MEASURING THE IMPACT OF WATER SUPPLY AND SANITATION FACILITIES ON

DIARRHOEA MORBIDITY: PROSPECTS FOR CASE-CONTROL METHODS

**Environmental Health Division
and
Diarrhoeal Diseases Control Programme**

**World Health Organization
Geneva**

Comments to be sent to:

Mr G. Schultzberg
Division of Environmental Health
World Health Organization
1211 Geneva 27
Switzerland

Dr J. Briscoe
Department of Environmental Science and Engineering
University of North Carolina
Rosena Hall 201H
Chapel Hill NC 27513
United States

Dr R.G. Feachem
Head, Department of Tropical Hygiene
London School of Hygiene and Tropical Medicine
Keppel Street (Gower Street)
London WC1E 7HT
United Kingdom

MEASURING THE IMPACT OF WATER SUPPLY AND SANITATION FACILITIES ON

DIARRHOEA MORBIDITY: PROSPECTS FOR CASE-CONTROL METHODS

Environmental Health Division
and
Diarrhoeal Diseases Control Programme

World Health Organization
Geneva

Document prepared by:

John Briscoe
University of North Carolina

Richard G. Feachem
London School of Hygiene and Tropical Medicine

Mujibur M. Rahaman
International Centre for Diarrhoeal Disease Research, Bangladesh

September 1985

TABLE OF CONTENTS

List of Tables

Lists of Figures

1. INTRODUCTION

In 1975 the World Bank, the major lending institution in the water supply and sanitation sector, convened an expert panel to advise Bank planners on reliable procedures for estimating the health effects of investments in the water supply and sanitation sector. The expert panel concluded that "long-term longitudinal studies of large size and expense are probably the only means through which there is any chance of isolating a specific quantitative relationship between water supply and health" and recommended that, given "the very high cost, limited possibility of success and restricted application of results", such studies not be undertaken (IBRD, 1976). A decade later, the need for reliable information has once again come to the fore. Within the water supply and sanitation sector there is a need to identify the effects of different levels of service on health, while within the health sector as a whole there is a need for information on the effectiveness of water supply and sanitation projects so that the cost-effectiveness of these and other investments in the health sector can be compared.

Participants at two recent scientific meetings in Cox's Bazaar, Bangladesh and Geneva, Switzerland (see "Notes" on page 49 of this paper) have deliberated on the methodologies that have been used to measure the impact of water supply and sanitation projects on health and have considered in detail the prospects for improving these methodologies. These meetings have taken account of the studies that have been initiated since the World Bank meeting in 1975 and have been influenced by the development of a methodology for evaluating the performance, functioning and utilization of water supply and sanitation projects (WHO, 1983).

Discussions at these meetings, which will be presented in detail in a forthcoming monograph (Briscoe, Feachem and Rahaman, 1985), focused on the following questions:
- Under what conditions should health impact evaluations (HIEs) be undertaken?
- What impact measures should be used in HIEs?
- What study designs should be used in HIEs?
- How should the results of HIEs be interpreted?

The scope of the present paper is more limited. It is assumed that a HIE of a water and sanitation project is to be undertaken, and that diarrhoea morbidity has been selected as the health impact measure. The options in choosing a study design for such an impact evaluation are outlined, the experience with different designs is discussed briefly, and a hitherto untried case-control method is proposed.

2. MEASURES OF DIARRHOEAL DISEASE MORBIDITY

In the ten years following the World Bank Expert Panel meeting, major advances in understanding of the epidemiology of diarrhoeal diseases have occurred. Whereas ten years ago diarrhoea was an "inscrutable syndrome" (Wall and Keeve, 1974) and pathogenic agents could be identified in less than 20 percent of episodes (WHO, 1979), today pathogens can be identified in up to 80 percent of diarrhoea cases reporting to a medical facility (Black, 1984).

If there were no constraints on obtaining and processing faecal samples, HIEs could evaluate the effect of water supply and sanitation facilities on infection and disease caused by each of the diarrhoeal pathogens. Typically, however, there are severe constraints and it is appropriate to limit aetiologic studies to those few pathogens that are known to be the major causes of diarrhoea in a particular age group, at a particular place and in a particular season. For a study of children under 5 years of age over a whole year, the appropriate pathogens for emphasis in many countries would be enterotoxigenic Escherichia coli (ETEC), rotavirus, Campylobacter jejuni and Shigella. However, there are also problems with this more limited approach because the isolation of rotavirus and ETEC requires moderately sophisticated laboratory facilities.

Many mild cases of diarrhoea are of doubtful public health significance and it will sometimes be appropriate to study only severe diarrhoeas (clinically defined), either with or without information on aetiology. Bacterial diarrhoeas (such as those due to ETEC and Shigella) peak in the warm season in many countries whereas viral diarrhoeas (such as rotavirus diarrhoea) peak in the cool season. While rotavirus has been implicated in some water-borne outbreaks in developed countries (Hopkins et.al., 1984), it appears that water supply and sanitation projects may have little impact on some viral diarrhoeas in developing countries (de Zoyza and Feachem, 1985). A study that is restricted to measuring the impact of water supply and sanitation projects on severe diarrhoeas only, in the season when bacterial diarrhoeas have their peak incidence rate, may maximize the chance of showing a significant impact for a given sample size.

3. STUDY DESIGNS

There is no universally accepted method for classifying epidemiologic designs, in large part because in practice "hybrid" (rather than "pure") designs are used. Nevertheless, it is useful to characterize the key differences between the major study types used in any particular type of application, so that the advantages and disadvantages of each particular type may be discussed. Figure 1 depicts the key distinctions characterizing the study designs discussed in this paper. Details of the method, role, advantages and disadvantages of each study design are available in standard epidemiology texts (such as MacMahon and Pugh, 1970); here the discussion is limited to the issues directly affecting the choice of a study design for an evaluation of the impact of water supply or sanitation facilities on diarrhoea.

