The Effect of Different Sensitivity, Specificity and Cause-Specific Mortality Fractions on the Estimation of Differences in Cause-Specific Mortality Rates in Children from Studies Using Verbal Autopsies

GILLIAN H MAUDE AND DAVID A ROSS

Maude G H (Tropical Health Epidemiology Unit, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK) and Ross D A. The effect of different sensitivity, specificity and cause-specific mortality fractions on the estimation of differences in cause-specific mortality rates in children from studies using verbal autopsies. International Journal of Epidemiology 1997; 26: 1097–1106.

Background. Verbal autopsies (VA) are increasingly being used in developing countries to determine causes of death, but little attention is generally given to the misclassification effects of the VA. This paper considers the effect of misclassification on the estimation of differences in cause-specific mortality rates between two populations.

Methods. The bias in the percentage difference in cause-specific mortality between two populations has been explored under two different models: i) assuming that mortality from all other causes does not differ between the two populations; ii) allowing for a difference in mortality from all other causes. The bias is described in terms of the sensitivity and specificity of the VA diagnosis and the proportion of mortality due to the cause of interest. Methods for adjustment of sample size and adjusting the estimate of effect are also explored.

Results. The results are illustrated for a range of plausible values for these parameters. The bias is more extreme as both sensitivity and specificity fall, and is particularly affected even by a small loss of specificity. The bias also increases as the proportion of all deaths due to the cause of interest decreases, and is affected by the size of the true change in mortality due to the cause of interest relative to the change in mortality from other causes. Calculations from existing data suggest prohibitively large sample sizes may often be required to detect important differences in cause-specific mortality rates in studies using existing VA.

Conclusions. Highly specific VA tools are needed before observed differences in cause-specific mortality can be interpreted. Loss of power due to misclassification may obscure real differences in cause-specific mortality.

Keywords: verbal autopsy, misclassification, sensitivity, specificity

Verbal autopsy (VA) is a tool for obtaining causes of death when medical diagnoses are not available. Symptoms, signs and circumstances which lead to a death are ascertained by interviewing a close relative or associate of the deceased, and a diagnosis is derived either from a review of the questionnaire by one or more physicians, or from a set of algorithms. Studies which need to measure changes in cause-specific mortality rates in young children are increasingly turning to VA techniques in situations where medically-certified causes of death are not routinely available. However, it is well recognized that diagnosis by VA, though probably the most encouraging method available in areas where death certification systems are weak, will be subject to misclassification errors, usually measured in terms of the sensitivity and specificity of individual causes.1,2

1. Measures of cause-specific mortality rates and differences in rates between populations, and the sample size requirements to measure these rates and differences, are conventionally calculated with the assumption that the incidence of mortality will be estimated correctly. The effect of misclassification error on the estimate of cause-specific mortality fraction using a VA has been discussed by Anker.3 She shows that the effects of
misclassification can result in serious over- and under-
estimates of cause-specific mortality fractions, even in
the presence of relatively high sensitivity and specific-
ity, and that the effects of misclassification are related
not only to the sensitivity and specificity of the VA, but
also to the cause of death structure.

In this paper, we consider the effect of misclassifica-
tion on the estimation of differences in cause-specific
mortality rates between two populations. Estimates of
this kind arise from the following situations: i) com-
parisons of cause-specific mortality rates between inter-
vention and control groups after an intervention targeted
to reduce deaths from a specific cause; ii) comparisons
of cause-specific mortality rates between children in
two distinct regions or countries; iii) comparisons of
cause-specific mortality rates within one population over
a period of time. Assuming that the intervention and
control groups have been selected at random in situa-
tion (i), the underlying cause-of-death structure (i.e.
before the intervention) can be assumed to be the same
in the two populations, but in the other two situations
this may not be the case.

After introducing some terminology, we make as-
sumptions for two different models and demonstrate
the implications of misclassification within each of these
models in turn. We then discuss approaches to adjust-
ment for the misclassification effects, firstly by adjust-
ing the sample size, and then by adjusting the estimate
of mortality difference. Full details of the underlying
algebra are given in the Appendix.

The cause-of-death structure may affect the sensitiv-
ity and specificity of the VA instrument for a particular
cause of death, depending on the presence or absence of
other causes which might be misclassified with it. The
implications of these operating characteristics of the
VA tool are also explored and are described throughout
in terms of assessing change in mortality between popu-
lation 1 and population 2.

