CHAPTER 27

Relative Risks, Benefits, and Costs of Intervention

Kathryn Bennett, Peter Tugwell, David Sackett, Brian Haynes

Those who provide, plan, or pay for health care must decide which health services should be provided to whom in order to effectively and efficiently reduce the burden of disease, disability, and untimely death. Such decisions demand a critical appraisal of existing evidence about the relative risks, benefits, and costs of intervention and, if required, a call for new evidence.

The approach described in this chapter, an outgrowth of work by Cochrane [1], Sackett [2], and Evans [3], provides a framework for assembling the specific subset of information that is most likely to help in reducing the burden of both morbidity (symptoms, physical, emotional, social, and functional impairment) and mortality for a disease. This is accomplished by subdividing the spectrum of health information inquiries into steps that constitute a logical progression—from quantifying the burden of illness, to identifying its likely causes, to validating interventions that prevent or ameliorate it and evaluating their efficiency, to monitoring the application of these interventions, and, coming full circle, to determining whether the burden of illness is reduced. Each of the six steps in the “measurement loop” (Fig. 27-1; Table 27-1) poses a different type of research or evaluation question and calls for a specific set of methodological standards applicable to either the critical appraisal of existing information or, if required, the generation of new evidence. The loop format emphasizes the importance of monitoring after implementing a health intervention to determine whether the planned reduction in the burden of illness is achieved. This process is iterative, since in almost all health care situations, the burden of illness is reduced by only a small proportion and repeated cycles of the “loop” are needed to eradicate even that portion for which effective interventions exist.

In the discussion that follows we show how an evaluation of relative risk, benefits, and costs can contribute to health decisions in less developed countries (LDCs). The measurement loop provides a practical guide to research methods of use to both (1) the “consumers” of research—health professionals and policymakers who wish to decide whether to apply the results of research investigations to health care decisions; and (2) the “doers” of research—those individuals involved in the planning and implementation of health research. Immunization for the prevention of measles and the use of oral rehydration salts (ORS) for the treatment of acute diarrhea of childhood are examples used to illustrate how the approach can be used to summarize the available evidence and make recommendations for action.

As will be discussed, application of the measurement loop approach can contribute to areas of concern to LDCs in at least four ways. First, the loop provides a framework for assembling and critically evaluating the evidence necessary to rank alternative health interventions in terms of both burden of illness and cost-effectiveness and to thereby determine priorities among interventions under consideration for implementation. Ranking of interventions within the context of health planning is of high relevance in situations of competing priorities and scarce resources. When the cost of a package of interventions outstrips the available resources, those actions with the greatest potential to reduce disease burden in terms of costs and effects given the available budget can be identified.

Second, the loop can contribute to the transfer of appropriate
technology to LDCs by assisting with the critical assessment and appropriate evaluation of medical innovations developed by industrialized nations, as the assumptions upon which the effectiveness of procedures developed in the west are based may not be valid in specific LDC settings.

Third, the loop emphasizes the importance of monitoring health programs after implementation. This is crucial to ensure that expected benefits are achieved and to reassess the health status of the population served. This is particularly relevant to large-scale community-based programs such as the provision of basic child and maternal health care. For example, monitoring of the quality of care provided by village and clinic health workers and of the availability of essential drugs is fundamental to the effectiveness of basic primary health care initiatives. If the cold chain is not monitored, the effectiveness of measles vaccination programs can be as low as zero [4,5]. If oral rehydration therapy is not used appropriately by mothers, the massive campaigns to promote it will not result in the levels of reduction in mortality that are possible by this simple intervention. In all these cases huge amounts of scarce resources could be wasted.

Fourth, the loop approach can assist in identifying gaps in the empirical data base, as will be illustrated with our two examples. Recent critical analyses to estimate the impact of a number of interventions relevant to LDC health problems have underlined this problem, with the authors needing to make a number of assumptions due to the lack of necessary data [6–9]. As is recognized by these authors, although evidence is available to support the efficacy (performance under ideal conditions) of many of the individual components of what is being recommended, (community) effectiveness under field conditions in actually reducing death and disability is untested in most instances. Further research under field conditions to validate recommendations is needed to ensure that the minimum magnitude of impact on health acceptable for a given intervention is achieved. This “essential national health research” [97,98] is the responsibility of researchers at the local or regional level and is a fruitful area for collaboration between individuals from industrialized and LDC countries.

**MEASUREMENT LOOP STEP 1: BURDEN OF ILLNESS**

The first step determines the current levels of morbidity (in terms of physical, emotional, and social function; symptoms) and mortality (death rates, healthy years of life lost). Good examples of studies carried out to assess the magnitude of health needs are: (1) a survey of health in seven countries directed by an international group of collaborators under the aegis of the World Health Organization [10]; and (2) a study of the major causes of illness and death in Ghana reported by the Ghana Health Assessment Project Team [11]. Both of these studies quantitatively assess functional health deficits as well as mortality; in fact, the latter study describes an interesting approach which quantifies the burden of illness by combining functional health deficits and mortality into units designated as “healthy days of healthy life lost.”

The magnitude of burden due to measles and acute diarrhea...
Table 27-1. Measurement iterative loop: Classification of health research and evaluation

1. **Burdens of illness:** Quantification of current levels of morbidity (functional impairment; physical, emotional, and social symptoms) and mortality (death rates; healthy days of life lost).
   a. **Avoidable burden of illness:** Those components of disability, symptoms, and mortality for which feasible, effective prevention or cure exists.
   b. **Unavoidable burden of illness:** Those components of disability, symptoms, and mortality for which no effective, feasible prevention or cure currently exists.

2. **Etiology:** Studies aimed at establishing the contribution of potential causal factors to the burden of illness.

3. **Community effectiveness:** How well an intervention known to be efficacious works when applied in the less than optimal conditions of clinical or field settings; community effectiveness is the product of five factors:

\[
\text{Community effectiveness} = \text{efficacy} \times \text{diagnostic accuracy} \times \text{provider compliance} \times \text{patient compliance} \times \text{coverage}
\]

**Efficacy** is the assessment of whether the maneuver, procedure, or service does more good than harm under conditions of optimal patient and provider compliance; **diagnostic accuracy** is the extent to which patients with the condition of interest are correctly discriminated from those without the condition; **provider compliance** is the extent to which the appropriate diagnostic and management (preventive, therapeutic, and/or rehabilitative) actions are complied with by the health provider; **patient compliance** is the extent to which patients comply with the health provider’s recommendations and treatment; and **coverage** is the assessment of whether a specific health intervention, known to be effective, is being appropriately utilized by all patients who could benefit from it, calculated by determining the proportion of people in need of the specific intervention who are offered it.

4. **Efficiency:** The assessment of the extent to which the maneuver, procedure, or service is being delivered to those who would benefit from it, with an optimal use of resources, in terms of the impact obtained for a specific cost.

5. **Synthesis and implementation:** The integration of burden of illness, clinical-community effectiveness, efficiency, and feasibility to make recommendations for implementation.

6. **Monitoring and reassessment:** Ongoing monitoring of the progress of the program using “markers” (short-term, intermediate, or long-term) to indicate progress selected according to the individual project; assessment of the success of the specific therapy or health intervention in achieving the predicted reduction in the burden of illness; and, if this is not achieved, identification of the specific factors responsible so they can be corrected.

of childhood is well documented, with both consistently ranked among the top five killers of the under-5 population in LDCs. The incidence of acute diarrhea in children under 5 is in the range of 2.2 to 3 episodes per year with an associated case fatality rate in the range of 0.5 to 1 percent, or approximately 4.6 million deaths per year worldwide [12]. The impact of multiple episodes of diarrhea on nutritional status and the ability to withstand infection further increases the associated burden in terms of increased risk of disease.

The risk of measles in the absence of immunization is essentially 100 percent. Although a relatively benign disease in industrialized nations, the case fatality rate in LDCs ranges between 1 to 5 percent due to the frequent and severe complications that include marked weight loss due to diarrhea, poor appetite, and general malaise [13,14]. Severity of complications and risk of death are significantly increased in malnourished children. Maternal antibodies, which protect against measles, begin to wane around 6 months of age in LDC populations; the optimal age for immunization now recommended is 9 months.

To make judgments about the quality of information concerning burden, three guidelines are applied. First, Is the attribute selected for the measurement relevant to the question being asked? The most appropriate and efficient method or indicator will vary according to the aspects of morbidity or mortality under consideration and the existing setup for collection of routine health statistics. For example, mortality rates indicate nothing about the magnitude of pain, mental suffering, or disability. These aspects of the burden of illness are equally as important as, or more important than, mortality rates for illnesses such as polio, malaria, schistosomiasis, leprosy, or trachoma, in which mortality contributes a relatively small proportion of the overall burden.

Second, Are the measurement methods accurate? This includes two components: (1) the ability of the indicator to reflect meaningful differences in the magnitude of the burden of ill-
ness and (2) the precision of the method. For example, mortality rates are assessed accurately in many (although by no means all) parts of the world where there is a legal requirement to report births and deaths, and these can be checked against other sources of existing information such as a census. Cause of death is rarely documented accurately (even in industrialized countries), and most components of morbidity will require specially designed surveys.

Cause-specific mortality rates for measles and acute diarrhea are particularly susceptible to problems of inaccuracy due to the need to rely on the report of family members or lay persons. Inaccuracy is also introduced by the likelihood that other diseases and chronic disorders such as malnutrition and infection may be present simultaneously. As a result, it is often impossible (and inappropriate) to attribute death to a specific cause. In fact, in Bangladesh, malnutrition is thought to be the cause of death in over 50 percent of all deaths, with concurrent diseases increasing the level of malnutrition and eventually precipitating the death [15].