3.1 QUASI-EXPERIMENTAL DESIGNS

In true experimental designs (of which vaccine and drug trials are common examples), subjects are assigned to treatment and control groups by some formal method of randomization. Since random assignment is impossible if the preventive measure can be applied only to an entire community (as in the case of water supply interventions), the "next best thing" is done, namely the treatment is applied to some (treated) communities and withheld from "similar" control communities. Because of the intuitive appeal of these quasi-experimental designs as the closest practical approximation to the classical experimental design, they have been the most popular design in HIEs of water supply and sanitation interventions.

The key issue in quasi-experimental designs is that of the comparability of the treatment and the control groups. Comparison may be "internal", with, for example, the incidence of diarrhoea in the group before the intervention being compared with the incidence after the intervention. Comparison may also be "external", with the diarrhoea incidence in the treatment group being compared to the diarrhoea incidence in the control group sometime after the intervention.

3.1.1 Problem 1: The comparability of treatment and control groups:

Rigorous statistical methods for the analysis of quasi-experimental designs were developed in the early 1960s (Campbell and Stanley, 1963). In the subsequent twenty years extensive experience with these designs has been accumulated, especially in the evaluation of social programmes in developed countries. It is now generally recognized that the approach is fraught with formidable methodological difficulties, with even the major developer of quasi-experimental designs publicly regretting the influence his work has had (Cook and McAnany, 1979). Experience in developed countries has shown that the treatment and control groups are seldom comparable and that it is extremely difficult to adjust for this lack of comparability using statistical methods (Cook and McAnany, 1979). In developing countries, as illustrated by the following examples, the same difficulties have been encountered. In the classic and well-designed studies of diarrhoea and nutrition in Guatemala (Scrimshaw, et.al., 1967), comparisons were both internal and external, yet, after years of observation and analysis the scientists found it impossible to determine how much of the difference in effects observed between villages was due to the different interventions, to general secular trends which were different between the villages, or to sudden unexpected events (such as epidemics) which affected only certain villages. Similarly, a recent comprehensive assessment of evaluations of nutrition programmes revealed that, because "the experimental context was unstable, unpredictable and unique in each case" (Drake et.al., 1983), the assumptions of comparability of treatment and comparison groups were violated in every single case. These problems are equally common in quasi-experimental studies of water supply and sanitation interventions (Blum and Feachem, 1983).

3.1.2 Problem 2: The sample sizes required

A second major concern affecting these designs (and the cohort designs to be discussed in the next section) is that the number of study subjects required to detect changes of public health significance in the outcome variables is very large. Assuming that on the average a child under the age of five years has 2.2 attacks of diarrhoea per year (Snyder and Merson, 1982), then if one-week recall data are obtained, the frequency of positive answers

to the question "has your child had an attack of diarrhoea which started in the past week?" will be

4.2 percent. Assuming that the study is designed to detect a 33 percent reduction in diarrhoeal incidence and assuming that a cluster sampling technique is used, then (see Table 1) approximately 6000 questionnaires will have to be administered to the group with improved water supply and 6000 questionnaires to the group without improved facilities. If only severe episodes of diarrhoeal disease are included in the study, the number of episodes is reduced to about 10 percent of the total number. If the recall period remains one week, then the frequency of positive answers falls to 0.4 percent and the number of questionnaires administered in each group rise to 60,000 (see Table 1).

Large as these numbers are, they are probably unrealistically low, because the recall period (of one week) is almost certainly too long. As a result of studies by Martorell et.al. (1976) and others which have shown that recall falls off sharply with length of recall, in studies of diarrhoeal disease it is now general practice to limit recall to 2 days. If the above calculations are repeated for a two-day rather than a one-week recall, the required sample sizes will be about three times as large.

3.1.3 Problem 3: Misclassification biases

A further methodological concern with experimental (and other) studies relates to the effect on measures of the association between disease and exposure (such as the odds ratio) of inevitable errors in classifying individuals as either "exposed" or "not exposed" to a risk factor and as either "diseased" or "not diseased".

A recent review of the methodological problems of HIEs of water supply and sanitation projects (Blum and Feachem, 1983) has shown that problems with defining the health indicator and with failing to record facility usage are common, but, as yet, no investigation of the effects of these misclassification errors on the results of such studies has been undertaken. In this section the direction and magnitude of biases emanating from misclassification errors which are plausible in quasi-experimental HIEs of water supply and sanitation programmes are investigated.

In acquiring information on an attribute of an individual (such as whether
or not the individual has had diarrhoea during the past week), two different
types of classification error can occur. First, individuals who did not have
diarrhoea may be classified as having had diarrhoea (false positives), and,
second, individuals who did have diarrhoea may be classified as not having had
diarrhoea (false negatives). These errors are represented in the disease
classification matrix below:

		TRUE DISEASE STATUS	
		With diarrhoea	Without diarrhoea
APPARENT DISEASE	with diarrhoea	1-A	B
STATUS	without diarrhoea	A	1-B

In standard terminology, the probability of a person with the disease being so
classified is known as sensitivity ((1-A) in the matrix) and the probability
of a person without the disease being so classified as specificity ((1-B) in
the matrix). Where the classification scheme is perfect there are no false
negatives (i.e. $A = 0$), no false positives ($B = 0$), and the sensitivity and
specificity of the measure are 100%.

