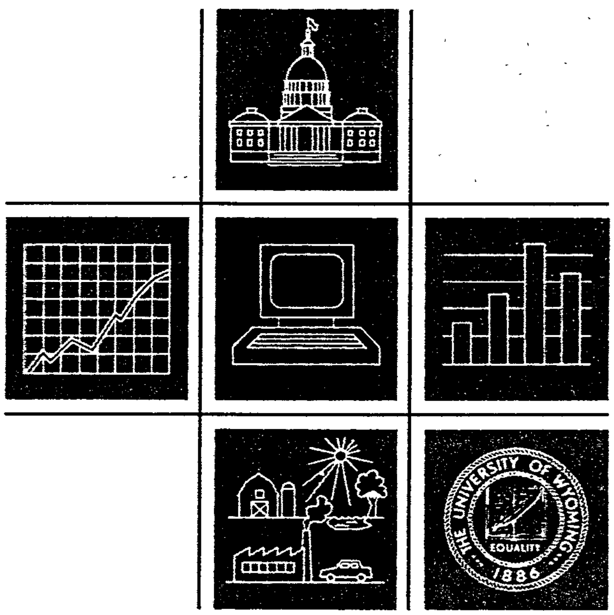
The Institute for Policy Research



...addressing public policy issues for the '80's.

Volume VII

ECONOMIC BENEFITS OF CONTROLLING THE EFFECTS OF ENVIRONMENTAL POLLUTION ON CHILDREN'S HEALTH

METHODS DEVELOPMENT IN MEASURING BENEFITS OF ENVIRONMENTAL IMPROVEMENTS

Volume VII

ECONOMIC BENEFITS OF CONTROLLING THE EFFECTS OF ENVIRONMENTAL POLLUTION ON CHILDREN'S HEALTH

by

Scott E. Atkinson
Thomas D. Crocker
Robert G. Murdock
University of Wyoming
Laramie, Wyoming 82071

Herbert L. Needleman
University of Pittsburgh
Pittsburgh, Pennsylvania 15213

USEPA Grant #CR808-893-01

Project Officer:

Dr. Alan Carlin
Office of Policy Analysis
Office of Policy, Planning and Evaluation
U.S. Environmental Protection Agency
Washington, D.C. 20460

OFFICE OF POLICY ANALYSIS
OFFICE OF POLICY, PLANNING AND EVALUATION
U.S. ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

DISCLAIMER

Although prepared with EPA funding, this report has neither been reviewed nor approved by the U.S. Environmental Protection Agency for publication as an EPA report. The contents do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

THE ECONOMIC CONSEQUENCES OF ELEVATED BODY LEAD BURDENS IN CHILDREN: A PROPSED STUDY FRAMEWORK

by

Scott E. Atkinson, Thomas D. Crocker, and Herbert L. Needleman

SECTION 1

INTRODUCTION

Efforts by economists to value environmental health effects have focused almost entirely upon adult populations' losses in productivity and their willingness-to-pay to avoid health risks, given their current occupations. The economic value of health impacts upon children has been neglected. Of all people, children and the elderly are generally considered to be the most susceptible to health damages from environmental pollutants. Children are thought to be particularly susceptible to neurological, neuromotor, and behavioral impacts from ambient lead concentrations [Needleman, et al. (1979); Provenzano, (1980)]. In this report, we outline an analytical framework suitable for estimating the economic losses that parents/guardians suffer from declines in their child's health status. In addition, given the effects of lead-induced changes in health status upon length of schooling and schooling success, we show how these health status changes can influence subsequent occupational choices and life-cycle incomes.

SECTION 2

PARENTAL INVESTMENT IN CHILD HEALTH

In accordance with Becker (1982, Chapter 2), presume that a household behaves as if one objective function were being maximized, given that the household head's objective function includes as arguments the utilities of the members and that the head has the ability to redistribute the benefits from each member's activities, be they additional earnings or produced service flows, to other members. In an assumed one-period, lifetime setting, the essence of the household's problem is to allocate scarce life-cycle resources between child-rearing and other activities, including market work; that is, parents can spend time and money on their own consumption and investments and/or they can use the same time and money to enhance the expected adult consumption efficiency and human capital stock of their children. For brevity, we refer to the child's expected adult consumption efficiency and capital stock as child-health.

If the economic value of public actions to control ambient lead levels is to be estimated, these actions must be connected with household decisionmaking about activities that influence child-health. In the chain of causation, public actions affect ambient lead levels, which in turn influence the child-health on which net benefits of the public actions depend. However, this simple chain is complicated by the obvious fact that parents are also able to influence child-health by devoting their time and money to its production. We conclude that increases in the child's body lead burden will increase this cost. In addition, we will show that increases in the costs of activities which have a positive impact on child health can increase *as* well as reduce the net benefits of ambient lead pollution control programs. In effect, when the activities which influence child-health are endogenous variables in the family decision process, benefits can result from increases in the marginal cost of producing a given level of child-health.

We adopt a household production formulation (Deaton and Muellbauer, 1980, Chapter 10) to structure the parental decision problem with respect to time and money investment in children. So as to eliminate the intertemporal issues that fertility decisions introduce, we presume the number of children in the household to be given. Household production formulations dominate the economics literature dealing with investments in health. Consistent with this literature, we divide into two stages the household's decision problem. First, the household, in its role as a producer, combines market-purchased goods and time to produce commodities that ultimately enter as arguments in its objective function. The household's problem in this first stage is to minimize the costs of producing any particular bundle of commodities. In the second stage, the household is characterized as having to select that commodity bundle from among the minimum cost set of bundles that maximizes the value of its objective function.

A. The Household's Cost-Minimization Problems

Let the production function for child-health (H) be:

$$H = H(x,\ t;\ g,\ r,\ \alpha), \tag{1}$$

where:

x is child-health related inputs, including other household children.

t is parental child-care time inputs.

g is the child's genetic stock.

r is a set of parental attributes such as the mother's education.

α is air pollution.

The terms in $H(\cdot)$ lying to the right of the semicolon are predetermined.

Label C as the opportunity cost of producing child-health, H. The household's cost-minimization problem is then:

$$\text{Minimize: } C = px + wt \tag{2}$$

subject to (1), where:

p is a price index for child-health related goods. The index is treated as being independent of the level of the child's body burden of lead.

x is a composite measure of child-health related goods. The measure is assumed to be independent of body lead burden.

w is the parental wage rate. This too is presumed independent of the child's body lead burden.

In order to reduce required notation, all terms in this and other expressions are treated as being scaler rather than vector-valued.

In addition to (1), the first-order conditions for an interior solution to the above problem are:

$$p - \lambda\left(\frac{\partial H}{\partial x}\right) = 0 \tag{3}$$

$$w - \lambda\left(\frac{\partial H}{\partial t}\right) = 0 \tag{4}$$

where λ is a Lagrangian multiplier representing the shadow price of making (1) more binding. The solution to (1), (3), and (4) is a cost function.

$$C = px(p,H,g,r,\alpha) + wt(w,H,g,r,\alpha) \tag{5}$$

It can be shown (Deaton and Muellbauer, 1980, Chapter 2) that (5) is positive linear homogeneous, concave in p and w, and nondecreasing in a. By Shephard's lemma, the derived demand for x and t is:

$$x^* = \frac{\partial C}{\partial x} = \frac{\partial x(\cdot)}{\partial p} \qquad (6)$$

$$t^* = \frac{\partial C}{\partial t} = \frac{\partial t(\cdot)}{\partial w} \qquad (7)$$

and the marginal cost of an increase in child-health is:

$$H^* = \frac{\partial C}{\partial H} = p(\partial x/\partial H) + w(\partial t/\partial H) \qquad (8)$$

The changes in the optimal values for x and t can be found by totally differentiating the first-order-conditions to obtain:

$$dp - \lambda H_{11}dx - \lambda H_{12}dt - \lambda H_{1\alpha}d_{\alpha} - H_1 d\lambda = 0 \qquad (9)$$

$$dw - \lambda H_{21}dx - \lambda H_{22}dt - \lambda H_{2\alpha}d_{\alpha} - H_2 d\lambda = 0 \qquad (10)$$

$$dH - H_1 dx - H_2 dt - H_{\alpha}d_{\alpha} = 0 \qquad (11)$$

where $H_1 = \frac{\partial H}{\partial x}$; $H_{11} = \frac{\partial^2 H}{\partial x^2}$; $H_{12} = \frac{\partial^2 H}{\partial x \partial t}$; etc.

Remember that, by assumption, changes in the child's body burden of lead do not influence the unit prices of either child-health related goods or the opportunity costs of child-care time; thus dp = dw = 0, when α changes. Upon putting (9), (10) and (11) in matrix terms and solving for the effects of changes in a upon x and t, one obtains:

$$\frac{dx}{d\alpha} = \frac{H_1 H_{\alpha} H_{22} + H_1 H_2 H_{1\alpha} - H_1 H_2 H_{2\alpha} - H_2 H_{\alpha} H_{12}}{2H_1 H_2 H_{12} - H_1^2 H_{22} - H_2^2 H_{11}} \qquad (12)$$

$$\frac{dt}{d\alpha} = \frac{H_1^2 H_{2\alpha} + H_1 H_{\alpha} H_{11} - H_1 H_{\alpha} H_{21} - H_1 H_2 H_{1\alpha}}{2H_1 H_2 H_{12} - H_1^2 H_{22} - H_2^2 H_{11}} \qquad (13)$$

By making the reasonable assumptions that $H_{1\alpha} = H_{2\alpha} = H_{12} = H_{21} = 0$, these expressions can be simplified to:

$$\frac{dx}{d\alpha} = \frac{H_1 H_{\alpha} H_{22}}{H_1^2 H_{22} + H_2^2 H_{11}} \qquad (14)$$

$$\frac{dt}{d\alpha} = \frac{H_2 H_{\alpha} H_{22}}{H_1^2 H_{22} H_2^2 H_{11}} \qquad (15)$$

Remembering the definition of C in (5), and differentiating (5) with respect to a:

$$\frac{\partial C}{\partial \alpha} = p\frac{\partial x}{\partial \alpha} + w\frac{\partial t}{\partial \alpha} \quad (16)$$

Substituting (14) and (15) into (16):

$$\frac{\partial C}{\partial \alpha} = -\frac{pH_1 H_\alpha H_{22} + wH_2 H_\alpha H_{11}}{H_1^2 H_{22} + H_2^2 H_{11}} \quad (17)$$

Presuming that child-health related goods have a positive but diminishing influence on child health ($H_1 > 0$; $H_{11} < 0$), that lead body burdens have a detrimental health influence ($H_\alpha < 0$),[11] and that parental child-care time inputs also have a positive but diminishing health impact ($H_2 > 0$; $H_{22} < 0$), then the right-hand-side of (16) will be positive in sign. Increases in a child's body burden of lead will increase the out-of-pocket costs (px) and the opportunity costs (wt) of producing a given level of child health.