Terminology
Differences in mortality from a given cause (A) will be
described in terms of the percentage difference in the
rate in population 2 relative to population 1. For sim-
plicity, the true rate in population 1 will be assumed to
be greater than the rate in population 2, and unless
otherwise specified the mortality rate will refer to the
mortality rate from a specific cause (A). The following
notation will be used:
\[\lambda = \text{all cause mortality rate (per 1000 child-years)}\]
\[c_A = \text{cause-specific mortality fraction for cause of interest (A) in population 1}\]
\[s_a = \text{sensitivity of VA diagnosis for cause A}\]
\[s_p = \text{specificity of VA diagnosis for cause A}\]
\[r_1 = \text{cause A mortality rate in population 1}\]
\[r_2 = \text{cause A mortality rate in population 2}\]
\[t_A = \text{true proportional difference in cause A mort-
ality between populations 1 and 2} = \frac{(r_1 - r_2)}{r_1}\]

thus
\[r_1 = c_A \lambda\text{ and } r_2 = c_A \lambda (1 - t_A).\]

The measure \((1 - t_A)\) is equivalent to the rate ratio \(r_2/r_1\)
\((<1)\).

Assumptions
The effects of misclassification will be described in
terms of bias in the measurement of the percentage dif-
fERENCE in cause-A mortality rate between populations
1 and 2 and will be demonstrated by showing the
estimated percentage difference, using a VA diagnosis,
for a given true difference in mortality rate. A simple
model is explored first, where the mortality from all
causes other than A is assumed not to differ between the
two populations (Model 1). This would be most likely
to occur in the situation of an intervention study com-
paring an intervention with a control group, where the
intervention was targeted at a specific cause of death,
for example a malaria vaccine trial.

The pattern of misclassification by a VA may also
be affected by the relative importance of other causes
of death in the populations of interest, and the relative
difference in mortality rates from other causes in popu-
lations 1 and 2. Thus in many situations Model 1 will be
an oversimplification. Therefore, the effects of misclas-
sification are further explored within a second model
(Model 2) which additionally allows for a difference
between the two populations in the mortality rates from
all other causes combined.

In a situation where mortality from other causes does
not change between the two populations, a comparison
of cause-specific mortality rates will always require a
smaller sample size than a comparison of all-cause mor-
tality rates, because the relative effect will be consid-
erably greater. This will not necessarily be the case
under Model 2.

MISCLASSIFICATION BIAS IN COMPARING
THE CAUSE A-SPECIFIC MORTALITY RATE
BETWEEN POPULATIONS 1 AND 2

Implications of Misclassification within Model 1
The bias in the estimate of a difference in cause-specific
mortality rate (e.g. a reduction in deaths from diarrhoea)
is determined by the fraction of the all-cause mortality due to the specific cause of interest (A) in population 1, the size of the true effect and the sensitivity and specificity of the VA for cause A. Using the terminology given above, the effect of misclassification is to estimate \( r_1 \) and \( r_2 \) by \( r_1' \) and \( r_2' \), where:

\[
\begin{align*}
 r_1' &= s_n c_A \hat{\lambda} + (1 - s_p) \lambda (1 - c_A) \\
 r_2' &= s_n c_A \lambda (1 - t_A) + (1 - s_p) \lambda (1 - c_A).
\end{align*}
\]

The estimated proportional difference in cause A mortality rate is

\[
(r_1' - r_2')/r_1' = \left\{ s_n c_A / \left[ s_n c_A + (1 - s_p)(1 - c_A) \right] \right\} \times t_A. \quad (1)
\]

Equation (1) demonstrates two important points: firstly, when the specificity is 100% (\( s_p = 1 \)) the estimated proportional difference is cause A mortality is always equal to \( t_A \), and is therefore estimated without bias, since equation (1) reduces to \( (r_1' - r_2')/r_1' = t_A \). Secondly, when the specificity is less than perfect, the estimated difference in cause A-specific mortality rate will always be underestimated by the VA, even if the sensitivity is 100%, because the additional term in the denominator of equation (1) is always greater than zero. For example, for a specificity of 80%, a sensitivity of 100% and a cause-specific mortality fraction of 25% for cause A, if the true percentage difference in mortality from cause A is 25% then the estimated difference is 16% (applying equation (1), \( \{1 \times 0.25/[1 \times 0.25 + (1 - 0.8) \times (1 - 0.25)]\} \times 0.25 = 0.16 \)). Sensitivity and specificity are both likely to be less than 100% for all VA instruments, and the bias in the observed cause A-specific mortality rate difference, relative to the true difference, becomes more extreme as both sensitivity and specificity fall. For example, if, in a controlled intervention study against diarrhoea mortality, diarrhoea accounts for 25% of all deaths in the control group, and the VA instrument has 60% sensitivity and 70% specificity for diagnosis of death due to diarrhoea, then a true intervention effect of 25% will be estimated as an effect of 10% from the VA. Furthermore, the bias in the observed cause A mortality rate difference also increases as the cause-specific fraction of deaths due to cause A decreases. If, in the above example, diarrhoea accounted for only 10% of all deaths in the control group, then a true intervention effect of 25% would be estimated as an effect of only 5%.