The problem of misclassification affects not only the disease indicator,
but the exposure indicator as well, as indicated in the following exposure
classification matrix:

		TRUE EXPOSURE STATUS	
		Exposed	Not exposed
		(not using improved water supply)	(using improved water supply)
APPARENT EXPOSURE	Exposed	1-C	D
STATUS	Not exposed	C	1-D

To understand how these misclassification errors affect the estimates of
the health impact of a water supply intervention, it is necessary to determine
how the apparent or observed relationships between exposure and disease differ
from the true relationships (see Annex 1).

The degree to which those who are exposed are more likely to have disease can be expressed by various measures. The measure used here is the odds ratio, which is defined in section 3.5 below and is roughly equivalent to the relative risk. The odds ratio emerging from the actual observations may be compared with the odds ratio emerging from the true values. This is done by computing the normalized (i.e. with a value varying between 0 and -1 for non-differential misclassification) bias in the odds ratio, where:

$$\text{Normalized bias} = \frac{\text{apparent odds ratio} - \text{true odds ratio}}{\text{true odds ratio} - 1}$$

Under ideal conditions, an investigation of the effects of misclassification errors would begin with estimates of the sensitivity and specificity of both disease and exposure measures. Some quantitative and much impressionistic data are available on the validity of data on diarrhoea collected through routine surveillance. In general it is believed that the validity of such information is low due to the problems of defining and confirming cases (Chen, 1980) and it has been shown (Martorell et.al., 1976) that sensitivity drops rapidly as recall period lengthens, with over 20% of diarrhoeas not being reported even if the household is visited by a trained worker every three days. Although investigators have frequently hypothesized that the quality of information on exposure status (e.g. use of an improved water source) is poor, no systematic investigations of the sensitivity and specificity of such information have been conducted.

Accordingly, an investigation of bias in the odds ratio arising from misclassification errors cannot be based on reliable estimates of the sensitivity and specificity of the disease and exposure variables, and it becomes necessary to investigate the magnitude of such biases over a plausible range of misclassification parameters.

a. Non-differential misclassification:

In conducting such an analysis it is initially assumed that misclassification is "non-differential," that is, the probabilities of misclassification of disease status are identical for both exposed and unexposed groups, and the probabilities of misclassification of exposure status are identical for both diseased and disease-free groups.

The first finding that emerges from this analysis is that, under the above assumption of non-differential misclassification, in all cases the observed odds ratio is biased towards unity (Newell, 1962). That is, wherever non-differential misclassification of either disease or exposure (or both) is present, the observed effect of, say, the use of improved water supply facilities will be less than the true effect.

The magnitude of such underestimation of the true effect depends on the sensitivity and specificity of the measures of disease and exposure, on the frequency of disease and on the true odds ratio. Under the assumption of non-differential misclassification, for particular values of disease frequency (10% in the unexposed population) and true odds ratio (1.5), the effects of simultaneous variations in the sensitivity and specificity of the disease measure are shown on Figure 2(a), and the effects of simultaneous variations in the sensitivity and specificity of the exposure measure shown on Figure 2(b).

In interpreting these results it is useful first to examine the effects of each of the four misclassification errors individually. From Figures 2(a) and (b) and Table 2 it is evident that (as has been shown by others (e.g. Copeland et.al. 1977)), the underestimation in the odds ratio is particularly sensitive to the specificity of the outcome measure. For the assumed values of disease frequency (10%) and true odds ratio (1.5), the odds ratio is underestimated by 47% (using the normalized measure of bias) if the specificity of the disease classification is 90% and all other sensitivities and specificities are 100%.

In practice, of course, errors in each of the categories might be expected to occur jointly. To illustrate the joint effects of such errors, for the same values of disease frequency and true odds ratio, Table 3 shows the underestimation in the odds ratio when all four factors are set at the same level. From Table 3 it is evident that even when the sensitivity and specificity of each of the measures is high (say 95%) the overall effect is to seriously underestimate (normalized bias of 41%) the impact of exposure on disease.

As discussed in the section on sample sizes, however, the frequency of diarrhoea in populations is not the 10% assumed in the above analyses, but is typically about 4% for all diarrhoeas and about 0.4% for severe diarrhoeas. As indicated on Figure 2(c), the normalized bias increases substantially as the frequency of disease decreases. Figure 2(c) also shows that, over the range of odds ratios likely to be of interest in water supply and sanitation studies, the bias is affected little by the true odds ratio (although, as shown by Gladen and Rogan (1979), biases increase substantially when true odds ratios are very large).

(b) Differential Misclassification

The above analysis is based on the assumption that misclassification is non-differential. Since violation of this assumption gives rise to estimates of the odds ratio which may be biased either towards or away from the null value, it is necessary to describe the types of differential misclassification which might arise in a HIE of a water supply and sanitation project and to assess the consequences of such biases.

(i) Differential misclassification producing POSITIVE biases:

Misclassifications which shift individuals towards "a" and "d" in the classification table.

Misclassifications which shift individuals diagonally from "d" to "a" and from "b" to "c".
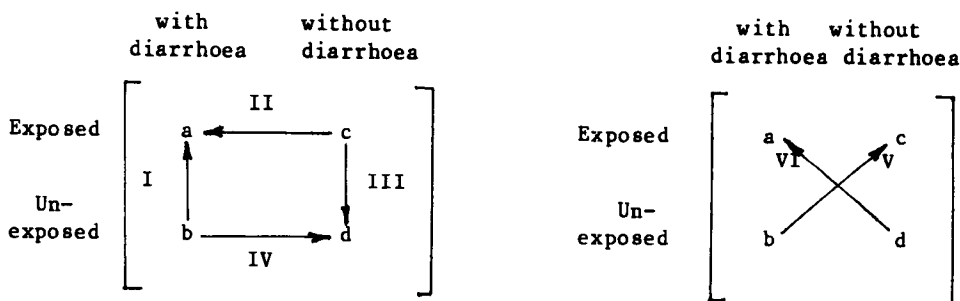
The causes and typical effects of the six different types of differential misclassifications which give rise to overestimates of the odds ratio are presented on Table 4.

