LEADING ARTICLE

Tropical Trials and Tribulations

A J HALL* AND P AABY†

Field trials are the cornerstone in the evaluation of new public health interventions in the developing world. There is a tendency to base the introduction of new public health measures on very few trials. This is understandable given the expense in terms of money and time in mounting these trials and the enthusiasm to do something about the persistently high mortality in many parts of the world. However in the conduct of vaccine trials in West Africa a number of problems of design and analysis have presented themselves. These problems do not seem to be widely recognized but they affect the design of new trials and the interpretation of those already published. Some of these points are relevant to other forms of intervention study, such as vitamin A supplementation, and to trials in any part of the world. They fall into three broad areas; randomization, speculative versus practical trials and outcome measurement.

RANDOMIZATION

There is no need, one would hope, to reiterate the arguments for randomization in field trials of interventions. If the randomization is at the individual level this presents few additional problems compared to clinical trials, except for size, unless the intervention has an effect on those who are not in the intervention group. For example, if an intervention against mosquitoes by the use of individual insecticide impregnated bednets reduced the total vectorial capacity in a village then some account would have to be taken of this protective effect on the control group. This is a situation in which randomization by social group or community may be appropriate. A much more common reason for adopting this design is the logistic requirements of delivering the intervention. If a change in treatment by the village health worker, or the introduction of a special form of health education, is the intervention it may be impossible to randomize within the village. In these situations the unit of randomization is often the village itself. This usually leads to a relatively small number of units of randomization.

It is when this group randomization design is used for trials that we believe problems may arise. In particular when total mortality is used as the primary outcome measure, and we argue below that this is the most appropriate in most trials. There are several major risk factors for childhood mortality and these may not be evenly distributed across the few units of randomization. To take some examples:

Twins carry a high risk of mortality, particularly in the first year of life. In some societies they may represent as much as 15% of all infant deaths (Authors’ unpublished observations). At the same time the rate of twinning varies from one ethnic group to another, in West Africa this variation may be as much as three-fold. Motherless children or children of mothers with failed lactation carry a high risk of mortality. So if maternal mortality varies markedly between the villages randomized then this may result in mortality differentials in children.

Children who have had blood transfusions, for example for malarious anaemia, in the areas of the world where HIV infection is common and screening of blood is not, will have a higher mortality. So if the villages have differing access to sophisticated health care there will be differential mortality. Household structure is now recognized as the major determinant of measles mortality. If a child catches measles from another person living in the same room they are at markedly increased risk of severe measles and of death from the infection. This is illustrated in Table 1 where in six studies the source of infection for each individual could be derived. It is clear that the risk of death is dependent on whether or not the person is infected in the household or not. The effect of this risk factor can also be seen in Table 2 where the proportion of secondary cases is closely correlated to the case-fatality rate in different geographical regions. Measles infection almost invariably produces clinical disease with highly

*London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK.
†Institute of Ethnology and Anthropology, University of Copenhagen, Denmark.
characteristic symptoms and signs. Other airborne viral infections are not so easily recognized and often produce subclinical disease in those infected. This makes similar observations for other viruses difficult. We suggest that the same risk of severity associated with secondary household cases is true for them as well. Since these are thought to be a major predisposing factor in bacterial pneumonia it seems logical that the number of secondary household cases may determine the number of infections that go on to pneumonia and death. In most tropical countries these pneumonias are the foremost cause of death in children under five years of age, so this possibility has implications for all interventions that use total mortality or respiratory mortality as endpoints. The point is illustrated by observations of respiratory epidemics in West Africa. One study from Senegal examined six rural outbreaks of measles and concluded that the case-fatality rate was higher during the rainy season and in villages with a low state of nutrition prior to the outbreak. However, the six outbreaks had occurred among two ethnic groups (three each); in one group (Wolof) there had been a mean number of 3.3 cases per compound and a case-fatality rate of 10.1% for children under four years of age. In the other group (Serer), the number of cases per compound was 8.9 and the case-fatality rate 27.7%. Since the Serer with the high case-fatality had a poor state of nutrition and experienced outbreaks during the rainy season it may well have been the ethnic family structure rather than these other factors that determined the level of mortality. However, heterogeneity may be found not only between ethnic groups in the same region, but also within one ethnic group. In Guinea-Bissau, there was an outbreak of whooping cough in two Mandinka villages, one of them a traditional one with very large families and the other closer to the town and with smaller families. In the traditional village there was a mean of four children under two years of age per compound and a case-fatality rate of 29% in this age group. In the other village, the mean number of children under two was only two and the case-fatality was 12% (Authors' unpublished observations).

