

Distribution: General

Original: English

DIARRHOEAL DISEASE CONTROL PROGRAMME

**A SOURCE OF BIAS IN STUDIES OF BI-DIRECTIONAL
ASSOCIATIONS: HYPOTHETICAL EXAMPLE OF
MALNUTRITION AND DIARRHOEA**

Simon N. Cousens¹

¹Maternal and Child Epidemiology Unit
Department of Epidemiology and Population Sciences
London School of Hygiene and Tropical Medicine
Keppel Street, London WC1E 7HT, United Kingdom

CONTENTS

	<u>Page</u>
SUMMARY	2
INTRODUCTION	3
THE ASSOCIATION BETWEEN DIARRHOEA AND MALNUTRITION	3
THE REVERSE CAUSALITY PROBLEM	4
THE MODEL	4
DISCUSSION	11
ACKNOWLEDGEMENTS	12
REFERENCES	13

SUMMARY

A potential source of bias in studies of the association between two variables A and B, which may be linked by causal pathways in both directions (A may increase the risk of B and B may increase the risk of A), is investigated. A simple probability model is developed with particular reference to the association between malnutrition and diarrhoea in young children in developing countries. It has been hypothesized that diarrhoea (A) contributes to poor nutritional status (B), which in turn leads to increased diarrhoea. Studies of the effect of poor nutritional status on diarrhoea incidence have typically compared diarrhoea rates in under-nourished children with those in the well-nourished. The model demonstrates that, in the presence of heterogeneous diarrhoea rates, such analyses may produce biased results, leading to an overestimate of the magnitude of any causal link from malnutrition (B) to diarrhoea (A). The results of a computer simulation to assess the potential magnitude of such bias are presented.

INTRODUCTION

In order to study whether a factor A increases the risk of developing an outcome B, epidemiologists frequently employ the following strategy:

1. Identify a group of individuals with A and a group of individuals without A;
2. Follow both groups of individuals over a period of time in order to observe who develops B;
3. Compare the rate at which B occurs in individuals with A with the rate at which it occurs in individuals without A;
4. If a higher rate of B is observed to occur in individuals with A (after taking account of confounding factors), then infer that A "causes" B.

This strategy is known as a cohort or follow-up study and one of its perceived advantages is that the temporal sequence between exposure (A) and outcome (B) "can be more clearly established" (Hennekens and Buring, 1987). The inference that "A causes B" may, however, be problematic. Rose and Shipley (1980), discussing an apparent link between low plasma cholesterol levels and increased risk of cancer, have pointed out that such an observation may arise if early, pre-clinical disease leads to low cholesterol levels - i.e., unobserved B causes A which is then followed by the observation of B. One infers that A causes B when, in fact, the reverse is true.

This difficulty in distinguishing between cause and effect is particularly acute when attempting to determine the direction and magnitude of a causal link between A and B when the association may be bi-directional. In many circumstances, A causing B will tend to confound the examination of whether B causes A, an effect which is frequently overlooked. In this paper a simple probability model is developed to demonstrate this confounding effect. The association between diarrhoeal diseases and malnutrition among young children in developing countries is used as a paradigm. The model shows how, in a hypothetical situation in which diarrhoea (A) causes malnutrition (B) but malnutrition (B) does not cause diarrhoea (A), and in which diarrhoea rates vary among children, standard epidemiological techniques can lead to the "observation" that malnutrition (B) causes diarrhoea (A).

THE ASSOCIATION BETWEEN DIARRHOEA AND MALNUTRITION

Among young children in developing countries, the association between poor nutritional status, as measured by anthropometry, and diarrhoea is well established (Tomkins and Watson, 1989). The precise nature of this association, however, remains less clear. It has been suggested that the relationship is bi-directional, episodes of diarrhoea having a negative impact on a child's nutritional status while children with poor nutritional status are at increased risk of diarrhoea (Scrimshaw et al., 1968). Biologically plausible mechanisms through which diarrhoea might lead to malnutrition and poor nutritional status might lead to increased rates of diarrhoea have been suggested (Chen, 1983; Chandra, 1986; Sirisinha et al., 1975).

