incidence based on the population from the Detroit Surveillance, Epidemiology, and End Results Program for women of the same ages and birth cohorts. Mothers and sisters of controls had the expected breast cancer incidence rates for the Detroit population.

In general, completeness and accuracy of reporting of family history of disease appear to depend on the disease, the relative, and the population under consideration. Because this bias is potentially devastating to genetic epidemiology, Go et al. (2) and Macera (3) addressed this problem for breast cancer several years ago. For approximately 800 relatives of breast cancer patients and unaffected women, we assessed recall bias by contacting or reviewing records for all relatives, whether or not they were reported to be affected. In that sample of Caucasian women in the United States, there was no difference between women with breast cancer and unaffected women in recall accuracy for breast cancer among their mothers and sisters. Both groups were extremely accurate: There were only two discrepancies among the approximately 800 relatives followed. However, reporting accuracy of breast cancer among mothers and sisters cannot be extended to more distant relatives, other diseases, other populations, or male respondents. For example, reporting of ovarian and endometrial cancers was not equally reliable by patients with these diseases and controls (Schwartz et al., unpublished data). As Mantel (4) indicates, it may be that reporting of dementia is similarly biased. For these and most other conditions, independent verification would be necessary. Reporting of cancer among more distant relatives is also likely to be incomplete. Finally, reporting by subjects in migrant and/or impoverished populations has not been consistently reliable, in our experience (3, 5, 6). Underreporting in these populations may be due to both a history of inadequate medical care and the geographic separation of migrants from their relatives.

We should also note that the only chromosome for which women are more closely related to their sisters than to their mothers is the X chromosome. Although the genetics of breast cancer susceptibility are complex, X-linkage of genes strongly influencing susceptibility to this disease can be excluded (2, 7). Based on genetics alone, there is no reason to expect breast cancer risk for sisters to be higher than the risk for mothers of patients.

REFERENCES

1. Schwartz AC, King M-C, Belle SH, et al. Risk of breast cancer to relatives of young breast cancer patients. JNCI 1985;75:665–8.
2. Go RC, King M-C, Bailey-Wilson J, et al. Genetic epidemiology of breast cancer and associated cancers in high-risk families. I. Segregation analysis. JNCI 1983;71:455–61.
3. Macera CA. Epidemiology and detection of breast cancer in a screening center population. PhD dissertation. Berkeley, CA: University of California, Berkeley, 1982.
4. Mantel N. Familial breast cancer and the awareness bias. Am J Epidemiol 1987;125:920.
5. Zunzunegui MV, King M-C, Coria CF, et al. Male influences on cervical cancer risk. Am J Epidemiol 1986; 123:302–7.
6. Yeung KS. Epidemiology of breast cancer among Chinese women in the San Francisco Bay area. PhD dissertation. Berkeley, CA: University of California, Berkeley, 1983.
7. Bailey-Wilson J, Cannon L, King M-C. Genetic analysis of human breast cancer: a synthesis of contributions to Genetic Analysis Workshop IV. Genet Epidemiol (in press).

Mary-Claire King
Program in Epidemiology
School of Public Health
University of California
Berkeley, CA 94720

INGESTED DOSE AND DIARRHEA TRANSMISSION ROUTES

In a stimulating theoretical paper, Briscoe (1) considered a hypothetical experiment, in which volunteers ingested various doses of an imaginary diarrhea-causing pathogen having a log-linear dose-response relation and showed that reductions in the ingested dose do not necessarily lead to proportionate reductions in the incidence of disease. Briscoe concluded that the elimination of one of several important fecal-oral transmission routes cannot be expected to cause measurable reductions in disease incidence.

Briscoe's analysis ignores the important factor of time. If the dose of a pathogen ingested from each transmission route is not constant through time, it is possible for each episode of diarrhea to be caused by a single transmission route, even though more than one route is important in the community. If this is the case, the elimination of an important route will reduce the incidence of disease in proportion to the number of episodes it causes. The validity of Briscoe's model therefore depends on two questions: 1) Over how long a period of time can an infectious dose be accumulated without reducing the probability of infection? 2) How widely does the ingested dose vary between periods of that length?