B. The Household's Utility Maximization Problem

The theory of household production, which developed from the work of Gorman (1956), has had considerable descriptive appeal in modelling the economic behavior of households. At this time, it has a near-monopoly as the framework used for analyzing the economics of non-market household activities.[27] The approach derives from the simple and intuitively appealing observation that households often acquire market goods which do not yield utility directly, but which are combined with other market goods and household time to produce commodities entering as arguments in the household's objective function. As Stigler and Becker (1977) argue at length, the fundamental advantage of the framework is that it distinguishes household tastes, which are not directly observable, from household technology, which can in principle be represented and estimated. However, many commentators consider Stigler and Becker (1977) to be too sanguine about the conceptual and empirical feasibility of distinguishing changes in behavior due to changes in tastes from changes in behavior due to changes in household technology. Pollak and Wachter (1975) show under fairly general conditions that the aforementioned distinction is in fact feasible if and only if the household production function is linearly homogeneous and if there is no jointness in production. Otherwise, implicit commodity prices will depend on both the household's tastes and its technology, causing these prices to be functions of the commodity bundle the household consumes rather than the parameters which the household confronts. In order to proceed with the household's utility maximization problem, we choose not to ignore the Pollak and Wachter (1975) criticism; we therefore presume that the household production for child-health in (1) exhibits constant-returns-to-scale and that it embodies no joint products. Plainly, these restrictions violate some reality, but the degree of violation is unclear at this time.

The constant-returns-to-scale premise implies that the marginal cost of producing child-health in expression (9) is also the average cost; that is:

$$H^* = \partial C/\partial H = C/H = Q\ (p,w,g,r,a) = px + wt, \qquad (18)$$

since $(\partial x/\partial H)$ and $(\partial t/\partial H)$ in (9) are now constants.

The household's "full-income" budget constraint can be constructed by initially considering separately the time constraint and the-budget constraint. For given values of p and w, define the. time constraint as:

$$tH + t_L + t_w = T, \qquad (19)$$

where:

- t continues to be parental child-care time inputs per-unit of child-health, H.
- t_w is parental time devoted to work. In order to simplify the exposition, mother's time and father's time is viewed as homogeneous.[3]
- t_L is parental time devoted to nonmarket activities, including leisure.

The budget constraint is:

$$pxH + M = V + wt_w, \qquad (20)$$

where:

- px is expenditures on child-health related goods per unit of child-health. The assumption of no joint products does not allow Rosenzweig's and Schultz's (1982) distinction between inputs acquired solely because of their contribution to child-health and goods (e.g., smoking) which simultaneously are inputs into child-health as well as sources of parental utility.
- M is parental consumption activities having no direct impact on child-health.
- V is predetermined income.
- wt_W is current labor income.

Assuming smooth substitutability between parental leisure and work, the "full income" constraint is obtained by first rearranging (19) so that $t_w = T - tH - t_L$, and then substituting for T_w in (20). Thus:

$$pxH + M = V + wT - wtH - wt_L$$

or:

$$(px + wt)H + (M + wt_L) = V + wT \qquad (21)$$

As noted in (18), (px + wt) is defined as Q, the marginal cost of child-health. This marginal cost is made up of the sum of expenditures on child-health-related inputs and the opportunity costs, as measured by their market wage rates, of parental child-care time. The term $(M + wt_L)$ is the resources the family has remaining for consumption activities after its expenditures on child-health inputs. The right-hand side of (21) is the household's total wealth during the period in question. Carrying the notation of (21) through the following exposition can be awkward. The burden is reduced by letting $R \equiv (M + wt_{,})$. Since we are uninterested in parental substitutions or complementaritres between leisure time and parental consumption activities unrelated to child-health, this simplification is achieved without sacrifice. Similar reasoning allows $S \equiv (V + wT)$.

With these notational simplifications, the Lagrangian for the household's utility maximization problem can be stated as:

$$\text{Maximize:} \quad U(H, R) \tag{22}$$

$$\text{subject to:} \quad QH + R = S \tag{23}$$

Upon substituting (1) into (22) and (23), the Lagrangian for the household's utility maximization problem becomes:

$$\mathcal{L} = U[H(x,t;g,r,\alpha),R] - \lambda[QH(x,t;g,r,\alpha) + R - S] \tag{24}$$

The simple problem specified in (24) has several features worthy of explicit comment. First, the household is unable to acquire child-health directly; instead, goods must be acquired and parental time must be used in order to influence child-health in the manner described by (1). Second, the appearance of H in the household's objective function means that child-health is valued in its own right. Finally, the introduction of R in the objective function, $U(\cdot)$, means that the parents are unwilling to sacrifice everything in order to secure an additional unit of child-health.

The first-order-conditions for the above problem are:

$$\frac{\partial \mathcal{L}}{\partial x} = \frac{\partial U}{\partial H}\frac{\partial H}{\partial x} - \lambda Q = 0 \tag{25}$$

$$\frac{\partial \mathcal{L}}{\partial t} = \frac{\partial U}{\partial H}\frac{\partial H}{\partial t} - \lambda Q = 0 \tag{26}$$

$$\frac{\partial \mathcal{L}}{\partial R} = \frac{\partial U}{\partial R} - \lambda = 0 \tag{27}$$

and (23). Expression (27), which applies to expenditures on the weakly separable composite commodity, R, is thoroughly conventional. Taken together, (25) and (26) state that the household will be maximizing its utility when it equates its subjective marginal-rate-of-substitution between child-health-related goods and child-care time to the marginal costs (= average costs) of producing child-health. Considering (27) and (26) or (25) together, note that if the mother works full-time outside the home, the marginal-rate-of-substitution between her consumption and self-investment activities and child-health must be less than the opportunity costs of her loss in leisure and/or child-care time. Similarly, if she is full-time at home, so that her $t_w = 0$, her time input into child-care cannot be enhanced unless she sacrifices leisure. In circumstances where her leisure time is invariant, the marginal-rate-of-substitution between her consumption activities and child-health must exceed the opportunity costs of not working.

Failures (corner solutions) to fulfill the second-order conditions for solving (24) can be readily dismissed since parents must have something left over for their own subsistence and since few, if any, families will be willing to sacrifice all child-health in order to enhance their own consumption and leisure and work-time. In short, there exist culturally-dictated nonzero minimal for both child-care time and child-health-related goods inputs.

The system (23) and (25)-(27) differs from the usual first-order-conditions in that Q is not a fixed price exogenously given to the household. It is instead a function of the household's decision variables and will vary over households to the extent that input prices, wage rates, genetic backgrounds, parental attributes, and child body lead burdens vary over households. A system of derived demand equations for H and R can be obtained by solving for H, R, and A in terms of the predetermined variables, Q and S.

$$H = H\ (Q,\ S) = H\ (p,\ w,\ V) \tag{28}$$

$$R = R\ (Q,\ S) = R\ (p,\ w,\ V) \tag{29}$$

These expressions state that parental demand for child-health H, and for consumption goods R, unrelated to child-health, depend upon the prices of child health-related goods, wage rates, and predetermined income. The effects of price, wage, and body burden lead upon the parents' demand for child-health can now be explored.

When the three-equation system in (25)-(27) is totally differentiated and terms are collected, the result is the bordered Hessian:

$$\begin{bmatrix} \frac{\partial^2 U}{\partial H^2} & \frac{\partial^2 U}{\partial H \partial R} & -Q \\ \frac{\partial^2 U}{\partial H \partial R} & \frac{\partial^2 U}{\partial R^2} & -1 \\ Q & -1 & 0 \end{bmatrix} \begin{bmatrix} dH \\ dR \\ d\lambda \end{bmatrix} = \begin{bmatrix} \lambda dQ \\ 0 \\ H^* dQ - dV - d(wT) \end{bmatrix} \tag{30}$$

remembering that $S \equiv V + wt$. Solving (30) for the vector of differentials yields:

$$\begin{bmatrix} dH \\ dR \\ d\lambda \end{bmatrix} = 1/Z \begin{bmatrix} Z_{11} & Z_{12} & Z_{13} \\ Z_{12} & Z_{22} & Z_{23} \\ Z_{13} & Z_{32} & Z_{33} \end{bmatrix} \begin{bmatrix} \lambda & 0 & 0 \\ 0 & 0 & 0 \\ H^* & -1 & -1 \end{bmatrix} \begin{bmatrix} dQ \\ dV \\ d(wT) \end{bmatrix} \tag{31}$$

where Z is the determinant of (30) and the Z_{ij} are elements of the matrix which is the adjoint of (30). Expression (31) now enables us to predict the impact of changes in any of the exogenous variables upon the signs of the changes they might induce in parental demand for child-health. In general, these demand changes can be expressed as:

$$dH = 1/Z\ [(\lambda Z_{11} + H^*Z_{13})dQ - Z_{13}dV - Z_{13}d(wt)] \qquad (32)$$

A change in body burden of lead will have the following effect upon parental demand for child-health:

$$\frac{dH}{d\alpha} = \frac{\partial Q}{\partial \alpha} \left[\lambda \frac{Z_{11}}{Z} + H^* \frac{Z_{13}}{Z}\right] < 0 \qquad (33)$$

Clearly, $\partial Q/\partial \alpha$ is the change in the marginal cost of producing child-health due to a change in the child's body burden of lead. For reasons previously explained, it is expected to be positive in sign, although, because of the child-health production function (1) which underlies it, its magnitude will vary inversely with the ease that parents have in substituting across child-health inputs. Thus, for example, $\partial Q/\partial \alpha$ will be greater when a mother has an inflexible outside work schedule than when she is a housewife with substantial discretion over the uses to which she puts her time.

The first term in the brackets on the right-hand side of (33) is analogous to the substitution effect of a price change. As with any substitution effect, it must be negative. If child-health is a normal good, then the second term in brackets, which is analogous to an income effect, must also be negative. Consequently, the entire right-hand side of (33) is negative, implying that an increase in the child's body burden of lead will reduce parental demand for child-health.

