley, New York.

Marais, M.L., Wecker, W.E., 1998. Correcting for omitted-variables and measurement-error bias in regression with an application to the effect of lead on IQ (with discussion). Journal of the American Statistical Association 93, 494–517.

Meier, P., 1986. Damned liars and expert witnesses. Journal of the American Statistical Association 81, 269–276.

Needleman, H.L., Gatsonis, C.A., 1990. Low-level lead exposure and the IQ of children: a meta-analysis of modern studies. Journal of the American Medical Association 263, 673–678.

Needleman, H.L., Gunnoe, C., Levitan, A., Reed, R., Maher, C., Barrett, P., 1979. Deficits in psychologic and classroom performance of children with elevated dentine lead levels. New England Journal of Medicine 300, 689–695.

Needleman, H.L., Geiger, S., Frank, R., 1985. Lead and IQ scores: a reanalysis. Science 227, 701–704.

Pocock, S.J., Smith, M., Baghurst, P., 1994. Environmental lead and children's intelligence: a systematic review of the epidemiological evidence. British Medical Journal 309, 1189–1197.

Scheines, R., Spirtes, P., Glymour, G., Meek, C., 1994. TETRAD II: Tools for Causal Modeling. User's Manual. Erlbaum, Hillsdale, NJ.

Scheines, R., Hoijtink, H., Boomsma, A., 1999. Bayesian estimation and testing of structural equation models. Psychometrika 64, 37–52.

Spirtes, P., Glymour, C., Scheines, R., 2000. Causation, Prediction, and Search, 2nd Edition. MIT Press, Cambridge, MA.

Waternaux, C., Petkov, E., DuMouchel, W., 1998. Comment: problems with using auxiliary information to correct for omitted variables when estimating the effect of lead on IQ. Journal of the American Statistical Association 93, 505–513.

Zeisel, H., Kaye, D.H., 1997. Prove It With Figures: Empirical Methods in Law and Litigation. Springer, New York.