DISTRIBUTION: GENERAL

WHO/CDD/EDP/90.2
WHO/ARI/RES/90.4

ORIGINAL: ENGLISH

DIARRHOEAL DISEASES CONTROL PROGRAMME
PROGRAMME FOR CONTROL OF ACUTE RESPIRATORY INFECTIONS

# OUTCOME MEASURES IN PROSPECTIVE STUDIES OF CHILDHOOD DIARRHOEA AND RESPIRATORY INFECTIONS: CHOOSING AND USING THEM

by
Simon N. Cousens
Betty R. Kirkwood

WORLD HEALTH ORGANIZATION

WHO/CDD/EDP/90.2
WHO/ARI/RES/90.4

# OUTCOME MEASURES IN PROSPECTIVE STUDIES OF CHILDHOOD DIARRHOEA AND RESPIRATORY INFECTIONS: CHOOSING AND USING THEM

by
Simon N. Cousens[1]
Betty R. Kirkwood[1]

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

## CONTENTS



## SUMMARY

Children in developing countries suffer repeated episodes of diarrhoea and respiratory morbidity. A number of outcome measures are available for use in prospective studies of these diseases: number of episodes, number of days of morbidity, average duration of each episode, severity of each episode. While the analysis of binary outcomes such as ill/not ill or survived/died is discussed extensively in the epidemiological literature, less attention has been paid to other types of outcome. In this paper we explain how the choice of an appropriate measure of outcome is governed by the research question posed, and we examine some of the basic statistical techniques available for analysing outcomes, such as number of episodes. A worked example is presented using data on diarrhoea morbidity.

WHO/CDD/EDP/90.2
WHO/ARI/RES/90.4

# OUTCOME MEASURES IN PROSPECTIVE STUDIES OF CHILDHOOD DIARRHOEA AND RESPIRATORY INFECTIONS: CHOOSING AND USING THEM

by
Simon N. Cousens[1]
Betty R. Kirkwood[1]

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

## CONTENTS



## SUMMARY

Children in developing countries suffer repeated episodes of diarrhoea and respiratory morbidity. A number of outcome measures are available for use in prospective studies of these diseases: number of episodes, number of days of morbidity, average duration of each episode, severity of each episode. While the analysis of binary outcomes such as ill/not ill or survived/died is discussed extensively in the epidemiological literature, less attention has been paid to other types of outcome. In this paper we explain how the choice of an appropriate measure of outcome is governed by the research question posed, and we examine some of the basic statistical techniques available for analysing outcomes, such as number of episodes. A worked example is presented using data on diarrhoea morbidity.

## 1. INTRODUCTION

Diarrhoeal and acute respiratory diseases are both major causes of morbidity and mortality among children in developing countries. Each is responsible for several million child deaths annually (UNICEF, 1990). In addition, they are epidemiologically similar in a number of ways. Young children may suffer from repeated episodes of diarrhoea and from repeated respiratory infections. Both are clinical manifestations of infections due to a range of pathogens rather than a single disease entity. Both are diseases of poverty. These epidemiological characteristics have a number of implications for the design, analysis, and interpretation of studies of childhood diarrhoea and respiratory infections. When studying chronic diseases, which are relatively rare in most settings and normally occur only once, a person's health status can usually be recorded as either diseased or not diseased. When conducting follow-up studies of common, recurrent diseases, such as childhood diarrhoea and acute respiratory infections, however, a simple classification such as this is frequently not sufficient since during a follow-up period of two years, almost every child will suffer from at least one episode of the disease. Thus, other measures of the disease burden, referred to hereafter as outcome measures, are needed in the study of these diseases. In this paper we discuss the different choices of outcome measure that are available in follow-up studies of diarrhoeal and respiratory illness and show how the appropriate choice is dependent upon the research question posed. Some of the measures available are binary, qualitative variables (e.g., survived/died). Binary outcomes are widely used in epidemiological studies and have been much discussed in the literature (e.g., Kleinbaum et al., 1982; Rothman, 1986). Others of the measures available are of types less frequently used, and there is consequently less literature on their analysis. In this paper we examine some of the basic statistical techniques available for the analysis of non-binary outcomes.