Consider now a change in the price of a good that is an input to child-health. In particular, because the mother's wage rate can be viewed as the opportunity cost of her child-care time, consider a change in her wage rate. From (1) and (32), we have:

$$\frac{dH}{dw} = \left[\lambda \left(\frac{Z_{11}}{Z}\right) + H^*\left(\frac{Z_{13}}{Z}\right)\right] \frac{\partial Q}{\partial w} - \left(\frac{Z_{13}}{Z}\right) \left(\frac{\partial (wT)}{\partial w}\right) \gtrless 0 \qquad (34)$$

The term $\partial(wt)/\partial w$ will clearly be positive, which, since Z_{13}/Z is negative, implies that the sign of last pair of terms on the right-hand side of (34) will be positive. As in (33), the term in brackets will be negative. Finally, since w is the opportunity cost of the mother's child-care time, a change in w will cause Q to change in the same direction, implying that the collection of terms to the left of the minus sign in (33) will be negative. This is the substitution effect of the change in the wage rate. The overall effect of the wage change is thus ambiguous. If there are good substitutes for the mother's child-care time, then the income effect will tend to dominate, and a wage increase may actually increase parental demand for child-health. On the other hand, if good substitutes for the mother's child-care time are unavailable, a wage increase for the mother can result in a reduction in the time she spends with the child, and the demand for child-health could actually decline.

In two-parent families, children tend to be female rather than husband time-intensive. An increase in the husband's wage can thus be treated as an increase in predetermined income, R; that is, full income is increased. Thus, from (32), assuming the mother's wage is unchanged:

$$\frac{dH}{dR} = -\frac{Z_{13}}{Z} > 0, \tag{35}$$

which implies that the relative of child-health declines with an increase in the husband's wage, and that the demand for child-health will increase. Child-health-related goods and female time will be substituted for husband's time. Moreover, given that the marginal product of female time in child-health care is positive, the husband's derived demand for female child-care time will increase as his wage rate increases. This implies that the labor supply of the mother will be inversely related to the husband's wage rate.

SECTION 3

EMPIRICAL IMPLEMENTATION

A. Specification

In Section 2, both the marginal costs of child-health; Q, and child-health itself, H, are endogenous. The difficulties that arise in estimating the above framework are therefore similar to the general problems of supply and demand estimation when both price and quantity are endogenous. There are at least two ways of overcoming this problem. First, one can assume Q to be constant and that there are no choice variables other than H; that is, R is simply whatever funds and time the parents have left over after having fulfilled some prior level of child-health. Obtain variation in prices sufficient to identify the demand for by supposing that families have different total cost fns and thus different constant marg costs. Then restrict parameters--consider, for example, linear approximations to demand (28) and supply (18) for the child-health service flow.

Demand for:

$$H = \alpha_1 + \alpha_2 S + \alpha_3 M + \alpha_4 Q + \varepsilon_1 \tag{36}$$

where M is a set of background "taste" variables.

Supply for:

$$Q = \beta_1 + \beta_2 Q + \beta_3 p + \beta_4 w + \beta_5 a + \beta_6 g + \varepsilon_2 , \tag{37}$$

Solving (36) and (37)--H and Q--in terms of the exogenous variables, we get the following reduced forms:

$$\left.\begin{aligned} H = {} & \frac{\alpha_1 + \alpha_4\beta_1}{1 - \alpha_4\beta_2} + \frac{\alpha_2}{1 - \alpha_4\beta_2} S + \frac{\alpha_3}{1 - \alpha_4\beta_2} M + \frac{\alpha_4\beta_3}{1 - \alpha_4\beta_2} p \\ & + \frac{\alpha_4\beta_3}{1 - \alpha_4\beta_2} w \frac{\alpha_4\beta_5}{1 - \alpha_4\beta_2} a + \frac{\alpha_4\beta_6}{1 - \alpha_4\beta_2} g + \frac{\alpha_4\varepsilon_2 + \varepsilon_1}{1 - \alpha_4\beta_2} \end{aligned}\right\} \tag{38}$$

$$\left.\begin{aligned} Q = {} & \frac{\beta_1 + \beta_2\alpha_1}{1 - \beta_2\alpha_4} + \frac{\beta_2\alpha_2}{1 - \beta_2\alpha_4} S + \frac{\beta_2\alpha_3}{1 - \beta_2\alpha_4} M + \frac{\beta_3}{1 - \beta_2\alpha_4} p \\ & + \frac{\beta_4}{1 - \beta_2\alpha_4} w + \frac{\beta_5}{1 - \beta_2\alpha_4} a + \frac{\beta_6}{1 - \beta_2\alpha_4} g + \frac{\varepsilon_2 + \beta_2\varepsilon_1}{1 - \beta_2\alpha_4} \end{aligned}\right\} \tag{39}$$

Now assume that air pollution increases the cost of producing a constant quality child-health services flow, thus implying that $\beta_5 > 0$. Further assume that there are decreasing returns to investments in child health, which means that $\beta_2 > 0$. Finally assume $\alpha_4 < 0$, or that increases in child health have positive utility. With these assumptions, an unambiguous definition of the effects of air pollution on the quantity demanded and the price of child-health service flows is obtained. Specifically,

$$\frac{\alpha_4\beta_5}{1 - \beta_2\alpha_4} < 0; \quad \frac{\beta_5}{1 - \beta_2\alpha_4} > 0 \, , \qquad (40)$$

(40) says increases in air pollution will reduce the quantity demanded of child-health service flows and increase the marginal cost of supplying these flows.

A second alternative is to collect enough information on exogenous parameters referring to genetic attributes and parental attributes--calculate number of exogenous variables needed by counting the number (k) of exogenous variables in each expression of the structural system--identification requires that at least k of the system's exogenous variables be excluded from each expression--thus, the system requires, at minimum, (k + 1) exogenous variables--moreover, because the arguments of the budget constraint help to determine Q, these k variables cannot appear in the budget constraint-- estimate the following system--in accordance with Barnett (1977).

$$x^* = \partial C/\partial x = x(\cdot)/\partial p = x\ (p,\ w_i,\ H;\ r) \qquad (6)$$

i = husband, wife

$$t_i^* = \partial C/\partial t = \partial t(\cdot)/\partial w = t\ (p,\ w_i,\ H;\ r) \qquad (7)$$

$$Q = \partial C/\partial H = Q\ (p,\ w_i,\ H;\ r,\ \alpha) \qquad (18)$$

$$H = H\ (x,\ t,\ Q) \qquad (8)$$

The problem with this alternative is that information on many of the relevant variables will be hard to get--moreover, is arguable whether the wage of the wife is exogenous

Obvious implication--child-health and family labor supply are jointly determined.

Since constant returns-to-scale have been assumed, (18) for Q will be independent of the level of child-health. Nevertheless, the system (6), (7), (8), and (18) does allow one to impose the restrictions--homogenity, symmetry, and negativity--available from the general theory of the consumer as applied to demand systems.

B. Handling Difficult-to-Measure Variables

Preferences for children.

Use indicators of family socail class--implies an hypothesis of socially-conditioned preferences--an hypothesis that competes with the econ model of price and income effects.

Perceptions of parental responsibility toward children differ by class.

Many components of child-health expenditures are joint with parental personal expenditures, e.g., housing.

Possible indicators of differences in tastes.

Usual social class measures.

Aspirations for children's education.

Contrary to much work that has been done, the $t(\cdot)$ and $x(\cdot)$ expressions, (6) and (7), include child-health as an endogenous variable--expression (8) represents the demand for child-health--entire treatment is couched in terms of lifetime labor supply, not short-term labor supply--how to get a measures of lifetime labor supply?

Employ instrumental variables techniques such that restirctions on the form of the relation between the observed and the unobserved variables are sufficient to identify the parameters to be estimated. Price of goods--likely to be very little varfation in overall prices if all individuals come from the same locale--but, because of various subsidy programs, effective prices of various child-health inputs may vary, e.g., day-care centers, school lunches, etc.

Wage rates.

Obtain for each period (age-specific) and then, as in Willis (1973), average over the periods of the life-cycle--make wage rates in each period a function of education, age, and family traits.

Or, use mother's wage prior to birth of first child; use husband's current wage.

Or, as in Nerlove and Razin (1981), use average values of the discounted wages per unit time for the prior-to-birth period, the child-rearing period, and the post child-rearing period--basically, need detailed information on mother's work history. Might not be able to observe mother's wage during post child-rearing period--make a function of experience and wages before first birth and during child-rearing period. Basic point is that opportunity cost of mother's child-rearing time is not only lost wages but lost experience (lost future productivity).

Price of goods inputs for child-health.

Will clearly depend in part on number and age structure of siblings. Could use Espenshade's (1973) or Lindert's (1978) estimates of the goods costs of raising children from birth to adulthood--but, as Muellbauer (1978) argues, these estimates are inherently full of analytical holes. Must otherwise worry about building a price index--or, on basis of findings such as Murane, et al. (1981), that goods inputs play a trivial role in children's achievements, could work only with parental time inputs and their opportunity costs--above some threshold level, marginal products of goods inputs are trivially small--would then have basically the same system as Nerlove and Razin (1981).

Or, could go to conditional cost or demand for literature, e.g., Pollak and Wales (1979).

Functional form to estimate demand system.

Could use the translog indirect utility function as set forth in Christensen, et al. (1975) --requires interior solutions for all goods but this is no problem here--Pollak and Wales (1980) illustrate how to handle family composition effects.

Estimators.

Must account for the fact that labor force participation is dichotomous, and that observed hours and weeks will be concentrated at aero.

SECTION 4

TYING RESULTS TO LIFE CYCLE EARNINGS

Am interested in $(\partial H/\partial Q)(\partial Q/\partial \alpha)$, where H is a school performance indicator, Q is the constant marginal cost of producing school performance, and a is the child's body burden of lead--presume $\partial H/\partial Q > 0$, and that $\partial Q/\partial \alpha > 0$.

Earnings function--more-or-less typical--see Mincer (1974) for arguments for semi-log.

$$\log W = u + bY + cH + sA, \tag{41}$$

where W is wages, Y is years of school, and A is age.

Implies that there exists complementarity between schooling and school performance --in particular, $(\partial^2 \log W/\partial Y \partial H) > 0$--the marginal product of more school years depends upon health (ability), as assessed by school performance.

Increases in health will not only increase the present value (W*) of earnings from a given number of school years--health increases will also improve the rate-of-return to additional schooling and increase the incentives for acquiring additional education--thus:

$$\Delta = \frac{dW^*}{dH} = \left.\frac{\partial W^*}{\partial H}\right|_{Y=\bar{Y}} + \frac{\partial W^*}{\partial Y}\frac{\partial Y}{\partial H}, \tag{42}$$

where Δ is the shadow price in terms of life-cycle earnings of one more unit of child health.

First term on the right-hand-side of (42) shows the life-cycle earnings gains of improved child health for a given number of schooling years.

Second term on the right-hand side shows the life-cycle earnings generated by the induced increase in years of schooling.

If years of schooling are the main determinant of the individual's occupation, then all work on the earnings impacts of pollution has neglected the second term on the right-hand side of (42)--has dealt only with the first term in which years of schooling are fixed. Further elaboration of (42).