Caution should be exercised when surrogate measures such as utilization or supply are used as indicators of burden, because they have been shown to be at variance with good survey data in some situations [10]. It is important to distinguish the burden of illness or need as assessed by health status information from three other types of measures often inappropriately substituted for need: want, utilization, or supply. Want refers to the public's perceptions of wants, expectations, and rights; it can be estimated by population surveys of self-perceived health status and expectations but more often is expressed by acts of individuals or groups (e.g., citizen groups formed to encourage the opening or discouraging the closing of local health facilities). The "seven countries" survey carefully assessed several such aspects of health in 12 study populations [10]. From Table 27-2 it is clear that self-perceived complaints (wants) may differ sharply from place to place irrespective of the prevalence of the disease of interest; for example, over twice as many Poles as Argentineans reported chest pain, while the prevalence of true angina pectoris in adults was the same.

Utilization data are abundant, as they are gathered in the course of providing most health services and are collated and reported by many governments. The data in Table 27-3 demonstrate the contrast between utilization data and the data for want and need shown in Table 27-2. Because of the abundance and low cost of routinely collected utilization data and under the assumption that use equals need, utilization data tend to be substituted for other measures in the evaluation of need. The shortcomings of substituting either want or use are important to emphasize since such information is frequently used as the basis for major health policy decisions (including targeting of research). Careful assessment of health status information should be insisted upon unless substitution of "want" or "use" data is judged acceptable.

The results of data collection must be easy to interpret and apply. Summary statistics for indicators, such as diarrhea-specific mortality rates, the proportion suffering lameness from polio, or the incidence rate for measles, are usually sufficient without any sophisticated mathematical transformation. Health indexes that combine several attributes or indicators of a disease into one number are often used to express disease burden. However, the assumptions and relative weights assigned to factors that make up the indexes need to be explicitly stated so that users can interpret the index and decide whether they can apply it in their setting.

The use of indexes that combine morbidity and mortality have the potential to be exceedingly helpful in quantifying the burden of illness and assessing the relative impact of different diseases. Such indexes quantify and then combine the healthy years of life lost due to both the disability and death associated with a specific disease. The result is a single measure of the overall impact of a disease providing a common unit of measure for comparison between different diseases. Weighting of the disability component according to severity can be used to increase the sensitivity of the index to the actual impact of an illness on an individual.

As mentioned above, the Ghana Health Assessment Project (Table 27-4) combined morbidity and mortality into a single index, "healthy days of life lost" (HDLLs) [11]. Details of the relative contribution of each are provided so that the results are meaningful to those utilizing them. Using this approach, the HDLLs due to 48 diseases or disorders were determined and ranked. Measles was the second-ranked disease, responsible for the loss of 23,400 HDLLs (96.6 percent due to premature death); acute diarrhea was ranked ninth, responsible for 14,500 HDLLs (93.3 percent due to premature death). However, the design, application, and use of such indexes is often not straightforward. Adequate data on disability days and years of life lost may not be easily obtainable, and deter-

<table>
<thead>
<tr>
<th>Country</th>
<th>Want</th>
<th></th>
<th></th>
<th>Need</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proportion</td>
<td>Proportion</td>
<td>Mortality</td>
<td>per 1000</td>
<td>from arteriosclerotic and degenerative heart disease</td>
<td></td>
</tr>
<tr>
<td></td>
<td>of adults</td>
<td>of adults</td>
<td>of adults</td>
<td>per 1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>reporting</td>
<td>with definite angina pectoris</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finland</td>
<td>45</td>
<td>1</td>
<td>2.5</td>
<td>1</td>
<td>3.4</td>
<td>2</td>
</tr>
<tr>
<td>Poland</td>
<td>41</td>
<td>2</td>
<td>1.7</td>
<td>3</td>
<td>1.4</td>
<td>6.5</td>
</tr>
<tr>
<td>Canada</td>
<td>36</td>
<td>3</td>
<td>0.8</td>
<td>7</td>
<td>2.7</td>
<td>4</td>
</tr>
<tr>
<td>Yugoslavia</td>
<td>30</td>
<td>4</td>
<td>0.8</td>
<td>6</td>
<td>1.4</td>
<td>6.5</td>
</tr>
<tr>
<td>United States</td>
<td>28</td>
<td>5</td>
<td>1.0</td>
<td>5</td>
<td>4.3</td>
<td>1</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>25</td>
<td>6</td>
<td>1.4</td>
<td>4</td>
<td>2.9</td>
<td>3</td>
</tr>
<tr>
<td>Argentina</td>
<td>16</td>
<td>7</td>
<td>1.7</td>
<td>2</td>
<td>1.9</td>
<td>5</td>
</tr>
</tbody>
</table>

Source: Adapted from [10].
Table 27-3. Utilization of health services in seven countries, %

<table>
<thead>
<tr>
<th>Country</th>
<th>Immunized within past 12 months</th>
<th>Taking prescribed medicine past 2 days</th>
<th>Consulted a physician within past 2 weeks</th>
<th>Consulted a dentist within past month</th>
<th>Admitted to hospital within past 12 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finland</td>
<td>17</td>
<td>28</td>
<td>12</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>Poland</td>
<td>59</td>
<td>19</td>
<td>16</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>Canada</td>
<td>23</td>
<td>30</td>
<td>15</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Yugoslavia</td>
<td>26</td>
<td>18</td>
<td>14</td>
<td>9.9</td>
<td>9.5</td>
</tr>
<tr>
<td>United States</td>
<td>36</td>
<td>34</td>
<td>16</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>14</td>
<td>24</td>
<td>18</td>
<td>8.5</td>
<td>7.9</td>
</tr>
<tr>
<td>Argentina</td>
<td>38</td>
<td>30</td>
<td>20</td>
<td>11</td>
<td>6.5</td>
</tr>
</tbody>
</table>


mining weights for the disability component is a complex and controversial task requiring a number of assumptions to be made. The Ghana team used expert opinion to estimate disability weights and recognized that special studies are needed to validate such assumptions. Application of this approach is also subject to value judgments regarding time preferences (a healthy day of life in the present has a greater value to an individual than one in the future) and the relative economic productivity of an individual at different ages. Prost and Prescott [6] and Barnum [7] have evaluated the effect of incorporating weighting for these factors into rankings using the HDLL approach. Their work has demonstrated that taking into account these factors can change the favored program(s) from those aimed at children to those aimed at "productive" adults.

BURDEN OF ILLNESS AND THE NATURAL HISTORY OF DISEASE

An important component to consider in determining the burden of illness is the natural history of a disease. The information gathered in natural history studies provides accurate information about the probability and magnitude of a change in health status over time in patients with specified conditions or in citizens with specified risk factors and predicts the burden of illness on a society. The study of tuberculosis reported by Doege [18] is an example of this type of study. In addition, the placebo group in randomized controlled trials can sometimes provide useful information on prognosis. Studies providing information on risk factors and the identification of

Table 27-4. Use of healthy days of life lost by Ghana Health Assessment Project: Specification of morbidity and mortality components

<table>
<thead>
<tr>
<th>Disease</th>
<th>Average age at onset</th>
<th>Average age at death</th>
<th>Disablement to death, %</th>
<th>Permanently disabled, %</th>
<th>Days of temporary disablement</th>
<th>Incidence</th>
<th>Days of life lost</th>
<th>Due to premature death, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Cholera</td>
<td>15</td>
<td>6.7</td>
<td>15</td>
<td>0</td>
<td>14</td>
<td>0.05</td>
<td>65</td>
<td>99.0</td>
</tr>
<tr>
<td>2. Typhoid</td>
<td>20</td>
<td>7.3</td>
<td>20</td>
<td>0</td>
<td>60</td>
<td>4.00</td>
<td>14,470</td>
<td>95.3</td>
</tr>
<tr>
<td>3. Gastroenteritis</td>
<td>2</td>
<td>1.0</td>
<td>2</td>
<td>0</td>
<td>14</td>
<td>70.00</td>
<td>11,065</td>
<td>94.6</td>
</tr>
<tr>
<td>4. Tuberculosis</td>
<td>20</td>
<td>35.0</td>
<td>25</td>
<td>0</td>
<td>200</td>
<td>2.00</td>
<td>11,065</td>
<td>94.6</td>
</tr>
<tr>
<td>5. Meningitis</td>
<td>10</td>
<td>20.0</td>
<td>10</td>
<td>0</td>
<td>30</td>
<td>1.25</td>
<td>1,227</td>
<td>99.3</td>
</tr>
<tr>
<td>6. Polio</td>
<td>3</td>
<td>2.5</td>
<td>3</td>
<td>95</td>
<td>25</td>
<td>0.22</td>
<td>1,227</td>
<td>77.4</td>
</tr>
<tr>
<td>7. Measles</td>
<td>2</td>
<td>3.0</td>
<td>2</td>
<td>0</td>
<td>21</td>
<td>39.00</td>
<td>23,358</td>
<td>96.6</td>
</tr>
<tr>
<td>10. Malaria</td>
<td>1</td>
<td>2.3</td>
<td>1</td>
<td>97.7</td>
<td>2</td>
<td>40.00</td>
<td>32,567</td>
<td>54.1</td>
</tr>
</tbody>
</table>

# of days of healthy life lost by the community attributable to the disease

Details of how days of healthy life lost have been determined

Source: Adapted from [11]. Table 1. “Disease problems of Ghana measured in terms of the days of healthy life which each costs the community (per 1000 persons per year).”
high-risk groups (for example, prediction of birth complications in pregnant women) are also included in this group.

Trend analysis is receiving increasing attention in the assessment of future health needs. The expanding elderly population and the AIDS crisis are two good subjects for trend analysis. A study has been done in China showing the social and economic consequences of the aging of the Chinese population in the twenty-first century and, accordingly, policy implications for old age support and health care needs [19].

AVOIDABLE AND UNAVOIDABLE BURDEN OF ILLNESS

The burden of illness can be usefully divided into “avoidable” and “unavoidable” to reflect whether efficacious methods of reducing the burden of illness are known. Unavoidable burden of illness includes those aspects of disability, symptoms, and mortality for which no feasible, efficacious prevention or cure currently exists. Clearly, allocation of resources to research into potentially reversible or curable etiological factors and the identification and testing of the efficacy of interventions are needed to reduce this component of the burden of illness.