## 2. CHOICE OF OUTCOME MEASURE: SOME OPTIONS

Four main categories of outcome measure are available for use in prospective studies of diarrhoeal and respiratory diseases:

(i) the number of episodes of illness experienced by each child (0,1,2,..),

(ii) the average duration of each episode experienced by the child (e.g., 3.2 days),

(iii) the total number of days of illness suffered by each child (0,1,2,..),

(iv) the severity of each episode.

This last category, severity, can be measured in a number of different ways. For example, it could be measured by the binary variable survived/died. Other examples are use of degree of dehydration (none/some/severe) as the outcome measure in a study of diarrhoea morbidity, or use of the presence/absence of chest indrawing as the outcome measure in a study of respiratory morbidity.

It should be noted that, while the first three of these measures (number of episodes, mean duration, number of days) are all quantitative (can be expressed as a number), the fourth (severity) is more qualitative. The nature of the outcome measure, quantitative or qualitative, has implications for the analysis of the data, a point to which we shall return.

Although the measures outlined above are closely related, they are not equivalent. For example, the number of days of illness experienced by a child is clearly closely related to the number of episodes suffered by that child. The two are, in fact, linked through the duration of each episode (number of days = number of episodes x average duration). There may, however, be factors which affect the duration of an episode but which do not influence the number of episodes suffered by a child, and vice versa. Each of the various outcome measures reveals something different about the disease under study and therefore one cannot say that one measure is "better" than the others. In some situations the number of episodes per child will be the most appropriate measure to use, whilst in others it will be the average duration of each episode, and in yet others some measure of the severity of the episode. Suppose, for example, that a child's nutritional (anthropometric) status affects the duration of respiratory infections (children with poor nutritional status suffering longer episodes), but does not affect the number of times the child suffers such an infection. Then, a well-designed study using average duration per episode or total days of illness as the outcome measure of interest would stand a good chance of detecting an association between nutritional status and respiratory morbidity. On the other hand, a study using number of episodes as its outcome measure could not expect to detect any association between nutritional status and respiratory disease.

In summary, there are a number of outcome measures that can be used in prospective studies of childhood diarrhoea and respiratory infections. In any particular situation, the measure that one chooses to use will be determined by the research question that one is trying to answer. In order to choose the most appropriate outcome measure it is therefore essential to understand and define clearly the research question(s) being posed.

## 3. SOME EXAMPLES OF RESEARCH QUESTIONS

To illustrate the way in which the research question to be answered determines the choice of outcome, and to provide examples of different questions that might be posed, consider the general question of whether or not a relationship exists between measles vaccination and respiratory morbidity and mortality. Some specific questions that might then be raised are discussed below.

Question 1

**"Can measles vaccination reduce the incidence of respiratory infections in young children?"**

In order to answer this question, one might follow a group of children, some vaccinated and others not, to see whether the children who were vaccinated suffered fewer episodes of respiratory morbidity than those who were not vaccinated. Clearly, in this case, the outcome measure of interest is the number of episodes suffered by each child. This measure is quantitative (a number: 0, 1, 2,...) and the unit of observation is the child. All children contribute information to the study, regardless of whether or not they become ill while it is under way.

Question 2

**"Can measles vaccination reduce the risk of mortality associated with an episode of respiratory morbidity?"**

This question is concerned with the course of the illness once the child has become ill. To study it, one might follow a group of children, identifying episodes of respiratory morbidity as they occur, and then monitoring the outcome, recovered or died, of each episode. (For the purposes of this discussion, we ignore any possible ethical problems surrounding such a procedure.) This procedure differs in two ways from that employed to answer Question 1: first, in the nature of the outcome measure used, and

second, in the unit of observation used. In this example the outcome measure is a binary, qualitative one, survived or died, rather than a quantity that can easily be added to or subtracted from other similar quantities. Other examples of qualitative outcomes already mentioned are: no dehydration versus some or severe dehydration, no chest indrawing versus chest indrawing. Some children, when followed over a period of time, will suffer several episodes of diarrhoea/respiratory morbidity. The qualitative nature of these outcomes makes it difficult to produce a summary measure for each child. For example, suppose a child had three episodes of respiratory morbidity during one of which s/he suffered chest indrawing. How does this child compare with another who suffered five episodes two of which were associated with chest indrawing, or a child who suffered one episode without chest indrawing? A more appropriate unit of observation than the child her/himself may be the individual episode. However, since the outcomes of different episodes in the same child may not be independent, the inclusion in the analysis, as separate observations, of two or more episodes occurring in the same child may introduce statistical problems. A further point to note is that whereas for Question 1 all children contribute information regardless of how many episodes of diarrhoea or respiratory morbidity they suffer, for this second question a child who does not suffer an episode of respiratory morbidity during the study provides no information to assist us in judging whether or not measles vaccination affects the risk of mortality from an episode of respiratory morbidity - i.e., in studies of this type of question, only children who become ill contribute useful information.