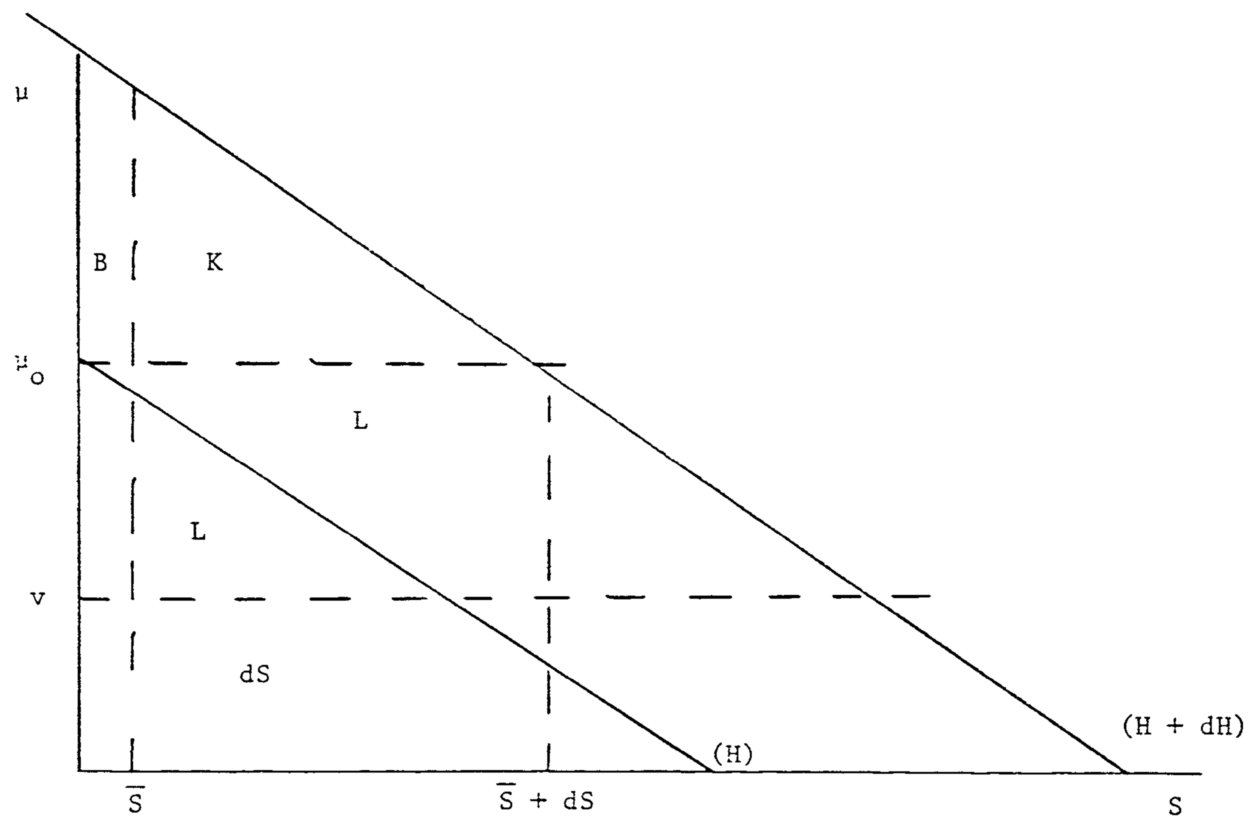
Let the (assumed) dimishing rate-of-return to additional schooling be

μ.

Presume that the household always equates its opportunity cost of capital to the marginal rate-or-return on additional schooling.

Figure 1

Adjustment in Years of Schooling Induced by a Decrease in Health



When the level of H increases, the rate-of-return schedule for education shifts upward. The household then adjusts its years of schooling so as to maintain the prior rate-of-return, μ_o--μ_o is determined by the household's cost of funds or by its discount rate.

Area B in Figure 1 shows the increase in the marginal rate-of-return on all intramarg schooling years--gives the first component of (42), namely $\partial W^*/\partial H$.

Area K is the individual's excess return on the additional induced schooling years, dY.

Let v be the society's cost of capital--if, in accordance with the risk pooling arguments of Arrow and Lind (1970), and Samuelson (1964), $v < \mu_o$, then LL gives the additional excess return to the society (over and above the excess return to the individual) of the additional induced years of schooling--if $v = \mu_o$, then LL would not exist.

If H is treated as school performance or as years of schooling, the estimated impact of changes in a child's body lead burden upon school performance are readily tied to the Haveman and Wolfe (1982) synthesis of the empirical literature which relates school performance to subsequent adult economic well-being.

SECTION 5

DATA REQUIREMENTS

Much of the data required to implement the model in Section 2 is already available in the study of Needleman, et al. (1979) in their study of body burden lead and intellectual deficits for 158 grade school children in Chelsea-Somerville, Massachusetts. Nevertheless, additional data would be required to give economic content to the Needleman, et a1.(1979) findings. The following listing describes the data already present in the Needleman sample, and the supplemental data an economic study must have (*), data one would fine very useful (**) but can do without, and data the absence of which would impose little loss in the reliability of estimates.

Data already available (when child was in Grades 1 or 2)

Child

- Tooth lead levels
- Frequency of negative reports on teacher's behavioral ratings
- Indices of intrinsic intelligence, visual and hearing acuity, and school performance
- Race
- Sex
- Pica history (blood)
- Completed immunizations
- Attended nursery school or day care center, age when started, hours per day, days per week
- Age
- Height
- Weight
- Head circumference
- Right arm skinfold
- Left arm skinfold
- Marital status of parents at time of testing
- Birth weight
- Takes medication
- Birth order
- Number of hospital admissions
- Length of infant hospital stay
- Length of pregnancy
- Grade school teacher's name

Father

Age at subject's birth
Education
Occupation
Natural father
IQ

Mother

Age at subject's birth
Number of pregnancies, successful and unsuccessful
Education
Occupation
Natural mother
IQ

Family

Index of aspirations for child
Index of home learning environment
Index of parental attitude toward school
Index of parental attitude toward child
Index of parental restrictions upon child

Supplemental Data: Child Dentine Lead Study

Identifying Material

Name of Subject
Identification number (same as in original study)*

Data Supplement to Original Study (all recall questions)

CHILD
Number of residental moves prior to time of original study?
**Including older siblings, any relatives who frequently spent one or more house a day taking care of the child?
If attended nursery school, who generally transported the child before and after school?
*Who generally transported child to and from grades 1 and 2?
**Who transported child when he/she visited a physician?
Price of nursery or day care center?
*Wash child covered by heatlh insurance?
Any chronic health problems that inhibited school attendance?

FATHER (or senior household male member)

*Chronic illness that inhibited work activity? Describe.
**Occupation at time of marriage?
*Age at time of marriage?
Time with employer?

Number of jobs simultaneously held?
*Highest paid position held prior to time of initial study?
*Number of months employed in highest paid position?

MOTHER (or senior female household member)

**Age when first entered labor force?
*Occupation at time of marriage that produced child subject?
  Full-time or part-time?
*Highest paid position between marriage and first birth?
Number of months in position?
*Normal work hours?
*Flexible times?
*For those who chose <u>not</u> to work--if you had worked during this period, what kind of job do you believe you could have obtained?
*Highest paid position held when any preschool child was in home? (Does not include nursery school or preschool)
Number of months in position?
*Normal work hours?
*Flexible times?
*For those who chose <u>not</u> to work--if you had worked during this period, what kind of job do you believe you could have obtained?
*Paid help with housework and child care in the home? Hours per week?
*Highest paid position held after all children reached school age?
Number of months in position?
*Normal work hours?
*Flexible times?
*For those who chose <u>not</u> to work--if you had worked during this period, what kind of job do you believe you could have obtained?
When you were 18 years old, what career did you hope to follow?
*Chronic illness that inhibited work activity? Describe.

Current Data

CHILD

*Current health status information (to be described by H. Needleman)
*School performance indices (to be described by H. Needleman)
Current marital status of biological parents?
**Name of school?
*Name of family physican, or other primary health care provider?
  Time required to get an appointment?
  Time consumed in combination of travel and waiting room?
*Expected out-of-pocket health care expenditures over next 12 months?
**Expected insured health care expenditures over next 12 months?
*Number of visits to primary health care provider during past year?
**Days of school missed last year?
**Name any special educational programs in which the child participates?
**Hours per week?
**Out-of-pocket cost?
*Chronic illnesses that inhibit school or special program attendance?

FATHER (or senior household male member)

*Same person as in original study?
*Employed full-time during last year? Part-time?
*Current occupation?
*Normal hours?
*Wage per unit time?
Hours per week having direct interactions with child?
**Hours per week in household chores?
*Chronic illness that inhibit work activity? Describe.
Additional education acquired since original study?

MOTHER (or senior household female member)

*Same person as in original study?
Additional education acquired since previous study?
*Employed full-time last year? Part-time?
*Current occupation?
*Normal hours?
*Wage per unit tine?
*If unemployed or part-time employed, how much do you believe you could command per hour or per week if you chose to work full-time? Kind of job?
**Any paid help for housework and child care in the home?
**Hours per week?
**Cost per week?
**Hours per week spent in housework?
Own microwave oven?
**Hours per week spent in direct interaction with the child?
*In 1983 dollars, how high would your weekly earnings have to be before you would seriously consider taking a full-time job?
*Currently? (Only if not employed full-time)
*When pre-school children were in the home?
*When all children are of school age?
*Chronic illnesses that inhibit work activity? Describe.
**If you and your husband had a choice between a cash bonus of $1,000 today and any one of the amounts on this card five years from now, what is the lowest card amount that would cause you to forego the $1,000 today? (Display card with $1,000, $2,000, $3,000, $4,000, $5,000, $7,000, $10,000, S13,000, $17,000, $20,000, printed upon it.)

FAMILY

*Ages and sexes of siblings?
Chronic health problems of siblings that inhibit school attendance? Describe.
*Percentage of family income from sources other than jobs?
Monthly expenditures by category (in rough percentage terms)
**Shelter?
**Food?
**Transportation?

**Entertainment?
**Clothing?
**Savings?
**Health and personal car?
**Other?

Key to relative importance of various questions.

* = must have.
** = can do without, but absence greatly weakens reliability.

The absence of any asterisks means that the data would be nice to have, but failure to acquire it will not cause great cries of anguish.

REFERENCES

1. See Bartel and Taubman (1979) for one of the most informative of this class of studies.

2. For an interesting deviation which nevertheless appeals to the household production framework, see Cauley and Sandler (1980).

3. This does not mean that we believe the criticisms of Pollak and Wachter (1975) to be universally valid. No one seems to have explored the circumstances (e.g., the weak complementarity theorem of Maler (1974) and Bradford and Hildebrandt (1977) under which the derived demand for goods used to produce the commodity might provide a measure of the value of the commodity.

4. A more detailed analysis would not only distinguish between mother's and father's work-hours but also the daily timing of these hours. On the possible relevance of the daily timing issue, see Presser and Cain (1983).