These effects of misclassification bias are demonstrated in Table 1 for selected cause-specific fractions of death (10%, 25%), selected true percentage reductions in the cause A-specific mortality rate, and a range of values of sensitivity and specificity. Table 1 illustrates two key features. Firstly the ratio of the estimated to true percentage difference in cause A-specific mortality rate remains constant for a given sensitivity, specificity and cause-specific mortality fraction (equation 1). For example, with a cause-specific mortality fraction in population 1 of 25% and sensitivity and specificity of 90% and 80% respectively, a true difference of 25% is estimated as 15%, while a true difference of 50% is estimated as 30%. Secondly, Table 3 shows that the misclassification bias is much greater when the specificity is poor than when the sensitivity is poor. For example, if the true cause A-specific mortality fraction in population 1 is 10%, and the true cause A percentage difference in mortality is 50%, then the percentage difference between the two populations will be estimated as 24% if the sensitivity is 80% and specificity is 90%, and as 17% if the sensitivity is 90% and the specificity is 80%.

**Implications of Misclassification within Model 2**

In many situations it is inappropriate to assume that mortality from other causes does not differ between the two populations being compared. When the model is extended to allow for changing mortality from all other causes combined, then the effect of the bias in the estimation of the proportional difference in cause A mortality rate between the two populations depends additionally on the magnitude of the percentage difference in mortality from other causes (\( t_O \)). The effect of the misclassification is to estimate \( r_2' \) by

\[
 r_2' = s_n c_A \lambda (1 - t_A) + (1 - s_p) \lambda (1 - c_A) t_O. \]

\( r_1' \) is estimated as in Model 1 above, and the estimated percentage difference in mortality rate is thus

\[
(r_1' - r_2')/r_1' = \left\{ s_n c_A / \left[ s_n c_A + (1 - s_p)(1 - c_A) t_O \right] \right\} / \left\{ s_n c_A + (1 - s_p)(1 - c_A) \right\}. \quad (2)
\]

Table 2 shows the estimated percentage difference in cause A mortality rate using this extended model for a range of values. For example, if the percentage of all deaths due to cause A in population 1 is 10% and the sensitivity and the specificity are both 90% then a true difference in cause A mortality of 50% is estimated as 37% if there is a reduction in mortality from all other causes of 25% (\( t_O = 0.25 \)) and is estimated as 12% if there is an increase in the mortality rate due to other causes of 25% (\( t_O = -0.25 \)).

When \( t_O = 0 \), Model 2 is identical to Model 1 and hence the entries in the 0 column of Table 2 are the same.
as those in the corresponding cells of Table 1. When \( t_O \) and \( t_A \) are equal the difference in mortality is estimated without bias, regardless of the other factors. If \( t_A > t_O \), the estimated difference in mortality rate is underestimated, while an overestimation is made when \( t_A < t_O \) (Equation 2, Table 2). One consequence is that, if there is little or no true difference in mortality rate from cause A between population 1 and population 2 (\( t_A \) close to 0), a substantial difference (e.g. >10%) may be estimated by the VA if the relative difference in mortality from other causes is greater than about 25% (Table 2).

**METHODS OF ADJUSTMENT FOR MISCLASSIFICATION BIAS**

**Adjusting the Sample Size**

Under Model 1, owing to the systematic underestimation of the size of the percentage difference in cause A mortality rate between populations 1 and 2 in the presence of misclassification, the power of a study which does not take sensitivity and specificity into account will be reduced, sometimes substantially. Conversely, if a required power is to be maintained, then the sample size should be increased. This is illustrated by the calculation of a multiplier (k) which shows the increase in sample size required, compared to a (hypothetical) study in which the cause-specific mortality rates were established without misclassification. The derivation of k and of the adjusted sample size is given in the Appendix. Lower rates of all-cause mortality demand larger sample sizes, but do not affect the multiplier k (Appendix equations (A1),(A2)). For perfect specificity, k is affected only by the sensitivity and is equal to \( 1/s_i \), while for less than perfect specificity k depends on the sensitivity, specificity, the cause-specific mortality fraction, and the percentage mortality reduction. For example, if the sensitivity and specificity of the VA instrument are 70% and 80% respectively and the true cause-specific mortality fraction is 25%, then

### Table 1 Estimated percentage difference in the cause A-specific mortality rate for selected values of true percentage difference in mortality rate due to cause A between populations 1 and 2, sensitivity and specificity, and selected cause-specific fractions of mortality due to cause A in population 1