Question 3

**"Can measles vaccination reduce the duration of episodes of respiratory morbidity?"**

This question is concerned with the course of an episode (how long it lasts) once it has begun. Thus, as with Question 2, only children who suffer at least one episode of morbidity contribute any information to the study. In contrast with Question 2, however, the outcome measure used this time (duration) is quantitative, and thus for each child we may easily add the durations of each episode and divide by the number of episodes to obtain a single measure for each child (the mean duration). In this instance we may use the child as the unit of observation (thus avoiding the problem of non-independence which may arise when each episode is treated as a separate observation) and the quantatitive outcome measure of mean duration.

Question 4

**"Can measles vaccination reduce respiratory mortality in young children?"**

At first sight this question appears to be the same as Question 2. Both are concerned with whether or not measles immunization can prevent respiratory mortality. There is, however, a difference between them.

Question 2 is concerned with the risk of mortality associated with an individual episode of morbidity - i.e., once the child has become ill, does measles immunization reduce the risk that the child will die from that episode? Question 4 is more general. Measles immunization may reduce a child's risk of respiratory mortality by reducing the risk of death associated with a particular episode of morbidity (the subject of Question 2) or by reducing the likelihood that the child will become ill in the first place (the subject of Question 1). To answer this question we might follow a group of (initially) healthy children to determine how many die from respiratory infections, and whether there is any difference in the mortality rates among those immunized against measles and those not. Note that all children contribute information regardless of whether or not they suffer any episodes of respiratory morbidity. The unit of observation is the child, and for each child the outcome measure is the binary, qualitative one of survived/died.

An additional point concerning these questions is worth noting. Questions 1 and 2 are subdivisions of Question 4 and it may be tempting to believe that by answering Questions 1 or 2 we will also be able to answer Question 4 - that if the answer to either

of the first two questions is yes, then the answer to Question 4 must also be yes. This is not so. For example, if measles immunization reduced respiratory morbidity by preventing the mildest episodes from occurring, but had no impact on the more severe, life-threatening episodes, the answer to Question 1 would be yes, but the answer to Question 4 no. While from a scientific viewpoint Questions 1 and 2 appear to be more precise, from a public health point of view the more general question, Question 4, may be the most important one to answer.

Question 5

**"Can measles vaccination reduce the prevalence of respiratory morbidity in young children?"**

The prevalence of respiratory morbidity is equivalent to the number of days of illness divided by the number of days of observation and depends upon both the number of episodes experienced and the duration of each episode. Thus, the appropriate (quantitative) outcome measure is the number of days of respiratory morbidity experienced by each child (adjusted if necessary for period of observation). All children contribute information and the unit of observation is the child.

## 4. STATISTICAL TECHNIQUES FOR THE ANALYSIS OF QUANTITATIVE OUTCOMES

Binary outcomes such as healthy/diseased or survived/died are widely used in epidemiological studies and the statistical techniques for analysing data of this type are well known (Kleinbaum et al., 1982; Rothman, 1986). The same applies to outcomes that are continuous and more or less normally distributed. Some of the outcome measures that we have discussed do not belong in either of these two categories and have received less attention in the epidemiological literature. We now examine some of the statistical techniques that are available for the analysis of three non-binary, non-normal outcomes:

(i) the number of episodes of illness experienced by each child (0,1,2,..),

(ii) the average duration of each episode experienced by the child (e.g., 3.2 days),

(iii) the total number of days of illness suffered by the child (0,1,2,..).

We illustrate approaches to the analysis of these measures with a worked example.

### 4.1 Description of the data

Throughout the worked example we use data on diarrhoeal morbidity among 600 children (Annex). All the children were aged 6-23 months at the start of the study and were visited weekly for a period of 12 weeks. In addition to the data on diarrhoea morbidity, data are available on the sex of the child, on her/his age at the start of the study, and on whether or not s/he was weaned before the age of 4 months.