BIBLIOGRAPHY

Arrow, K.J., and R. Lind, "Uncertainty and the Evaluation of Public Investment Decisions," The American Economic Review 60(1970), 364-378.

Barnett, W.A., "Pollak and Wachter on the Household Production Function Approach," J. of Political Economy 85 (October 1977), 1073-1083.

Bartel, A., and P. Taubman, "Health and Labor Market Success: The Role of Various Diseases," The Review of Economics and Statistics 61 (February 1979), 1-8.

Becker, Gary S., A Treatise on the Family, Cambridge, Mass.: Harvard University Press (1981).

Bradford, D.D., and M. Hildebrandt, "Observable Preferences for Public Goods," J. of Public Economics 8 (June 1977), 106-114.

Cauley, J., and T. Sandler, "A General Theory of Interpersonal Exchange, Public Choice 35 (1980), 587-606.

Christensen, L.R., D.W. Jorgenson, and L.J. Lau, "Transcendental Logarithmic Utility Functions," The American Economic Review 65 (June 1975), 367-383.

Deaton, A., and J. Muellbauer, Economics and Consumer Behavior, New York: Cambridge University Press (1980).

Espenshade, R.J., "The Price of Children and Socio-Economic Theories of Fertility," Population Studies 26 (July 1972), 207-222.

Gorman, W.M., A Possible Procedure for Analyzing Quality Differentials in the Egg Market, Ames, Iowa: Iowa State College, mimeo. (1956).

Guyton, A.C., Basic Human Physiology: Normal Function and Mechanisms of Disease, Chicago, Illinois: W.B. Saunders Co. (1971)

Haveman, R., and B. Wolfe, Education and Economic Well-Being: The Role of Non-Market Effects, Madison, Wisconsin: Institute for Research on Poverty Discussion Paper #716-82, Univ. of Wisconsin (January 1983).

Lindert, P.H., Fertility and Scarcity in America, Princeton, N.J.: Princeton University Press (1978).

Maler, K.G., Environmental Economics: A Theoretical Inquiry, Baltimore, Maryland: The Johns Hopkins University Press (1974).

Mincer, J., "The Distribution of Labor Incomes: A Survey with Special Reference to the Human Capital Approach," J. of Economic Literature 8 (March 1970), 1-26.

Muellbauer, J., "Household Production Theory, Quality and the 'Hedonic Technique,'" The American Economic Review 64 (December 1974), 977-994.

Muellbauer, J., "Testing the Borten Model of Household Composition Effects and the Cost of Children," Economic J. 87 (September 1977), 460-487.

Murane, R.J., R.A. Maynard, and J.C. Ohls, "Home Resources and Children's Achievement," The Review of Economics and Statistics 63 (August 1981), 369-377.

Needleman, H.L., C. Gunnoe, A. Leviton, R. Reed, H. Peresie, C. Maher, and P. Barrett, "Deficits in Psychologic and Classroom Performance of Children with Elevated Dentine Lead Levels," The New England J. of Medicine 300 (March 1979), 689-695.

Nerlove, M., and Razin, "Child Spacing and Numbers: An Empirical Analysis," in A. Deaton, ed., Essays in the Theory and Measurement of Consumer Behavior, New York: Cambridge University Press (1981), 297-324.

Pollak, R.A., and M.L. Wachter, "The Relevance of the Household Production Function and Its Implications for the Allocation of Time," J. of Political Economy 83 (April 1975), 255-277.

Pollak, R.A., and. N.J. Wales, "Comparisons of the Quadratic Expenditure System and the Translog Demand Systems with Alternative Specifications of Demographic Efforts," Econometrica 48 (1980), 595-612.

Pollak, R.A., and N.J. Wales, "Welfare Comparisons and Equivalence Scales," The American Economic Review: Papers and Proceedings 69 (May 1979), 216-221.

Presser, H.B., and V.S. Cain, "Shift-Work among Dual-Earner Couples with Children," Science 219 (February 18, 1983), 876-879.

Provenzano, G., "The Social Costs of Excessive Lead Exposure during Childhood," in H.L. Needleman, ed., Low-Level Lead Exposure: The Clinical Implications of Current Research, New York: Raven Press (1980), 299-315.

Rosenzweig, M.R., and T.P. Schultz, "The Behavior of Mothers as Inputs to Child Health: The Determinants of Birth Weight, Gestation, and Rate of Fetal Growth," in V.R. Fuchs, ed., Economics Aspects of Health, Chicago, Illinois: University of Chicago Press (1982), 53-92.

Samuelson, P.A., "Discussion," The American Economic Review: Papers and Proceedings 54 (1964), 93-96.

Stigler, G.J., and G.S. Becker, "De Gustibus Non Est Disputandum" The American Economic Review 67 (March 1977), 76-90.

Willis, R.J., "A New Approach to the Economic Theory of Fertility Behavior," J. of Political Economy 81 (March/April 1973) Part II, S14-S64.

CHAPTER II

HAVE PRIORS IN AGGREGATE AIR POLLUTION EPIDEMIOLOGY DICTATED POSTERIORS?

by

Scott E. Atkinson, Thomas D. Crocker, and Robert G. Murdock

SECTION 1

INTRODUCTION

Most problems encountered by environmental policymakers involve sampling as well as model specification uncertainty. Given the difficulties of conducting controlled experiments to improve understanding of environmental phenomena, statistical inference is frequently employed. This research usually involves a statistical model that is dependent on the investigator's prior beliefs about the relationship between the dependent variable of interest and a list of explanatory variables. When the policymaker uses these research results to select a course of action, he employs the combined result of the investigator's prior beliefs and data. Unfortunately, the investigator's prior beliefs and the implications of different model specifications, i.e., the degree of model specification uncertainty, are not often reported completely.

This paper focuses on statistical information generated by the Lave and Seskin [1,2,3] and the Chappie and Lave [4] studies of the human health impacts of urban air pollution. After having admittedly engaged in substantial pretesting, the authors of these studies report a selected set of results. However, they provide little information about the role in selection that their prior beliefs played; that is, they do not report the robustness of their reported results with respect to key parameters of interest (focus variables) as the set of included explanatory variables (doubtful variables) changes. Since different sets of doubtful variables may be equally plausible *a priori*, the investigator should report the sensitivity of the estimates of the signs and magnitudes of the focus variables to changes in the list of included doubtful variables. A failure to consider and report results for the full range of alternative model specifications which could be "true" means that the opportunity for the policymaker to select whatever mix of possibly "true" specifications best suits his objectives has been censored. The selection of one or a few models conforming to the investigator's priors can be misleading when several models that differ in their policy implications have some prior credibility. All available information bearing on the robustness and general validity of the alternative models should be provided the policymaker. Because the reported results of Chappie and Lave [4] and Lave and Seskin [1,2,3] have been so widely cited, we apply Learner's [5] procedure to estimate the specification uncertainty of their models.

SECTION 2

A BRIEF HISTORY OF AGGREGATE AIR POLLUTION EPIDEMIOLOGY

Although its influence on policy is unclear, the sequence of papers and books produced by Lave and his colleagues on the human health effects of air pollution has been the most frequently referenced work in urban environmental economics over the last two decades. The basic approach has remained that adopted in the path-breaking effort of Lave and Seskin [1]. Using aggregate, cross-sectional data for 114 U.S. metropolitan areas, they employed single equation, ordinary-least-squares methods to regress 1960 total, infant, and disease-specific mortality rates in each of 114 U.S. metropolitan areas upon average ambient sulfate and particulate concentrations, and assorted demographic and socioeconomic variables. They concluded that the total mortality elasticity with respect to ambient sulfates was 0.05; with respect to ambient particulates, the estimated elasticity was 0.04. In a subsequent paper, Lave and Seskin [2] increased the sample size to 117, introduced additional air pollution, demographic, and socioeconomic variables, and tested specifications that were nonlinear in the original variables. The conclusions of their 1970 paper were unaltered, however. Finally, the detailed and carefully written Lave and Seskin [3] book evaluated 1969 as well as 1960 metropolitan area data, employed a,variety of cross-sectional, time-series, and pooled models, and yielded nearly the same conclusions.

Recently, Chappie and Lave [4] have reconfirmed the results reported in Lave and Seskin [1,2,3]. They reestimated earlier models with 1974 data for 104 U.S. metropolitan areas. The only new important result was the increased sulfate elasticity (now 0.13) and the reduced particulates elasticity (now 0.006). Additional general confirmations are provided by several authors who have been inspired to adopt the Lave-Seskin techniques and to apply them to different aggregate epidemiology data sets.[1/]

These confirmations have nevertheless failed to deter numerous skeptics who, as Chappie and Lave [4] note, question the aggregate nature and the poor quality of the data, and raise issues of omitted variable bias, incorrect functional forms, and the presence of simultaneity. The skeptics' general procedure has been to use the same or similar data and to find a model which provides air pollution coefficients undermining the Lave-Seskin conclusions. According to Freeman [6], Viren [7] proceeds by adding assorted explanatory variables to the Lave and Seskin [1,2] regressions until a combination is found that reduces the air pollution coefficients to statistical insignificance. Thibodeau, et al. [8] remove a set of "outliers" from the Lave and Seskin [1, 2] data, reestimate, and conclude that the magnitude of the air pollution-mortality association, though positive, is obscure. By positing a reciprocal relationship between mortality incidence and physicians per capita, Gerking and Schulze [9] obtain statistically significant negative air pollution coefficients. These skeptics agree with the latter authors who conclude that "... small changes in model specification appear to produce comparatively large changes in implications."[2] Neither "small" nor "large" is defined, however. Whatever these definitions, the obvious thrust of the skeptics' stance is that it "... may be unwarranted..."[3] to employ Lave-Seskin type data and methods to infer a consistent link between air pollution and mortality incidence. No hint is provided the reader about how difficult it would have to be to produce these exceptions before the skeptics could believe that inferences of a consistent link are warranted.

Lave and Seskin [3] and Chappie and Lave [4] have responded in kind *to* the issues the skeptics raise. They add and delete combinations of explanatory variables, partition their data sets, and experiment with different functional forms, equation systems, and estimators. The estimates for several alternative specifications are reported. For example, for each choice of a mortality dependent variable and its density function, and for each choice of an equation system and functional form, Chappie and Lave [4] have 53 measures which they or their critics consider to be plausible candidates for explaining variations in 1974 metropolitan area mortality incidences. Of the (2^{53}) possible inclusion-exclusion combinations of these candidate explanatory variables, 9 ordinary-least-squares single equation regressions are reported in which the unadjusted total mortality rate is the dependent variable. Another 12 similar regressions with the nontraumatic mortality rate as the dependent variable are also reported.[4] This dependent variable also appears in 2 single equation, generalized-least-squares regressions. Finally, four two-stage-least squares regressions that consider the possible simultaneity between physicians per capita and mortality incidence are reported. Chappie and Lave [4] do not exhaust the alternative regressions which might have been reported. Without even having to resort to simultaneous equation systems, nonlinear forms, or restrictions on coefficient signs and magnitudes, anyone who wishes to obtain a contradictory set of results can most likely find them among the (2^{53}) single equation linear model choices.