<table>
<thead>
<tr>
<th>Cause-specific fraction of deaths due to cause A in population 1</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
<td>90</td>
</tr>
<tr>
<td>True difference = 25%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>25</td>
<td>13</td>
</tr>
<tr>
<td>90</td>
<td>25</td>
<td>12</td>
</tr>
<tr>
<td>80</td>
<td>25</td>
<td>12</td>
</tr>
<tr>
<td>70</td>
<td>25</td>
<td>11</td>
</tr>
<tr>
<td>60</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>0.25</td>
<td>25</td>
<td>19</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>19</td>
</tr>
<tr>
<td>90</td>
<td>25</td>
<td>18</td>
</tr>
<tr>
<td>80</td>
<td>25</td>
<td>18</td>
</tr>
<tr>
<td>70</td>
<td>25</td>
<td>17</td>
</tr>
<tr>
<td>True difference = 50%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>50</td>
<td>26</td>
</tr>
<tr>
<td>90</td>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td>80</td>
<td>50</td>
<td>24</td>
</tr>
<tr>
<td>70</td>
<td>50</td>
<td>22</td>
</tr>
<tr>
<td>60</td>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td>0.25</td>
<td>50</td>
<td>38</td>
</tr>
<tr>
<td>100</td>
<td>50</td>
<td>38</td>
</tr>
<tr>
<td>90</td>
<td>50</td>
<td>36</td>
</tr>
<tr>
<td>80</td>
<td>50</td>
<td>35</td>
</tr>
<tr>
<td>70</td>
<td>50</td>
<td>33</td>
</tr>
<tr>
<td>60</td>
<td>50</td>
<td>33</td>
</tr>
</tbody>
</table>

NB: For simplicity, the true cause A-specific mortality rate in population 1 has always been assumed to be greater than that in population 2.
to obtain a P-value of 0.05 for a true percentage difference in cause A mortality rate between populations 1 and 2 of 25%, the sample size must be increased by 2.8 times in comparison with the sample size required to achieve the same power with perfect sensitivity and specificity. Further examples are shown in Table 3 for selected values of these factors. Clearly, the multiplier is affected most by lack of specificity, particularly when the cause-specific mortality fraction is low. It is greater than 2 for most of the situations presented, and considerably so in many situations.

The sample size requirements under Model 2 are further affected by the percentage change in mortality due to other causes (tO) in addition to the factors described above (Equations (A3),(A4)). When tO = 0, Model 2 is identical to Model 1 above. When tO = tA there is no bias in the estimation of the percentage change in mortality, however there is an effect on the sample required. In this case the multiplier depends on the cause-A specific fraction of mortality and the sensitivity and specificity, but not on tA. The multiplier, k, is less than 1 when \((1 - s_p)(1 - c_a)\) is greater than \((1 - s_n)c_a\). For example, when the true difference is 25%, the sensitivity and specificity is 90%, and the fraction of mortality due to cause A is 10%, a true change in cause-A mortality rate from 12 per 1000 child-years to 9 per 1000 is estimated by the VA as a change from 21.6/1000 to 16.2/1000, which requires a smaller sample size (k = 0.56).

Specific Illustrative Examples from Developing Country Populations
Table 4 gives the cause-specific and all-cause mortality rates, and the cause-specific mortality fractions for diarrhoea, acute respiratory infections, measles, and malaria reported for three populations of children in developing countries; one in Bangladesh,4 one in The Gambia,5 and one in northern Ghana.6 Although these data were themselves determined using (different) VA which were constrained to yield a single cause of death, the reported all-cause and cause-specific mortality rates from the original sources have been taken for illustrative purposes as the ‘true’ mortality rates. Table 5 shows the number of child-years of surveillance per group (i.e. in population 1 and in population 2) which would be required to have a 90% power to detect a 25% or 50% true difference in the cause-specific mortality rate as significant at the 5% level. This is shown both for the hypothetical situation with no misclassification of cause-specific mortality, and also using the actual operating characteristics of VA instruments used within two validation studies, one in the Philippines2 and the other in Kenya.7 To calculate the total population-years

<table>
<thead>
<tr>
<th>True % difference in cause A mortality rate (tA)</th>
<th>% of all deaths due to cause A in popn. 1</th>
<th>Sensitivity (%) and specificity (%) for cause A</th>
<th>Percentage difference in mortality from all other causes (tO)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25</td>
<td>90</td>
<td>(6)</td>
</tr>
<tr>
<td>10</td>
<td>90</td>
<td>70</td>
<td>(12)</td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>90</td>
<td>12</td>
</tr>
<tr>
<td>10</td>
<td>90</td>
<td>70</td>
<td>3</td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>90</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>90</td>
<td>70</td>
<td>15</td>
</tr>
</tbody>
</table>

* The percentages not in () indicate that the mortality rate is lower in population 2 than in population 1, whereas those in parentheses [e.g. (14)] indicate that the mortality rate is higher in population 2.