(ii) Differential misclassification producing NEGATIVE biases:

(a) Misclassifications which
    shift individuals towards
    "b" and "c" in the
    classification table.

(b) Misclassifications which shift
    individuals diagonally from "a"
    to "d" and from "c" to "b".
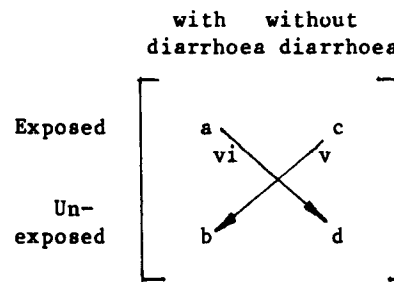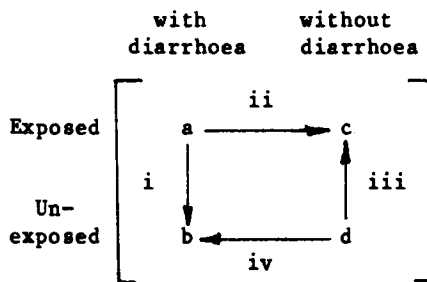
The causes and typical effects of the six different types of differential misclassifications which give rise to low odds ratios are presented in Table 5.

(iii) Discussion

In the presence of the differential misclassification it is possible to have an estimate of the odds ratio which is biased either up or down. From an examination of the causes of these differential misclassifications in Tables 4 and 5 it appears that there is a limited set of related differential misclassifications which might arise with some regularity in actual field studies of the impact of water supply and sanitation conditions on diarrhoeal disease. Specifically, it appears that the following four conditions might be met in some practical situations:

Condition A: Poor people tend to use poor quality facilities and rich people use better facilities;

Condition B: Poor people tend to have more diarrhoea than rich people;

Condition C: Those who use poor facilities will tend to report using better facilities; and

Condition D: Poor people, being more used to experiencing diarrhoea, will tend to under-report diarrhoea.

It is thus possible that differential misclassifications of Types (i), (ii) and (vi) might occur in actual field studies. Of particular importance is the observation that, while differential misclassifications which lead to **overestimates** of the OR can not, in general, be ruled out, for the most plausible types of differential misclassification the estimated odds ratio will, again, **underestimate** the true odds ratio.

(c) Summary on Misclassification

In summary, for the types of misclassification, the levels of specificity and sensitivity of the disease and exposure measures, and the levels of disease and exposure frequency likely to be encountered in practice, the odds ratio will usually be substantially underestimated. Since virtually all published studies on the effect of water supply and sanitation conditions on diarrhoeal disease are quasi-experimental or cohort studies, it is impressive that a review of the literature shows that in many cases positive impacts have been demonstrated (Esrey et al. 1985).

3.1.4. Problem 4: Ethical problems

In addition to these methodological problems, there are other serious concerns with the use of quasi-experimental designs where they involve conscious manipulation of the availability of water supply and sanitation facilities. Although there are sharp disagreements about the magnitude of the effect of water supply and sanitation conditions on health, there is general agreement that such positive effects exist. Insufficient attention has been paid to the ethics of conducting "trials" with treatments of accepted efficacy such as water supply and sanitation. Certainly by the standard criterion applied to the ethics of drug and vaccine trials -- namely that a clinical trial is ethical only if the proposed treatment is promising and if there is a reasonable doubt about its efficacy under field conditions -- these "trials" would be considered unethical. While advantage may be taken of the fact that large water supply and sanitation programmes are necessarily carried out in phases, it is evident that under such conditions the allocation of communities to "treatment" and "control" groups would be on the basis of political and other criteria which are different from the scientific procedures required for valid quasi-experimental designs.

3.1.5. Problem 5: Time and resources required for the study

A final constraint on the use of the most valid of the quasi-experimental designs (those which rely on both internal and external comparisons) is that the evaluations cannot be initiated only after it has been verified that a particular project is performing well and is being utilized. Rather, such a study has to be initiated prior to the initiation of the project itself in order to establish that the diarrhoea rates in the intervention and control groups were not different prior to the project. These studies often are of projects which are neither performing well nor being utilized, they take several years to complete and are extremely expensive, with costs sometimes running over a million dollars per study (IBRD, 1976).

3.2 CONCURRENT COHORT DESIGNS

A concurrent cohort study (sometimes called a "prospective" or "longitudinal" study) involves identifying a population in which there are individuals or groups with differing levels of exposure (for instance to contaminated water), and following the population forward in time to determine and compare disease incidence. This method has seldom been used in analyzing the health impacts of water supply and sanitation projects.

Except for the method for controlling for confounding, concurrent cohort studies are similar to the quasi-experimental studies discussed above, and suffer from many of the same problems. First, the required sample sizes are very large (being identical to those required in the quasi-experimental designs). Second, with regard to the problem of misclassification, it is instructive to examine separately the likelihood of a given level of misclassification and then the consequences of that level of misclassification. The information on disease status is, as in a quasi-experimental study, collected through household surveillance. The likelihood of misclassification of disease status is thus similar in concurrent cohort and quasi-experimental studies. In cohort studies data on exposure is collected through household surveys. While the sensitivity and specificity of such information is often not high (Blum and Feachem, 1983), it is probably generally somewhat better than the exposure information in quasi-experimental studies (in which the exposure status of individuals is often assumed). Since

the consequences of a given level of misclassification are (for the same
disease frequency and odds ratio) identical in concurrent cohort and quasi-
experimental studies, the bias in the odds ratio due to misclassification may
be slightly less in concurrent cohort studies as carried out in this field
than in quasi-experimental studies. Finally, because these are purely
observational studies (with no manipulation of availability of water supply or
sanitation services), the ethical dilemmas faced in quasi-experimental studies
are reduced.