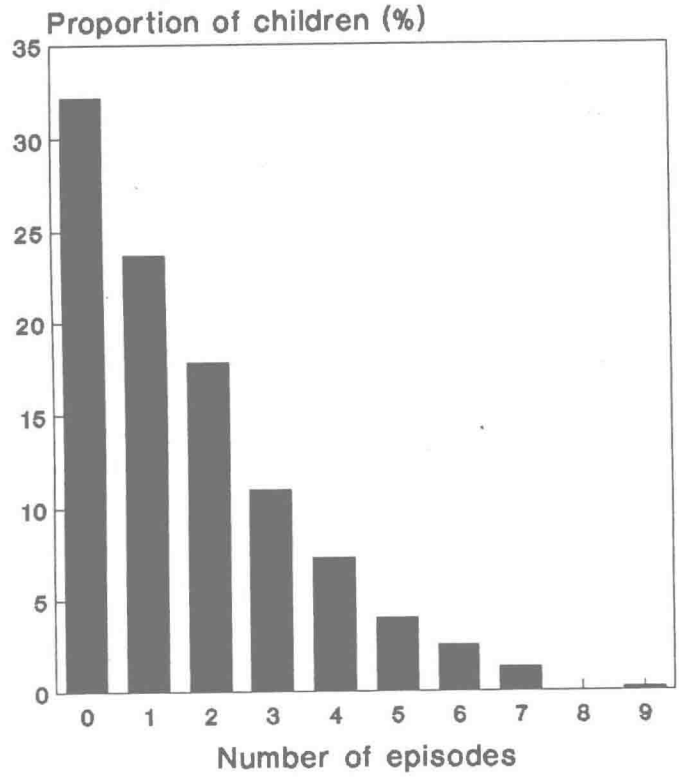
Diarrhoea (and respiratory) morbidity among young children is, typically, strongly associated with age. Usually, diarrhoea morbidity is greatest between the ages of 6 months and 2 years and decreases thereafter (Snyder and Merson, 1982). Our data set includes only children in this high-risk age group. By restricting the children to this relatively narrow range we reduce (but do not remove entirely) the age-dependent variability of diarrhoea rates in our study population and, for the sake of simplicity, we ignore the effect of age in the following analyses.

## 4.2 Analysis of the number of episodes

Consider the question "Does early weaning lead to an increased incidence of diarrhoea morbidity?". The outcome measure appropriate to answering this question is clearly the number of episodes suffered by each child.

Over a period of 12 weeks, a total of 1005 episodes were recorded among the 600 children included in this analysis, an average of 1.675 episodes per child. The distribution of episodes among these children is presented in Figure 1. It is clear from Figure 1 that the number of episodes suffered by each child does not follow a normal distribution: the distribution is not symmetrical but has a mode at one end (0 episodes) and a long right tail (up to 9 episodes). This has certain statistical consequences to which we shall return later.

FIGURE 1: Distribution of episodes of diarrhoea among 600 children




### 4.2.1 Categorized data

Analyses of data should always begin with simple techniques before going on to more complex and sophisticated procedures. In our exploration of the association between diarrhoea and early weaning we begin, therefore, by examining the data categorized according to the number of episodes suffered by the child. Before deciding on the categories to use, we examine more closely the distribution of episodes of diarrhoea among the children (Figure 1 and Table 1).

**TABLE 1. Distribution of episodes of diarrhoea among 600 children aged 6 to 23 months**

| | Episodes of diarrhoea | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Total |
| Children | 193 | 142 | 107 | 66 | 44 | 24 | 15 | 8 | 0 | 1 | 600 |

In deciding what categories to use, we bear in mind some simple principles:

- Children who suffered no diarrhoea should be kept separate from children who did suffer diarrhoea (or a respiratory infection).
- Each category should include a "reasonable" number of children.
- The number of categories should not be too large.
- The width of each interval/category should, whenever possible, be equal.

In Table 2 we present one way of categorizing the data. Other categorizations are possible and would be reasonable. For example, we might treat children suffering three episodes of illness as one category and have as our last category children suffering four or more episodes.