With regard to the first question, it is clear that if a dose can accumulate over a protracted period of time, this will smooth out short-term variations in the numbers of pathogens ingested. However, in the case of diarrheal diseases with incubation periods of 1–3 days, organisms which have not colonized the intestine within 24 hours of ingestion are unlikely to be affected by others that reach the intestine the following day. Among pathogens sensitive to gastric acidity, which affects most bacteria (2), the vast majority die within a short time of ingestion, reducing the maximum period of accumulation to much less than one day. Indeed, there is reason to believe that most infections caused by doses less than the median infective dose (ID_{50}) are initiated by a single organism (3) which has survived the mortality of the "decisive" initial stage (4). Two bouts of diarrhea caused by the ingestion of infectious doses on different days can, of course, overlap, but the chance of this is small in practice.

If pathogen doses cannot accumulate from one day to the next, the question of the daily variation in the dose arises (question 2). It can be assumed that the dose of any diarrheal pathogen received via a particular route varies from day to day no less than the

corresponding dose of fecal indicator organisms, for the following reasons.

Three types of factors determine the ingested dose of a specific fecal organism: 1) elimination or regrowth in the environment; 2) the amount of fecal matter ingested; and 3) the concentration of the organism in that fecal matter at defecation.

Regarding the first factor, regrowth of pathogens may be greater or less than that of fecal indicator bacteria. (This discussion excludes viral and protozoal pathogens, which cannot multiply in the environment.) However, indicator species are chosen to survive longer in the environment than do pathogens. The second factor, the amount of fecal matter ingested, affects pathogens and indicators equally. With regard to the third, the number of fecal indicator bacteria per gram of human feces normally varies within a range of one or two $\log_{10}$ cycles (2). On the other hand, the concentration of a particular pathogen ranges from zero in uninfected persons to thousands of millions per gram in some clinical cases of diarrhea (5). The three factors together argue that the ingested dose of pathogens will vary no less than, and probably more than, the dose of fecal coliforms.

Consider first the dose ingested from the waterborne transmission route. Fecal coliform concentrations in water from open wells typically vary by about one cycle of logarithms between consecutive samples, and the highest readings are usually isolated events followed by much lower values (6). Concentrations of fecal bacteria in streams are also likely to attain short-lived peaks during periods of high flow, as excreta are washed into them or flushed from the banks (7). Spira et al. (8), quoted by Briscoe (1), found concentrations of *Vibrio cholerae* varying from zero to more than 10,000 colony-forming units per 100 ml in surface water sources and in household cooking water.

With regard to food contamination, the usual situation is that only a minority of cooked food samples are detectably contaminated at all. When fecal contamination has occurred, the possibility of regrowth increases the range of bacterial concentrations. This is illustrated by the data in table 1, collected in a village in The Gambia, which show fecal coliform concentrations varying over four logarithmic cycles. As might be expected, fecal pathogens occur less frequently on food samples than do fecal coliforms; Spira et al. (8) found *V. cholerae* in only 0.13 per cent of food samples.

The foregoing data suggest that a person does not ingest a constant daily dose of pathogens from each transmission route, but rather ingests a number of pathogens which may often be zero and which on other days vary by one or more factors of 10. While this range may be small in comparison with experimental dose-response relations (9–11), it is sufficient to ensure that the doses ingested from different routes on a given day are not normally of the same order of magnitude as one another. On any given day, the dose from one route will be far greater than all the others, so that each episode of diarrhea will be caused by a single transmission route and not by an additive accumulation of different doses as postulated in Briscoe's model.

TABLE 1

*Concentrations of Escherichia coli in various foods in Gambian households**

	Boiled foods	Steamed foods
Escherichia coli not found	60	95
E. coli present in 0.1 g	27	26
E. coli present in 0.001 g	5	20
E. coli present in 0.00001 g	7	8
No. of samples examined	99	149

* Unpublished data collected in 1977 by R. A. E. Barrell at the Medical Research Council Laboratories, Keneba, The Gambia.

