

# People are not Passive Acceptors of Threats to Health: Endogeneity and its Consequences

JOHN BRISCOE,\* JOHN AKIN\*\* AND DAVID GUILKEY\*\*

Briscoe J (The World Bank, Washington DC 20433, USA) Akin J and Guilkey D. People are not passive acceptors of threats to health. Endogeneity and its consequences. *International Journal of Epidemiology* 1990, 19: 147–153. The effect of behaviour on health is a major area of contemporary epidemiological enquiry. Most epidemiological studies of the effect of behaviour on health assume that the levels of the behaviour-related variables are determined by factors other than those under study. However, in many instances, obvious examples are breastfeeding and smoking, not only do behaviours affect health but, conversely, individuals take into account their (observable and non-observable) health conditions when making behavioural decisions. In models which allow for the joint determination of health and behaviour, both health and behavioural variables are 'endogenous', that is, determined by forces acting within the model. Through some simple didactic examples it is shown that estimates of the effect of behaviour on health are biased if endogeneity is ignored. Review of the small empirical literature on this subject shows perverse results, such as a negative relationship between the use of prenatal care and infant mortality, when endogeneity is ignored. Standard procedures for taking account of the effects of endogeneity are described briefly.

## AN ILLUSTRATIVE MODEL OF HEALTH AND HEALTH-RELATED BEHAVIOUR

Consider an epidemiological study of the effect of breastfeeding on health and assume, for the sake of illustration, that the following relationships are true: *Biology*: A baby is less likely to become ill if it is breast-fed. *Behaviour*: The culture is one in which breastfeeding of young children is common. Mothers who have sickly children are more likely to stop breastfeeding their babies.

Further, assume that these relationships can be represented by the following equations:

$$\begin{aligned} S &= \alpha_1 + \beta_1 * F + \gamma_1 * L + \epsilon_1 \\ F &= \alpha_2 + \beta_2 * S + \gamma_2 * M + \epsilon_2 \end{aligned} \quad (1)$$

where, S is the probability of sickness in a particular period, F indicates whether the child is breastfed, L is the level of environmental sanitation and M indicates whether the mother is educated.

There are many epidemiological studies of the effect of breastfeeding on health (for a recent review of the studies on breastfeeding and diarrhoea, see Feachem and Koblinsky).<sup>1</sup> Many of these studies pay careful attention to the standard sources of bias considered by

epidemiologists (such as selection bias, misclassification and confounding). Virtually all studies ignore the fact that the level of breastfeeding not only affects health, but that the decision to breastfeed is also a response to the perceived health (or susceptibility) of the child.

In this paper we posit that studies of the effect of behaviours (including breastfeeding, prenatal care, smoking, exercise and many others) on health ignore an important and pervasive source of potential bias which arises because these behaviours not only affect health, but are also determined, in part, by (observable and non-observable) health attributes of the individual. Specifically we will show that:

- ignoring this endogeneity (in which behaviour affects health and health affects behaviour) leads to biased measures of the effect of behaviour on health;
- these biases are large in realistic conditions;
- in the few empirical studies which address this issue, the direction of the biases is as predicted, and the policy implications profound;
- there are standard statistical techniques for correcting for the effect of endogeneity.

For illustrative purposes, assume that the effect of breastfeeding (F) on sickness (S) is:

\*The World Bank, Washington DC 20433, USA.

\*\*Department of Economics and Carolina Population Center, University of North Carolina, Chapel Hill, NC 27514, USA.



$$S = 0.2 - 0.3 * F - 0.2 * L$$

and that the effect of sickness (S) on the probability of breastfeeding (F) is:

$$F = 0.8 - 1.0 * S - 0.1 * M$$

Following a didactic scheme developed by Wonnacott and Wonnacott,<sup>2</sup> let us simulate what happens when a statistician estimates the effect of breastfeeding on sickness by observing data on F and S.

In this highly simplified model, there is only one combination of values for breastfeeding and sickness that satisfies both equations simultaneously (for given values of L and M). For a given level of sanitation (L), allowing education to shift from 0 ('no education') to one ('educated') traces out the sickness line. (A similar graph could be used to show how shifts in L trace out the breastfeeding line.) The implication is that all coefficients in such a model can be estimated (in technical terms, both equations are identified).