Because of the large sample sizes required and the likelihood of bias,
concurrent cohort studies are generally not appropriate for evaluating the
impact on diarrhoea of water supply and sanitation projects. However, there
are situations in which study designs of this sort might be used. For
instance, where well-designed concurrent cohort studies are being carried out
for other purposes, and where it is possible to broaden the scope of such
studies to include water supply and sanitation considerations at modest
expense, such opportunities obviously should be exploited.

Finally it should be noted that in both quasi-experimental and cohort
designs the requirement for large sample sizes may be relaxed by studying
groups having an exceptionally high frequency of diarrhoea, such as families
in which there is an identified index case. The internal validity of such
studies is often high (i.e. the conclusions are valid for the particular study
question in the particular population), and useful findings have emerged from
studies of this type, such as those in Bangladesh of the effect of handwashing
on secondary transmission of shigellosis (Khan, 1982) and of the effect of
household water treatment on the secondary transmission of cholera (Khan
et.al., 1985). The great drawback of such studies is that the external
validity of such studies is generally low (i.e. it is difficult to extrapolate
the findings to address broader questions in the general population). In the
present context, for instance, the effect of contaminated water on
transmission of diarrhoea within a community is likely to be quite different
from the effect of contaminated water on transmission of diarrhoea within a
family.

3.3 HISTORIC DESIGNS

Sometimes it is possible to use existing records to determine the status of members of a population with respect to some exposure at some time in the past, and also to determine the frequency of diarrhoea at some subsequent time. A "historic quasi-experimental" study is one in which a comparison is made between the diarrhoea experience in served and unserved communities, on the assumption that the communities are similar in all respects other than the intervention. Similarly, a "historic cohort" study is an analysis of the effect of past water supply and sanitation conditions on health outcomes which have already been recorded, with the effect of confounding variables, on which information must also be available, taken into account by statistical means. Both studies are thus similar in principle to the "concurrent" versions discussed above, in that they proceed from exposure to subsequent development of disease.

Because information on exposure is generally collected from secondary sources, the validity of exposure data would not usually be high. Since the consequences of a given level of misclassification are (for the same disease frequency and odds ratio) as shown in Tables 2 - 5 and Figure 2, the estimate of the odds ratio in a historic quasi-experimental or cohort study will usually be more biased that the odds ratio in the concurrent versions of these designs.

The obvious and great constraint on such studies is the availability and validity of the necessary records. It is striking, however, that over the past eight years several such studies, addressing, among other questions, the effects of improved water supplies and sanitation in nineteenth century Europe and North America, have been conducted by demographers and economic historians (e.g., Condron and Cheney, 1982; Preston and Van de Walle, 1976; Higgs and Booth, 1979). While similar opportunities in developing countries are limited (IBRD, 1976), there are certain settings where rich sets of longitudinal data on diarrhoea morbidity are available. Khan's study of tubewell use and cholera in Bangladesh (Khan et al., 1981) is one example of the use of such a data set; there may be other such opportunities to be tapped at modest cost.

3.4. CROSS-SECTIONAL DESIGNS

All of the above studies require observations at more than one point in time. In cross-sectional studies, by contrast, measurements of exposure and disease status are made at a single, common point in time. Because of the simultaneous nature of the measurements of exposure and disease, in most settings cross-sectional studies are restricted to the generation of hypotheses and cannot be used for testing hypotheses. However, where the exposure status of an individual is more or less permanent (as is generally the case with exposure to inadequate water supply and sanitation conditions) then an individual's current exposure status is an adequate measure of previous exposure status, and a cross-sectional study can be used to test causal hypotheses (MacMahon and Pugh, 1970).

With regard to misclassification biases, it it is instructive, as before, to consider separately the likelihood of a given level of misclassification and the consequences of that misclassification on the estimate of the odds ratio. Since disease information is usually collected similarly in quasi-experimental, concurrent cohort and cross-sectional studies, the likelihood of a given level of misclassification of disease status is similar in all three designs. However, because in a cross-sectional study it is assumed that present and past exposures are identical, exposure information is usually considered to be less valid in cross-sectional studies than in concurrent cohort studies. In the case of an investigation of the effect of water supply and sanitation facilities on diarrhoeal disease, however, the relevant exposures occurred in the very recent past and thus the recall-period for exposure is short. In this particular case, then, the validity of exposure information in cross-sectional studies is similar to the validity of exposure information in a concurrent cohort study. As before, for the same population, the consequences of different levels of misclassification are as shown in Tables 2 - 5 and Figure 2. The problem of biased odds ratios due to misclassification in cross-sectional studies is thus similar to the problem in concurrent cohort studies.

A characteristic of cross-sectional studies is that the outcome variable, viz. disease prevalence, is affected not only by the incidence of the disease but also by duration. Because the objective of most epidemiologic studies is to measure changes in incidence, this is generally considered to be a

disadvantage. In the particular case of water supply and sanitation, however, it has been argued that the use of prevalence is an advantage since both incidence and duration may be reduced by environmental improvements and therefore prevalence may be a more responsive measure of impact than incidence (Esrey et al, 1985).

These differences aside, a cross-sectional study is similar in many respects to the quasi-experimental and cohort designs described earlier. Specifically, cross-sectional studies have similar sample size requirements, and the problems of misclassification and confounding are similar.