**TABLE 2. Distribution of episodes of diarrhoea according to weaning history among 600 children aged 6 to 23 months**

| | No. of episodes of diarrhoea | | | | |
|---|---|---|---|---|---|
| Group | 0 | 1 | 2 | 3+ | Total |
| Weaned early | 31<br>(16%)<br>(26%) | 19<br>(13%)<br>(16%) | 23<br>(21%)<br>(19%) | 47<br>(30%)<br>(39%) | 120<br>(20%) |
| Not weaned early | 162<br>(84%)<br>(34%) | 123<br>(87%)<br>(26%) | 84<br>(79%)<br>(18%) | 111<br>(70%)<br>(23%) | 480<br>(80%) |
| | 193 | 142 | 107 | 158 | 600 |

In Table 2 we present two sets of percentages. The top set are column percentages. Thus the 16% in the top left-hand corner of the table indicates that 16% of children who suffered no episodes of diarrhoea had been weaned early. In contrast, 30% of children who suffered three or more episodes had been weaned early. These percentages thus describe the risk of exposure given the number of episodes of illness. The lower set of figures represent row percentages. Thus, the 26% in the top left-hand corner indicates the probability of suffering no episodes of diarrhoea given that the child was weaned early. The figure of 23% in the bottom right-hand corner indicates the risk of suffering three or more episodes given that the child was not weaned early. These figures, therefore, indicate the proportion of children in each exposure category who suffer a given number of episodes of illness. Looking at the top set of percentages across the

top row of Table 2, there appears to be a trend: the more episodes of diarrhoea a child suffers the more likely it is that s/he was weaned early. The lower set of percentages are not particularly helpful when looking for a trend, but are important in describing the distribution of disease.

There are several possible explanations for the trend we have observed:

(i) early weaning increases a child's risk of diarrhoea;

(ii) children who suffer a lot of diarrhoea are more likely to be weaned early;

(iii) another factor (or factors) associated with early weaning increases a child's risk of diarrhoea (confounding);

(iv) bias in the selection of children for the study or in the recording of information on exposure (early weaning) and/or disease (diarrhoea) led to the observed association;

(v) there is no underlying association between early weaning and diarrhoea, and the one we have observed is due to chance.

The last of these explanations is the most straightforward to assess. We may perform a statistical test to determine how likely it is that we would observe such a trend by chance (Kirkwood, 1988). For the data in Table 2, testing for a trend results in a chi-squared statistic

$$X^2 = 11.83.$$

In the absence of any underlying trend, this statistic is distributed as a chi-squared random variable with 1 degree of freedom. The probability of obtaining such a result by chance is less than one in a thousand ($p<0.001$). This is, therefore, strong evidence that an underlying trend towards increased diarrhoea in children weaned early exists. We conclude that it is very unlikely that (v) above explains the association we have observed.

Table 2 also enables us to estimate <u>approximately</u> two important epidemiological measures: the incidence rate ratio of diarrhoea in the two groups and the proportion of all episodes of diarrhoea that are associated with early weaning. First we estimate the total number of episodes that occurred in each group. In order to do this we need to make an assumption regarding the average number of episodes suffered by children in the last category (3+ episodes) (this is where the approximation comes in). We assume an average figure of four episodes. Then we obtain the following estimates:

Number of episodes among children weaned early = 19x1 + 23x2 + 47x4
= 253

Number of episodes among children not weaned early = 123x1 + 84x2 +111x4
= 735

These estimates produce a total of 988 episodes, slightly below the 1005 actually observed. Using them, we can estimate the <u>incidence rate ratio</u> of diarrhoea in the two groups:

$$\text{Incidence rate ratio} = \frac{\text{incidence rate in exposed}}{\text{incidence rate in unexposed}}$$

$$= \frac{253/120}{735/480} = 1.38$$

We interpret this as indicating that children who were weaned early suffered 1.38 times as many (38% more) episodes of diarrhoea as (than) other children.

In order to estimate the proportion of all episodes of diarrhoea that might be associated with early weaning, we also estimate the number of episodes we would have expected to observe if all children had suffered episodes of diarrhoea at the same rate as children who were not weaned early.