A number of epidemiological studies have attempted to investigate the association between diarrhoea and malnutrition. (For a recent bibliography, see Tomkins and Watson, 1989.) Typically, these studies have followed a group of children over a period of time, monitoring morbidity and taking regular anthropometric measurements. Then, in order to examine whether diarrhoea causes malnutrition, growth velocity over a period of time is compared with diarrhoea morbidity over the same time period. Authors observing an association between increased diarrhoea and decreased weight velocity have interpreted their results as evidence that diarrhoea causes malnutrition. In order to determine whether malnutrition causes diarrhoea, children are classified as malnourished or not according to their anthropometric status at the beginning of the follow-up period. The two groups of

children are then compared with regard to their diarrhoea experience over the subsequent time interval. Higher rates of diarrhoea in the malnourished are interpreted as evidence that malnutrition causes diarrhoea.

THE REVERSE CAUSALITY PROBLEM

A number of these studies may be criticized on a variety of grounds: failure to take account of confounding variables; the division of the follow-up period into several sub-intervals so that each individual child contributes several "independent" observations to the analysis; and failure to consider the potentially confounding effect of a causal link operating in the opposite direction to the one under consideration. It is with the last of these problems that this paper is concerned. The general question of interest is:

"Can the fact that A causes B result in an association which makes it appear that B causes A (or overestimates the strength of the causal link between B and A)?"

To avoid becoming lost in abstractions, we consider the more specific question:

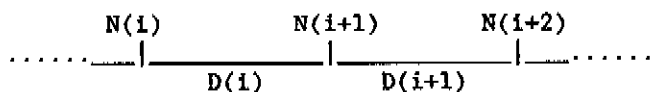
"Can the fact that diarrhoea (A) causes malnutrition (B) in a particular setting result in an association which makes it appear that malnutrition (B) causes diarrhoea (A) when it does not?"

In order to investigate this question a simple probability model is developed.

THE MODEL

In constructing a model we follow the standard approach to the design and analysis of studies investigating whether malnutrition causes diarrhoea. In such studies children are weighed and measured at the start of a period of observation. Each child is then classified as "well-nourished" or "malnourished" according to whether her/his anthropometric status falls above or below a cut-off point. The children are then followed for a period of time and a record of diarrhoea morbidity is kept. To assess whether malnutrition causes diarrhoea, the rate of diarrhoea in the malnourished is compared with that in the well-nourished. The children may be reweighed and remeasured periodically throughout the study, each reweighing constituting the start of a new observation period.

Consider an individual child participating in such a study of the association between malnutrition and diarrhoea. The child's progress through the study may be represented in the following way:



where $N(i)$ represents the child's anthropometric status at the beginning of the i th observation period, and $D(i)$ indicates whether or not the child suffered an episode of diarrhoea during the i th interval. It is assumed for the sake of simplicity that a child does not suffer more than one episode of diarrhoea in any given observation period.

Let $N(i) = 1$ if the child is malnourished at the start of the i th interval,

$N(i) = 0$ otherwise,

$D(i) = 1$ if the child suffers diarrhoea during the i th interval,

$D(i) = 0$ otherwise,

and assume the following:

1. that the child's anthropometric status (B) at the beginning of the i th interval does not affect the probability that s/he suffers an episode of diarrhoea (A) during the i th interval - i.e., B does not cause A;
2. that the probability that the child suffers an episode of diarrhoea in any interval is p , independent of the occurrence or not of diarrhoea in preceding intervals -

$$\text{i.e. } P\{D(i) = 1 | p, n(i-1), d(i-1), d(i-2), \dots\} = p;$$

3. that the child's nutritional status at the end of the i th interval (beginning of the $i+1$ th interval, $N(i+1)$) depends only upon her/his nutritional status at the beginning of the i th interval ($N(i)$) and upon whether or not the child suffers from diarrhoea during the i th interval ($D(i)$) - i.e., A "causes" B.