Table 2 Estimated percentage difference* in the mortality rate due to cause A for selected true percentage reductions in cause A mortality rate, percentage of all deaths due to cause A, sensitivity (%) and specificity (%) for cause A and percentage difference* in mortality from all other causes combined
TABLE 3 Increase in sample size required to achieve the same power as in a study with no misclassification, for selected values of true percentage difference in the mortality rate due to cause A between populations 1 and 2, sensitivity, specificity, and selected cause-specific fractions of mortality due to cause A in population 1

<table>
<thead>
<tr>
<th>Cause-specific fraction of deaths due to cause A in population 1</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
<td>90</td>
</tr>
<tr>
<td>True difference = 25%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>90</td>
<td>1.1</td>
<td>2.4</td>
</tr>
<tr>
<td>80</td>
<td>1.2</td>
<td>2.9</td>
</tr>
<tr>
<td>70</td>
<td>1.4</td>
<td>3.5</td>
</tr>
<tr>
<td>60</td>
<td>1.7</td>
<td>4.5</td>
</tr>
<tr>
<td>0.25</td>
<td>1.0</td>
<td>1.3</td>
</tr>
<tr>
<td>90</td>
<td>1.1</td>
<td>1.5</td>
</tr>
<tr>
<td>80</td>
<td>1.2</td>
<td>1.8</td>
</tr>
<tr>
<td>70</td>
<td>1.4</td>
<td>2.1</td>
</tr>
<tr>
<td>60</td>
<td>1.7</td>
<td>2.6</td>
</tr>
<tr>
<td>True difference = 50%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>1.0</td>
<td>2.2</td>
</tr>
<tr>
<td>90</td>
<td>1.1</td>
<td>2.6</td>
</tr>
<tr>
<td>80</td>
<td>1.2</td>
<td>3.1</td>
</tr>
<tr>
<td>70</td>
<td>1.4</td>
<td>3.9</td>
</tr>
<tr>
<td>60</td>
<td>1.7</td>
<td>5.0</td>
</tr>
<tr>
<td>0.25</td>
<td>1.0</td>
<td>1.4</td>
</tr>
<tr>
<td>90</td>
<td>1.1</td>
<td>1.6</td>
</tr>
<tr>
<td>80</td>
<td>1.2</td>
<td>1.9</td>
</tr>
<tr>
<td>70</td>
<td>1.4</td>
<td>2.2</td>
</tr>
<tr>
<td>60</td>
<td>1.7</td>
<td>2.8</td>
</tr>
</tbody>
</table>

NB: For simplicity, the true cause A-specific mortality rate in population 1 has always been assumed to be greater than that in population 2.

TABLE 4 Cause-specific mortality rates and mortality fractions in three developing country populations

<table>
<thead>
<tr>
<th>Setting</th>
<th>Age group (months)</th>
<th>Cause of death</th>
<th>Mortality rate in population 1 (per 1000 child-years)</th>
<th>Cause-specific mortality fraction (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matlab, Bangladesh(^4)</td>
<td>6–35</td>
<td>Diarrhoea</td>
<td>13.6</td>
<td>53.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ARI</td>
<td>4.5</td>
<td>17.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measles</td>
<td>0.2</td>
<td>0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Malaria</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All causes</td>
<td>25.3</td>
<td>100.0</td>
</tr>
<tr>
<td>Farafenni, The Gambia(^5)</td>
<td>0–83</td>
<td>Diarrhoea</td>
<td>9.7</td>
<td>17.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ARI</td>
<td>7.9</td>
<td>14.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measles</td>
<td>1.8</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Malaria</td>
<td>7.4</td>
<td>13.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All causes</td>
<td>54.1</td>
<td>100.0</td>
</tr>
<tr>
<td>Navrongo, Ghana(^6)</td>
<td>6–95</td>
<td>Diarrhoea</td>
<td>8.8</td>
<td>29.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ARI</td>
<td>2.7</td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measles</td>
<td>4.4</td>
<td>14.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Malaria</td>
<td>5.1</td>
<td>17.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All causes</td>
<td>29.5</td>
<td>100.0</td>
</tr>
</tbody>
</table>

For sources see the reference numbers given as superscripts.
of surveillance needed, as a general rule of thumb, one may assume that approximately 10% of the population in Bangladesh will be within the 6–35 month age group, 25% of the Gambian population within the 0–83 month age group, and 25% of the Ghanaian population within the 6–95 month age group.

Adjusting the Estimate of Mortality Difference

Adjustment of sample size may overcome the effects of misclassification on significance tests, but does not address the issue of bias in the estimation of the effect. Kalter has pointed out that if the sensitivity and specificity of the VA for a particular cause of death are known, then one can obtain the proportion of mortality due to that cause from the VA estimate of proportional mortality from the equation

$$c_A = (c_A' + s_p - 1)/(s_n + s_p - 1)$$ (3)

where $c_A'$ is the fraction of mortality due to cause A as estimated by the VA. Using suffices 1 and 2 for the two populations, and substituting for $c_{A1}$ and $c_{A2}$ from equation (3), assuming that the mortality rate from all other causes is unchanged (Model 1) and that the sensitivity and specificity for cause A diagnosis are the same in the two populations, the proportional difference in cause A mortality between the two populations can be obtained (see Appendix) as:

$$t_A = 1 - [(c_{A2}' + s_p - 1)(s_n - c_{A1}')/ (c_{A1}' + s_p - 1)(s_n - c_{A2}')]$$ (4)