Avoidable burden of illness includes those aspects of disability, symptoms, and mortality for which feasible, efficacious prevention or therapy exists. Demonstrated efficacy should be a requirement for the allocation of resources to study community effectiveness and efficiency and should guide the subsequent decisions by clinical or community health services.

To distinguish avoidable and unavoidable burden requires knowledge about possible causes of the burden and whether efficacious preventive, therapeutic, or rehabilitative strategies are available. The next two steps of the loop address these issues.

MEASUREMENT LOOP STEP 2: ETIOLOGY

The second step focuses on determining the causes of the health problems contributing to the burden of illness identified in step 1. Elucidation of causal factors requires a careful review of all the biological and behavioral attributes that might contribute to the specific health problem of interest; the “hypothesis space” (Fig. 27-2) is explored, and the most likely potential causes assessed. This involves a wide array of techniques ranging from the laboratory to the clinic and the community. For multiple causal factors, the relative contribution of each to a health problem is difficult to estimate but is crucial given the consequent commitment of time and money to developing interventions to reverse their effects. This is particularly relevant to the care of the under-5 population of LDCs, in whom malnutrition and infection interact to increase the morbidity and mortality associated with other diseases of childhood [13].

Figure 27-2. The hypothesis space.

Guidelines for assessing evidence for causation relationships originated with the work of Robert Koch in the late 1900s, looking at infections [20]. Virologists subsequently modified Koch’s guidelines, introducing the concept of using an experimental approach wherein the investigator controls exposure to the putative cause. These guidelines have been modified for use in noninfectious disease by Bradford Hill [21] and further developed [22] by the Department of Clinical Epidemiology and Biostatistics at McMaster University (Table 27-5).

The cause of measles is well-established, an RNA parovirus that affects only humans [13, 14]. Acute diarrhea of childhood, on the other hand, is known to be caused by numerous infectious and parasitic organisms. In only 50 to 70 percent of cases can an underlying cause be identified [23, 24].

The application of these commonsense principles involves two stages. First, the reader should scan the report to see whether the research methods used were strong or weak. Second, the reader should apply the set of “diagnostic tests” for causation to the methods and to the results. These two steps are elaborated upon below.

STEP 1: THE RESEARCH METHODS

Suppose, for example, an investigator wanted to find out whether vitamin A deficiency is a cause of acute diarrhea in the under-5 population in LDCs. What would be the most powerful sort of study to be found in the literature? Most investigators, it is hoped, would select a true experiment in humans; a study method in which children under 5 years of age with vitamin A deficiency would be randomly allocated (by a system analogous to tossing a coin) to receive or not to receive vitamin A supplements and then followed to see if the incidence of diarrhea was reduced in the children who received supplements. Evidence from such a randomized trial (class 1 evidence) is the soundest that can be obtained, whether it concerns etiology, therapeutics, or any other causal issue. The basic architecture of the randomized trial is shown in Fig. 27-3.

A randomized controlled trial of vitamin A supplementation has been implemented in Sumatra and shows a reduction in overall mortality rates among children aged 12 to 71 months
Table 27-5. Development of a set of guidelines for assessing evidence for causation

<table>
<thead>
<tr>
<th>OUTCOME</th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excessive Acute Diarrhea*</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Not (or Less) Exposed to Putative Causal Factor</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

* i.e. >2.2 episodes per year

Figure 27-3. A randomized trial.

two cohorts would then be followed forward in time to determine the diarrhea-related morbidity and mortality that occurred in each (see Fig. 27-4). If the rates were higher in the xerophthalmic cohort, this would constitute reasonably strong evidence that vitamin A deficiency led to increased diarrhea-specific morbidity and mortality.

However, such a cohort analytic study (class 2 evidence) is not as strong as a randomized trial: the reason becomes apparent if we consider the potential for lack of comparability of the study groups. The cohort study could provide a distorted answer to the causal question if children at high risk of diarrhea for extraneous reasons were not equally distributed between the cohorts of those who were and those who were not xerophthalmic. Are xerophthalmic children more likely to utilize the local health care facilities as often as nonxerophthalmic children? Are xerophthalmic children more likely to have mothers who are less well educated? Are xerophthalmic children generally less well and living under conditions more conducive to vitamin A deficiency? Yes. Therefore, a nonexperimental study such as the cohort study should be viewed with some caution. Documentation of the similarity of the two cohorts on potentially confounding factors can help in dealing

[25]. Unfortunately, diarrhea-specific morbidity and mortality was not assessed. Some controversy exists in the literature regarding the results of this trial, with questions raised regarding potential confounding associated with the uncertain baseline comparability of the control and supplemented villages and the lack of placebos and blind outcome assessments [26–32]. Further demonstrations of this relationship are needed.

Although the true experiment (randomized trial) would give the most accurate or valid answer to a question of causation, it cannot always be used for reasons of ethics and feasibility. For example, no researcher would ever consider conducting an experiment that would deliberately cause urinary tract schistosomiasis in a random half of a group of people to see whether they were more prone to developing bladder cancer. Thus, nonexperimental studies into issues of causation are much more likely to be encountered.

The next most powerful study method would identify two groups (or cohorts) of children, one mildly xerophthalmic (as an indicator of vitamin A deficiency) and the other not. These

<table>
<thead>
<tr>
<th>OUTCOME</th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed to Putative Causal Factor</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Not (or Less) Exposed to Putative Causal Factor</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

* i.e. >2.2 episodes per year

Figure 27-4. Cohort and prospective analytic surveys.
with this problem, but usually even with this information uncertainty will still exist. A prospective cohort study of the relation between mild xerophthalmia and the incidence of diarrhea has been done, showing a threefold increase in the incidence of diarrhea in mildly xerophthalmic children when compared to children with normal eyes [32]. This association was independent of both age and anthropometric status (weight for length).

Before-after studies have similar problems with comparability. For example, if diarrhea-related morbidity and mortality rates are compared before and after the implementation of a vitamin A supplementation program, it is difficult to control contemporaneous changes in other relevant factors that might also influence diarrhea incidence, such as a measles epidemic.

Finally, the case-control study (class 3 evidence) also deserves caution in interpretation (Fig. 27-5). In a case-control study, the investigator identifies two groups of children, one with and one without the outcome of interest: “cases” (children with more than the expected number of bouts of diarrhea per year) and “controls” (children with equal to or less than the expected number of bouts of diarrhea per year). In this case, the direction of inquiry is backward in time. Both groups of infants are then assessed for vitamin A deficiency, if children with excessive bouts of diarrhea are found more often to have vitamin A deficiency, this would constitute evidence, although not very strong evidence, that vitamin A deficiency might contribute to the risk of diarrhea.

Why is the case-control study low on the scale of strength of evidence? It is because it is so liable to bias. In addition to the biases noted with the cohort study, it is also susceptible to several other biases [34]. For example, given that diarrhea is associated with an increased risk of death, some children will not survive long enough to be included in a case-control study, much less to be assessed for vitamin A deficiency. As a result, an increased risk from vitamin A deficiency could appear to be absent when, in fact, deficiency was lethal for some children. Furthermore, if the diarrhea altered a child’s eating habits or ability to absorb vitamin A, then vitamin A deficiency would follow the diarrhea, not cause it. In fact, this issue is still the source of some debate [26]. Accordingly, the results of case-control studies are not as compelling as the more rigorous designs.

One final type of study deserves mention. This is the case-series or descriptive study (class 4 evidence), in which it is noted that a substantial proportion of children with repeated cases of acute diarrhea are vitamin A-deficient. No comparison group of any sort is provided, although there will often be reference to other studies by other investigators using different eligibility criteria (often called historical controls). About all that can be concluded is that higher diarrhea-related morbidity and mortality rates might (but will not necessarily) follow vitamin A deficiency in children. In terms of strength, such studies are best used to stimulate other, more powerful investigations.

In summary, then, readers of reports purporting to show etiology or causation should decide whether the research design used was strong or weak. Class 1 evidence from randomized trials provides the best basis for making decisions about causation, providing the trial is implemented reasonably well. A cohort study, although weaker than a randomized trial, is always to be preferred over a case-control study and can sometimes be trusted. For example, useful knowledge about the effectiveness of oral rehydration therapy comes from a large cohort study conducted in two communities in Bangladesh [35]. A cohort study conducted in Taiwan has provided us with the best evidence currently available on the risk of liver cancer due to hepatitis B virus [36]. The case-control study is a weak design, and has led to erroneous conclusions in the developed world, such as the now discarded link between reserpine and breast cancer [37]. However, for some extremely rare disorders (especially rare adverse drug reactions) case-control studies may be the only ones available. Finally, case reports should clearly be passed over when stronger evidence is available.

<table>
<thead>
<tr>
<th>OUTCOME</th>
<th>CASES</th>
<th>CONTROLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXPOSED TO PUTATIVE CAUSAL FACTOR (Vitamin A deficiency)</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>NOT (OR LESS) EXPOSED TO PUTATIVE CAUSAL FACTORS (No Vitamin A deficiency)</td>
<td>C</td>
<td>D</td>
</tr>
</tbody>
</table>

*Fig. 27-5. Case-control studies.*

STEP 2: APPLYING THE NINE DIAGNOSTIC TESTS FOR CAUSATION

As noted above, in making a causal decision information should be sought relative to some commonsense rules of evidence, herein called diagnostic tests and listed in Table 27-5. These are discussed below in order of decreasing importance. Table 27-6 indicates the impact of each upon causal decisions.

### Appropriateness and strength of study design

As discussed above, any consideration of causation should begin with a search for the best evidence. The appropriateness of the design for answering the question posed and, accordingly, the strengths and weaknesses in terms of the potential for bias should be identified.
Table 27-6. Importance of individual diagnostic tests in making the causal decision

<table>
<thead>
<tr>
<th>Diagnostic test</th>
<th>Test result positive</th>
<th>Test result neutral</th>
<th>Test result reverse of hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human experiments</td>
<td>+ + +</td>
<td>- - -</td>
<td>- - -</td>
</tr>
<tr>
<td>Cohorts</td>
<td>+ +</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Case comparison</td>
<td>+</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Strength:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>From experiment</td>
<td>+ + +</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>From cohort study</td>
<td>+ +</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>From case-comparison study</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Consistency</td>
<td>+ + +</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>Temporality</td>
<td>+ +</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Gradient</td>
<td>+ +</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Epidemiological sense</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Biological sense</td>
<td>+</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Specificity</td>
<td>+</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Analogy</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Symbols: + = Causation supported; - = causation not supported; 0 = causal decision not affected. The number of +’s or –’s indicates the relative contribution of the diagnostic test to the causal decision.