Then let $P\{N(i+1)=1 | N(i)=1, D(i)=1\} = t_{11}$.

$$P\{N(i+1)=1 | N(i)=1, D(i)=0\} = t_{10}$$

$$P\{N(i+1)=1 | N(i)=0, D(i)=1\} = t_{01}$$

$$P\{N(i+1)=1 | N(i)=0, D(i)=0\} = t_{00}$$

A situation in which malnutrition at time i increases the risk of malnutrition at time $i+1$, and diarrhoea during the i th interval increases the risk of malnutrition at time $i+1$, may be expressed mathematically as follows:

$$t_{11} > t_{10} \quad \text{and} \quad t_{01} > t_{00} \quad (1)$$

$$\text{and } t_{11} > t_{01} \quad \text{and} \quad t_{10} > t_{00} \quad (2)$$

The first pair of inequalities (1) expresses the fact that diarrhoea predisposes to malnutrition. The second pair indicates that a child who is already malnourished is more likely to be malnourished at the next observation point than a previously well-nourished child.

In summary, this model may be described as follows: diarrhoea predisposes to poor nutritional (anthropometric) status, but poor nutritional (anthropometric) status does not predispose to diarrhoea.

Under the above model, the child's nutritional status over the course of the study ($N(1), N(2), \dots$) may be considered as a Markov process, since the distribution of $N(i+1)$ given $N(i)$ and p is independent of $N(i-1), N(i-2), \dots$. The transition matrix of this process may be represented as:

		N(i+1)	
		0	1
N(i)	0	(1-U)	U
	1	V	(1-V)

where $U = pt_{01} + (1-p)t_{00}$.

and $V = p(1-t_{11}) + (1-p)(1-t_{10})$.

Here, U is the probability that during any interval the status of the child changes from well-nourished to malnourished. V is the probability that during any interval the reverse occurs, the child changing from being malnourished to well-nourished. Note that the pair of inequalities (1) implies that U is a monotonically increasing function of p and that V is a monotonically decreasing function of p . That is, as p increases U increases and V decreases. Then, by a well-known result (Cox and Miller, 1972), in the long run, the probability that at the beginning of any interval the child is malnourished is given by:

$$R(p) = \frac{U}{U+V}$$

At an intuitive level this result says that the long-run probability that a child is malnourished at any point in time is proportional to the probability that in any interval s/he becomes malnourished. Similarly, the probability that at any time s/he is well-nourished is proportional to the probability that s/he becomes well-nourished. The denominator (U+V) ensures that these two probabilities add up to 1 since the child is either well-nourished or malnourished.

Since U is a monotonically increasing function of p and V is a monotonically decreasing function of p, R(p) is a monotonically increasing function of p. In other words, as the child's risk/incidence of diarrhoea (p) increases, so does the probability that at any given moment s/he is malnourished. This is what we would expect intuitively given that, in defining the model, we specified that an episode of diarrhoea predisposes to malnutrition.

Now consider a population of children among whom the incidence rate of diarrhoea varies - i.e., different children have different values of p. Let the probability of diarrhoea occurring in an interval be denoted by the random variable P and let f(P) represent the probability density function (pdf) of P. In studies investigating whether malnutrition predisposes to diarrhoea, the risk ratio of diarrhoea associated with prior malnutrition is frequently estimated by comparing the risk of diarrhoea among those who are malnourished at the beginning of the observation period with the risk of diarrhoea among those who are not malnourished. The expectation of this estimator is:

$$E[RR] = \frac{P(D(i)=1 | N(i)=1)}{P(D(i)=1 | N(i)=0)}$$

$$\text{Now, } P(D(i)=1 | N(i)=1) = \frac{P(D(i)=1, N(i)=1)}{P(N(i)=1)}$$

$$\begin{aligned} \text{where } P(D(i)=1, N(i)=1) &= E[D(i) \times N(i)] \\ &= E_p[E[D \times N | P]] \\ &= \int pR(p)f(p)dp \\ &= E[P \times R(P)], \end{aligned}$$

$$\text{and } P(N(i)=1) = E[R(P)].$$

$$\text{Similarly, } P(D(i)=1 | N(i)=0) = \frac{E[P \times (1-R(P))]}{1-E[R(P)]}$$

$$\text{and thus } E[RR] = \frac{E[P \times R(P)]}{E[R(P)]} \cdot \frac{(1-E[R(P)])}{E[P \times (1-R(P))]}.$$