REFERENCES

1. Briscoe J. Intervention studies and the definition of dominant transmission routes. Am J Epidemiol 1984;120:449–55.
2. Drasar BS, Hill MJ. Human intestinal flora. London: Academic Press, Inc., 1974.
3. Meynell GG, Meynell EW. The growth of micro-organisms *in vivo* with particular reference to the relation between dose and latent period. J Hyg (Camb) 1958; 56:323–46.
4. Meynell GG, Maw J. Evidence for a two-stage model of microbial infection. J Hyg (Camb) 1968;66:273–80.
5. Feachem RG, Bradley DJ, Garelick H, et al. Sanitation and disease: health aspects of excreta and wastewater management. Chichester: John Wiley & Sons, 1983.
6. Barrel RAE, Rowland MGM. The relationship between rainfall and well water pollution in a West African (Gambian) village. J Hyg (Camb) 1978;83:143–50.
7. Feachem RG. Faecal coliforms and fecal streptococci in streams in the New Guinea highlands. Water Res 1974;8:367–74.
8. Spira WM, Khan MU, Saeed YA, et al. Microbiological surveillance of intra-neighbourhood El Tor cholera transmission in rural Bangladesh. Bull WHO 1980;58:731–40.
9. Hornick RB, Greisman SE, Woodward TE, et al. Typhoid fever: pathogenesis and immunologic control. N Engl J Med 1970;283:686–91.
10. Levine MM, DuPont HL, Formal SB, et al. Pathogenesis of *Shigella dysenteriae* 1 (Shiga) dysentery. J Infect Dis 1973;127:261–70.
11. Cash RA, Music SI, Libonati JP, et al. Response of man to infection with *V cholerae*. I. Clinical, serological and bacteriological responses to a known inoculum. J Infect Dis 1974;129:42–52.

Sandy Cairncross
London School of Hygiene and Tropical Medicine
London WC1E 7HT
England

THE AUTHOR REPLIES

The letter from Cairncross (1) concerns the omission of a particular factor (time) in my paper (2). In this reply, I will show that inclusion of this factor does not affect the major conclusion of the original paper, namely that it is incorrect to "... assume that the ratio 'number of cases eliminated:number of residual

cases' measures the relative importance of the eliminated route *vis-à-vis* the residual transmission route" (2, p. 449). I will conclude with a discussion of the criteria by which simple didactic models of the sort presented in the original paper should be judged.

Including time in the model

As Cairncross correctly points out, "the dose of a pathogen ingested from each transmission route is not constant through time" (1). The simple model presented in the original paper can be adapted to incorporate this modification. Assuming, as in the original paper (2), a log-linear dose-response curve, a version of the model in the original paper (now with just two transmission routes) which incorporates frequency distributions (rather than constant transmission) can be constructed. The baseline situation shown in table 1 includes transmission through a "red route" (with frequencies based on the frequency of fecal coliforms in traditional water supplies in rural Africa (3)) and a "blue route" (with frequencies derived from the food contamination data for rural Africa presented by Cairncross (1)).

Now let us consider the case in which an improvement is made in the red route. (Specifically, the frequency distribution for the red route is changed from that of a traditional rural African water supply to that of an improved piped water supply in rural Africa (based on data in Young and Briscoe (3).) The "post-intervention" situation is presented in table 2.

If tables 1 and 2 were presented to an epidemiologist, he might say, "There has been a major improvement in the red route (with the mean number of organisms transmitted through this route reduced by 83 per cent), yet the overall reduction in disease is only 42 per cent. I therefore conclude that the blue route (the only other route) must be at least as important as the red route and expect that a similar improvement in the blue route would have a similar effect on disease."

Ideally, however, we should test the importance of the blue route directly (by actually reducing transmission through the blue route) rather than indirectly (by inferring the importance of the blue route from the results of improving the red route). If we make an improvement in the blue route similar to the improvement made in the red route in the first simulation (reducing the average number of organisms transmitted through this route by 83 per cent), then the situation is as represented in table 3.

From tables 1 and 3, it can be seen that reducing the mean number of organisms transmitted through the blue route by 83 per cent did not result in a reduction in disease of about 42 per cent (as our epidemiologist had inferred from tables 1 and 2), but in a reduction of just 14 per cent.

The major point of the original analysis (the so-called "residual fallacy") was that

> ... if the dose-response relationship is nonlinear, if there are several transmission routes, and if the effects on disease incidence of eliminating one transmission route are known, then it is

TABLE 1

Baseline situation

	Red	Blue	Total
No. of times/100 days that count equals			
0	15	60	
10	0	20	
100	15	10	
1,000	70	10	
Mean count	715	112	827
No. of infections	$(15 \times 0) + (0 \times 0.33)$ $+ (15 \times 0.67) + (70 \times 1)$ $= 80$	$(60 \times 0) + (20 \times 0.33)$ $+ (10 \times 0.67) + (10 \times 1)$ $= 23$	103

TABLE 2

Situation after intervention in red route

	Red	Blue	Total
No. of times/100 days that count equals			
0	30	60	
10	40	20	
100	20	10	
1,000	10	10	
Mean count	124	112	236
No. of infections	$(30 \times 0) + (40 \times 0.33)$ $+ (20 \times 0.67) + (10 \times 1)$ $= 37$	$(60 \times 0) + (20 \times 0.33)$ $+ (10 \times 0.67) + (10 \times 1)$ $= 23$	60