In practice, however, relationships are never so perfectly defined. Rather, as in equations (1), there are error terms ( $\epsilon$ s) in each equation. When these errors are introduced, the figure becomes a little more complicated, with the relationships becoming 'bands' (as in

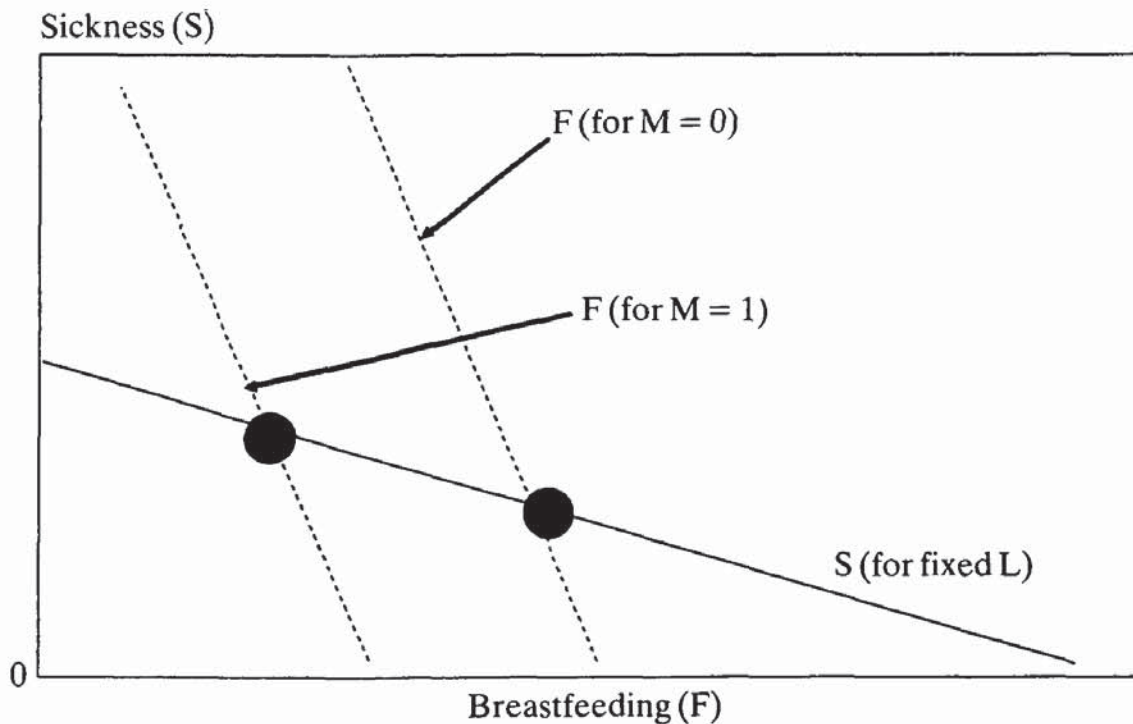
Figure 2) rather than lines. When data are collected in a world described by Figure 2, all observations obeying the biological relationship (those in the band around the 'S' curve) cannot be observed, because of the constraints imposed by the behavioural relationship. Rather, the eligible observations are only those in the two 'overlapping areas' illustrated in Figures 3 and 4.

#### THE BIASES WHEN ENDOGENEITY IS IGNORED

If one 'eyeballs' a regression line through the eligible data points on Figure 4, it is evident that the estimated effect of breastfeeding on health will be substantially different from the true relationship. This is so because, as a result of the constraints imposed by the behavioural relationships:

- when the explanatory variable (F) is large, the error term ( $\epsilon_1$ ) is likely to be negative;
- when the explanatory variable is small, the error term is likely to be positive.

In technical terms, the problem is that when an explanatory variable is correlated with the error term



Where: M indicates level of mother's education and L indicates level of environmental sanitation.

FIGURE 1 The true biological relationship.