But R(p) is a monotonically increasing function of p and therefore the covariance of R(P) and P is greater than zero. This implies that:

$$\begin{aligned} E[P \times R(P)] &> E[P] \times E[R(P)] \\ \longrightarrow E[P \times R(P)] - E[P \times R(P)] \times E[R(P)] &> E[P] \times E[R(P)] - E[P \times R(P)] \times E[R(P)] \\ \longrightarrow E[P \times R(P)] \times (1-E[R(P)]) &> E[R(P)] \times E[P \times (1-R(P))] \\ \longrightarrow E[RR] &> 1 \end{aligned}$$

Thus, in this model, in which malnutrition does not lead to diarrhoea, a comparison of diarrhoea rates in malnourished and adequately nourished children suggests that the risk/rate of diarrhoea is greater in the malnourished. This is indeed the case, since diarrhoea increases the risk of malnutrition and thus those children with the highest diarrhoea rates are at greatest risk of malnutrition. However, because each child's nutritional status is measured at the start of the observation period and the subsequent incidence of diarrhoea is recorded (i.e, malnutrition appears to precede diarrhoea), such results may be misinterpreted as evidence that malnutrition leads to diarrhoea. Similarly, if some children tend to suffer longer or more severe episodes of diarrhoea than others and this increases their risk of malnutrition, we may observe a link between malnutrition and increased severity or duration of diarrhoea in the absence of any causal association. It should be noted that this "bias" arises from the combination of reverse causality (diarrhoea causes malnutrition) and the heterogeneity of diarrhoea rates in the community. In the absence of such heterogeneity ($P(P=p)=1$), $E[P \times R(P)] = E[P] \times E[R(P)]$ and $E[RR] = 1$; i.e, there is no bias.

Two recent studies in Sudan (El Samani et al., 1988) and Mexico (Sepulveda et al., 1988) performed analyses which take account of diarrhoea in the interval preceding the period of observation. For these analyses observation periods were divided into two categories; those with an episode of diarrhoea in the previous period, and those without. Analyses similar to that outlined above, but with the data stratified according to history of diarrhoea in the previous period, were then performed. While the technique, as used by these authors, is problematic, since each child appeared in the analyses several times as supposedly independent observations, it might be anticipated that such a strategy would tend to separate children with high diarrhoea rates from those with lower rates, thus reducing the heterogeneity within each subgroup. It is therefore of interest to examine the effectiveness of this strategy in controlling the bias arising from reverse causality and the heterogeneity of diarrhoea rates in the population.

Observation periods preceded by an episode of diarrhoea

Denote the risk ratio for periods preceded by an episode of diarrhoea by $RR_{D=1}$. Then:

$$E[RR_{D=1}] = \frac{P(D(i)=1 | N(i)=1, D(i-1)=1)}{P(D(i)=1 | N(i)=0, D(i-1)=1)}$$

$$= \frac{P(D(i)=1, N(i)=1, D(i-1)=1)}{P(D(i)=1, N(i)=0, D(i-1)=1)} \times \frac{[1 - P(N(i)=1, D(i-1)=1)]}{P(N(i)=1, D(i-1)=1)}$$

$$P(D(i)=1, N(i)=1, D(i-1)=1) = \int p^2 (R(p)xt_{11} + (1-R(p))xt_{01})xf(p)dp$$

$$= t_{01}E[P^2] + (t_{11}-t_{01})E[P^2R(P)]$$

$$= r \text{ (say)}$$

$$P(D(i)=1, N(i)=0, D(i-1)=1) = E[P^2] - r$$

$$\text{and } P(N(i)=1, D(i-1)=1) = t_{01}E[P] + (t_{11}-t_{01})E[R(P)]$$

$$= s \text{ (say)}$$

So, $E[RR_{D=1}] = \frac{r}{E[P^2]-r} \times \frac{(EP-s)}{s}$

Observation periods not preceded by an episode of diarrhoea

Denote the risk ratio for periods not preceded by an episode of diarrhoea by $RR_{D=0}$.
Then:

$$E[RR_{D=0}] = \frac{P(D(i)=1 \mid N(i)=1, D(i-1)=0)}{P(D(i)=1 \mid N(i)=0, D(i-1)=0)}$$