Strength of association

Strength here means the odds favoring the outcome of interest with, as opposed to without, exposure to the putative cause. The higher the odds, the greater the strength. There are two strategies for estimating the strength of the association. Both randomized trials and cohort analytical studies permit direct calculations of relative risk (strength) by comparing outcome rates in exposed and unexposed persons (Table 27-7). In case-control studies, however, relative risk can only be indirectly estimated by calculation of relative odds (Table 27-7).

Etiological fraction (EF) (Table 27-7) can be used to quantify the magnitude of the contribution of the putative cause to the outcome. It tells us the portion of outcomes observed in a community or population that are related (or “due”) to exposure to the putative cause, which would not have occurred in its absence (if it were, indeed, the cause). The important advantage of etiological fraction is that it quantifies the maximum reduction in the burden of illness that can be achieved in a specific population by interventions designed to modify the factor being assessed. For example, suppose a cohort analytic study of 200 children was conducted (of whom 100 had received vitamin A supplements and 100 had not) to investigate the relation between vitamin A deficiency and acute diarrhea. The results obtained are displayed in Table 27-7, the relative

Table 27-7. Strategies for estimating strength of association

<table>
<thead>
<tr>
<th>Exposure to putative cause</th>
<th>Outcome of interest</th>
<th>Yes</th>
<th>No</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>For class 1 &amp; 2 studies:</td>
<td>Relative risk (RR) = ( \frac{a(a + b)}{c(c + d)} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Etiological fraction (EF) = [ \frac{\theta(\text{RR} - 1)}{\theta(\text{RR} - 1) + 1} ]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>For class 3 studies:</td>
<td>Relative odds (RO) = ( \frac{\theta(\text{RO} - 1)}{\theta(\text{RO} - 1) + 1} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Etiological fraction (EF) = [ \frac{\theta(\text{RO} - 1)}{\theta(\text{RO} - 1) + 1} ]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( \theta = \) prevalence of exposure in the population being considered.

Table 27-8. Relation between vitamin A supplements and excessive acute diarrhea*

<table>
<thead>
<tr>
<th>Excessive diarrhea</th>
<th>Yes</th>
<th>No</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vitamin A supplements</td>
<td>No</td>
<td>20</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>10</td>
<td>90</td>
</tr>
<tr>
<td>RR = [ \frac{20}{100} \div \frac{0.2}{0.1} = 2 ]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EF = [ \frac{0.2 - 1}{0.2 - 1 + 1} = \frac{0}{0 + 1} ]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In this population \( \theta \) is 50% since half of the children were not "exposed" to vitamin A supplements (100/200)

\[ .5 \div .5 + 1 = .33 \]

*For example, > 2.2 episodes per year.
risk is 2.0, and the etiological fraction 33 percent (0.33 × 100). From this it can be concluded that children receiving no vitamin A supplementation are more than twice as likely to experience excessive bouts of diarrhea as children receiving them, and that 33 percent of the excessive diarrhea found in this study was "due to" vitamin A deficiency and therefore possibly avoidable.

Consistency

Consistency refers to the repetitive demonstration of an association between exposure to the putative cause and the outcome of interest, using different strategies and different settings. For example, the link between hepatitis B and liver cancer has been reported consistently in studies carried out in Africa and Asia [36, 38]. The consistency of these findings in a Caucasian population [38] lends strength to the causal relationship, greatly reducing the likelihood that other factors, whether toxic, genetic, racial, or nutritional in nature, may be responsible.

Temporality

Temporality refers to a consistent sequence of events—exposure to the putative cause followed by the occurrence of the outcome of interest. For example, as outlined above, if one were studying the relationship between vitamin A deficiency and diarrhea in children in a case-control study, it would not be possible to sort out whether the vitamin A deficiency predisposed to acute diarrhea or vice versa. Although this diagnostic test looks easy to apply, it is not. What if a predisposing factor is responsible both for low vitamin A levels and for diarrhea?

Gradient

Gradient refers to demonstrable increasing risk or severity of the outcome of interest in association with an increased "dose" or duration of exposure to the putative cause. Increasing rates of acute diarrhea associated with progressively increasing severity of xerophthalmia would satisfy this guideline. Studies of schistosomiasis investigating the relationship between the magnitude of worm burden and clinical disease show that light infections produce little impact on the individual while heavy worm loads are associated with significant clinical consequences [40]. Reverse gradients are useful, too; for example, if rates of diarrhea progressively decreased with increasing amounts of vitamin A supplementation, this would be a reverse gradient.

Epidemiological sense

This criterion refers to agreement with our current understanding of the distribution of causes and outcomes. The reason for this criterion being low on the list is that it does not require evidence that the exposure and outcomes affected the same individual. Vitamin A deficiency is not found in areas with a ready, natural supply of vitamin A. Indeed, initiation of investigations of the link between hepatitis B and liver cancer began based on observations that the incidence of liver cancer seemed to be associated with high levels of hepatitis B infection [41].

Biological sense

Biological sense refers to agreement with our current understanding of the responses of cells, tissues, organs, and organisms to stimuli. It is by this yardstick that nonhuman experimental data should be measured. Laboratory or animal studies have been responsible for many of the major advances in medicine, and such basic research plays a vital role in our understanding of the cellular basis for disease. For example, specific defects in cellular immune response have been documented in malnourished persons, suggesting a link between infection and nutrition [42].

Some biological observations are compelling [20, 43]. However, biological sense is relatively low among the criteria because, while it frequently provides essential explanatory information, virtually any set of observations can be made biologically plausible (given the ingenuity of the human mind and the vastness of the supply of contradictory biological facts).

Specificity

Specificity refers to the limitation of the association to a single putative cause and a single effect. One of the minor diagnostic tests, this is only moderately useful, and then only when it is present. The weakness of this test is underscored when considering teratogens, which commonly produce multiple effects in several organ systems.

Analogy

Analogy refers to the similarity of the association to another previously demonstrated causal relationship. The last and least of the diagnostic tests, an appropriate example in relation to diet and disease would be the possible relationship between the diarrhea found in pellagra due to deficiency of another vitamin, niacin [44].

Summary

When confronted by a question of causation, these nine diagnostic tests can be used to distill one's prior knowledge and, with the assistance of judgments such as those shown in Table 27-6, reach a causal conclusion. The diagnosis of causation is not simply arithmetic, and the strategies and tactics for making this judgment are still primitive. The diagnostic tests presented here are a start. It is suggested that their use, particularly when clearly specified before a review of relevant data, will lead to more rational—albeit less colorful—discussions of causation in human biology and health care.
This model can also be applied when assessing the effectiveness of interventions as well as causal agents in general. The emphasis in assessing effectiveness will be upon the first few criteria, since it is usually possible to find information from more rigorously designed studies concerning an intervention than concerning causal agents.

Although knowledge of causes is very important for identifying possibly efficacious interventions, it is not always necessary to insist upon a complete understanding of the aetiology of disease prior to intervention (John Snow took off the pump handle 50 years before the etiological agent for cholera was discovered).

**MEASUREMENT LOOP STEP 3:  COMMUNITY EFFECTIVENESS**

Step 3 looks at information on how well an intervention with potential for reducing burden will work when applied in the community. Disability and disease for which effective, feasible interventions exist can then be identified as avoidable.

Community effectiveness is determined by five factors: efficacy, screening and diagnostic accuracy, health provider compliance, patient compliance, and coverage. Evidence of efficacy may be transportable from country to country, but patient compliance, health provider compliance, and coverage are more likely to differ both between and within countries. Definitions and methodological guidelines for evaluating the quality of evidence relating to each of these five components are discussed below.

**Efficacy**

Efficacy asks the question, Can it work? It is defined as the extent to which a specific health intervention does more good than harm to patients (citizens) who are diagnosed correctly and appropriately cared for and who fully comply with recommendations or treatment. That is, the evaluation of efficacy assumes optimal diagnostic accuracy, health provider compliance, and patient compliance. Careful attention to these factors is required when deciding whether a study is truly assessing efficacy. Efficacy is the “anchor point” when estimating the benefit of interventions applied in a community setting and quantifying the upper bound or maximum benefit that can be achieved; it is regarded to be more stable and less liable to fluctuation in different circumstances than the four other components of community effectiveness.

Methodological standards for the assessment of efficacy are summarized in Table 27-9. It is unusual for a single study to fully comply with all seven standards for reasons such as the nature of the disorder in question, the limitations introduced by the experimental setting, or the cost of the assessments required. However, each of these standards should be considered when assessing or designing studies in order to determine whether the failure to satisfy a given standard invalidates the trial.

<table>
<thead>
<tr>
<th>Table 27-9. Methodological standards for studies of the efficacy of therapeutic or preventive health interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Is the research design appropriate?</td>
</tr>
<tr>
<td>a. Were the major sources of bias avoided or, if present, measured?</td>
</tr>
<tr>
<td>b. Were the methods used for sampling, assessment of exposure, and analysis acceptable?</td>
</tr>
<tr>
<td>2. Were all relevant outcomes reported?</td>
</tr>
<tr>
<td>3. Were the study patients or population recognizably similar to your own?</td>
</tr>
<tr>
<td>4. Were both clinical-community and statistical significance considered?</td>
</tr>
<tr>
<td>5. Is the therapeutic maneuver feasible in your setting?</td>
</tr>
<tr>
<td>6. Were all patients who entered the study accounted for at its conclusion?</td>
</tr>
<tr>
<td>7. Are the study results consistent with those of others?</td>
</tr>
</tbody>
</table>

Do we have evidence for the efficacy of measles immunization and ORS? We believe that the efficacy of both are well established [45,47]. The major evidence for the efficacy of measles vaccine comes from a randomized trial conducted in Great Britain by the British Medical Research Council, which showed an 85 percent reduction in the incidence of measles due to measles vaccine [46]. Other trials are also available focusing on seroconversion and reporting rates of up to 95 percent.

