Should verbal autopsy results for malaria be adjusted to improve validity?
From ALEXANDER K ROWE

Sirs—Verbal autopsies performed in community settings are a key data source for determining probable causes of death in areas where vital registration data are unreliable.1,2 However, validation studies of the verbal autopsy method in which verbal autopsy diagnoses are compared with ‘gold standard’ diagnoses (based on clinical and laboratory data) show that verbal autopsies have low sensitivity (range: 24–75%) and moderate specificity (range: 77–100%) for identifying deaths caused by malaria (i.e. verbal autopsies often miss true malaria deaths and misclassify non-malaria deaths as malaria).1,3–8

Theoretically, results from verbal autopsy validation studies could be used to adjust the proportion of deaths attributable to malaria (PDAM) estimated by an unvalidated verbal autopsy study, and if done correctly, the adjusted PDAM should be less biased. The expression for adjusting verbal autopsy results is shown below.9

Adjusted PDAM

\[
\text{PDAM from an unvalidated verbal autopsy study + specificity - 1} \over \text{sensitivity + specificity - 1}
\]

In practice, adjusting verbal autopsy results with validation studies is not so easy. The first major problem is that verbal autopsy validation studies themselves have important limitations.

- Study populations are typically hospitalized patients and are probably different from community populations (i.e. potential selection bias).9 There are at least two reasons why differences may exist: (i) people who visit hospitals may be different from those typically enrolled in studies using verbal autopsy (e.g. research sites typically enrol rural populations, and people visiting hospitals may be more likely to be urban or peri-urban residents); and (ii) even if underlying populations were similar, hospital settings may change relatives’ understanding of a patient’s disease and thus may bias verbal autopsy results (e.g. compared with the community setting, in hospitals, relatives learn more about the true cause of the death).
- The ‘gold standard’ clinical diagnosis may misclassify the cause of death,9 especially in malaria-endemic areas where asymptomatic parasitaemia is common.
- The relatively small sample size of some validation studies leads to sensitivity and specificity estimates with low precision.

These limitations are all relevant when deciding whether or not to use the adjustment formula. Indeed, according to Chandramohan et al.,10 the formula to adjust verbal autopsy results can be used ‘assuming that all sources of error in measures of sensitivity and specificity are negligible’. For malaria, this assumption does not seem valid.

The second major problem with adjusting verbal autopsy results with validation studies is that the specific verbal autopsy methods (e.g. time elapsed between death and post-mortem interview, questionnaire, interviewing method, and diagnostic algorithms) ‘tested’ in validation studies are probably different from many of the verbal autopsy methods used in the unvalidated verbal autopsy studies that are to be adjusted. In other words, results from even a well-done validation study might not apply to a verbal autopsy study that used a different verbal autopsy method.

The third major problem is that, even if the validation study has no bias and its method is the same as all verbal autopsy studies that are to be adjusted, the validation process may still introduce bias if the results to be adjusted were obtained in areas with cause-specific mortality patterns that were different from that of the validation study. According to Chandramohan et al., it is not possible to use sensitivity and specificity estimates derived from a location-specific validation study to adjust for misclassification in verbal autopsy data from populations with substantially different patterns of cause-specific mortality.10 This statement is supported by an analysis that found that when verbal autopsy results from one site (Morogoro, Tanzania) were adjusted using validation study results from three different sites that all used the same verbal autopsy methods (Ifakara, Tanzania; Jimma, Ethiopia; and Bawku, Ghana), the three sets of adjusted results were considerably different (see table 5 of Ref. 10). For example, the unadjusted proportion of deaths attributable to diarrhoea from Morogoro was 14.9%; the Ifakara-adjusted proportion was 13.3%, the Jimma-adjusted proportion was 19.9%, and the Bawku-adjusted proportion was 54.3%.

The fourth problem is deciding how to implement the adjustment formula shown above. There are seven reasonably good published validation studies for malaria,1,3–8 and it is not clear how sensitivity and specificity results from these studies should be used (e.g. Should results be combined? Should results from the single ‘best’ study be used? How can uncertainty in the sensitivity and specificity estimates be incorporated into the adjustment process?).

The fifth problem is that the adjustment process may lead to impossible results. For example, from Kenya,11 the unadjusted PDAM is 4.3% (15/348). If this result is adjusted using the sensitivity and specificity values from the verbal autopsy validation study by Todd et al.,7 then the following values would be used in the adjustment formula: sensitivity = 58.3% [i.e. 21/ (21 + 15)] and specificity = 79.2% [i.e. 42/(42 + 11)]. With these values, the formula gives an adjusted PDAM of −44.0% [i.e. (0.043 + 0.792 − 1)/(0.583 + 0.792 − 1)]. Impossible results arise because estimates of sensitivity and specificity are used in a formula that is only mathematically correct for the true sensitivity and specificity. This problem was so serious that Williams et al.12 did not use this adjustment method in their article on the global burden of acute respiratory infections, which used verbal-autopsy-based studies.

A recent review of childhood malaria mortality trends tried to overcome this obstacle by letting specificity be a function of the observed PDAM (i.e. validation studies were used to develop a model of specificity, and PDAM values from unvalidated
community verbal autopsy studies were fed into the model to estimate specificity.\textsuperscript{13} This approach, as implemented, had several potential problems.

- The PDAM in the hospital setting, where validation studies were conducted, may not have been the same as the PDAM in the community. (Indeed, if the hospital-based PDAM equalled the community-based PDAM, then one could simply measure the hospital-based PDAM, and community research sites would not be necessary to study causes of death.)
- The age range of children in the validation studies was not always the age range targeted by the review (0–4 years). For example, neonates were excluded from about half of the validation studies used in the review. As malaria is unlikely to be the direct cause of death for neonates, the sensitivity, specificity, and proportion of deaths attributed to malaria by verbal autopsy would differ for studies including or excluding neonates. If the exclusion of neonates does affect validation study results and the verbal autopsy studies to be adjusted include neonates, it would seem problematic to use results from validation studies that excluded neonates in the adjustment process.
- Validation studies did not always analyse deaths with unknown causes (DUCs) (either by verbal autopsy or by the ‘gold-standard’ clinical method) in the same way. If validation studies analysed DUCs differently, then the causes implicitly attributed to these deaths probably differed. For example, when DUCs are included as a separate ‘cause of death’ category, the deaths are implicitly assumed to have not been caused by malaria. When DUCs are excluded from the analysis, it is implicitly assumed that they have a distribution of causes similar to deaths with known causes. Thus, results from validation studies that analyse DUCs differently may not be comparable and probably should not be combined.
- The validation studies included in the review had a wide range of values for sensitivity (mean: 56%; range: 45–75%); however, only the mean sensitivity was used in the adjustment formula.
- From an empirical standpoint, two of the adjusted values of the PDAM were greater than 64% (unadjusted values were ~45%). Although the true PDAM is not known for these two verbal autopsy studies, a value of 64% seems too high. Indeed, the very fact that the method was designed to adjust a PDAM of 45% to a value of 64% (see figure 1b of Ref. 13) suggests a problem with the method.

In conclusion, empirical results from any method for determining causes of death should probably only be adjusted if the adjustment method has clearly been shown to improve the validity of the empirical results. Although it is theoretically possible to improve the validity of verbal autopsy results by adjusting with results of a validation study, it is not practical to do so at present. Adjustment with existing validation studies may very well reduce some biases of verbal autopsy studies, but then introduce new biases. Without knowing the true causes of death and the differences in causes of death in populations where the adjustment is derived vs where it is applied, it is impossible to know the direction or magnitude of any of these biases.

References