Neither the adherents nor the skeptics of the Lave-Seskin type methods have the means to close the debate; they are unable to provide convincing coverage of the range of plausible models. Both defenders and skeptics have been quick to point out that the source of the difficulty lies in the lack of a priori information with which to curb the numerous aspiring models. In Koopmans' [10] terms, the estimation exercise therefore becomes an hypothesis search rather than an hypothesis test. The tests being applied are not independent of the information embodied in the sample. One is looking for hypotheses which best fit the data without being able to specify the alternative hypotheses that might find greater or lesser support. According to whether one is a defender or a skeptic of Lave-Seskin type methods, multiple regression analysis is used to browse for significant or insignificant t-statistics.[5] With the highly aggregated data the Lave-Seskin methods use, there is little prior knowledge either to guide the search for the model that best fits the data or that best uncovers causal relationships.[6] In the absence of more structural information, an informed decision about which models to dismiss and which to accept requires a complete and communicable method of model searching and a compact format for reporting the results of the entire search.

In spite of the large number of papers using Lave-Seskin methods, only Smith [11] and Page and Fellner [12] supply charts that allow the reader to duplicate their work. The latter authors employ factor analysis and canonical correlation techniques. Each of these techniques follows a mechanical yet communicable statistical format to form scalar indices of groups of variables. Standard hypothesis tests are then employed to assess the associations among the groups. However, the mechanical nature of the statistical format makes it difficult to introduce restrictions provided by prior information; moreover, the relationship of the indices to any real phenomenon is frequently unclear.[7] Thus, while the Page and Fellner [12] procedures reduce the temptation to arrive at a "final" form for a model by repeated application of hypothesis tests to the same set of data, they do not obviate it.

Smith [11] chose to apply the Ramsey [13,14] tests for specification error to 32 models he regarded as "fairly representative" of those most often accepted as "final" in the Lave-Seskin type literature of the 1970's. His stated purpose was to ascertain whether the "final" models others had arrived at via the pretesting procedures common to the Lave-Seskin type literature were acceptable on the basis of the Ramsey [13,14] tests for incorrect functional form, omitted independent variables, simultaneity, and heteroscedasticity. His remarks contain a hint of surprise that most of the models performed quite creditably according to the tests.

Smith's [11] as well as Page and Fellner's [12] results are consistent with the Lave and Seskin [3] estimates of the association of air pollution and human mortality. However, it is unclear how to evaluate the alternative specifications with which Smith [11] and Page and Fellner [12] work. The prior beliefs of the researchers who originally specified the alternative "final" models are unknown. One therefore has to accept or to reject each separate model, with its unknown priors embedded.

SECTION 3

A BAYESIAN APPROACH TO SPECIFICATION ANALYSIS

With 53 or even as few as 6 or 7 explanatory variables available for use and with a number of alternative functional forms, mortality measures, and density functions for each mortality measure, aggregate epidemiology researchers have numerous ways to impose their prior beliefs about the impact of air pollution upon human mortality. Though Smith [11] and Page and Fellner [12] are mindful of the role that priors have played in reported estimates, they are incapable of assessing the range of priors other investigators might have employed. Leamer's [5] SEARCH method (Seeking Extreme and Average Regression Coefficient Hypotheses) provides this assessment and portrays it with compact summary statistics. The SEARCH method is fully described elsewhere.[8/] In this section, we try only to convey enough of the flavor of SEARCH to allow the reader to form judgments about the informativeness of the inferences that our subsequent air pollution aggregate epidemiology estimates furnish.

In accordance with Leaner and Leonard [15], consider the following simple linear regression:

$$Y_t = \beta x_t + \gamma_1 z_{1t} + \gamma_2 z_{2t} + \mu_t, \qquad (1)$$

where Y_t is mortality incidence, t indexes a set of T observations, μ is an independently and identically distributed random normal error term with mean zero and unknown variance, σ, x is the focus variable, air pollution in our case, and the z_{it} (i = 1,2) are uncertain variables. Air pollution is a focus variable because it is the center of research concern and will therefore be included in every specification the investigator tests. He wants to know the sign and the magnitude of the unknown parameter, β. Doubtful variables are the z_{it} $(i=1,2)$, because the prior necessity of their presence in (1) is uncertain. These are the variables, in the absence of orthogonality, whose introduction confronts the researcher with a tradeoff between increasing the bias and reducing the variance of his estimates. In air pollution aggregate epidemiology, physicians per capita, percentage college-educated, and percentage over 65-years old are traditional examples of doubtful variables. Alternatively, if one has prior belief that percentage over 65-years old obviously belongs in any regression that purportedly explains mortality incidence, he would insist it be an additional focus variable.

Only 2 doubtful variables are included. It might therefore be feasible to estimate and report the four regression specifications resulting from decisions to include or exclude z_{1t} and/or z_{2t}. This would sharpen the reader's judgments about the robustness of the estimates;

however, the procedure does not allow the investigator to employ any prior restrictions he suspects might apply to the signs and magnitudes-of γ_1 and γ_2. Learner and Leonard [15] suggest that the investigator employ these priors and thereby enlarge the search. Specifically, they urge him to define a composite variable

$$w_t(\theta) = z_{1t} + \theta z_{2t}, \qquad (2)$$

where θ is a variable which reflects the investigator's priors. For each value of θ combined with the sample data, there is a unique regression specification, and therefore a different estimate, $\beta(\theta)$, for the air pollution coefficient. Because θ can be continuous over the real line, the set of alternative prior specifications of (1) need no longer be limited only to the four combinations of z_{1t} and z_{2t} based on their exclusion and/or inclusion. An obvious measure of specification uncertainty is then the difference in the extreme values of B(6). If the interval $[\beta min, \beta max]$ is small relative to the sampling uncertainty, or if decisions are insensitive to variations in the values of β over this interval, then the specification is relatively unambiguous. Sampling uncertainty is defined as 4 times the standard error of the coefficient for the focus variable, which corresponds to a 95 percent confidence interval. A large difference between βmin and β max relative to sampling uncertainty implies that specification uncertainty plays a large role in the overall uncertainty about the value of the focus coefficient, β. SEARCH evaluates specification uncertainty by searching out the extreme values of β that occur over all possible covariance matrices.

Learner [5] demonstrates that the set of all possible values of (γ_1,γ_2) generated by varying θ over the real line is an ellipse of constrained estimates. Each value of θ represents a different constraint, a different point on the ellipse, and thus a different tradeoff between bias and variance. However, the sample data may make some of these points appear to be extremely unlikely. For example, if γ_1 is the coefficient for percentage of the population 65 years old or more, a coefficient value which allowed 99.9 percent of the population to exceed this age would be unlikely to appeal to the user of aggregate epidemiology data. The set of points to be considered on the ellipse of constrained estimates can be bounded by defining an a percent ($0 \le \alpha \le 100$) sample confidence ellipse.[9/] This point set, which is defined by the intersection of the points in the interior of the locus of constrained estimates and the α percent sample confidence ellipse, represents all possible posterior pairs of (γ_1,γ_2) that can result from some prior distribution, given that only sample points lying in the α percent confidence ellipse are to be considered. For each confidence ellipse, minimum and maximum values of $\beta(\theta)$ can be generated; that is, one can show how different weights on the prior and the sample distributions cause specification uncertainty to vary. Figure 5.1 in Learner [5] is helpful in fixing these ideas.

Learner [5] provides a role for the precision of the prior distribution by constructing an "information contract curve" completely analogous to the Edgeworth-Bowley contract curve used in the economic theory of exchange for pairs of consumers. In this case, the sample data, which is analogous to one of the consumers, conveys its information via a likelihood function. The other consumer is a researcher who communicates his information by means of a prior distribution. Learner's [5] Figure 5.8 and his surrounding discussion show how this contract curve, which is the locus of tangencies between the conflicting information represented by prior ellipses and sample ellipses, is the locus of informationally efficient points that are jointly preferred by the prior and the data. As with any contract curve, one cannot discriminate among points on it unless more structure is introduced. Thus the distance along the curve can be used as another measure of specification uncertainty. Of course, since the curve is a locus of tangencies between prior and sample ellipses, one could restrict attention to an interval of the curve lying within some α percent confidence level of the data.

Learner [5] shows that more structure with which to choose among points on the contract curve is provided by a measure of the relative precisions of the prior and the sample distributions. For example, if the sample information has low relative variance, one would be more interested in that part of the contract curve closer to the least-squares point. Alternatively, if the prior information is more precise, points on the contract curve in the vicinity of the prior point would be preferred. The difficulty is that the precision of the prior distribution is frequently vague. Learner [5] proposes to overcome this difficulty with a procedure which identifies the standard deviation a normally distributed prior must have ("prior sigma") in order to be simultaneously on the contract curve and within a particular confidence ellipse. If, for example, the prior σ is very informative and one is dealing, say, with the 95 percent confidence ellipse, he may infer that the contract curve point is quite unlikely, since the prior would have had to be quite small in order to generate it.

The discussion has concentrated upon a single prior; however, Leaner [5] shows that the same procedures may be extended to linear combinations of focus variables. Thus, when different researchers have quite different combinations of priors, the specification uncertainty inherent in each of the combinations may be fully described.

SECTION 4

AN APPLICATION

After having made the explorations reviewed in Section II, Chappie and Lave [4] (pp. 365, 371) conclude that the combination of their and others' earlier results and the results from their 1974 data show that:

> "A strong, consistent, and statistically significant association between sulfates and mortality persists . . . . When related to the EPA's [16] estimate of abatement costs, these results support and strengthen the conclusions of Lave and Seskin [3] that stringent abatement of sulfur oxides and particulates would produce social benefits (based on health effects alone) greatly exceeding social costs. We regard the evidence for stringent abatement as compelling...."