Expected number = (123x1 + 84x2 + 111x4) x 600/480

= 735 x 600/480 = 918.75

Comparing this number with the (estimated) number of episodes observed (988) shows that we observed 69.25 more episodes than we would expect to have observed if the children who were weaned early had suffered diarrhoea rates similar to those in other children. This figure constitutes 7% (69.25/988) of the total number of episodes which occurred and suggests that if early weaning is responsible for the extra episodes of diarrhoea (explanation [i] above), then by successfully educating mothers to delay weaning until after 4 months of age it might be possible to reduce the number of episodes of diarrhoea in the population by about 7%. This estimate is of great value to us in assessing the possible public health importance of early weaning.

In Table 2 we presented the data in a very simple fashion, categorizing children according to the number of episodes of diarrhoea they suffered. We used four categories. We could have presented these data in an even simpler form, a 2 x 2 table, by categorizing children according to whether or not they suffered any episodes of diarrhoea (Table 3). In doing so, we have to sacrifice some of the information contained in Table 2, a sacrifice that is not without cost.

**TABLE 3. Risk of diarrhoea according to weaning history among children aged 6 to 23 months**

| | Episodes of diarrhoea | | |
|---|---|---|---|
| Group | None | One or more | Total |
| Weaned early | 31(26%) | 89(74%) | 120(100%) |
| Not weaned early | 162(34%) | 318(66%) | 480(100%) |
| | 193 | 407 | 600 |

Table 3 reveals that, of 120 children who were weaned early, 89 (74.2%) suffered at least one episode. On the other hand, among children who were not weaned early, 318 (66.3%) suffered one or more episodes of diarrhoea. Thus, in our sample, children who were weaned early were more likely to suffer an episode of diarrhoea than children who were not weaned early.

Before performing any further statistical test we compare the "commonsense" strength of the evidence presented in Tables 2 and 3. In Table 3 there appears to be a difference between the two groups of children but, because the data are presented in such a simple way, we are unable to look for any pattern. In Table 2, on the other hand, we are able to observe a pattern, the trend. At an intuitive level, data that follow a pattern present a more convincing case for an association than data in which no pattern can be observed.

We may also compare the statistical strengths of the two methods of data presentation by performing a chi-squared test of the data presented in Table 3 and determining the statistical significance of the association. For Table 3,

$$X^2 = 2.41.$$

Comparing this with a table of values for a chi-squared statistic with 1 degree of freedom, we obtain a p-value of 0.12. The data, as presented in Table 3, therefore do not provide strong evidence of an association between early weaning and diarrhoea. This result is in contrast to that obtained from Table 2, which did provide strong evidence of an association. Thus the analysis of the data as presented in Table 2 is both intuitively and statistically more powerful than the analysis of the data as presented in Table 3. On the basis of an analysis of Table 3, we would conclude that these data do not provide any strong evidence of any underlying association between early weaning and diarrhoea, and that explanation (v) above is a plausible explanation for the results we have observed. The sacrifice we made in the detail of the data in order to present them in a simpler form has removed our ability to look for patterns and reduced the statistical power of our analysis.

The data as presented in Table 3 allow us to estimate one important epidemiological measure. A common way of quantifying the association between an exposure (early vs late weaning) and a disease (diarrhoea) is by a measure known as the risk ratio. This is defined in the following way:

$$\text{Risk ratio} = \frac{\text{risk in exposed}}{\text{risk in unexposed}} = \frac{89/120}{318/480} = 1.12$$

This estimate of the risk ratio indicates that, in our sample, children who were weaned early were 1.12 times as (12% more) likely to suffer at least one episode of diarrhoea as (than) children who were not weaned early.

### 4.2.2 Untransformed, continuous data

An alternative approach to analysing these data is to treat the number of episodes of diarrhoea as a continuous variable and to compare the mean number of episodes among children weaned early with the mean number of episodes among other children. One way in which we may do this is by means of a t-test (Kirkwood, 1988).