3.5 CASE-CONTROL DESIGNS

Unlike all of the above study designs, the case-control study (also known as a "case-history" or "retrospective" study) proceeds not from cause to effect but from effect to cause. For example, in a community which has improved and unimproved sources of water, individuals who report to a clinic with diarrhoea (the cases) may be compared with individuals who report to the clinic with respiratory infections (the controls). Cases and controls are compared with respect to the sources of water which they have used. The odds of cases using unimproved water may be divided by the odds of controls using unimproved water to obtain an odds ratio. The significance of this odds ratio may be tested and it may be used to estimate the relative risk of diarrhoea among users of unimproved water compared to users of improved water. For rare diseases, the odds ratio is a good estimate of the relative risk. If, for instance, the prevalence of severe diarrhoea amongst those using poor quality water is 0.4%, if the numbers using poor and good quality water supplies are equal, and if the relative risk is 1.500, then the odds ratio is 1.502.

3.5.1. Potential Advantages of the Case-Control Method

In evaluating the impact of water supply or sanitation facilities on diarrhoea morbidity, the case-control approach has several advantages over the quasi-experimental, cohort and cross-sectional alternatives.

First, the sample sizes are smaller, since the ratio of cases to non-cases is fixed by the investigator, making the method more efficient than other designs when the frequency of the outcome is relatively rare. For instance, if 40 percent of the study population uses, say, an improved water supply and if, as before, the study is designed to detect a 33 percent reduction in diarrhoea morbidity (i.e. the odds ratio to be detected is 1.5), then, as shown in Table 6, only 460 cases and a similar number of controls are needed in a case-control study. The numbers required in the study are independent of the frequency of occurrence of the disease in the community, and are thus the same whether mild diarrhoea, severe diarrhoea or etiology-specific diarrhoea is studied. (A more detailed discussion of sample sizes in case-control studies is presented in Section 3.5.3.)

A second attraction of the case-control method is that (as discussed in more detail on page 23) the sensitivity and specificity of the disease measure used are substantially higher than these measures in quasi-experimental or cohort studies which rely on surveillance for detection of cases. Furthermore, given the much smaller sample sizes in case-control studies, much closer attention can be paid to the quality (and thus the validity) of information on exposure. As discussed in detail later in this section, under a set of assumptions which reflect the conditions under which health impact evaluations of water and sanitation facilities would usually be undertaken, misclassification biases would generally be smaller in a case-control study than in a quasi-experimental, cohort or cross-sectional study.

A third advantage of the case-control method is that an impact evaluation using this method need be initiated only after a prior evaluation (using the Minimum Evaluation Procedure of the WHO [WHO, 1983] or some similar methodology), has demonstrated that the system is functioning adequately and that the improved facilities are being used appropriately. Fourth, as with the functioning and utilization evaluation procedures, the case-control method requires the collection of only a single round of data. Finally, results are available rapidly and the ethical problems associated with some quasi-experimental designs are avoided.

Despite the promise which the case-control method seems to hold in resolving some of the most serious problems faced in HIEs of water and sanitation facilities, these advantages remain _potential_ rather than _realized_ since the method has yet to be applied to this purpose.

3.5.2. Some Reasons for Neglect of the Method

In view of the attractiveness of the case-control method for the analysis of the health impact of water supply and sanitation programmes, why is it that the method has not been applied to this set of problems?

First, where the case-control approach is being used and more than one outcome measure (say diarrhoeal disease and nutritional status) is being monitored, then separate studies have to be conducted for each of the outcome measures. In a cohort study, by contrast, the impact on more than one outcome measure can be analyzed using a single study design.

Perhaps more important is the fact that case-control studies have long been regarded as scientifically unsound, producing results which cannot be trusted (Sartwell, 1980). Over the last twenty five years, however, as the method has come to be used extensively in analyzing chronic disease problems in developed countries (between 1956 and 1976 the number of case-control studies reported in major journals increased fourfold while the number of cohort studies declined by 50% (Sackett, 1979)), the major methodological problems have been identified and some solutions to these problems have been developed. Only recently have articles exploring the possibilities of applying the case-control method to problems of endemic infectious diseases appeared in the epidemiological literature (Hogue et.al., 1983 and Smith et.al., 1984).

It is thus not surprising that the case-control methodology has not yet been applied to HIEs of water supply and sanitation programmes or that, in the deliberations of the World Bank Expert Committee in 1975 (IBRD, 1976), it was implicitly assumed that only studies of the quasi-experimental or concurrent cohort design were appropriate in this field.

3.5.3. Sample Sizes in Case-control Studies

For the simplest case, namely when there are just two exposure categories and when there is a 90% chance of detecting the specified reduction in the odds ratio at the 5% significance level, Table 6 is used to calculate sample sizes. In practical applications in the water supply and sanitation sector, certain adjustments have to be made in specifying the sample sizes in a case-control study.

First, even though an odds ratio of 1.5 may be declared to be "statistically different from 1.0", the lower end of the confidence interval will be close to 1.0, thus casting doubt on whether any real risk had been detected. An alternative procedure for specifying sample sizes, then, might be to require that the lower end of the, say, 95% confidence interval be a specified level above the value of 1.0. Annex 2.1 and Table 7 show how the sample size needs to be increased as the permissible confidence interval is tightened.

A second modification is necessary because in most practical applications interest is not confined to just two exposure categories. Rather, interest is in the effect of several different levels of service. Annex 2.2 shows the procedure to be followed where the exposure variable is not simply dichotomous, and shows, for a realistic case in which the population uses either a traditional water supply, public standposts or household taps, that sample sizes may typically be about 20% larger than those required for the case of two exposure categories.

3.5.4. Some Problems of Applying the Case-Control Method

When estimating the odds ratio from a sample of a population, error is introduced in the estimate as a result of sampling error and bias. "Sampling error" is a consequence of variations in the possible samples of the population, and can be minimized by appropriate sampling procedures which make use of known similarities and differences within the population. Because "bias", a term used to describe the phenomenon whereby the results of a study differ systematically from the truth (Sackett, 1979), is generally more serious than sampling error in developing country settings (Casley and Lury, 1981), strenuous efforts must be taken to identify and account for possible sources of bias in a case-control study in a developing country.