TABLE 3

Situation after improvement in the blue route

	Red	Blue	Total
No. of times/100 days that count equals			
0	15	85	
10	0	4	
100	15	10	
1,000	70	1	
Mean count	715	20	735
No. of infections	$(15 \times 0) + (0 \times 0.33)$ $+ (15 \times 0.67) + (70 \times 1)$ $= 80$	$(85 \times 0) + (4 \times 0.33)$ $+ (10 \times 0.67) + (1 \times 1)$ $= 9$	89

> fallacious to assess the relative importance of the eliminated and residual transmission routes by comparing the reduction in incidence due to elimination of the one route, on the one hand, to the residual incidence of the disease, on the other (2, p. 451).

In this analysis, it has been shown that, even when allowance is made for the fact that the dose of a pathogen ingested from each transmission route is not constant through time, the danger of the residual fallacy remains.

THE USE OF SIMPLE DIDACTIC MODELS

Cairncross' concerns raise a more general problem about the use of simple didactic models such as that presented in the original paper. Epidemiologic phenomena are inevitably complex. Attempts have been made to develop detailed simulation models which include the multitude of factors which affect the epidemiology of real-world diseases (4). It has been argued that, while such elaborate modeling exercises are useful for some purposes, they are not effective in giving a reader an intuitive grasp of the central dynamics of the disease. It has also been argued that simple models, the function of which are to give the reader a good understanding of the "woods" but not of the "trees," can be useful in transmitting this intuitive grasp (5).

Accordingly, the criterion by which simple didactic models must be judged is not the degree of "completeness" (as discussed in the original paper (2, p. 455), a more complete model " . . . would probably be stochastic and would certainly include other epidemiologically-significant phenomenon [sic] such as infection:case ratios and the effect of acquired immunity"), but rather 1) the degree to which the model explains the macroepidemiologic features of interest, and 2) the ease with which readers grasp and maintain an understanding of the underlying structure of the model.

With regard to the first criterion, in the particular case at hand the conclusions emanating from the original model (and the modification presented here) concur with an emerging body of empirical data. A few examples follow.

1) International comparisons. It has been shown that the impact of improved water supply has been greater in middle-income countries than in poor countries (6), and it has been hypothesized that this is because in the middle-income countries prior reductions in certain other transmission routes have taken place (2).

2) National experience with infant mortality over time. In Chile, it has been noted (7) that almost all of the early interventions to reduce disease transmission appeared to be ineffective, while during the last decade it appears that many interventions have been successful (presumably because these recent interventions were eliminating the residual routes of transmission).

3) Specific diseases as sanitation improves. For both polio (8) and typhoid (9), it has been shown that initial improvements from "appalling" to "bad" sanitation were associated with an increase in the prevalence of severe disease, and that it was only once sanitary conditions became "good" that the anticipated declines in severe disease were observed.

With regard to the second criterion, the simple model presented in the original paper appears to be easily understandable and to give readers an intuitive grasp of the interactions (which arise from many causes) between different transmission routes.

It thus appears that, while the model certainly omits important factors (just one of which Cairncross has discussed) and can never be considered a definitive model of all aspects of the transmission of fecal-oral diseases, it satisfies the criteria by which simple didactic models should be judged.

REFERENCES

1. Cairncross S. Ingested dose and diarrhea transmission routes. Am J Epidemiol 1987;125:921–2.
2. Briscoe J. Intervention studies and the definition of dominant transmission routes. Am J Epidemiol 1984;120:449–55.
3. Young BA, Briscoe J. Water and health in rural Malawi: aspects of the performance, utilization and health impact of the Malawi Self-Help Rural Water Supply Project. Lilongwe, Malawi: US Agency for International Development, 1986.
4. Cvjetanovic B, Grab B, Uemera K. Dynamics of acute bacterial diseases. Geneva: World Health Organization, 1978.
5. Briscoe J. On the use of simple analytic mathematical models of communicable diseases. Int J Epidemiol 1980;9:265–70.
6. Shuval HI, Tilden RL, Perry BH, et al. Effect of investments in water supply and sanitation on health status: a

threshold saturation theory. Bull WHO 1981;59:243–8.
7. Brunser O In: Briscoe J, Feachem RG, Rahaman MM, eds. Evaluating health impact: water supply sanitation and hygiene education. Ottawa, Ontario, Canada: International Development Research Center, 1986:256.
8. Gregg MB. Poliomyelitis. In: Last JM, ed. Maxcy Rosenau public health and preventive medicine. 11th ed. New York: Appleton Century Crofts, 1980:149–52.
9. Hornick RB. Selective primary health care: strategies for control of disease in the developing world. 20. Typhoid fever. Rev Infect Dis 1985;7:536–46.