This can be extended for Model 2, with the result that $t_A$ is estimated by

$$1 - [(c_{A2}' + s_p - 1)(s_n - c_{A1}')/ (c_{A1}' + s_p - 1)(s_n - c_{A2}')] (1 - t_O).$$

Thus, for example, if there is no change in mortality from other causes between the populations, a VA with 90% specificity and 70% sensitivity for a particular cause, which estimates cause-specific fractions of mortality as 0.20 and 0.15 in populations 1 and 2 respectively, will lead to an estimate of the true percentage difference in mortality of $t_A = 0.54$, rather than the uncorrected value of 0.25. If $s_n = c_{A2}'$ or $c_{A1}' + s_p = 1$ then equation (4) is undetermined: it is most unlikely that $s_n = c_{A2}'$ since no single cause of mortality is likely to account for more than 40% of the deaths; $c_{A1}' + s_p = 1$ is more likely to occur (e.g. $c_{A1}' = 0.1$, $s_p = 0.9$, and highlights a limit to this approach.

<table>
<thead>
<tr>
<th>Cause of death</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Number of child-years</th>
<th>Bangladesh (6–35 months)</th>
<th>Gambia (0–83 months)</th>
<th>Ghana (6–95 months)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>25%</td>
<td>50%</td>
<td>25%</td>
<td>50%</td>
</tr>
<tr>
<td>Diarrhoea</td>
<td>100</td>
<td>100</td>
<td>21 600</td>
<td>4600</td>
<td>30 300</td>
<td>6500</td>
</tr>
<tr>
<td></td>
<td>78^2</td>
<td>79^2</td>
<td>34 900</td>
<td>7700</td>
<td>93 800</td>
<td>22 100</td>
</tr>
<tr>
<td></td>
<td>36^7</td>
<td>96^7</td>
<td>66 400</td>
<td>14 460</td>
<td>133 400</td>
<td>30 300</td>
</tr>
<tr>
<td>Acute respiratory</td>
<td>100</td>
<td>100</td>
<td>65 600</td>
<td>14 100</td>
<td>37 000</td>
<td>7900</td>
</tr>
<tr>
<td>infections</td>
<td>59^2</td>
<td>77^2</td>
<td>341 700</td>
<td>81 500</td>
<td>224 600</td>
<td>53 900</td>
</tr>
<tr>
<td></td>
<td>28^7</td>
<td>91^7</td>
<td>634 800</td>
<td>150 300</td>
<td>413 400</td>
<td>98 600</td>
</tr>
<tr>
<td>Measles</td>
<td>100</td>
<td>100</td>
<td>1 659 700</td>
<td>355 650</td>
<td>164 600</td>
<td>35 300</td>
</tr>
<tr>
<td></td>
<td>98^2</td>
<td>90^2</td>
<td>29 710 500</td>
<td>7 367 100</td>
<td>742 100</td>
<td>179 500</td>
</tr>
<tr>
<td></td>
<td>90^7</td>
<td>96^7</td>
<td>15 131 700</td>
<td>3 717 100</td>
<td>455 200</td>
<td>107 300</td>
</tr>
<tr>
<td>Malaria</td>
<td>100</td>
<td>100</td>
<td>NA</td>
<td>NA</td>
<td>39 700</td>
<td>8500</td>
</tr>
<tr>
<td></td>
<td>46^7</td>
<td>89^7</td>
<td>NA</td>
<td>NA</td>
<td>234 600</td>
<td>55 600</td>
</tr>
</tbody>
</table>

NB: For sources see the reference numbers given here as superscripts.
DISCUSSION

While studies which have used the VA technique may have acknowledged that the VA tool is imperfect, little attention has been paid to the implications of misclassification of causes of death when drawing conclusions about cause-specific mortality rates. This paper has concentrated on the effects of misclassification due to imperfect sensitivity and specificity on the estimation of change in cause-specific mortality between two populations or in the same population at different times. Substantial bias in the percentage difference in mortality rate can arise, and is more affected by loss of specificity than loss of sensitivity. If misclassification has been ignored, studies may have drawn incorrect conclusions. For example, the observed difference in cause-specific mortality rate between the two populations may be substantially less than the true difference, and no significant difference between the observed mortality rates may be detected when there is a true difference.

Two models have been explored, and their effects have been described for a range of values of the operating characteristics. When mortality from all causes other than the cause of interest (A) is assumed to remain unchanged between the two populations (Model 1), a VA with less than perfect specificity will give rise to underestimation of the true difference in mortality rates, and this bias may be substantial even if the specificity is higher than 90%. Examples have been given for cause-specific fractions of 10% and 25% to reflect a range of values of interest. The effect of the specificity becomes increasingly large as the cause-specific fraction falls, since for low values of $c_A$, most of the deaths fall in the ‘other’ category. Thus, high specificity is an important requirement for a potential VA tool.