Case-control studies are subject to three major categories of potential bias: due to distortions from measurement error or misclassification of subjects with respect to disease and exposure status ("misclassification bias"), due to distortions resulting from the manner in which the subjects are selected into the study ("selection biases") and due to distortions if the effect of the study factor is mixed with the effects of extraneous variables ("confounding") (Kleinbaum et.al., 1982). While none of these is simple to deal with, misclassification and selection biases will receive most attention in this discussion because they present unique problems which have to be dealt with at the design (rather than the analysis) stage.

3.5.4.1. Problem 1: Misclassification Bias

In the specific case of a study of the impact of water supply and sanitation conditions on infectious diarrhoeal disease in young children, it is possible to compare the validity of data on disease and exposure in quasi-experimental, cohort and cross-sectional studies as normally conducted in this field, on the one hand, and case-control studies, as they are envisaged, on the other. We consider the two study types in turn.

QUASI-EXPERIMENTAL, COHORT AND CROSS-SECTIONAL STUDIES
Disease Status:
   As discussed by Martorell et. al. (1976) and Chen (1980), in surveys of diarrhoeal diseases information on diarrhoea is collected by recall and there are typically a large number of false negatives. The likelihood that the sensitivity of the information is poor is therefore "very high".
   In surveys of diarrhoeal disease, a substantial proportion of mild diarrhoeas may not be caused by enteric infections (Black, 1984). Since the measure of disease status is intended to capture only those diarrhoeas due to enteric infections, there may be a large number of false positives. The likelihood that the specificity of the information on disease status is poor is therefore "very high".
Exposure status:
   Using these study designs, sample sizes are large and it is thus difficult to obtain high-quality information on actual facility use. It may therefore be expected that substantial numbers report not being exposed (i.e. using the improved facilities) when in fact they continue to use unimproved facilities. The likelihood that the sensitivity of the information on exposure status is poor is therefore "high".
   It seems unlikely that there would be many who would report not using improved facilities when in fact they are using such facilities. It is therefore probable that there are few false positives. The likelihood that the specificity of the information on exposure status is poor is therefore "low".

CASE CONTROL STUDY
   Disease status
      In a clinic-based case-control study, all prospective cases and
   controls are examined by a health professional and, in the course of
   that examination, asked whether the child is suffering from diarrhoea.
   It is highly unlikely that a child who does not have diarrhoea will be
   reported as having diarrhoea. The likelihood that the sensitivity of
   the information on disease status is poor is therefore "low".
      In surveying severe diarrhoeas which are presented at a clinic, there
   are some which are not caused by enteric infections, but the proportion
   is much smaller than for mild diarrhoeas (Black, 1984). Therefore the
   number of false positives is relatively small, and the likelihood that
   the specificity of the information on disease status is poor is
   "moderate".
   Exposure status
      In a case-control study, as in a quasi-experimental, cohort or cross-
   sectional study, data on exposure are obtained primarily through
   questionnaires administered to the mother in the home. There will thus
   be the same tendency to over-report use of improved facilities.
   However, since far fewer mothers are interviewed in a case-control
   study, it is possible to pay closer attention to getting valid
   information. The likelihood that the sensitivity of the information on
   exposure status is poor is therefore "moderate".
      As in the follow-up study, false positives are unlikely. The
   likelihood that the specificity of the information on exposure status is
   poor is therefore "low".

This information is summarized on Table 8, in which the likelihood of poor

validity of the disease and exposure measures is presented for, first, quasi-

experimental, concurrent cohort and cross-sectional studies as they are

normally conducted in this field, and, second, clinic-based case-control

studies as envisaged in this document. As shown by Table 8, one of the major

attractions of case-control studies as envisaged here, over quasi-

experimental, concurrent cohort or cross-sectional studies as normally

conducted in the field, is the reduction in the likelihood of

misclassification.

   With regard to the consequences of a given level of misclassification, it

is tempting to believe that the consequences of misclassification, especially

of poor specificity in the disease variable, might be to cause less bias in a

case-control study than a cohort study. In a cohort study, if 10% of the very

numerous undiseased individuals are wrongly classified as diseased (a

specificity of 90%) they will overwhelm numerically the small number of

diseased individuals and produce a high false-positive proportion and a large

bias in the odds ratio (Table 2). In a case-control study, a

misclassification of 10% of controls as cases will, if cases and controls are equal in number, produce a false-positive proportion of only 9.9% and consequently a much smaller bias. However, in a case-control study, cases are sampled from those apparently diseased and controls are sampled from those apparently undiseased. If the method of diagnosis were to lead to 90% specificity and a consequently high false-positive proportion in a cohort study, the cases in a case-control study would display the same high false-positive proportion. Thus, for a common diagnostic method and true disease frequency in the community, the false-positive proportion and consequent bias would be the same in a case-control and a cohort study.

In the conduct of a clinic-based case-control study of the sort envisaged in this document, it is essential that every effort be made to reduce the likelihood of misclassification. With respect to disease status, this requires that the personnel responsible for determining whether a child is eligible as a case or as a control be medically trained, have substantial diagnostic experience and be fully informed of the criteria for selecting cases and controls (see Section 3.5.4.2). With regard to exposure status, anecdotal evidence suggests that there might be some tendency for individuals to report using improved water supply and sanitation facilities when in fact they are not doing so, but only one study in the literature has examined this issue rigorously. In this study (in rural Bangladesh, Curlin et.al., 1977) 175 families were surveyed, and in not a single case did a family falsely report using tubewell water.