Although the comparative efficacy and safety of ORS for rehydration relative to intravenous fluid (IV) replacement has been clearly established in randomized trials [47,48], ethical considerations rule out randomized trials of ORS efficacy against no treatment, as such a design would necessitate withholding treatment from a dehydrated child. However, the efficacy of ORS compared to no treatment can be derived from descriptive studies, particularly those of severely ill cholera patients, many of whom would be expected to die in the absence of treatment. Accordingly, as will be discussed below, ORS represents an unusual case wherein sufficiently dramatic results have been observed in descriptive studies to draw conclusions about the efficacy of ORS. Based on the available evidence, when applied to patients with mild to moderately severe dehydration, ORS is considered to be 95 to 100 percent efficacious [47,49]. In cases of severe dehydration, IV treatment is recommended for initial rehydration with ORS administered subsequently for maintenance.

Is the research design appropriate?

**Random Allocation.** As was the case for determining causation, the well-executed randomized trial provides the strongest evidence for efficacy. In any disease that is not uniformly fatal, improvement unrelated to the intervention can best be taken into account by identifying and following a control group of patients who are similar in as many ways as possible to those receiving the intervention.
The comparability of the control subjects is absolutely crucial if the results are to be valid. The best way of achieving comparability is to ensure that every patient who enters the study has the same, known probability (typically 50 percent) of receiving one or the other of the treatments being compared; thus, assignment to one treatment or another should have been carried out by a system analogous to flipping a coin. It is usually easy to decide whether this was done, for key terms such as randomized trial or random allocation should appear in the abstract, the methods section, or even the title of such articles. A formal explanation for this strict rule is lengthy, but its conclusion is straightforward. As discussed above in measurement loop step 2 (causation), random allocation eliminates many of the biases that lead to false results in nonrandomized trials. Instances are numerous in which the developed world has been misled by accepting evidence from nonrandomized trials. For example, clofibrate was growing in popularity in the developed world before publication of a randomized clinical trial showed that it actually increased mortality rates; the drug was subsequently banned in several countries [50]. Furthermore, it has been estimated that 2500 gastric freezing machines had been purchased to treat tens of thousands of peptic ulcer patients by the time that a randomized clinical trial demonstrated the lack of efficacy of this procedure [51]. Similarly, a randomized trial was required in which angina patients were randomly allocated to undergo or not undergo internal mammary ligation only after the arteries had been surgically exposed to demonstrate how often symptomatic improvement can follow placebo medications and procedures [52]. Finally, the risk of relying on poorly designed trials is underlined by the work of Sacks, Chalmers, and Smith [53]. Their review of 50 articles evaluating six different therapies revealed that poorly designed, uncontrolled studies claimed important benefit 79 percent of the time, while well-designed, controlled trials claimed important benefit only 20 percent of the time.

The review of Bacillus Calmette-Guérin (BCG) vaccination by Clemens provides an excellent example relevant to LDC health problems. It is particularly interesting in that it addresses this same issue from the other direction—namely, poor methodology leading to rejection of a potentially useful intervention [54]. Their review concludes that the best available evidence (keeping in mind that these studies were conducted in industrialized settings) for the efficacy of BCG vaccination suggests a high efficacy for the vaccine for clinical tuberculosis. Undoubtedly, additional instances relevant to the major health problems in the developing world will arise as more information from controlled trials becomes available.

The efficacy of polio immunization [55] is accepted by virtually all providers and consumers of health care, as is the efficacy of antimicrobial prophylaxis in the secondary prevention of acute rheumatic fever [56]. On the other hand, the long-term efficacy of health education programs remains hotly contested [57]. The major reason for consensus in the first two cases and its absence in the third concerns the nature of the available evidence on the efficacy of these preventive maneuvers. The claims of the efficacy of polio immunization and rheumatic fever prophylaxis arise from placebo-controlled, randomized trials. Neither the subjects nor their clinicians knew who had received the active agent and who had received the placebo (thereby making the trial double-blind). In each case, the intervention was judged efficacious when it found that the target disorder was far less likely to befall individuals who received the intervention than those who did not.

It is widely acknowledged that noneperimental evidence (the cohort, before-after, case-control, and case-series designs described previously) can provide important information about etiology and the adverse effects of therapy. However, those most familiar with these noneperimental approaches agree that frequently they are not suited to the demonstration of efficacy, and this conclusion is substantiated in the requirements for experimental validation of new drugs and, increasingly, new surgical procedures. This conclusion is also true for the validation of health care services; the realization of the pitfalls of research designs has led to the recognition that noneperimental evidence is often insufficient in judging the efficacy of these services as well.

Skeptics insist upon evidence from true experiments in which, for example, individuals at risk for a given health problem are randomly allocated either to undergo or not to undergo early diagnostic and subsequent treatment procedures. Such experiments are now commonplace and have documented both the clear-cut advantages of early detection and treatment of moderate and severe hypertension [58], and the failure of the hospital admission laboratory screening for this problem to benefit either patients or those who care for them [59].

There are three exceptions to this requirement for experimental evidence. First, when a disorder is associated with a uniformly fatal outcome, any maneuver that saves lives is efficacious, and no randomized trial is necessary. Thus, experimental evidence was not required to validate the efficacy of treatment of tuberculosis meningitis with streptomycin [60]. Similarly, as noted above, initial experience with ORS in severely ill cholera patients showed a dramatic reduction in the case-fatality rate that could only be attributed to this simple intervention [49,61]. Second, when a disorder producing substantial mortality rates is uniformly cured, a controlled trial is also unnecessary—as, for example, with the use of penicillin for treatment of pneumonia in patients without other morbidity [62].

The third exception to the requirement for experimental evidence on the benefit of a health service occurs when random allocation to receive or not receive the maneuver is judged unethical or would be politically unacceptable. However, this latter exception is closely tied to geography and local practice. For example, in the developed world randomized trials to eval-
ute coronary care units were considered unethical in the United States at the same time that such trials were in progress in the United Kingdom.

In summary, then, although the randomized trial can sometimes produce an incorrect conclusion about efficacy (especially when it is a small trial), it is by far the best tool currently available for identifying the clinical maneuvers or health services that do more good than harm.

**Comparability of Study Groups and Prognostic Stratification.** Although random allocation is the method most likely to produce comparable study groups, this method does not guarantee similarity of groups with respect to all important variables. The use of prognostic stratification prior to randomization is one way to ensure that study groups are comparable with respect to known important prognostic factors.

For example, the underlying risk of developing and dying from the target condition should be identified for members of the study population. If subgroups in the study population vary widely in risk, consideration should be given to stratification. For example, if periodic screening tests for bladder cancer were being performed upon individuals with schistosomiasis, it would be important to stratify these subjects in terms of their tobacco use, a previously identified risk factor for bladder cancer. The investigator should also look for stratification in primary prevention trials of a new vaccine if members of the study population vary widely in their risk of developing the target infection; in fact this particular problem may have influenced the results of the major randomized trial of BCG vaccine conducted in India [63]. In the presentation of the results of such trials, the effect of this stratification upon the subsequent outcomes should be described in detail.

**Were all relevant outcomes reported?**

**Mortality and Morbidity.** The outcomes assessed should include (or predict) all those components of health status relevant to the intervention being assessed, including quality of life and patient preference. The World Health Organization’s definition of health (optimal physical, emotional, and social well-being) can be broken down to cover the spectrum of health in measurable terms utilizing the 10 D’s: death, disease, distress, discomfort, disability, dysfunction, disharmony (family impact), dissatisfaction, disposition (risk factors), and debt [64]. For example, when looking at measles immunization it is important (although more difficult) to determine the incidence of measles rather than simply seroconversion rates. This was done in the Medical Research Council trials conducted in Great Britain, which is a major strength of that data. Increasing attention is being directed toward the importance of including an assessment of the impact of the intervention on the patient’s quality of life. Such assessments address health status, changes in functional abilities, and patient preferences for alternative treatments or health states [65–67].

**Explicit, Objective Outcome Criteria.** Outcome criteria should be reproducibly defined, and for each outcome the following issues should be considered when deciding whether the results are likely to be meaningful:

1. Credibility of the assessment method used: For outcomes other than death that require a set of criteria for definition or an index or other instrument for measurement, credibility refers to the extent to which the method appears to measure what it is intended to measure and is therefore acceptable to those health professionals who might want to use the results of the study.
2. Reproducibility: This concerns the extent to which the measurement instrument produces the same result on repeated applications.
3. Whether the results are assessed by blind observers: Assessments performed by individual(s) unaware of specific characteristics of the patient that might influence the assessment (e.g., the study group to which the patient belongs, previous diagnosis, and/or aspects of the patient’s history) are necessary to minimize the likelihood of bias.
4. Responsiveness to change: This refers to evidence that the instrument or measurement approach used can detect a change in outcome when it is known to exist by other accepted assessment methods.

**Were the study patients (population) recognizably similar to your own?**

This criterion has two elements. First, the study patients must be recognizable; that is, how they were selected, what diagnostic criteria were used, and the patients’ clinical and socio-demographic status must be described in sufficient detail for you to be able to recognize the similarity between the study patients and your own patients. Second, the study patients must be similar to patients in your practice or community. To put it another way, you should ask yourself, Are the patients in the study so different from those I am concerned with that I could not apply the study results? When both recognizability and similarity are satisfied, you will be able to predict, with confidence, the outcomes to be expected from the application of the specific therapy or program to specific patients or populations.

The reports of the community-based studies of measles immunization and ORS [35,68] could each have benefited from inclusion of further details about the populations studied. Information on the age, sex, and socioeconomic status of the population is needed to allow readers to recognize the populations studied and judge the applicability of the results to the settings in which they work.