Since this simple model may not always be applicable, for example where an intervention targeted against malaria also affects mortality from gastro-enteritis, the second model explored allows for a change in mortality from all other causes combined. No bias is observed in the percentage difference in mortality rates when the true change in mortality due to cause A is equal to the true change for all other causes combined, and the effect is over- or under-estimated according to whether $t_0$ is greater than or less than $t_A$.

Since there are few individual causes of death which account for over 10% of all mortality, values of $c_A$ will typically be low, so both these over- and underestimates are typically large.

In practice, both models explored are an oversimplification of the true situation. In reality, some mortality rates from causes other than the cause of special interest may fall while others rise or remain stable between the two populations. Furthermore, the chances of misclassification of deaths due to cause A will depend on the cause of death structure, in particular the relative importance of the other causes with which A might be confused. For example, classification of deaths due to gastro-enteritis may be more or less valid according to whether malaria contributes to the underlying causes of death in that population. This means that, while it is convenient to attach a sensitivity and specificity to a VA diagnosis for a given cause (which are normally derived from a validation study), in practice these are only valid in a population with a cause of death structure similar to that in which the validation took place. Each setting for a validation study will have its own mix of causes of death, and bias associated with the possibility of a different mix of causes in the hospital (reference) population than in the community. In addition, the extent to which sensitivity and specificity are influenced by cultural factors for some causes of death is not widely known. In an attempt to obtain a tractable model, the sensitivity and specificity in this paper have explicitly been kept the same for populations 1 and 2, but a range of values of sensitivity and specificity have been explored. It would be possible to extend these models to include varying sensitivities and specificities using, for example, simulation techniques. However, prior knowledge of the true cause of death structure is likely to be poor wherever VA are being used, since VA will only usually be used where other more valid methods are unavailable.

Misclassification by the VA has major implications for the power of studies designed to detect differences in cause-specific mortality rates between the two populations. Where the difference in the cause-A specific mortality rate is underestimated there may be a substantial loss of power. The effect on power depends on a complex relationship between sensitivity, specificity, cause-specific fraction of mortality and $t_0$ and $t_A$, and where there is no bias or the effect is overestimated, a smaller sample size will usually be sufficient to achieve the desired power. If the sensitivity, specificity and cause-specific fraction are known, or can be estimated validly for the population under study, then a multiplier can be used to adjust the sample size. However, the likely pattern of mortality changes from all other causes needs to be incorporated, and the experience from four developing countries presented above suggests that the adjusted sample sizes would be prohibitively large in all but a few situations. Adjustments to the estimate of the difference in mortality rate to take account of misclassification effects also require good estimates of the sensitivity and specificity and cause-specific fraction. Theoretically, adjustments both to sample size and to the estimate are possible, but in practice they lead
to circular arguments where the information required to apply the adjustments is precisely that which is not well known.

The above considerations have been focused specifically on the estimation of changes in cause-specific mortality rates when the cause of death is subject to misclassification by VA. However, many of the theoretical considerations and calculations above apply to other types of misclassification such as comparisons of morbidity; laboratory testing where individuals are classified into one of a number of groups, for example, various antigenic strains of bacteria or parasites, or malaria species classification; or differences in the reported use of several alternative health services. Where a degree of misclassification is acknowledged or expected, comparisons of outcomes should not normally be accepted at face value. In the case of VA, highly specific VA instruments (e.g. questionnaire with algorithms) are required before changes in cause-specific mortality can be measured with acceptable accuracy.

ACKNOWLEDGEMENTS
The authors thank Martha Anker and Richard Hayes for their technical contributions and support for this work; acknowledge the financial support of WHO and UNICEF, and thank Stefan Wedner for checking the accuracy of the equations. Gillian Maude was supported by the UK Medical Research Council.

REFERENCES

(Revised version received March 1997)

APPENDIX
Sample Size Calculations
Allowing for change in mortality from cause A but from no other causes (Model 1).

The power of a test is obtained by referring $z_2$ to the Normal distribution tables where

$z_2 = \{\sqrt{n/(r_1 + r_2)}/(r_1 - r_2)\} - z_1$

and $n$ = the number of person-years required in each group when there is no misclassification, and $z_1$ corresponds to the significance level of the test (taken from the Normal distribution, $z_1 = 1.96$ for $P = 0.05$). Hence, to achieve the same power with misclassification, a sample size $n'$ is required in each group, where

$\{\sqrt{n/(r_1 + r_2)}/(r_1 - r_2)\} - z_1 = \{\sqrt{n'//(r_1 + r_2')}/(r_1' - r_2')\} - z_1$

Thus $n' = \{(r_1 - r_2)^2(r_1' - r_2')^2/(r_1 + r_2)\}n$

$= \{(s_n c_A \lambda t_A)^2(s_n c_A \lambda (2 - t_A) + 2(1 - s_p)\lambda(1 - c_A))\}$

$/(s_n c_A \lambda t_A)^2\}$

$= \{(s_n c_A (2 - t_A) + 2(1 - s_p)(1 - c_A))/$

$s_n^2 c_A (2 - t_A)\}$

$= \{(s_n c_A (2 - t_A) + 2(1 - s_p)(1 - c_A))/$

$s_n^2 c_A (2 - t_A)\}n$ (A1)

when $s_p = 1$, this reduces to $n' = n/s_n^2$.