The conservative approach to ensuring high validity of exposure information is not to rely on interviews in the clinic but to carry out home visits to all study individuals so that reliable data on water and sanitation usage can be obtained. In the initial field tests of this methodology such a conservative approach should be taken. However, given the very great logistic advantages if all data can be collected through interviews at the clinic (and no home visits made), priority should be given to the collection of data on the validity of exposure information collected through questionnaires at the clinic. The procedure will be simple where the improved and unimproved sources can be distinguished through an objective test (such as conductivity in the Bangladesh study referred to above). In settings in which the validity of exposure information collected through questionnaires at the clinic is found to be high, in subsequent studies it will be sufficient to collect exposure data through questionnaires administered at the clinic.

In either case, care has to be taken not to introduce other sources of information bias. Specifically there is a danger that there may be a subconscious bias to score the sanitary conditions of the "cases" worse than comparable conditions in the house of a "control". Such bias is best minimized by use of a well-designed, structured questionnaire, and by emphasizing the importance of objective questioning during interviewer training. If home visits are conducted, the field workers should be unaware of whether they are visiting the home of a case or the home of a control.

3.5.4.2. Problem 2: Selection Bias

In addition to the distortions which may arise from misclassification biases, the estimates of effect (such as the odds ratio) may also be biased because of the manner in which subjects are selected and because of confounding. Although the effects of selection biases sometimes appear to be similar to those of confounding, these are logically different problems and should be treated differently. Accordingly it is useful to first define the differences between selection biases and confounding.

To produce confounding a variable must be associated, in the subjects actually studied, with the exposure under study and, independent of this association, must also be a risk factor for disease. Thus, "confounding in a case-control study is the same phenomenon as confounding in a follow-up study. It arises from associations in the causal network in the underlying population and cannot be removed by appropriate study design alone. An essential part of the analysis is examination of possible confounding effects and how they may be controlled" (Breslow and Day, 1980).

Selection bias, by contrast, is not a bias which arises because of underlying causal relationships which exist among the variables in a population, but is a bias which arises because of the way in which cases and controls are recruited into a study. The case-control methodology assumes that under the null state (of unit true odds ratio) cases and controls would have an equal chance of having been exposed to the risk factor of interest. Avoiding systematic violations of this assumption, the problem of selection bias, is the "truly large problem of the case-control study" (Cole, 1979).

(a) Selecting Cases and Controls

A first issue in defining eligibility for recruitment of cases and controls is how to deal with individuals who report to the clinic with non-diarrhoeal diseases which are known to be associated with water supply and sanitation conditions. In Annex 3 an illustrative example is presented for assessing the effect of different procedures for dealing with water- or sanitation-related diseases. From this example the following general rules can be derived.

Children who come to the clinic with diarrhoea as the primary complaint are eligible as cases. If other water- and sanitation-related diseases (such as typhoid fever, hepatitis A and a variety of nematode infections) are secondary complaints, this does not disqualify the child as a case. In recruiting controls, those children who come to the clinic with a water- or sanitation-related disease as the primary complaint are not eligible for recruitment as controls. However, children who come to the clinic primarily because they are suffering from one of the eligible control diseases (such as an acute respiratory infection) and are found to have a water- or sanitation-related disease as a secondary complaint, should be included in the control group.

A second issue is whether individuals can be recruited into the study more than once. There are four questions which need to be answered, viz. can cases become cases again, can controls become controls again, can a case become a control and can a control become a case. It has been shown (Greenland and Thomas, 1982) that to obtain unbiased estimates of the odds ratio, the traditional "cumulative density" sampling procedure (in which controls are selected from those still unaffected by the disease at the end of the study period) should, where possible, be replaced by the "incidence density" sampling procedure (in which controls are selected for each case from those at risk at time of onset of the case, and in which a person can be sampled several times during the course of the study). In the present context this means that cases can become cases again, and controls can become controls again, providing the clinic visit which leads to the second recruitment is not simply a second visit for the same episode of illness. (In cases of repeat recruitment it is essential that careful clinical histories be taken so that subsequent decisions can be made about inclusion or exclusion of the second event. These repeats should not constitute more than a very small proportion of all recruitments.)

With rare diseases (such as diarrhoea severe enough to be brought to the clinic), the problem of how to treat controls who later become cases, or cases who are later selected as controls, is of academic rather than practical interest. However, on intuitive grounds it appears more logical to delete controls who become cases from the control group and include them only in the case group, and to exclude cases from eligibility for later selection as controls. Since this scheme has the advantage that it will tend to enhance the odds ratio measured (if it is greater than unity) and will thus increase the power of the study, it is recommended that this procedure be followed (P.G. Smith, personal communication).

(b) Other sources of selection bias

There are other potential sources of selection bias which are specific to, and particularly important in, case-control studies of the impact of water supply and sanitation conditions on diarrhoeal morbidity. The problem arises when the probability that a child with diarrhoea will be brought to the clinic is affected by whether or not the individual is exposed to, say, a poor water supply. This will happen when, first, the probability of reporting is affected by the level of a particular variable (such as distance from the clinic or socio-economic status) and, second, the particular variable (such as distance or socio-economic status) is not uniformly distributed amongst exposed and unexposed.

To anticipate the argument which will be made over the next several pages, it transpires that this type of selection bias is (in most cases) not difficult to deal with in practice, since it simply requires that the people who are recruited as controls meet certain eligibility criteria. The argument, however, is fairly complex and will be developed in stages. First we will deal with the most straightforward of this class of problems, that posed by the relationship between distance and the reporting of diarrhoea, on the one hand, and distance and exposure, on the other. Subsequently we will expand the analysis to deal with other variables which pose additional problems.