In summary, answers to the following three questions are needed to fulfill this criterion:

1. Were study subjects drawn from a random population sample or from a population of patients at some health facility?
2. Exactly how were they diagnosed or detected, including diagnostic criteria?
3. What were the patients’ sociodemographic status and clinical status (e.g., severity and duration of condition)?

Were clinical-community and statistical significances considered?

**Clinical-Community Significance.** Clinical-community significance here refers to the importance of a difference in health outcomes between treated and control patients. Differences observed between treatment and control patients are considered significant clinically or in regard to the community when they lead to changes in health provider or planner clinical or community-related behavior (i.e., decision making).

Clinically significant changes are usually reported in terms of relative risk reduction. For example, in the measles immunization trial carried out in the Kasongo [68], the relative risk reduction was 55 percent ([36 – 16]/36); that is, the risk of measles for a vaccinated individual would be less than half that of an unvaccinated individual. On the other hand, changes significant to the community are usually reported in terms of absolute risk reduction. For example, in the Kasongo report, the absolute risk of measles was 36 percent in the unvaccinated group and 16 percent in the vaccinated group, reflecting an absolute risk reduction for a vaccinated child of 20 percent.

It is important to note that although these assessments are often made from the perspective of the clinician or those making decisions about specific communities of individuals, clinical and community significance could also be defined from the patient’s perspective in terms of “important differences in the quality of life.” Utility measurement techniques, which quantify the strength of an individual’s preference for alternative health outcomes or interventions is an approach that has been used to address this issue in a number of diseases [66,67] but has not yet been widely applied in developing country settings.

A new perspective on clinical significance, the number needed to be treated (NNT), has recently been developed [69]. NNT is determined as the reciprocal of the absolute risk reduction (the difference in event rates between treated and untreated groups of patients). The advantage of this approach is that it combines the baseline risk of the patient for a target event with the consequences of treatment, which is not the case with either the relative risk reduction or the odds ratio. It is useful to the health provider in that it provides a yardstick for determining the number of patients who must be treated in order to prevent one adverse event. For example, using the measles data from the Kasongo project summarized above, the absolute risk reduction was 20 percent, or 0.2. Taking the reciprocal (1/0.2) yields an NNT of 5; that is, five children must be immunized to prevent one case of measles.

**Statistical Significance.** By contrast, statistical significance merely tells us whether a difference is likely to be real, not whether it is important or large. More precisely, the statistical significance of a difference is nothing more than a statement about the likelihood that this difference is due to chance alone.

The determinants of clinical and community significance are therefore the determinants of change in clinical or community action. If the results of a study lead you to manage your patients differently or to abandon an established program for a new one, the difference in the effects of these programs is thereby of clinical or community significance. The determinants of the statistical significance of any given result rises (that is, the p value falls) when the number of subjects in the study is increased, when the health outcomes show less fluctuation from day to day or from patient to patient, and when the measurement of this health outcome is both accurate and reproducible.

**Adequacy of the Sample Size or β Error.** On the basis of the foregoing, the busy reader can develop quick yardsticks to use when reading therapeutic articles. First, is the reported difference of clinical or community significance? Readers must scrutinize the difference in clinical outcomes in the studies to see whether they are of potential significance. If so, is the difference statistically significant—if yes, then the results are both real and worthy of implementing.

Second, if the difference is not statistically significant, is the number of patients large enough to show a significant difference clinically or in the community if it should occur? As discussed in the previous paragraph, the number of patients in a study is one of the determinants of statistical significance. Thus, if a study is huge, the difference in health outcomes can be statistically significant ("real") even when its magnitude is not clinically significant. Conversely, however, if a study is too small, even large differences of enormous potential clinical or community significance may not be statistically significant. For example, in the recent report on measles vaccination effectiveness by the Kasongo project team [68], no difference in overall mortality was found. However, examination of the death rates in the two groups suggests that an important difference may exist which was not found to be statistically significant due to insufficient sample size. The observed death rates were 4.5 percent in vaccinated group 1 and 7 percent in unvaccinated group 2, a risk reduction of 35 percent. Consideration of such a difference in the planning stages of this study would have revealed that to be 80 percent confident that a risk reduction of this size or greater would not be missed would require about 460 children per group (p = .05, one-tailed) rather than the approximately 250 children per group studied [68].

Is the maneuver or health intervention feasible in your setting?

**Replicable Description of Maneuver or Health Intervention.** The health care maneuver or intervention has to be described in sufficient detail for readers to replicate it with precision.
How were the patients diagnosed? Who did what to whom, with what formulation and dose, administered under what circumstances, with what dose adjustments and titrations, with which searches for and responses to side effects and toxicity, for how long and with what clinical criteria for deciding that therapy should be given, increased, tapered, or terminated? This issue presents special problems in trials evaluating the impact of a "package" of interventions. Such is the case in a community-based trial of ORS where the intervention consisted not only of ORS packets but also of a training program for volunteer depot holders and promotional activities related to the availability of ORS and how to use it [35].

Contamination and Coinervention. When reading the description of the maneuver in a published report, readers should note whether the authors avoided two specific biases in its application: contamination (in which control patients accidentally receive the experimental treatment)—this results in a spurious reduction in the difference in outcomes between the experimental and control groups) and coinervention (the differential application of additional diagnostic or therapeutic acts to either experimental or control patients that could influence clinical outcomes and thereby bias the magnitude of difference observed between experimental and control groups). It should be apparent that coinervention is prevented by "blinding"—both study patients and their clinicians as to who is receiving which treatment.

Coinervention due to unequal access of the two treatment areas to health care was a possibility in the Rahaman study of community-based distribution of ORS [35]; the fact that BCG vaccination leaves a scar may have alerted health workers to look more or less intensively for the occurrence of tuberculosis [65].

Were all patients who entered the study accounted for at its conclusion?

The astute reader will note how many patients entered the study (usually the numbers of experimental and control patients will be almost identical) and will tally them again at its conclusion to make certain that they correspond.

Again, the major trial of BCG in India provides an example [62]. Although the authors tell us exactly how many individuals received each of the three vaccination schedules being tested, it is not clear how many of these individuals were followed up and included in the results. Clearly, in a study such as this there are many practical difficulties in meeting this criterion, but it is crucial to the results. Perhaps representative samples (of adequate number statistically) could be identified after the maneuver is administered and carefully followed, thereby ensuring the validity of the results and reducing the number of intensive, expensive follow-ups required.

What can the reader do when outcomes are not reported for missing subjects? One approach (admittedly conservative and therefore liable to lead to the β error) is to arbitrarily assign a bad outcome to all missing members of the group with the most favorable outcomes. If this maneuver fails to shift the statistical or clinical significance of the results across a decision point, the reader can accept the study's conclusions.

Are the study results consistent with those of others?

This guideline concerns whether or not the study results agree with those of others. As is illustrated by Clemens' review of the trials of BCG conducted to date, different results can often be explained by the strength of the research methods used [54].

With the foregoing seven guidelines, both planners and providers of health care should be able to critically assess and judge the validity, applicability, and gaps in our knowledge about the efficacy of a specific health intervention.

SCREENING AND DIAGNOSTIC ACCURACY

Information on diagnostic accuracy can be obtained from studies of the extent to which patients with the condition of interest are correctly discriminated from those without the condition. This is not confined to technological or laboratory tests and includes the assessment of the accuracy of clinical signs (history, physical examination) or other paraclinical investigations (laboratory, x-ray, etc.). Detection or screening of patients with remedial health needs is necessary to establish the denominator of the population at risk. Screening of children under 1 year of age for measles susceptibility might be carried out in settings where it is judged that the cost of such screening is outweighed by the reduction in the number of children subsequently vaccinated and, accordingly, cost.

Unavoidable health needs (i.e., those for which no therapy is currently available) should not be the focus of screening, not only because of expense but also because of the negative consequences to health of the "labeling" that results. Hypertension, a health problem receiving increasing attention in developing countries, provides an excellent example of the potential impact of labeling. A study involving careful follow-up of steelworkers screened for hypertension revealed that absenteeism rose among previously unaware hypertensives labeled as a result of the screening process [70]. This led the authors and three Canadian task forces on hypertension to recommend that detection of hypertension only be carried out in settings where adequate therapy and long-term follow-up was ensured. In this way the disadvantages associated with being labeled hypertensive could be countered by the long-term benefits of blood pressure control.

Diagnostic accuracy with regard to measles is usually not an issue given that intervention in LDC settings is most often in the form of mass immunization programs with all children in a specified age range vaccinated. In the case of diarrhea among the under-5s, in some settings it may be necessary to discriminate acute from dysenteric and chronic diarrhea since the efficacy of ORS in these indications is not established [9].
EVALUATION OF HEALTH PROVIDER COMPLIANCE

Health provider compliance is assessed by whether the appropriate diagnostic and management actions (prevention, therapy, and rehabilitation) are performed by the health provider. The literature on quality of care calls these actions “clinical process” to differentiate them from “structure” (the supply of facilities and qualified personnel) and “outcome” (patient’s health status). Studies of health provider compliance should be restricted to situations in which the causal relationship between the process of care and patient outcome is established, i.e., for a demonstration of efficacy.

Information from evaluations of the ability of different health professionals to apply efficacious interventions is included in this category. A study carried out in Thailand to compare the health care and outcomes of women receiving postpartum tubal ligations by either specially trained nurse midwives or doctors provides an example of this type of evaluation [71].

For both measles immunization and acute diarrhea, provider compliance can drastically alter the impact of treatment. Crucial to measles immunization is maintenance of the cold chain. As previously noted, studies have shown that vaccine effectiveness can be as low as zero due to lack of care by providers in ensuring that the vaccine is maintained at the appropriate temperature [4,5].

Turning to ORS, provider compliance concerns not only appropriate management of diarrhea by the provider (i.e., the decision to treat and the vigor of treatment applied) but also the ability to effectively train others responsible for administering the ORS to the sick child, particularly mothers and other members of the child’s household. Studies are available that suggest that adequate levels of ORS preparation skills can be achieved in special training programs [47].