Epidemics present a particular difficulty in community randomized trials. They tend not to attack villages uniformly. In an outbreak of meningitis in The Gambia there were adjacent villages in which some were untouched and others had large numbers of cases. This difficulty may not be overcome by observing epidemics during the study as those that occur before the study may continue to exert an influence on mortality. There is now considerable evidence that measles infection contributes to mortality for months, if not years, after the child has been infected.

These difficulties do not arise in individually randomized trials, nor if there is a suitable period of pre-trial observation to allow randomization within strata of mortality. Recording the risk factors that are known, such as those above, and ensuring that they are equally distributed between the limbs of the trial will also overcome some of these problems. If they are not, for example for twins, then these are best excluded from the study since the number of events are likely to be too small to allow statistical adjustment. In a trial of Edmonson Zagreb measles vaccine in Guinea-Bissau there were 13 twins out of 240 children in one limb and three out of 187 in the other. Such a difference may seem small but even in large trials with large numbers of units the outcome in terms of deaths will be based on relatively small numbers, perhaps 20 to 30 in the non-intervention group. It is important in the analysis of these group randomized trials not to use individual data but to analyse by the randomized groups. Analysis should be by group. There are other problems such as standardization of health care within the village, uniformity of diagnoses and so on but these are well recognized. We do not believe that those above are.

**SPECULATIVE AND PRACTICAL TRIALS**

In a text on clinical trials the terms explanatory and intervention group. It is important in the analysis of these group randomized trials not to use individual data but to analyse by the randomized groups. Analysis should be by group. There are other problems such as standardization of health care within the village, uniformity of diagnoses and so on but these are well recognized. We do not believe that those above are.

### TABLE 1 Community studies of measles infection according to exposure.

<table>
<thead>
<tr>
<th>Site</th>
<th>Case-fatality rate</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Index cases</td>
<td>Secondary cases</td>
</tr>
<tr>
<td>Guinea Bissau</td>
<td>12%</td>
<td>37%</td>
</tr>
<tr>
<td>Guinea Bissau</td>
<td>2%</td>
<td>29%</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>1%</td>
<td>18%</td>
</tr>
<tr>
<td>Sunderland, England 1885</td>
<td>8%</td>
<td>22%</td>
</tr>
<tr>
<td>Senegal</td>
<td>7%</td>
<td>17%</td>
</tr>
<tr>
<td>Kenya</td>
<td>3%</td>
<td>10%</td>
</tr>
</tbody>
</table>

NB Superscripts are reference numbers.

### TABLE 2 Frequency of secondary cases and case-fatality ratio in measles, community studies.

<table>
<thead>
<tr>
<th>Country</th>
<th>Ratio of secondary cases (secondary/total cases)</th>
<th>Case-fatality ratio (deaths/no. ill)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guinea Bissau</td>
<td>58</td>
<td>27</td>
</tr>
<tr>
<td>Senegal</td>
<td>56</td>
<td>20</td>
</tr>
<tr>
<td>England</td>
<td>46</td>
<td>14</td>
</tr>
<tr>
<td>Guinea Bissau</td>
<td>45</td>
<td>14</td>
</tr>
<tr>
<td>Guatemala</td>
<td>38</td>
<td>5</td>
</tr>
<tr>
<td>Kenya</td>
<td>22</td>
<td>6</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>USA</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>Gambia</td>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>

NB Superscripts are reference numbers.
pragmatic were introduced to differentiate two types of trial. An explanatory trial is one in which the objective is to determine as accurately and scientifically as possible the effects of an intervention when given under experimental conditions. Pragmatic trials are those in which the intervention is introduced as nearly as possible as it would be in routine practice. This terminology has been criticized and we have chosen to use here those of Aristotle in defining the sciences. Speculative trials refer to those in which 'the end is to know and only to know'. The object of practical trials 'is also to know, but also to turn knowledge to account in devising ways of successful interference with the course of events.'