**TABLE 4. Results of a t-test comparing the mean number of episodes among 120 children weaned early with the mean number of episodes among 480 children not weaned early**

| Group | Mean number of episodes | Standard deviation | Number |
|---|---|---|---|
| Weaned early | 2.23 | 2.01 | 120 |
| Not weaned early | 1.54 | 1.63 | 480 |

$$t = 4.00, \quad p < 0.0001$$

This analysis provides strong evidence that children who are weaned early suffer more episodes of diarrhoea than other children ($p<0.0001$). Over the period of the study, children weaned early suffered, on average, 0.70 more episodes of diarrhoea than other children (95% confidence interval 0.36,1.04). We may also calculate the (incidence) rate ratio of diarrhoea in the two groups. This is done as follows:

$$\text{Rate ratio} = \frac{\text{rate in the exposed}}{\text{rate in the unexposed}} = \frac{2.23}{1.54} = 1.45$$

The rate ratio indicates that children weaned early suffered 1.45 times as many episodes of diarrhoea as children who were not weaned early or, expressed in another way, children weaned early suffer, on average, 45% more episodes of diarrhoea than other children. This (exact) estimate of the rate ratio is larger than the (approximate) estimate that we obtained from Table 2 (1.38). The approximation (and hence mismeasurement) involved in the calculations used with Table 2 has led to an underestimate of the rate ratio. While the continuous analysis gives a better estimate of the rate ratio, the analysis of the categorized data allows us to look for trends that are more convincing than simple differences.

It is also instructive to contrast the estimate of the rate ratio with the estimate of the risk ratio (= 1.12) which we obtained from Table 3. The rate ratio is larger than the risk ratio and, for common recurrent diseases like diarrhoea and respiratory infections, is the better measure. To understand why rate ratio is to be preferred to risk ratio consider a hypothetical example. Suppose that we are studying a population in which early weaning leads to increased incidence of diarrhoea, and that diarrhoea is very common in our study population. To take an extreme situation, suppose that every child suffers at least one episode of diarrhoea each year. Then, if we follow the population for one year, the risk in both the exposed and unexposed populations will be 1.0 and so the risk ratio will be one; i.e., we will not observe any association between early weaning and risk of diarrhoea. If, instead, we consider the number of episodes (rate) in each group, we shall observe a higher rate in the children weaned early and thus we still detect the association. The relationship between risk ratio and rate ratio may be summarized as follows: over a short follow-up period during which the disease of interest is rare, the risk ratio is approximately equal to the rate ratio, but as the period of follow-up becomes longer the risk ratio will move away from the rate ratio towards 1.0.

From the data in Table 4 we may once again estimate the proportion of all episodes of diarrhoea that are attributable to early weaning. The mean number of episodes among children not weaned early was 1.54. Therefore, in a population of 600 children none of whom were weaned early, we would expect to observe a total of 921.25 (600 x 1.5354) episodes. In practice, we observed 1005 episodes. This suggests that 8.3% of all episodes of diarrhoea may be attributable to early weaning. This (exact) estimate is greater than the (approximate) estimate obtained from Table 2.

### 4.2.3 Transformed, continuous data

Strictly speaking, a t-test is only valid when comparing the mean values of two samples from normally distributed populations with equal standard deviations. The number of episodes of diarrhoea in each group of children is clearly not normally distributed (Figure 1). Fortunately, when the two samples are reasonably large (each greater than 30, say), the t-test is "robust" in the presence of departures from normality. Thus, this is not likely to be a serious problem in the analysis of this data set. There is, however, some evidence that the standard deviations of the two groups are different (2.01 versus 1.63, F=1.52, p=0.02). Again, for large samples, this does not constitute a serious problem since a modified version of the t-test may be used (Kirkwood, 1988).

One possible approach to dealing with the problems of non-normality and unequal standard deviations when they do arise is to transform the data. By transforming the data it is hoped to make the distribution more normal (in the statistical sense) and/or the standard deviations more equal. However, with a distribution like that in Figure 1, with a mode occurring at one end of the distribution (zero episodes), no sensible transformation will produce a normal-looking distribution. When dealing with data with a long right tail (Figure 3) the log transformation (log[1+number of episodes]) is commonly used. Performing a t-test on the log-transformed data produces the following results (Table 5).

**TABLE 5. Results of a t-test comparing the mean of log(1 + number of episodes) among 120 children weaned early with the mean among 480 children not weaned early**

| Group | Mean of log(1+episodes) | Standard deviation | Number |
|---|---|---|---|
| Weaned early | 0.42 | 0.29 | 120 |
| Not weaned early | 0.32 | 0.27 | 480 |

$t = 3.49, \quad p < 0.0005$

While the transformation has not resulted in two normally distributed groups of children, it has been effective in rendering the standard deviations more equal. There is now no evidence that they differ between the two groups (F=1.2, p=0.20). The results of this analysis confirm those of Table 4, although with a somewhat reduced t-statistic. In fact the t-statistic obtained using the transformed data (3.49) is very similar to that which would have been obtained from the original data using the modified t-test (3.53). It is not, however, easy to derive any simple and meaningful epidemiological measures of the association between diarrhoea and early weaning from this analysis.