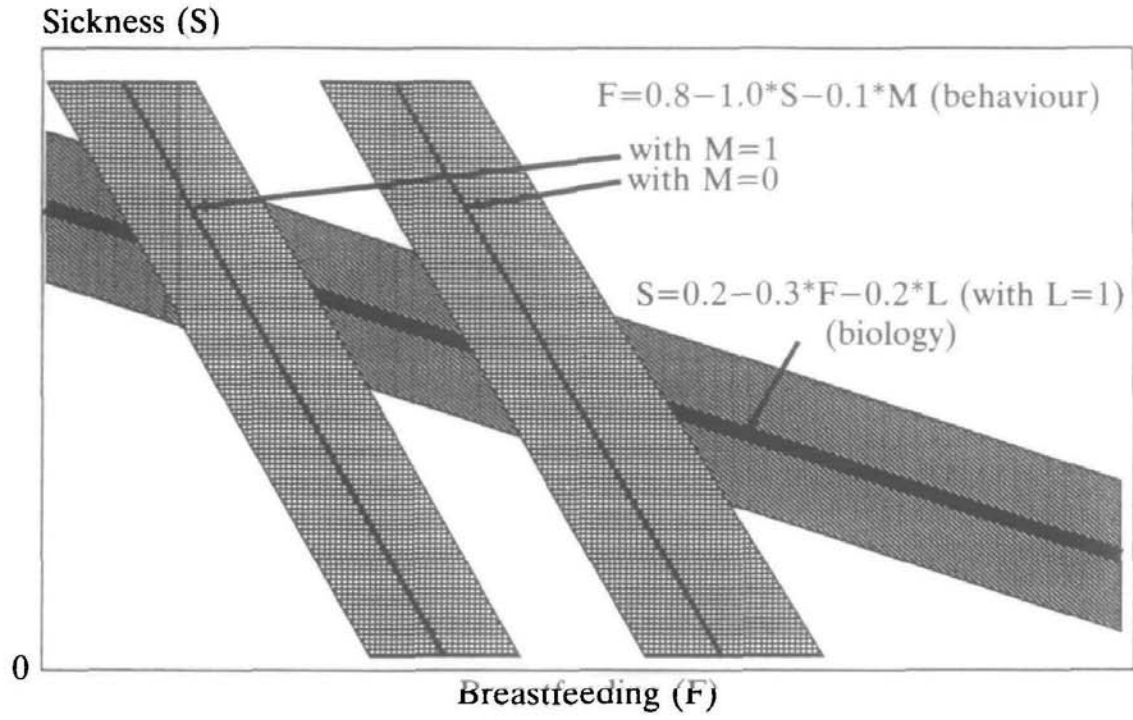


FIGURE 2 *Biology and behaviour—observations.*

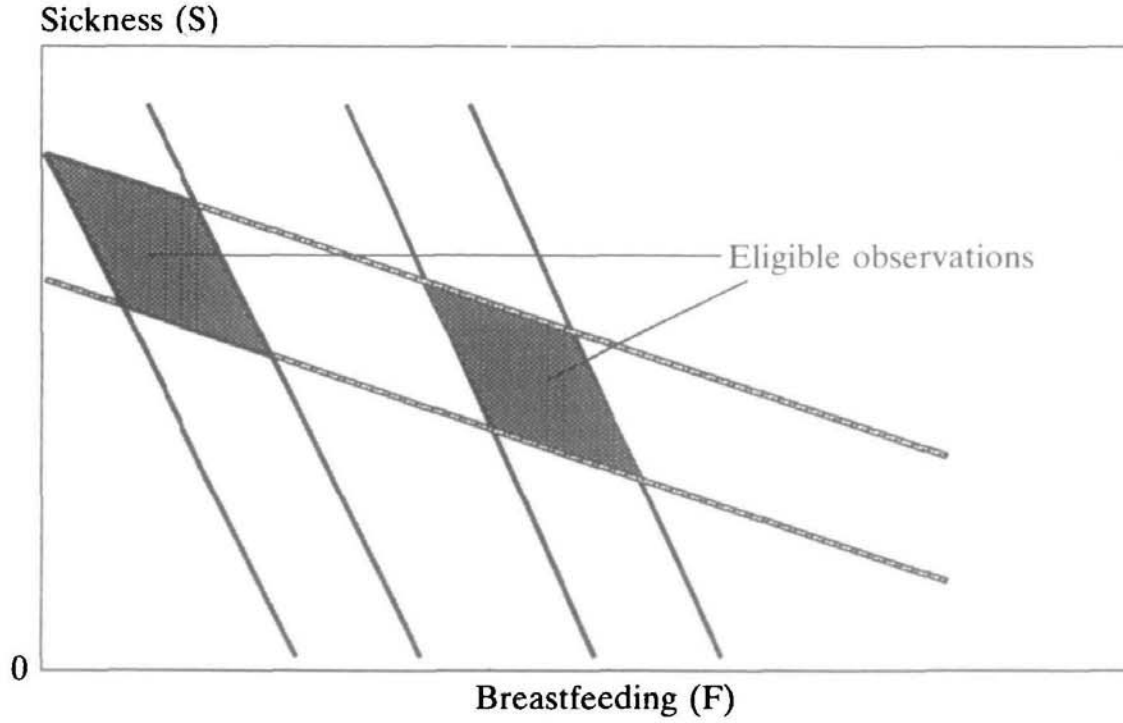


FIGURE 3 *The eligible observations.*

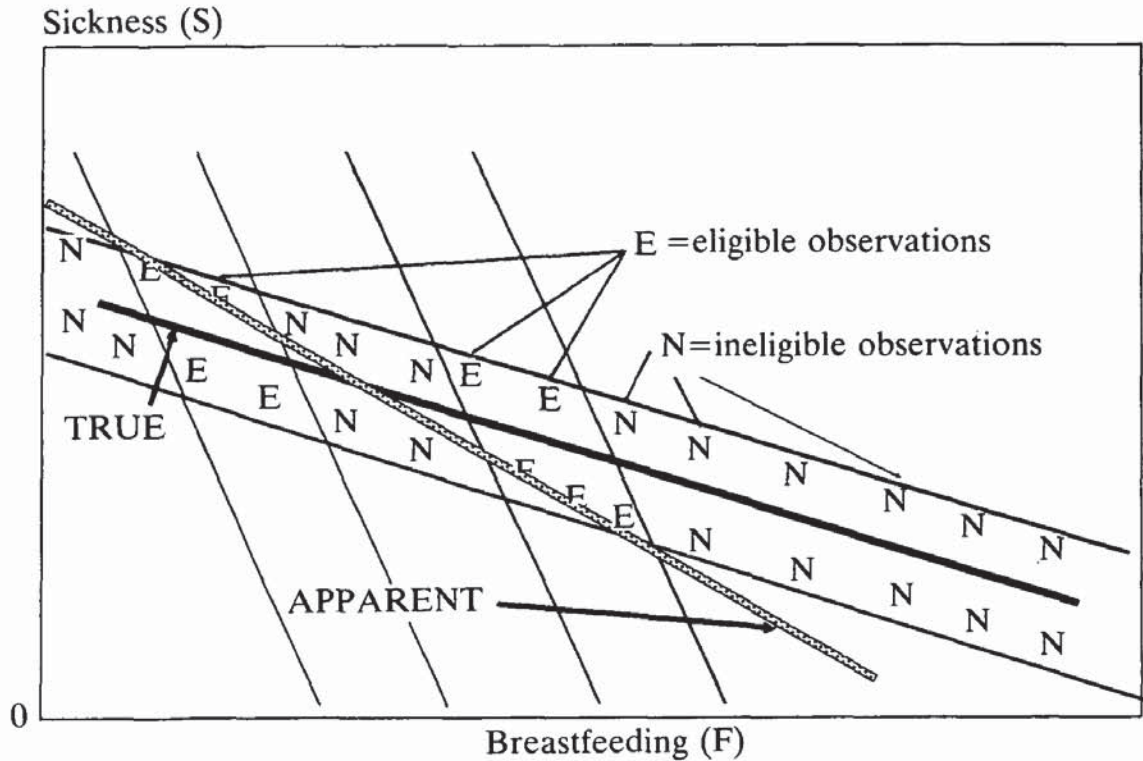


FIGURE 4 Observations and true and apparent relationships.

this constitutes a violation of a basic assumption of regression theory. As shown in elementary statistics texts, this systematically leads to biased estimates of the effect of  $F$  on  $S$ . Intuitively this is so because regression procedures give as little credit as possible to the error ( $\epsilon$ ), and as much credit as possible to the explanatory variable ( $F$ ). Since  $E$  and  $F$  are correlated, some of the effect of the error is wrongly attributed to the explanatory variable ( $F$ ).

It is evident from Figure 4 that the direction of the bias depends on the relative slopes of the biological and behavioural relationships. By repeating Figure 4 for a variety of values of  $\beta_1$  and  $\beta_2$ , the effect of endogeneity can be derived for all combinations of  $\beta_1$  and  $\beta_2$ . The results are presented in Figure 5.