Immunization is a good example. Virtually all trials of vaccines are done using fixed dose intervals and within the trials the doses are actually given on the dates intended. When these vaccines are entered into a public health programme the vaccine is delivered when the children are presented for immunization. This will not usually be precisely on the target date. In the Gambian Hepatitis Intervention Study hepatitis B vaccine has been integrated into the Expanded Programme on Immunisation. The target ages for the four doses of vaccine are in the first month of life, at two months of age, at four months and at nine months. The actual proportion of vaccinees who receive their dose in the intended month is 70% for the first dose, 44% the second, 33% the third and 40% the fourth. Yet there are no trials with this vaccine examining these varying intervals. It is possible to examine within this study the effects of this variation on response. Since so few major field trials are done we believe that they should adopt this practical design although clearly the intervention will have to show its value in smaller speculative trials before the large practical trials are embarked on.

Speculative trials also require intensive training, supervision and motivation of field staff to minimize the errors of delivery and measurement. This input may, of itself, have an impact on disease and may interact with the intervention to produce more of an effect than the intervention alone. This extra input and potential interaction will not be present in the routine programmes. This effect is frequently seen in new programmes. There is an initial phase of training and staff enthusiasm with resultant high coverage and disease impact. Then another new programme comes along and the old one withers.

The reason for this preference for the practical trial is that many interventions are evaluated in the speculative manner and then immediately transferred to the routine health services. No formal examination is made of problems of implementation or whether the intervention will be effective under routine delivery services. Clearly trials cannot be carried out in all of the varied public health systems in the tropics. Case-control studies have a part to play here. Nevertheless there is a need to carry out some practical trials of new interventions and this must be done early or it will be deemed unethical.

The policy of measles immunization is an illustration of the difference between the two types of trial and the need for practical trials. The policy of immunization at nine months of age was based mainly on studies of seroconversion in different age groups. It was implicitly assumed that seroconversion was equivalent to protection and that non-conversion was the same as susceptibility. Based on month-of-age specific attack rates in a community study in Kenya it was calculated that immunization at eight or nine months of age would give the same number of prevented cases. This was based on the concept that immunization at eight months would result in fewer young cases but more vaccine failures and the reverse at nine months. Measles in immunized children was assumed to be as severe as in the non-immunized and that this would undermine confidence in the EPI so a nine-month target was chosen. However, subsequent studies have found that immunized children develop milder infection. Rather than undermining confidence 'mild measles' may be a powerful argument for the immunization. Furthermore, several studies have found both high acute and subsequent delayed mortality in children infected in early life. It is therefore likely that a practical trial of measles at seven or eight months of age would have shown a better survival compared with immunization at nine months. Since measles is a major cause of death the implications of such a small difference in age of immunization are not minor.

The practical trial design is another reason for community randomization since individual randomization does not mimic the public health programme.

OUTCOME MEASUREMENT
It is tempting only to look at the specific outcome which it is expected that the intervention will affect. Thus in a trial of haemophilus vaccine one would only measure haemophilus meningitis and septicaemia. The outcome might even be a biological intermediary rather than disease itself. A study of hepatitis vaccine might only look at carriage rates of the hepatitis B virus rather than the incidence of liver disease and hepatocellular carcinoma. This approach is less than ideal from two points of view. First the intervention may have unexpected beneficial effects. These may be of sufficient magnitude to tip the cost-effectiveness
balance towards favouring implementation. There was not a trial of measles immunization on total mortality before it was introduced as a public health measure. Such a trial would now be unethical. Yet there is now evidence from case-control studies of a much larger effect on mortality than might be expected from the prevention of acute measles deaths alone. 24,25 This finding is supported by the evidence of the long-term effects of measles on respiratory, nutritional and diarrhoeal deaths. 17,18 Had acute measles been less of a public health problem this intervention might have been rejected. Conversely, diseases, even infectious diseases, may have a more complex causation than we expect. The initial trials of BCG showing a good protective effectiveness against tuberculosis were extrapolated to other geographical areas and other ages of administration. The results, as the South India BCG trial showed, have not always been as good as one might have expected.