Ordinary-least-squares regression number 2-5 in Chappie and Lave [4] embodies nearly all their maintained hypotheses about the relation between mortality and air pollution. Most important, its coefficients for the arithmetic mean air pollution measures are very similar to those in their other reported regressions and thus form the basis for the above-quoted conclusion. The fitted equation is:

$$\begin{aligned}\text{1974 TMR} = {} & \underset{(6.19)}{528.819} - \underset{(-0.57)}{3.043}(\text{MINS}) + \underset{(2.87)}{13.866}(\text{MEANS}) - \underset{(-2.34)}{1.774}(\text{MAXS}) \\ & + \underset{(0.73)}{1.234}(\text{MINP}) - \underset{(-1.19)}{1.008}(\text{MEANP}) + \underset{(1.25)}{0.191}(\text{MAXP}) + \underset{(16.27)}{58.417}(\%65+) \\ & + \underset{(3.21)}{2.412}(\%\text{NW}) - \underset{(-1.39)}{0.009318}(\text{MEDINCM}) + \underset{(1.05)}{18.813}(\text{LOGDENS}) \\ & - \underset{(-1.51)}{26.236}(\text{LOGPOPN}) - \underset{(-4.56)}{10.092}(\%{>}4\text{YRCOLL}),\end{aligned}$$

where t-statistics are in parentheses and the variables are defined in Table 1. With a sample of 104 metropolitan areas, the unadjusted R^2 for this expression is 0.888. Most of the coefficients are intuitively reasonable in both sign and magnitude, and several achieve high degrees of statistical significance.

We now apply Learner's [5] SEARCH procedure to this equation. Initially, we take MEANS to be the only focus variable. All other candidate explanatory variables are doubtful in the sense that we doubt that their coefficients differ from zero or from small numbers. The upper

TABLE 1

Definition of Variables*

1974 TMR	--	The unadjusted 1974 mortality rate per 100,000 population from all causes of death,
MINS	--	Smallest 24-hour sulfate reading in micrograms per cubic meter.
MEANS	--	Arithmetic mean of 24-hour sulfate readings in micrograms per cubic meter.
MAXS	--	Largest 24-hour sulfate reading in micrograms per cubic meter.
MINP	--	Smallest 24-hour total suspended particulate reading in micrograms per cubic meter.
MEANP	--	Arithmetic mean of 24-hour suspended particulate readings in micrograms per cubic meter.
MAXP	--	Largest 24-hour total suspended particulate reading in micrograms per cubic meter.
%65+	--	Percentage of area population at least 65 years old.
%NW	--	Percentage of nonwhites in area population.
MEDINCM	--	Median income of families in area in dollars.
LOGDENS	--	The logarithm of population density per square mile in the area.
LOGPOPN	--	The logarithm of total population in millions.
%>4YRCOLL	--	Percentage of area population at least 25 years old who are college graduates.

*All definitions, sources, and data are identical to those in Chappie and Lave [4].

and lower bounds of the estimated coefficient for MEANS are therefore the range of estimates that can be produced by examining all alternative weighted average combinations of the regressions formed by omitting or not omitting each of the doubtful variables. Thus, the regression results that Chappie and Lave [4] report, and all results they could have reported, must lie within these bounds.

The upper and lower bounds in Table 2 are the extreme values of the coefficients for MEANS with various levels of the data confidence ellipse referred to as "data confidence" in the table. These correspond to the extreme values within the ellipse of constrained estimates referred to in Section III. At the extreme left of the table are the least-squares estimates. The contract curve traces the value of the coefficient for MEANS along the locus of tangencies between the prior ellipses and the sample ellipses, given the researcher's choice of the length of the prior confidence intervals. The t-value of the coefficient for the pooling of the sample and the prior evaluated at a particular point on the contract curve is represented by the posterior-t. The value of the standard deviation of the prior distribution one would have to select to obtain the same point on the contract curve is given by the prior sigma. Specification uncertainty is simply the difference between the upper bound and the lower bound of the MEANS coefficient at the indicated levels of confidence in the data.

For all values of the data confidence in Table 2, the specification uncertainty exceeds the sampling uncertainty. At the prior (prior sigma = 0), the specification uncertainty exceeds the sampling uncertainty by more than a factor of 5 and the lower bound of the MEANS coefficient is -35.9. Moreover, except for a data confidence of 0.250 or less, the lower bound of the MEANS coefficient is negative throughout. Further, its extreme bounds increase dramatically as the data confidence interval increases, i.e., as the importance of the prior increases. Although the average of the upper and lower bound is more-or-less constant, the increased range can prove costly to the policymaker. If he considers the sample information to be far more precise than the prior information, the positive association between MEANS and mortality incidence is clearcut. However, if he does not hold this belief, these results fail to make a compelling case for a statistically significant association between arithmetic mean ambient sulfate concentrations and mortality incidence.

One might justifiably argue that some of the variables we have treated as doubtful while constructing Table 2 should really be focus variables. The addition of these new focus variables could cause the conclusions-drawn from Table 2 to be altered. We possess strong priors, for example, that increasing the number of people more than 65-years old, will, _ceteris paribus,_ increase mortality incidence. Most air pollution epidemiologists have strong prior beliefs that total suspended particulates, especially their "fine" particulate versions, have undesirable health impacts. Better education supposedly makes one a more efficient producer of health, while higher income increases the demand for health and also reduces the relative price of access to health-producing services. The influences these and other priors have upon the upper and lower bounds of the coefficients for

TABLE 2

Extreme Bounds and Uncertainty Measures for the Coefficient of Mean Sulfates (MEANS)

Standard error (Sample Sigma) of MEANS = 4.826

Data confidence	0.0	.250	.500	.750	.950	.990	1.000
Upper bound	13.9	27.8	30.0	32.3	36.0	38.7	70.0
Lower bound	13.9	.170	-1.97	-4.23	-7.71	-10.3	-35.9
Specification Uncertainty	–	27.970	31.97	36.53	43.71	49.0	105.9
Contract curve	13.9	8.11	8.13	8.23	8.48	8.73	20.2
Posterior t-value	2.87	3.76	3.88	4.02	4.26	4.46	13.7
Prior Sigma (σ_0)	∞	9.53	8.23	7.23	6.12	5.50	0.0

Sampling Uncertainty = 18.92

TABLE 3

Extreme Bounds on Mean Sulfates (MEANS)
When MEANS and Another Variable are Focus

Focus Combination		Data Confidence 0.0	.250	.500	.750	.950	1.00
MEANS	U	13.9	27.7	29.9	32.3	35.9	68.7
and MEANP	L	13.9	.180	-1.96	-4.23	-7.70	-35.7
MEANS	U	13.9	26.9	28.6	30.4	32.8	35.7
and %65+	L	13.9	.403	-1.54	-3.51	-6.25	-10.6
MEANS	U	13.9	27.8	30.0	32.3	36.0	70.0
and %NW	L	13.9	.178	-1.96	-4.22	-7.69	-35.7
MEANS	U	13.9	27.8	30.0	32.3	35.9	65.5
and MEDINCM	L	13.9	.227	-1.89	-4.12	-7.53	-31.2
MEANS	U	13.9	27.7	29.9	32.2	35.9	69.8
and LOGDENS	L	13.9	.360	-1.74	-3.95	-7.34	-33.2
MEANS	U	13.9	27.8	30.0	32.3	36.0	69.2
and LOGPOPN	L	13.9	.254	-1.86	-4.10	-7.52	-33.6
MEANS	U	13.9	27.3	29.3	31.5	34.7	51.7
and %>4YRCOLL	L	13.9	.187	-1.96	-4.23	-7.70	-28.5

U ≡ extreme upper bound.

L ≡ extreme lower bound.

MEANS at alternative levels of sample data confidence are presented in Table 3. Although the bounds on the MEANS coefficients are nearly always reduced by these priors, the reduction is very small with the sole exception of the lower bound for %65+. As in Table 2, specification uncertainties continue to exceed the MEANS sampling uncertainty of 18.92 for all levels of data confidence down to 0.250. Similarly, the lower bound of the MEANS coefficient for all priors remains negative down to this same data confidence. The lower bound becomes barely positive if one chooses to confine the data to a small confidence ellipse and to place a high variance on the prior. This exception will hardly be sufficient to convince most people that Chappie and Lave's [4] (p. 365) data rather than their priors generate "... a strong, consistent, and statistically significant association ..." between sulfates and mortality. Instead, the range of inferences about the impact of air pollution on mortality incidence remains wide under a variety of alternative models.

The high degree of specification uncertainty that the MEANS coefficient exhibits in Tables 2 and 3 could, of course, be due to the aggregate nature of the data employed. As earlier noted, some of the candidate explanatory variables, such as %65+, are obvious focus variables for any expression intended to explain mortality incidence. If the coefficients for these variables also display so much specification uncertainty that they are uninformative, then one might reasonably conclude that little can be learned from this aggregate epidemiology data set. Table 4 presents the extreme bounds for other focus variables, each in pair-wise combination with the focus variable, MEANS. With the sole exception of %65+, the range in the extreme bounds is great. Except for the extreme bounds of %65+ and %>4PRCOLL, the signs of the upper and lower bounds usually differ; however, even for these two variables, specification uncertainty exceeds sampling uncertainty at high levels of data confidence, i.e., broad confidence intervals. One might reasonably conclude that there are a large number of explanatory variables not included in this data set that would exhibit no more specification uncertainty than is exhibited by the variables in Table 4.

The preceding discussion is limited to the single equation specifications with mortality incidence as the sole endogenous variable that comprise nearly all the published work in aggregate air pollution epidemiology. Chappie and Lave [4] recognize that simultaneities may exist between mortality and certain of their explanatory variables such as %65+. At the same time they admit that their single equation results could be biased due to the omission of medical care and life-style variables. Perhaps because the plausible reciprocity between medical care and health status has been a frequent target for critics of earlier work,[107] they estimate by two-stage least squares a linear system in which physicians per capita and mortality incidence are endogenous. Because of the absence of data on alcohol consumption in two areas, they reduce the sample size from the 104 metropolitan areas of Table 2 to 102 areas. The structural expression that they estimate (their regression number 6-5) includes all the right-hand-side variables of Table 1, plus per capita smoking expenditures, per capita alcohol expenditures, and the endogenous variable, patient care physicians per 10,000 people. We fully concur in their

TABLE 4

Extreme Bounds on Other Variables When Mean Sulfates (MEANS) and Other Variables are Focus

Focus Combination		0.0	Data Confidence .250	.500	.750	.950	1.00
MEANP	U	-1.01	1.46	2.28	2.93	3.42	9.18
and MEANS	L	-1.01	-3.41	-4.17	-4.77	-5.22	-9.34
		Sampling Uncertainty of MEANP = 6.76					
%65	U	58.42	64.5	66.4	67.7	68.6	70.1
and MEANS	L	58.42	52.8	51.3	50.3	49.8	49.3
		Sampling Uncertainty of %65+ = 14.40					
%NW	U	2.41	3.98	4.46	4.84	5.12	7.29
and MEANS	L	2.41	.732	.170	-.286	-.634	-5.44
		Sampling Uncertainty of %NW = 3.01					
MEDINCM	U	-.0093	.0054	.0099	.0134	.0159	.0320
and MEANS	L	-.0093	-.0254	-.0308	-.0351	-.0385	-.0795
		Sampling Uncertainty of MEDINCM = .0268					
LOGDENS	U	18.81	23.7	28.5	32.2	34.9	54.6
and MEANS	L	18.81	-8.65	-14.3	-18.9	-22.4	-70.8
		Sampling Uncertainty of LOGDENS = 71.67					
LOGPOPN	U	-26.24	4.36	9.34	13.2	16.1	40.6
and MEANS	L	-26.24	-27.7	-33.2	-37.6	-40.9	-80.2
		Sampling Uncertainty of LOGPOPN = 69.50					
%>4YRCOLL	U	-10.09	-7.42	-6.79	-6.37	-6.13	-5.78
and MEANS	L	-10.09	-14.2	-15.6	-16.8	-17.6	-30.0
		Sampling Uncertainty of %4YRCOLL = 8.85					

U ≡ extreme upper bound.