The appropriate use of antibiotics is a major concern in the treatment of acute diarrhea with their use considered to be appropriate in less than 10 percent of cases [24,72]. Inappropriate use of antibiotics for undifferentiated acute watery diarrhea can lead to harmful overgrowth of organisms such as Clostridium difficile and may actually prolong duration in some cases [24].

EVALUATION OF PATIENT COMPLIANCE

Patient compliance is determined by whether patients follow the health provider’s recommendations and treatment. Immunization programs that involve more than one painful injection in children can fail for this reason alone. Oral rehydration salts for acute diarrhea of childhood provides another good example. Effective treatment requires compliance by both the mother and infant. The effectiveness of most community-based programs depends upon the mother accepting ORS as a treatment and learning to correctly prepare and administer the ORS liquid. The extent to which mothers accept that ORS fluids will stop diarrhea, as opposed to causing it, may influence whether mothers comply with recommendations for its use. Studies of mothers’ knowledge and behavior related to ORS preparation and administration have reported varying levels of compliance ranging from 80 to 5 percent [45]. An unannounced assessment of fluid composition revealed that solutions prepared by village workers were on the average satisfactory, while those prepared by mothers tended to be diluted [73].

EVALUATION OF COVERAGE

Coverage refers to the proportion of the target population, i.e., all patients (citizens) in need of a specific, efficacious intervention who are offered it. Coverage should be differentiated from patient compliance: coverage describes whether or not an individual in need of a specific health intervention makes contact with the health professional, while patient compliance encompasses the adherence by the patient to the subsequent advice. Moreover, as coverage estimates do not take into account the other components of community effectiveness, such estimates do not reflect the proportion of all patients in need who are effectively treated [74].

For example, coverage for ORS will depend upon whether the delivery system is health center or home-based. Clearly, home-based distribution systems should result in close to 100 percent coverage. Raitman’s study of health center–based distribution found that between 60 to 70 percent of children with diarrhea attended a clinic for care of an episode [25]. He also showed that coverage declined with distance from the health facility; 90 percent of diarrheal cases within 1 mile attended the clinic, while less than 70 percent within 2 miles did.

Coverage evaluation requires that the use of health services be related to the need for them in a defined population during a specified time period. Utilization of services in the form of activity-to-population ratios, although very popular, is rarely accurate as a measure of availability since it fails to incorporate information about need. For example, mass measles immunization campaigns in LDCs usually target the specific age group of infants (usually 7 to 9 months of age) considered most at risk (i.e., most likely to be susceptible to measles). Since the proportion of susceptibles is expected to be very close to 100 percent, no actual assessment of susceptibility is usually obtained. It is worth noting that coverage estimates for measles immunization usually reflect how many children were actually immunized rather than how many children were offered an injection, and therefore incorporate patient compliance. The report by McBean on a measles immunization campaign for children 6 to 36 months of age in the Cameroons illustrates the impact that need has on measles immunization coverage of susceptible children [75]. Overall 78 percent of children in the target age group were taken to the measles immunization sessions. However, this study found that only 40 percent of children in this age range were susceptible to measles, and, of these, only 51 percent were given vaccination during the campaign. In fact, children not susceptible were
more likely to present for vaccination than those who were (53 percent of those who remained at home were susceptible versus 36 percent of those who attended). Thus, although 78 percent of children attended, coverage of those in need was only 51 percent.

Availability and acceptability of effective health services

Utilization of effective health services by those in need depends upon availability and acceptability. Availability concerns whether efficacious health services are accessible to those in need and they are aware of their being available. This can be measured by estimating the supply of services (the resource-population ratio), taking into account the distribution. Awareness of the availability of these services by those in need is also relevant here. This component is important in screening studies to ensure that there is appropriate linkage of identified patients with the condition of interest to the health provider, so that they get treated rather than just "labeled." as in the hypertension example above [70]. For both measles immunization and ORS, promotional campaigns are required to enhance levels of coverage.

Quantification of acceptability can best be obtained from surveys and should not be confined to users of health services, since the latter statistic does not tell us how many individuals in need of an intervention fail to receive it. Measurement of the acceptability of health services can usefully be divided into patients' perceptions of (1) the resources or facilities, (2) the behavior of health professionals and their staff, and (3) the benefits expected from the health service. The acceptability to mothers of ORS as described above provides a useful example.

PREDICTION OF THE MAGNITUDE OF COMMUNITY EFFECTIVENESS

It is important to be able to estimate the impact of specific treatment interventions when assessing whether a program is achieving its full potential, or when assessing the economic efficiency of the whole program or looking at alternatives for improving the program. All of the economic approaches described in the next step of the loop except for cost-minimization require an estimate of community effectiveness. It is important to point out that the lack of rigorous and valid evidence on clinical-community effectiveness limits the validity of economic evaluations [78]. Given that assumptions must be made in conducting cost-effectiveness, cost-benefit, and cost-utility analyses, it is important to minimize sources of uncertainty and imprecision in estimates of impact.

The relationship between the five factors that determine community effectiveness is most accurately estimated by using a multiplicative conditional probabilities model (see boxed text). Unfortunately, the necessary information on the conditional probabilities is rarely available. However, an acceptable alternative [77] is to use a simple multiplication formula. This assumes that all the factors are independent. It is unlikely that the factors are highly correlated, (e.g., many patients given optimal care do not comply), but research is needed to confirm the robustness of the simple multiplicative formula. When community effectiveness is expressed in terms of the burden of illness (for example, healthy days of life lost), the percent reduction in the total overall burden of illness in a community (identified in step 1) that might be achieved by a specific health intervention can be estimated.

### Conditional Probabilities

The multiplicative law of combining conditional probabilities [77] states that the chance of two events, $x$ and $y$, both happening is $p_{xy}$ and $p_{xy} = p_x p_y$, assuming both events are independent. When one event is dependent upon the other, such as $y$ is dependent upon $x$, this is expressed as $p_{xy} = p_y | x = p_y | x p_x$. Community effectiveness can therefore be stated as follows:

\[
\text{Community effectiveness}^* = p_{\text{coverage}}
\times p_{\text{diagnostic accuracy/coverage}}
\times p_{\text{health provider compliance/coverage and diagnostic accuracy}}
\times p_{\text{efficacy of treatment/coverage and diagnostic accuracy and health provider compliance}}
\times p_{\text{patient compliance/coverage and diagnostic accuracy and health provider compliance and efficacy of treatment}}
\]

Under the assumption of independence this can be simplified to:

\[
\text{Community effectiveness}^* = \text{Efficacy} \times \text{diagnostic accuracy} \times \text{health provider compliance} \times \text{patient compliance} \times \text{coverage effectiveness}
\]

*Determinants of community effectiveness are organized according to the sequence in which they occur in the process of delivering and receiving health care.
Table 27-10. Sample calculations for community effectiveness: Measles immunization

<table>
<thead>
<tr>
<th>Type of estimate</th>
<th>Efficacy</th>
<th>Diagnostic accuracy</th>
<th>Provider compliance</th>
<th>Patient compliance</th>
<th>Coverage</th>
<th>Community effectiveness (measles incidence)</th>
<th>Percent of efficacy achieved in the community</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Best evidence&quot; estimate</td>
<td>69% reduction in measles incidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>35% ↓</td>
<td>(\frac{35}{85} = 41%)</td>
</tr>
<tr>
<td>Range of evidence estimate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>51-54% ↓</td>
<td>(\frac{25}{85} = 29-41%)</td>
</tr>
<tr>
<td>From studies combining all components</td>
<td>49.5-69% reduction in measles incidence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25-35% ↓</td>
<td>(\frac{25}{85} = 29-41%)</td>
</tr>
<tr>
<td>From studies of individual components</td>
<td>85% reduction in measles incidence</td>
<td>100%</td>
<td>0-95%</td>
<td>100%*</td>
<td>0-44% ↓</td>
<td>(\frac{0.44}{85} = 0-52%)</td>
<td></td>
</tr>
<tr>
<td>Under improved conditions of provider compliance and coverage</td>
<td>85%</td>
<td>100%</td>
<td>90%</td>
<td>100%*</td>
<td>90%</td>
<td>69%</td>
<td>(\frac{69}{85} = 82%)</td>
</tr>
</tbody>
</table>

*Assuming coverage incorporates compliance (see text).

Table 27-10 shows some sample calculations for our two examples using the simple multiplication formula. The purpose of these sample calculations is to illustrate the effect on efficacy estimates (obtained under ideal circumstances) of the other factors that influence community effectiveness. These calculations are "best estimates" based on currently available evidence, and we fully appreciate that they may be subject to error. However, we feel that they serve the purpose of illustrating the relative magnitude of the difference between efficacy estimates for an intervention and its impact when implemented under community conditions. In addition, by examining each of the individual components of community effectiveness we can identify the ones that, if improved, would have the greatest impact on increasing community effectiveness.

There are several approaches for deriving best estimates based on the results of multiple studies. These include voting (based on a tabulation of study results and assessment of study quality), delphi and nominal group techniques, consensus, and statistical approaches (meta-analysis, overview analysis) for combining results from two or more studies [78]. For our purposes, meta-analysis constitutes the approach of choice, as this quantitative technique would provide us with the best estimate based on the currently available evidence in the literature [79]. However, we were not able to employ this technique because for both of our examples the current data base did not meet the requirements and assumptions necessary to do so [45,47]. Instead, as will be discussed in detail below, we have used a systematic approach to reviewing the literature using the methodological criteria discussed above. We have defined as our best estimate the results of the study considered to be the strongest according to the methodological standards applied. We have also summarized the range of results, including all of the studies reviewed.