The second reason for the cause-specific outcome being less than ideal is one of safety. The first rule in preventive medicine must be to do more good than harm. There may be interactions in tropical environments which are unexpected. Iron supplementation in pregnancy has been transferred from non-tropical parts of the world in the expectation that it would counter anaemia and result in improved birthweights. There is considerable evidence that raised iron stores make malaria worse 26 and also other infections which are combatted by cell-mediated immunity. It is not clear whether iron supplementation in malarious areas does confer benefit or not.

Investigators must not fall prey to causal arrogance. The only way to avoid this and to evaluate the full spectrum of consequences of an intervention is to include total as well as cause-specific mortality in a study. Mortality is emphasized because it is the high levels of mortality which must be the target of public health programmes in tropical countries. At the same time measures of morbidity are of scientific importance in explaining failures and successes of programmes. The ideal trial would include all of the potential outcomes as outlined in Table 3.

The sample size of trials gets smaller as one moves from the top to the bottom of this table. This is because the dilutional effect of events not related to the target of the intervention gets smaller and smaller. Cause-specific mortality in many tropical settings has inherent misclassification. The cause is not attributed by a physician who attends the patient during their final illness but often by some form of oral postmortem. This involves interviewing people present during the final illness to try to ascribe the cause. The resulting misclassification may disguise true benefit or harm from an intervention—total mortality will help to counterbalance this. Syndrome-specific morbidity gives more biological plausibility and scientific information about the effects of the intervention. As does cause-specific morbidity and the use of biological intermediaries. One problem which arises from their use is that they demand some special surveillance system, and in the case of cause-specific morbidity, a sophisticated system. This will inevitably require the investigators to see a large proportion of ill subjects to investigate the cause. The best possible treatment must obviously be offered these patients. This should prevent deaths. It is therefore impossible to evaluate the impact on mortality if these measurements are made.

This conflict is often not apparent to funding agencies with little practical experience of health care systems in poor countries. One solution is to carry out two trials in parallel with a large field trial to assess
mortality and a smaller one for morbidity. This design is being used in the vitamin A supplementation trial in northern Ghana (personal communication: Ross, Dollimore and Smith). In this situation it will be essential that both trials end simultaneously or the results of one may make the continuation of the other unethical.

Finally, an intermediate biological measure may be used to evaluate the intervention such as an antibody response. This situation is useful in that it gives an early indication of the probable effects of the intervention but is also open to the dangers of unexpected events. Thus the appearance of a possible new hepatitis B virus has complicated the evaluation of immunization programmes in West Africa. Abnormal western blot patterns mimicking HIV infection have complicated the early assessment of HIV vaccines. The intervention may interfere with the biological measure without influencing disease. It is feasible that a vaccine might suppress septicaemia without preventing deaths from the infection. In the MRC trial of BCG the skin test conversion rate from the vaccine bore no relationship to disease protection.

CONCLUSION
These points of controversy in the design of trials are not trivial. Few large-scale field trials are performed. Major, even global, public health strategies may be determined by these few trials. Once a trial has been carried out which shows a positive effect it may be considered unethical to mount another even though this is in a different continent with a different spectrum of disease. In this small time window in which practical field trials can be done it is essential that the maximum information is gained from them. All of the points above must be carefully considered to ensure this.

ACKNOWLEDGEMENT
We are grateful to Professor Peter Smith for comments on an earlier draft of this paper.

REFERENCES