$$= \frac{P(D(i)=1, N(i)=1, D(i-1)=0)}{P(D(i)=1, N(i)=0, D(i-1)=0)} \times \frac{[1 - P(N(i)=1, D(i-1)=0)]}{P(N(i)=1, D(i-1)=0)}$$

$$P(D(i)=1, N(i)=1, D(i-1)=0) = \int P(1-p)(R(p))x_{t_{10}} + (1-R(p))x_{t_{00}} \times f(p) dp$$

$$= t_{00}E[P(1-P)] + (t_{10}-t_{00})E[P(1-P)R(P)]$$

$$= r \text{ (say)}$$

$$P(D(i)=1, N(i)=0, D(i-1)=0) = E[P(1-P)] - r$$

$$\text{and } P(N(i)=1, D(i-1)=0) = t_{00}E[1-P] + (t_{10}-t_{00})E[(1-P)R(P)]$$

$$= s \text{ (say)}$$

So, $E[RR_{D=0}] = \frac{r}{E[P(1-P)] - r} \times \frac{(1-E[1-P]-s)}{s}$

It can be shown algebraically that each of the above expectations is greater than 1. Thus, stratifying observation periods according to whether or not diarrhoea occurred in the previous period does not eliminate the problem of bias arising from heterogeneous diarrhoea rates.

Example

In order to illustrate the problem of reverse causality in the presence of heterogeneous diarrhoea rates, and in order to obtain some idea of the potential magnitude of the bias thus arising, we consider a numerical example. Parameter values which look "reasonable" have been chosen, but it should be noted that these are not derived from any hard data.

First, let the probabilities that a child is malnourished at the start of an observation period, conditional on the child's status at the beginning of the preceding period and whether or not diarrhoea occurred during the interval, be as follows:

$$t_{11} = 0.99, t_{10} = 0.7, t_{01} = 0.3, t_{00} = 0.05.$$

Then, let the distribution among the population of the risk of diarrhoea in a single observation period be as shown in Table 1. This distribution has been chosen to provide a reasonably wide range of diarrhoea rates in the population. For a child with a given risk of diarrhoea p , the risk of malnutrition at the start of any interval $[R(p)]$ may then be calculated (Table 1). With this distribution the expectations of P , R , and $P \times R$ are:

$$E[P] = 0.40,$$

$$E[R(P)] = 0.45,$$

$$E[P \times R(P)] = 0.21.$$

TABLE 1. Distribution among the population of risk of diarrhoea during a single observation period and probability of malnutrition

Risk of diarrhoea p	Proportion of population E(p)	Probability of malnutrition R(p)
0.1	0.10	0.22
0.2	0.15	0.29
0.3	0.15	0.37
0.4	0.20	0.45
0.5	0.15	0.53
0.6	0.15	0.61
0.7	0.10	0.70
0.8	0.0	0.79
0.9	0.0	0.88

The following features of this example should be noted.

- (i) on average, 40% of children suffer from diarrhoea in any one observation period;
- (ii) on average, 45% of children are malnourished at the beginning of any observation period;
- (iii) the nutritional status of a child who is malnourished at the start of an observation period and who suffers an episode of diarrhoea during the period is unlikely to improve ($p=0.01$);
- (iv) a child who is adequately nourished at the start of the observation period and who does not suffer an episode of diarrhoea is unlikely to become malnourished ($p=0.05$);
- (v) a malnourished child who does not suffer an episode of diarrhoea has a 30% chance of improving nutritionally;
- (vi) an adequately nourished child who suffers an episode of diarrhoea has a 30% chance of becoming malnourished;
- (vii) the probability of malnutrition ($R(p)$) increases as p increases;
- (viii) poor nutritional status does not increase a child's risk of diarrhoea which depends only upon p .