For the illustrative example of the relationship between breastfeeding and ill health, both  $\beta_1$  and  $\beta_2$  are negative and  $\beta_1\beta_2 < 1$ . Figure 5 shows that in this case  $\beta_{1, \text{estimate}} < \beta_{1, \text{true}}$ . That is, if endogeneity is ignored, the effect is to overestimate the protective effect of breastfeeding.

#### DO EMPIRICAL RESULTS ACCORD WITH THEORETICAL EXPECTATIONS?

To our knowledge there are only two empirical inves-

tigations in the health literature of the distortionary effects of ignoring endogeneity.

#### Example 1: The Effect of Prenatal Care on Infant Health in the US

Schultz<sup>3</sup> has investigated the effect of prenatal care on infant health in the US. His analysis explicitly took into account the probability that mothers-to-be make conscious decisions on the level of prenatal care to be sought, and that this decision is affected by the 'health endowment' of the foetus (which is unobservable to a researcher). Specifically, when the woman has either had a problematic prior pregnancy or difficulties with the current pregnancy, she is more likely to seek prenatal care. Accordingly, the hypothesized biological and behavioural relationships are as illustrated in Figure 6.

In terms of the general scheme of biases,  $\beta_1$  is negative and  $\beta_2$  is positive. As shown in Figure 5, under such circumstance  $\beta_{1, \text{estimate}} > \beta_{1, \text{true}}$ . That is, if endogeneity is not taken into account, either (a) the beneficial effect of prenatal care would be underestimated, or (b) prenatal care would even appear to adversely affect child health (ie  $\beta_{1, \text{estimate}}$  would be positive when  $\beta_{1, \text{true}}$  is negative).

In his investigation of the determinants of infant



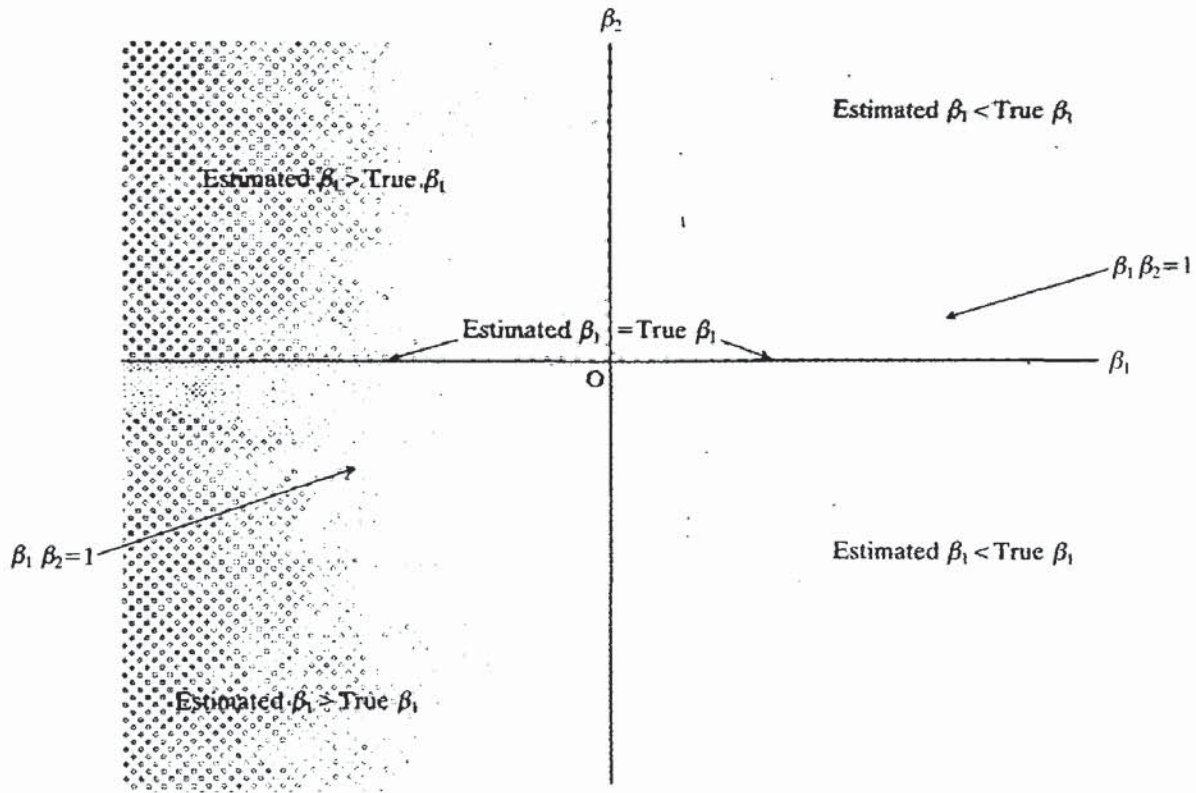


FIGURE 5 Bias in  $\beta_1$  due to endogeneity.