### 4.2.4 Non-parametric analysis

An alternative approach to the problem of comparing the means of samples from non-normal populations or samples with unequal standard deviations is to perform a non-parametric test. Such tests do not make any assumptions about the distributions from which the samples are drawn. An appropriate non-parametric test for comparing two unmatched, independent samples is the Wilcoxon rank sum test (Kirkwood, 1988). The results of performing this test are presented in Table 6.

**TABLE 6. Results of a Wilcoxon rank sum test comparing the number of episodes of diarrhoea among 120 children weaned early with the number of episodes among 480 children not weaned early**

| Group | Number | Sum of ranks | Expected | Standard deviation |
|---|---|---|---|---|
| Weaned early | 120 | 41 790 | 36 060 | 1651.96 |
| Not weaned early | 480 | 138 510 | 144 240 | 1651.96 |

$Z = 3.47, \quad p = 0.0005$

This result confirms the statistical significance of the association between diarrhoea and early weaning observed when t-tests were conducted. The value of the Z statistic is very close to that of the t-statistic resulting from the analysis of the log-transformed data.

The Wilcoxon rank sum test makes no assumptions about the distribution of the number of episodes of diarrhoea per child and is almost as powerful as the parametric t-test. It is therefore excellent for hypothesis-testing in this situation. Unfortunately, however, it does not enable us to estimate any useful epidemiological measure(s) with which to assess the biological and/or public health importance of the association. Thus, while it may be useful as a confirmatory test, used alone it is not very revealing.

In Table 2 we presented the data in a simple form with children categorized according to the number of episodes of diarrhoea they experienced. We were able to look for pattern (trend) in the data and to conduct a powerful statistical analysis. The analysis of Table 2 provided strong evidence of an underlying association between early weaning and increased diarrhoea rates ($p<0.001$). In addition, we were able to estimate two important epidemiological parameters: the rate ratio of diarrhoea and the proportion of episodes of diarrhoea associated with early weaning. In trying to present the data even more simply (Table 3), we sacrificed much of the information in the data. We were no longer able to look for trends in the data and our statistical analysis failed to provide strong evidence of an association between early weaning and diarrhoea. The epidemiological parameter estimated from Table 3, the risk ratio, was less appropriate than the rate ratio estimated from Table 2.

An alternative and complementary approach to the analysis involved the comparison of the mean number of episodes of diarrhoea in each group, treating the number of episodes as a continuous outcome variable (Table 4). This approach also provided strong evidence of an association between early weaning and diarrhoea and we were able to estimate the rate ratio of diarrhoea and the proportion of all episodes associated with early weaning. While the use of the t-test is, strictly speaking, only valid when comparing samples drawn from normally distributed populations with equal standard deviations, the results obtained applying a t-test to the raw data were similar to those obtained from a t-test of the log-transformed data (Table 5) and from a non-parametric test (Table 6).

### 4.3 Analysis of the duration of episodes

Consider the question "Does early weaning lead to increased duration of diarrhoeal episodes?". This question is concerned with the course of an episode (how long it lasts) once it has begun. Only children who suffer at least one episode of morbidity contribute any information towards answering this question. Among the 600 children studied, 407 suffered at least one episode of diarrhoea. We restrict our attention to these children. Altogether, 1005 episodes of diarrhoea occurred (a mean of 2.47 episodes per child experiencing diarrhoea), lasting a total of 4222 days (a mean of 4.20 days per episode). The mean duration per child is shown in Figure 2.

While still very skewed to the right, this distribution appears somewhat closer to a normal distribution than that of number of episodes of diarrhoea. The minimum mean duration observed was 1.0 days (28 children), and the maximum 17.0 days (2 children), with a median of 3.5 days. Mean duration was slightly correlated with incidence rate (number of episodes divided by days at risk), the correlation co-efficient of this association being 0.14 ($p<0.01$). Thus, in this population, there is some statistical evidence that children who suffer more episodes of diarrhoea suffer longer episodes. The correlation itself is, however, weak.