John Briscoe
Operations Policy and Research Division
Water and Urban Development Department
World Bank
Washington, DC 20433

RE: "BINOMIAL REGRESSION IN GLIM: ESTIMATING RISK RATIOS AND RISK DIFFERENCES"

Wacholder has described the use of the program package GLIM in the analysis of cumulative incidence type of data (1). Here we present an analysis of case-control data based on the same procedures.

Logistic regression in case-control studies uses the logarithm of the odds of the probability P of being a case. With I strata, with K variables in the model, and on the assumption of constant odds ratios (OR) over the strata, the model is

$$\ln(P_i/(1 - P_i)) = a_i + b_1X_1 \ldots + b_KX_K \qquad i = 1, \ldots, I$$

where $\exp(b_k)$ is interpretable as the ratio of the odds associated with a one unit change in the kth variable X_k, i.e., for a dichotomized variable $\exp(b_k)$ is the ratio of the odds for the exposed to the unexposed. Maximum likelihood estimates of the b's are easily obtained by the use of GLIM with the macros described by Wacholder.

To illustrate the application of these for case-control data, we have considered the data from a study on the relation between coffee consumption and myocardial infarction used as an example by Rothman and Boice (2). The GLIM program to analyze these data using logistic regression is given in the Appendix. Each of the five data lines starting with READ gives the data for one stratum. The columns give numbers of cases and controls, respectively, for three different exposure levels beginning with the unexposed.

The data printout gives parameter estimates of the b's with corresponding standard errors. The odds ratios for medium consumers of coffee compared with nonconsumers and for heavy consumers of coffee relative to nonconsumers were consistently estimated at 1.60 with a 95 per cent confidence interval of 1.22–2.09 and at 2.19 with a 95 per cent confidence interval of 1.55–3.11, respectively.

The chi-square value based on the scaled deviance was $\chi^2_8 = 9.73$ which corresponds to a p value of 0.28. For evaluation of the model, GLIM also provides observed and expected numbers in each cell (data not shown). Several alternative models are possible, including models with interaction terms and those analyzing the trend (3).

REFERENCES

1. Wacholder S. Binomial regression in GLIM: estimating risk ratios and risk differences. Am J Epidemiol 1986; 123:174–84.
2. Rothman KJ, Boice JD. Epidemiologic analysis with a programmable calculator. Washington, DC: US GPO, 1979 (NIH publication no. 79–1649).
3. Breslow NE, Storer BE. General relative risk functions for case-control studies. Am J Epidemiol 1985;122:149–62.

Lars Alfredsson
Department of Occupational Medicine
Karolinska Hospital
S-104 01 Stockholm
Sweden

Anders Ahlbom
Department of Epidemiology
National Institute of Environmental Medicine
S-104 01 Stockholm
Sweden

APPENDIX

```
$INPUT 51 72 $C READ W'S MACROS FROM UNIT 51 (LOCALLY DEFINED) $
$ECHO
$UNITS 15
$DATA D C
$READ 16  1596  62  4424  10  791
       8   492  65  1602  13  335
      16   323  63   940  13  112
      19   182  52   423  16   66
      11   153  56   349  15   50 $
$CALCULATE STRAT=%GL(5,3):COF=%GL(3,1)
$CAL N=D+C
$LOOK D N STRAT COF
$FACTOR STRAT 5 COF 3
$YVAR D
$USE OR
$FIT STRAT+COF $DIS E
$STOP
```

THE AUTHOR REPLIES

Alfredsson and Ahlbom correctly point out that GLIM is a useful program for analyzing case-control data (1). In their monograph on case-control studies, Breslow and Day (2) sketched the mechanics of a logistic analysis of case-control data using GLIM. Indeed, the authors of at least two recent papers in this *Journal* used GLIM for this purpose (3, 4). No macros are required, since the built-in GLIM command $ERROR B N $ can be used to indicate logistic regression with binomial denominator N. Thus, a sim-