It is of interest to note that in the Sudanese study (El Samani et al., 1988) about 36% of all child observation periods included an episode of diarrhoea, while in the Mexican study (Sepulveda et al., 1988) the proportion was about 53%. These figures are comparable with that of 40% ($E(P)$) in the model. In addition, the overall prevalence rates of mild and moderate malnutrition in the two studies were: Sudan 66%, Mexico 36%. In the model the expected prevalence of malnutrition is 45% ($E(R(P))$), which is between these two figures. If we perform a "crude" analysis of a study conducted in such a setting, taking no account of whether or not diarrhoea occurred in the preceding period, then the expectation of the estimate of the risk ratio that will be obtained is given by:

$$E[RR] = \frac{0.21 \times (1-0.45)}{0.45 \times (0.40-0.21)}$$

$$= 1.30.$$

Thus, in this situation, in which diarrhoea predisposes to malnutrition but malnutrition does not predispose to diarrhoea, a "traditional" analysis of the data might lead one to believe that malnutrition increases a child's risk of diarrhoea by 30%.

If the data are classified into two groups, periods preceded by an episode of diarrhoea, and periods not preceded by an episode of diarrhoea, then the expectations of the estimates that would be obtained from a stratified analysis are given by:

$$E[RR_{D=1}] = 1.15$$

and $E[RR_{D=0}] = 1.22.$

Taking account of a history of diarrhoea in the preceding interval has thus reduced, but not eliminated, the bias arising from the heterogeneity of the risk of diarrhoea. For periods preceded by an episode of diarrhoea, malnutrition appears to be associated with a 15% increase in the risk of diarrhoea. For periods following a diarrhoea-free interval, malnutrition appears to be associated with a 22% increase in the risk of diarrhoea.

In order to examine the variability of these estimates, computer simulations (written in BASIC) of an epidemiological study similar to those conducted in Sudan and Mexico were performed. In each simulation a population of 400 children was followed over 6 observation periods. At the start of each simulation, using a random number generator, each child was allocated a risk of diarrhoea (p_i) according to the distribution $f(P)$ shown in Table 1. The probability that the child was malnourished [$R(p_i)$] was then calculated and another randomly-generated number, uniformly distributed on the interval (0,1), was used to determine the child's nutritional status at the start of the study. Each child was then followed through 6 observation periods, the occurrence of diarrhoea and malnutrition being determined with randomly-generated numbers applied to the probabilities defined above. Three hundred computer simulations were performed and a summary of the results obtained is presented in Table 2. The mean values of the various estimators of risk ratio are very close to the theoretical expectations derived above. The "90% c.i." indicates the range within which 90% of the estimates lay.

TABLE 2. Results of 300 computer simulations of a population of 400 children followed over 6 observation periods

	"Crude" relative risk	Relative risk given diarrhoea in preceding interval	Relative risk given no diarrhoea in preceding interval
Mean	1.291	1.136	1.224
Minimum	1.084	0.924	0.949
Maximum	1.498	1.486	1.557
Standard deviation	0.081	0.101	0.104
90% c.i.	(1.16,1.42)	(0.98,1.31)	(1.04,1.38)

DISCUSSION

In this paper we have developed a simple probability model to examine the problem of distinguishing between cause and effect when investigating potentially bi-directional associations. This model shows how, if A causes B and the risk of A varies among individuals, commonly used epidemiological methods may lead to the "observation" that B causes A. It must be emphasized that the model is a hypothetical one and tells us nothing about the real nature of the association between diarrhoea and malnutrition. Many investigators have attempted to unravel this association. The model simply demonstrates that, in the presence of heterogeneity of diarrhoea rates in the study population, the approach commonly used to investigate this association is unable to distinguish properly whether poor nutritional status precedes diarrhoea or vice versa. If diarrhoea leads to poor nutritional status and diarrhoea rates in the community are heterogeneous, it may appear that poor nutritional status results in increased diarrhoea when this is not in fact the case.

It must also be remembered that the model used is very simplistic and, in the context of the example of diarrhoea and malnutrition, can be criticized on a number of grounds:

- it assumes that the population is in a "steady state" with regard to nutritional status;
- it assumes that, given an individual's underlying diarrhoea rate (p), the occurrence of diarrhoea during a given observation period is independent of previous history of diarrhoea;
- it assumes that risk of malnutrition is homogeneous in the population (apart from the effect of heterogeneous diarrhoea rates).

The effects of relaxing these assumptions in favour of more realistic scenarios require further research.