mortality in the US, Schultz<sup>3</sup> showed that if prenatal care was treated as an exogenous variable then the effect of prenatal care appeared to be to *increase* infant mortality. When prenatal care was treated as an endogenous variable, however, prenatal care was found to *reduce* infant mortality significantly. The direction of bias resulting when endogeneity was ignored, then, was as predicted.

*Example 2: The Effect of Medical Care on Respiratory Illness in the US*

An Environmental Protection Agency study<sup>4,5</sup> analysed the determinants of respiratory illness in the U.S. In an initial analysis, the availability of physicians appeared to have no effect on respiratory illness, even after confounders are taken into account. In a more realistic formulation (illustrated in Figure 7), allow-

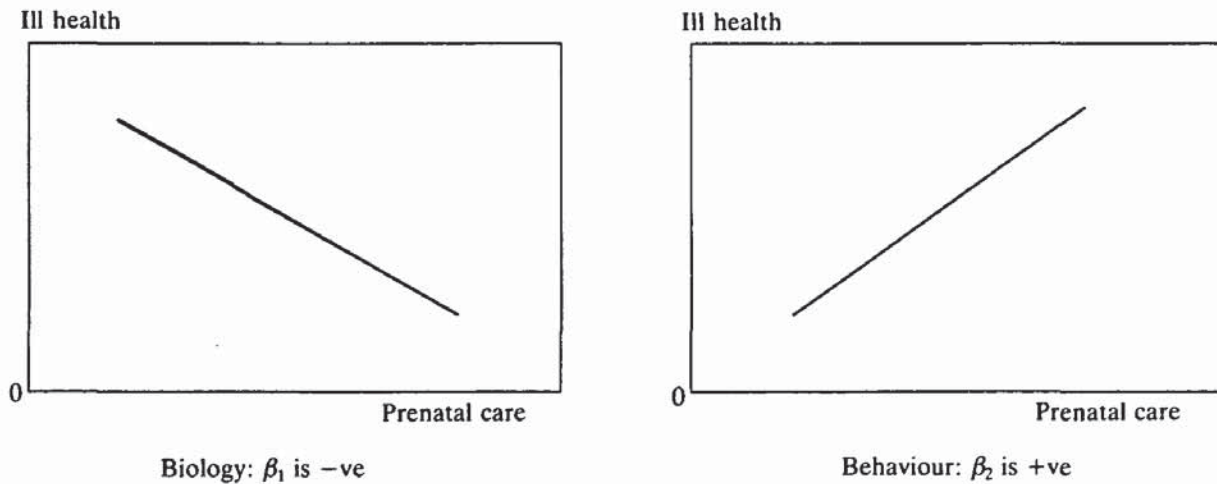


FIGURE 6 The effect of prenatal care on child health.