L ≡ extreme lower bound.

conclusion (p. 365) that: "Neither the addition of a medical care variable ... nor the use of a simultaneous equation framework has much effect on the estimated air pollution coefficients." Table 5 reports the results for MEANS of an application of the SEARCH procedure to the Chappie and Lave [4] simultaneous system. Only MEANS and MEANP are focus variables. A comparison of this table with our Table 2 makes evident the basis of our agreement with them. Table 5 provides no reason whatsoever to alter the conclusions we earlier drew from Table 2.[11]

SECTION 5

CONCLUSIONS

In this paper, we have examined the role that the priors of investigators have played in aggregate air pollution epidemiology. We do not dispute the possibility of a significant relationship between urban air pollution and human mortality. Our sole purpose has been tc demonstrate the crucial role that priors play in attempts to infer this relationship from aggregate epidemiological data. Because we lack strong priors with which to choose among the candidate explanatory variables in Chappie and Lave [4], we conclude that their results are most likely dominated by their choice of "doubtful" variables, i.e., variables of doubtful significance. We have shown that this specification uncertainty causes their estimates to be fragile. Only if one considers their sample information to be very precise (that is, by examining a confidence interval less than .50) relative to the prior information, can he assert a significant positive association between air pollution and mortality. As the precision of the prior information increases relative to that of the sample information, the precision of the air pollution - mortality association declines and even includes negative values.

In spite of our results, we recognize that the painstaking and original work of Lave and his colleagues has focused a great deal of academic and regulatory interest on the existence and the size of an air pollution - human mortality relationship. What is now needed is a means of reducing the specification uncertainty associated with this relationship. To accomplish this, we suggest that further air pollution epidemiology research employ data on individuals, thus allowing the use of a limited set of stronger Bayesian priors.[12]

TABLE 5

Extreme Bounds and Uncertainty Measures for the Coefficient of Mean Sulfates (MEANS) in a Simultaneous Equation System Involving 2 Focus and 12 Doubtful Variables

Standard error (Sample Sigma) of MEANS = 4.7302

Data Confidence	0.0	.250	.500	.750	.950	.990	1.000
Upper Bound	14.5	29.3	31.5	33.8	37.4	40.1	71.1
Lower Bound	14.5	-.203	-2.34	-4.61	-8.09	-10.7	-37.7
Specification Uncertainty	–	29.5	33.8	38.4	45.5	50.8	108.8
Contract Curve	14.5	18.1	18.6	19.0	19.7	20.3	19.0
Posterior t Value	3.06	4.39	4.59	4.81	5.16	5.42	13.3
Prior Sigma (σ_o)	∞	8.02	7.27	6.61	5.79	5.28	0.0

Sampling Uncertainty of MEANS = 18.37

REFERENCES

1/ See, for example: Koshal and Koshal [17]; McDonald and Schwing [18]; Mendelsohn and Orcutt [19]; and Lipfert [20]. This does not exhaust the list.

2/ Gerking and Schulze [9], 229.

3/ Gerking and Schulze [9], 233.

4/ The nontraumatic mortality rate excludes ICDA Codes 000-999, that is, accidents, homicides, suicides, and other external causes.

5/ Lave and Seskin [2],(p. 286), are explicit about the hypothesis search technique they employed. They arrived at their "best" model in the following way:

> "Variables whose coefficients were greater than their standard error were retained and the others were eliminated, subject to two qualifications. Since interest centered on the air pollution variables, at least one was retained from each set.... Sometimes the retained air pollution variable still contributed nothing to the statistical significance of the regression. Such variables were eliminated, subject to the restriction that at least one air pollution variable was retained in the final equation."

As Atkinson and Crocker [21] note, this pre-test approach in which numerous variables are "tried on" and only the "final" or "best" results are reported fails to minimize mean squared error or other reasonable loss function criteria. The tradeoff the researcher makes between increases in bias due to incorrect priors and reductions in variance is unclear.

6/ Sargent [22] provides an interesting guide to searching for models that uncover causes as opposed to searching for models that best fit the data.

7/ As are all the Lave-Seskin type studies, the "raw" data used by Page and Fellner [12] are measures of central tendency taken over metropolitan areas. In effect, their techniques therefore form indices of indices.

8/ See Leamer [5], Cooley and LeRoy [23], and Leamer and Leonard [15]. The latter expository paper is quite thorough while also being very accessible. Leamer [24] presents a rather whimsical treatment.

Dhyrmes [25] gives a critical commentary on the overall philosophy of the method. Leamer [26] admits that the method retains some opportunity for the investigator to disguise his priors. Roberts [27] and Thiel [28] are early treatments of ideal criteria for reporting scientific results.

9/ In the simple bivariate case, an isoprobability ellipse is the contour in 2-space representing all combinations of the variables which have identical probability.

10/ See, for example, Gerking and Schulze [9], and Freeman [6].

11/ An application of SEARCH to the endogenous physicians per capita variable in the structural expression for mortality incidence revealed specification uncertainties of .627, 1.07, and 1.98 respectively at data confidence levels of .250, .990, and 1.000. The sampling uncertainty for the endogenous physicians per capita variable is .647. The simultaneous system thus appears to pay a price in increased variance for a questionable gain in reduced bias.

12 See Atkinson and Crocker [21] for a detailed discussion of our views on where potentially useful research directions in air pollution epidemiology might now lie.

BIBLIOGRAPHY

1. Lave, L.B., and E.P. Seskin, "Air Pollution and Human Health," Science 169 (Aug. 21, 1970), 723-733.

2. Lave, L.B., and E.P. Seskin, "An Analysis of the Association Between U.S. Mortality and Air Pollution," J. of the American Statistical Assoc. 68 (June 1973), 284-290.

3. Lave, L.B., and E.P. Seskin, Air Pollution and Human Health, Baltimore, Md.: The Johns Hopkins University Press (1977).

4. Chappie, M., and L.B. Lave, "The Health Effects of Air Pollution: A Reanalysis," J. of Urban Economics, 12 (November 1982), 346-376.

5. Leamer, E.E., Specification Searches: Ad Hoc Inference with Nonexperimental Data, New York: John Wiley & Sons (1978).

6. Freeman, A.M. III, Air and Water Pollution Control: A Benefit-Cost Assessment, New York: John Wiley & Sons (1982).

7. Viren, J.R., Cross-Sectional Estimates of Mortality Due to Fossil Fuel Pollutants: A Case for Spurious Association, A paper prepared for the U.S. Department of Energy, Washington, D.C. (1978) -- as cited in Freeman [6].

8. Thibodeau, L.A., R.B. Reed, Y.M. Bishop, and L.A. Kammerman, "Air Pollution and Human Health: A Review and Reanalysis." Environmental Health Perspectives 34 (1980), 165-183.

9. Gerking, S., and W.D. Schulze, "What Do We Know About Benefits of Reduced Mortality from Air Pollution Control?" The American Economic Review: Papers and Proceedings 71 (May 1981), 228-234.

10. Koopmans, T.J., "A Reply," The Review of Economics and Statistics 31 (Feb. 1949), 86-91.

11. Smith, V.K., The Economic Consequences of Air Pollution, Cambridge, Mass.: Ballinger Publishing Co. (1977).

12. Page, W.P., and W. Fellner, "Exploratory Techniques for the Determination of Potential Dose-Response Relationships Between Human Health and Air Pollution," J. of Environmental Economics and Management 5 (Dec. 1978), 376-389.

13. Ramsey, J.B., "Tests for Specification Errors in Classical Least Squares Regression Analysis," J. of the Royal Statistical Society B 31 (1969), 350-371.

14. Ramsey, J.B., "Classical Model Selection Through Specification Error Tests," in P.Zarembka, Ed., Frontiers in Econometrics, New York: Academic Press (1974), 123-151.

15. Leamer, E.E., and H. Leonard, "Reporting the Fragility of Regression Estimates," The Review of Economics and Statistics 64 (May 1983), 306-317.

16. U.S. Environmental Protection Agency, The Cost of Clean Air and Water, Washington, D.C.: USGPO, (August 1979).

17. Koshal, R.K., and M. Koshal, "Environments and Urban Mortality: An Econometric Approach," Environmental Pollution 4 (June 1973), 247-259.

18. McDonald, G.C., and R.C. Schwing," Instabilities of Regression Estimates Relating Air Pollution to Mortality," Technometrics 15 (Aug. 1973), 463-481.

19. Mendelsohn, K., and G. Orcutt, "An Empirical Analysis of Air Pollution Dose-Response Curves," J. of Environmental Economics and Management 6 (June 1979), 85-106.

20. Lipfert, F.W., "Differential Mortality and the Environment: The Challenge of Multicollinearity in Cross-Sectional Studies,: Energy Systems and Policy 3 (Dec. 1980), 367-400.

21. Atkinson, S.E., and T.D. Crocker, "On Scientific Inquiry into the Human Health Effects of Air Pollution: A Reply to Pearce, et al.," J. of the Air Pollution Control Assoc. 32 (Dec. 1982), 1121-1126.

22. Sargent, T.J., "Interpreting Economic Time Series," J. of Political Economy 89 (April 1981), 213-248.

23. Cooley, T.F., and S.F. LeRoy, "Identification and Estimation of Money Demand," The American Economic Review 71 (Dec. 1981), 825-844.

24. Leamer, E.E., "Let's Take the Con Out of Econometrics," The American Economic Review 73 (March 1983), 31-43.

25. Dhyrmes, P.J., "Comment," Econometric Reviews 1 (1982), 129-132.

26. Leamer, E.E., "Comment," Econometric Reviews 1 (1982), 141-145.

27. Roberts, H.V., "Reporting of Bayesian Studies," in S.E. Feinberg and A. Zellner, eds., Studies in Bayesian Econometrics and Statistics, Vol II, Amsterdam: North-Holland Publishing Co. (1977), 155-174.

28. Thiel, H., Economic Forecasts and Policy, Amsterdam: North-Holland Publishing Co. (1961).