In the model used, the degree of bias introduced into the estimate of the risk ratio of the association between malnutrition and diarrhoea depends upon a number of parameters: the various transition probabilities between different nutritional states in the presence and absence of diarrhoea, and the distribution of diarrhoea rates within the population. In the simulations, values of these parameters that resulted in a plausible prevalence of malnutrition and a plausible risk of diarrhoea were used. When compared with the estimated risk ratios obtained from studies of the association between diarrhoea and malnutrition, the estimates of the risk ratio obtained with the model suggest that reverse causality could have substantially biased the results of those studies. By choosing different parameter values one could have decreased or increased the degree of bias introduced into the estimate of the risk ratio. It would be of interest to estimate some of these parameters from existing datasets to see how the estimates compare with the parameters used in the simulations. These parameter estimates could then be applied in the model to assess the level of bias that might have occurred in those settings. This would provide a better idea of whether or not the bias arising from reverse causality is a serious practical problem in studies of malnutrition and diarrhoea.

The bias that we have observed to occur in the model arises from the combination of reverse causality (A causes B) and heterogeneity of the rate of A (diarrhoea) in the community. If all individuals had identical risks of A, or A did not cause B, then the estimate of the risk ratio obtained from a "traditional" analysis would be unbiased. This suggests that the problem could be approached, at least in the case of diarrhoea and malnutrition, by performing "within individual" analyses of the data in which each individual acts as his/her own control. This approach has been adopted by Cole (1989), who used "within-subject regression" to investigate the effects of disease on weight velocity (does A cause B?). The effectiveness of different techniques for performing within-individual analyses in eliminating bias due to reverse causality, particularly when independence between observation periods cannot be assumed, has still to be evaluated.

In summary, this paper has examined the problem of disentangling associations that may be bi-directional in nature. It has shown that, in the presence of heterogeneous disease rates, commonly used epidemiological methods can produce biased results. Alternative approaches to the design and analysis of studies to investigate such associations need to be developed and evaluated.

ACKNOWLEDGEMENTS

This work was supported by the Diarrhoeal Disease Control Programme of the World Health Organization and the Overseas Development Administration, United Kingdom. Thanks are due to Jean-Pierre Habicht, Betty Kirkwood, Chessa Lutter, and Jose Martines for their constructive comments and suggestions regarding earlier drafts of this paper.

REFERENCES

- Chandra, R.K. (1986) Nutritional regulation of immunity and infection: from epidemiology to phenomenology to clinical practice. Journal of Pediatric Gastroenterology and Nutrition, 5:844-852.
- Chen, L.C. (1983) Interactions of diarrhoea and malnutrition: mechanisms and interventions. In: Chen L.C. and Scrimshaw, N.S.(eds). Diarrhoea and malnutrition. New York, Plenum Press.
- Cole, T.J. (1989) Relating growth rate to environmental factors - methodological problems in the study of growth-infection interaction. Acta Paediatrica Scandinavica, Supplement, 350:14-20.
- Cox, D.R. and Miller H.D. (1972) The theory of stochastic processes. 2nd ed. London, Chapman and Hall, p80.
- El Samani, E.F.Z. Willett, W.C. and Ware, J.H. (1988) Association of malnutrition and diarrhoea in children aged under five years. A prospective follow-up study in a rural Sudanese community. American Journal of Epidemiology, 128:93-105.
- Hennekens, C.H. and Buring, J.E. (1987) Epidemiology in medicine. Boston, Little, Brown and Company, p153.
- Rose, G. and Shipley, M.J. (1980) Plasma lipids and mortality: a source of error. Lancet, i:523-526.
- Scrimshaw, N.S. Taylor, C.E. and Gordon, J.E. (1968) Interactions of nutrition and infection. (Monograph Series No. 57). Geneva, World Health Organization.
- Sepulveda, J. Willett, W. and Munoz, A. (1988) Malnutrition and diarrhoea. A longitudinal study among urban Mexican children. American Journal of Epidemiology, 127:365-376.
- Sirisinha, S. Suskind, R. Edelman, R. et al. (1975) Secretory and serum IgA in children with protein-calorie malnutrition. Pediatrics, 55:166-170.
- Tomkins, A. and Watson, W. (1989) Malnutrition and infection: a review. London, Clinical Nutrition Unit, London School of Hygiene and Tropical Medicine.