The sample size, for a power corresponding to $z_2$, required in each group is given by:

$n' = (1.96 + z_2)^2(r_1 - r_2)/\{(1.96 + z_2)^2\}$

$= (1.96 + z_2)^2(s_n c_A \lambda (2 - t_A) + 2(1 - s_p)\lambda(1 - c_A))$

$/(s_n c_A \lambda t_A)^2\}$

$= (1.96 + z_2)^2(s_n c_A (2 - t_A) + 2(1 - s_p)(1 - c_A))/$

$s_n^2 c_A (2 - t_A)$

(A2)
Allowing for true differences in mortality rate due to both cause A and due to all other causes combined (Model 2).

To achieve the same power compared with a study of size \( n \) with no misclassification, the sample size is calculated as before by obtaining a multiplier, a substituting for \( r_1' \) and \( r_2' \) as derived in the text:

\[
\begin{align*}
n' &= \{ (r_1' - r_2')^2 (r_1' + r_2') / (r_1' - r_2')^2 (r_1' + r_2') \} n \\
&= \{ (c_A \lambda t_A) [s_n c_A \lambda (2 - t_A) + (1 - s_p) \lambda (1 - c_A)(2 - t_O)] / [(s_n c_A \lambda t_A) + (1 - s_p) \lambda (1 - c_A) t_O]^2 (c_A \lambda (2 - t_A)) \} n \\
&= \{ c_A t_A^2 [s_n c_A (2 - t_A) + (1 - s_p)(1 - c_A)(2 - t_O)] / [(s_n c_A t_A) + (1 - s_p)(1 - c_A) t_O]^2 (2 - t_A) \} n \quad (A3)
\end{align*}
\]

The sample size, for a power corresponding to \( z^2 \), required in each group is given by:

\[
\begin{align*}
n &= (1.96 + z^2)^2 (r_1' + r_2') / (r_1' - r_2')^2 \\
&= (1.96 + z^2)^2 [s_n c_A \lambda (2 - t_A) + (1 - s_p) \lambda (1 - c_A)(2 - t_O)] / [(s_n c_A \lambda t_A) + (1 - s_p) \lambda (1 - c_A) t_O] \\
&= (1.96 + z^2)^2 [s_n c_A (2 - t_A) + (1 - s_p)(1 - c_A)(2 - t_O)] / [(s_n c_A t_A) + (1 - s_p)(1 - c_A) t_O]^2 \\
&= (1.96 + z^2)^2 [s_n c_A (2 - t_A) + (1 - s_p)(1 - c_A)(2 - t_O)] / [(s_n c_A t_A) + (1 - s_p)(1 - c_A) t_O]^2 \lambda
\end{align*}
\]

Adjusting the estimate of mortality rate difference.

Under Model 1,

\[
\begin{align*}
r_1 &= c_{A1} \lambda \\
r_2 &= c_{A2} [(1 - c_{A1}) \lambda + (1 - t_A) c_{A1} \lambda] \\
&= c_{A2} (1 - c_{A1}) \lambda / (1 - c_{A2})
\end{align*}
\]

Using equation (3)

\[
\begin{align*}
r_1 &= (c_{A1}' + s_p - 1) \lambda \\
r_2 &= (c_{A2}' + s_p - 1) [(1 - (c_{A1}' + s_p - 1)) \lambda / (1 - c_{A2}' + s_p - 1)] \\
&= (c_{A2}' + s_p - 1)(s_n - c_{A2}') \lambda \\
&= (c_{A2}' + s_p - 1)(s_n - c_{A1}') \lambda \\
&= (c_{A2}' + s_p - 1)(s_n - c_{A2}') \\
&= (c_{A2}' + s_p - 1)(s_n - c_{A1}') \\
&= (c_{A2}' + s_p - 1)(s_n - c_{A2}')
\end{align*}
\]

For Model 1, since \( t_A = (r_1 - r_2) / r_1 = 1 - r_2 / r_1 \)

\[
\begin{align*}
t_A &= 1 - (c_{A2}' + s_p - 1)(s_n - c_{A1}') \\
&= (c_{A1}' + s_p - 1)(s_n - c_{A2}')
\end{align*}
\]

This can be extended for Model 2 in a similar way.