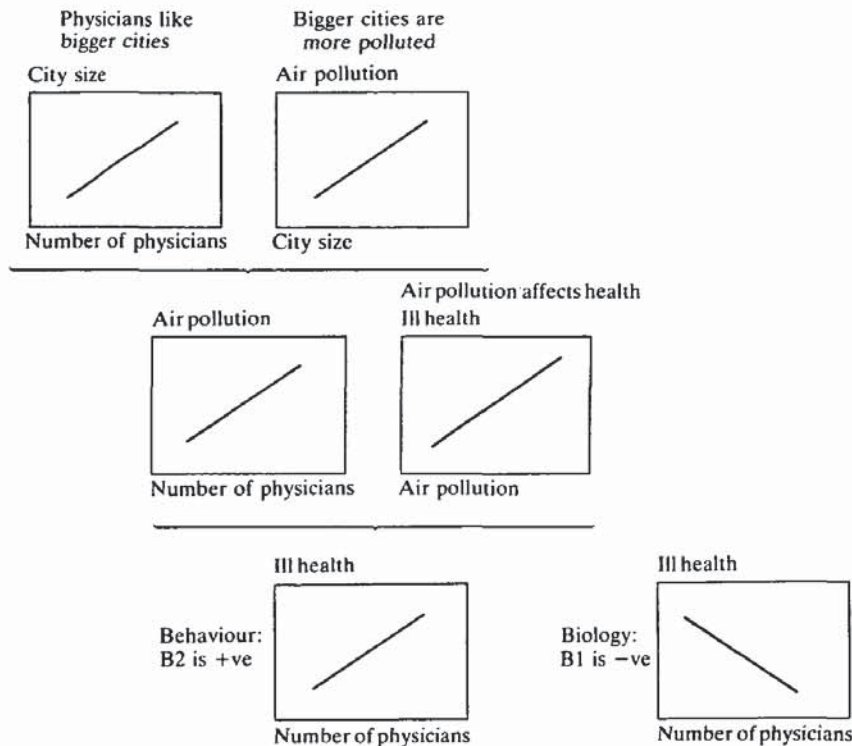


FIGURE 7 The effect of medical care on respiratory illness.

ance was made for the fact that physicians choose to live in areas where air pollution levels are relatively high. When this behaviour is taken into account (in technical terms, when the level of medical care was treated as endogenous), the effect of medical care was to significantly *reduce* the levels of respiratory illness. Again the bias found in this empirical analysis is consistent with the bias predicted in Figure 5—with  $\beta_1$  negative and  $\beta_2$  positive,  $\beta_{1, \text{estimate}} > \beta_{1, \text{true}}$ .

#### HOW DOES ONE CORRECT FOR ENDOGENEITY?

Although the epidemiological literature on endogeneity is sparse (as the above 'review' suggests), the problem is one which has long been recognized by economists and one for which econometricians have devised standard procedures (eg reference 2). From the earlier discussion, it should be apparent that the central task is to purge the model of the correlation between the explanatory variable and the error term. The standard procedure for doing this is to substitute the explanatory variable with a proxy (known as an 'instrumental variable') which is correlated with the explanatory variable but which is not correlated with the error term. When such an instrumental variable is used in place of the original explanatory variable, then there is no longer any correlation between the explanatory variable (now the instrumental variable) and the error term, and thus the estimate of effect is unbiased.<sup>2</sup>

In econometrics this procedure is widely understood and routinely applied. In describing the procedure to epidemiologists and others unfamiliar with the technique an important task is to convince them that it is correct and necessary to substitute the known, true value of a variable (often obtained at great effort) with an estimate (the instrument) which is necessarily different from the true value of the variable. A major purpose of this paper is to help explain this seemingly-irrational procedure!

#### CONCLUSIONS

Most studies of the effect of behaviour on health assume that the levels of the behaviour-related variables are determined by factors other than those under study, that is that they are 'exogenous'. In fact, however, there is widespread evidence that people are not passive acceptors of risks to their health, but that they adjust their behaviour because of their perceptions of their health and the risks to their health. In technical terms, these health-related behaviours are determined by forces acting within the model under study and are therefore 'endogenous' to the model. In this paper some simple methods for tracing the effects of this endogeneity are presented. It is shown that the estimated effects of behaviour on health are in general erroneous when endogeneity is not taken into account. In the case of the few published epidemiological stud-

ies which take account of this effect, it is shown that the biases are as predicted by the theoretical models. Finally, statistical procedures for obtaining correct estimates of the effects of behaviour on health are described.

## REFERENCES

- <sup>1</sup> Feachem R G, Koblinsky M A. Interventions for the control of diarrhoeal diseases among young children: promotion of breast-feeding'. *Bull WHO* 1984; **62**(2): 271-91.
- <sup>2</sup> Wonnacutt T H, Wonnacutt R J. *Econometrics* New York, Wiley, 1970.
- <sup>3</sup> Schultz T P. Studying the impact of household economic and community variables on child mortality. *Pop Dev Rev* 1984; **10**: 215-36.
- <sup>4</sup> Crocker T D, Schulze W, Ben-David S, Kneese A V. Experiments with Economics of Air Pollution epidemiology, volume 1 of Methods Development for Assessing Air Pollution Benefits, *Report EPA-600/5-79-00 1a*. Washington DC, US Environmental Protection Agency, 1979.
- <sup>5</sup> Kneese A V. *Measuring the Benefits of Clean Air and Water*. Washington DC. Resources for the Future, 1986.

*(Revised version received April 1989)*