For measles immunization, our estimate of efficacy comes from randomized trials conducted in developed country settings. From these data, we estimate the efficacy of measles immunization in reducing the incidence of measles to be 85 percent. Community-based studies of measles immunization conducted in LDC settings are available, reflecting the collective impact of the other factors that determine effectiveness at the community level with the exception of coverage. None of the studies available include an assessment of coverage—that is, the proportion of children susceptible to measles who received a vaccination. Therefore, in using these studies to estimate community effectiveness, an adjustment must be made to take into account the effect of less than 100 percent coverage on the overall impact in the community. For others interested in deriving estimates for use in their own setting, it will often be worthwhile to carry out local checks on compliance and coverage to ensure accuracy (given that the cost of the expansion of health services warranted the expense of conducting the check). Efficacy should be reasonably stable and therefore suitable for generalization to most populations [76].

The Kasongo study provides the best estimate of community effectiveness, according to the methodological guidelines discussed above, reporting a 69 percent reduction in the incidence of measles in children 7 to 35 months of age. Adjusting this for coverage, estimated to be in the range of 50 percent [75,80], suggests the reduction in measles incidence due to an immunization program is approximately 35 percent. It is important to point out that a comprehensive review of the literature reveals a wide range of results, suggesting that the community impact of measles immunization programs may vary significantly [45]. As is evident from the literature, provider compliance and coverage can have a major impact, causing suboptimal levels of community effectiveness. For example, cold chain problems have been shown to have resulted in the
administration of a completely ineffective vaccine [4-5]. Regarding coverage, there is clearly room for improvement, and successful programs will need to include strategies to maximize coverage in their overall activity plan.

As is shown in Table 27-10, the potential for a major loss of measles immunization efficacy is great. The Kasango results reflect a 59 percent loss, and taking into account other available data suggests the loss could be as great as 100 percent. If provider compliance and coverage could be raised to a more optimal level (90 percent for each), and it seems reasonable that this could be achieved given the appropriate motivation and commitment of those involved, community effectiveness could be improved to 69 or 82 percent of the maximum possible impact based on efficacy estimates.

For oral rehydration salts (Table 27-11), our estimate of efficacy comes from several initial studies; in particular, the report of the experience under extraordinarily adverse conditions in a Bangladesh refugee camp where the dramatic reduction of the cholera case fatality rate to 3.6 percent was observed [67]. It is hence reasonable, as has previously been discussed, to conclude that efficacy under ideal conditions is at least as great as 95 percent. Estimates for community effectiveness and the other components of the equation are based on community-based field trials of ORS and studies targeting the specific individual components [47]. Again, for others interested in deriving estimates for use in their own setting it will often be worthwhile to carry out local checks on compliance and coverage to ensure accuracy. Also, as has been pointed out in the literature, the overall impact of ORS on a community may depend upon the proportion of acute watery diarrhea compared to dysentery or chronic diarrhea for which the effectiveness of ORS is not yet established [9].

The impact of ORS has been evaluated in several community-based studies. Accordingly, the results of these trials reflect the collective impact of all five factors. The best evidence, according to the guidelines outlined above, comes from the trial conducted by Rahaman in Bangladesh [35]. The results of this community-based trial suggest that we can expect a 79 percent reduction in diarrhea case fatality rates with the implementation of health center-based distribution programs. The magnitude of the impact observed should be interpreted with some caution and taking into account that this study represents an extremely intensive community-based effort that may not be easily replicated in other settings because of resource limitations. As well, total mortality was not reported, leaving some uncertainty as to the accuracy of the cause-specific rates reported by unblinded observers. Once again it is important to point out that a comprehensive review of the literature reveals an extremely wide range of results suggesting that the impact of ORS programs can vary significantly [47]. As is evident, patient compliance and coverage can have a major impact, causing suboptimal levels of community effectiveness. For example, it is estimated that home-based distribution programs may actually be less effective than health center-based programs in terms of reduction in case fatality rates [47]. Patient compliance, in particular, the attitudes and behavior of the mother, is the most likely source of this loss of efficacy. Accordingly, as is shown in Table 27-10, although there is evidence that up to 83 percent of the maximum efficacy of ORS may be achieved at the community level, reflecting a loss of less than 20 percent of efficacy, impact could be as low as 3 percent of efficacy, reflecting a virtual loss of efficacy. If patient compliance and coverage could be raised to a more optimal level (again, 90 percent for each seems reasonable) community effectiveness could be improved to 66 or 70 percent of the maximum possible impact based on efficacy estimates. (Given the results of Rahaman, where 83 percent of efficacy was maintained, this estimate of 70 percent should be viewed as the minimum target under improved conditions of health service delivery.)

<p>| Table 27-11. Sample calculations for community effectiveness: Oral rehydration salts |
|---------------------------------|-----------------|----------------|-----------------|-----------------|-----------------|----------------|</p>
<table>
<thead>
<tr>
<th><strong>Type of estimate</strong></th>
<th><strong>Efficacy</strong></th>
<th><strong>Diagnostic accuracy</strong></th>
<th><strong>Provider compliance</strong></th>
<th><strong>Patient compliance</strong></th>
<th><strong>Coverage</strong></th>
<th><strong>Community effectiveness (case fatality rate)</strong></th>
<th><strong>Percent of efficacy achieved in the community</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Best evidence’ estimate</td>
<td>79% ↓</td>
<td>79% reduction in case fatality rate</td>
<td>79% ↓</td>
<td>79% reduction in case fatality rate</td>
<td>79% ↓</td>
<td>79% reduction in case fatality rate</td>
<td></td>
</tr>
<tr>
<td>Range of evidence estimate</td>
<td>56–79% ↓</td>
<td>56–79% ↓</td>
<td>56–79% ↓</td>
<td>56–79% ↓</td>
<td>56–79% ↓</td>
<td>56–79% ↓</td>
<td></td>
</tr>
<tr>
<td>From studies combining all components</td>
<td>56–79% ↓</td>
<td>56–79% ↓</td>
<td>56–79% ↓</td>
<td>56–79% ↓</td>
<td>56–79% ↓</td>
<td>56–79% ↓</td>
<td></td>
</tr>
<tr>
<td>From studies of individual components</td>
<td>95–100% ↓</td>
<td>95–100% ↓</td>
<td>95–100% ↓</td>
<td>95–100% ↓</td>
<td>95–100% ↓</td>
<td>95–100% ↓</td>
<td></td>
</tr>
<tr>
<td>Under improved conditions of provider compliance and coverage</td>
<td>95</td>
<td>95</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>66</td>
</tr>
</tbody>
</table>

The table shows the calculations for community effectiveness and the percent of efficacy achieved in the community.
It should be noted that for simplicity we have ignored the iatrogenic complications of therapy in these examples. For example, measles vaccination can include complications of both morbidity and mortality, although the risk is considered to be very small. ORS, when inappropriately administered, can result in severe electrolyte imbalances, particularly hyper- and hyponatremia.

These estimates also do not tell us to what extent the community as a whole would be better off if an intervention were implemented. The reduction in the overall burden of illness experienced by the community will depend upon two factors. First, the proportion of the total burden accounted for by a specific disorder will determine the overall influence on the health of the community. For example, it is well known in LDC settings that children who avoid measles may simply succumb to some other disease.

**MEASUREMENT LOOP STEP 4: EFFICIENCY**

Step 4 provides information on whether the intervention is being delivered to those who would benefit from it with an optimal use of resources, and involves the relationship between the costs and effects.

\[ \text{Efficiency} = \frac{\text{patient benefit (outcome)}}{\text{cost}} \]
\[ \text{or} \]
\[ = \text{net costs ($)} - \text{net benefits ($)} \]

Efficiency is expressed as effects (the number of lives saved, number of disability days avoided) obtained for a specific cost (expressed in dollars). Again, the assessment of efficiency should not be conducted in the absence of evidence of efficacy and community effectiveness, although in research, the study of efficiency may be combined with that of efficacy and community effectiveness. Other issues that should be considered in the assessment of efficiency have been discussed by Stoddart [81,82]. The most widely used approaches to evaluating efficiency are as follows: (1) Cost-effectiveness: effectiveness is measured in a common unit of health impact such as lives saved, levels of function restored, or proportion of patients in whom symptoms are controlled. This may be derived from efficacy studies in tightly controlled situations or from studies incorporating variable numbers of the other components that make up community effectiveness. (2) Cost-benefit: effectiveness (benefit) is measured in the appropriate unit (lives saved, functional improvement, symptom control) and then converted into monetary units (i.e., dollars and cents). (3) Cost-utility: effectiveness (utility) is measured in the appropriate unit (lives saved, functional improvement, symptom control) and then converted into “utility units” that are measures of the relative social value (importance) of these outcomes.

The cost-effectiveness of measles vaccination has been investigated in a number of developing country settings [83–87]. This work has considered the relative costs and consequences of measles immunization alone as well as with other immunizations within the context of the WHO expanded program of immunization (EPI). Overall, these studies suggest measles immunization is cost-effective. As these studies show, cost-effectiveness is influenced by the number of children vaccinated and whether measles vaccine is given alone or is incorporated into other vaccination schedules. In contrast, although some information is available on the cost of ORS [88,89], a formal evaluation incorporating the methodological issues relevant to conducting an economic evaluation alluded to above has not yet been carried out.

**MEASUREMENT LOOP STEP 5: SYNTHESIS AND IMPLEMENTATION**

The fifth step integrates feasibility with the estimates of community effectiveness and efficiency obtained in the previous steps to make recommendations for action. This step identifies: (1) the possible limiting constraints on the effectiveness and efficiency estimates in the setting where the intervention will be implemented; and (2) whether and to what extent these constraints can be removed or reduced. Conclusions can then be drawn as to the likely impact of the intervention or program on the burden of illness. Constraints include social, cultural, and political barriers as well as adequacy and availability of facilities and manpower and budgetary considerations.

Inherent in this step is the setting of goals, important both for defining realistic objectives and for assessing success. Estimates of community effectiveness and efficiency can be used to evaluate the success of the intervention over time.

**MEASUREMENT LOOP STEP 6: MONITORING**

The sixth step is ongoing monitoring of the impact of a health program. Monitoring needs to be tailored to the individual program and may consist of short, intermediate, and long-term criteria of success selected to profile progress. Monitoring methods usually focus on one or more of the following categories: (1) structure: buildings built and equipped and qualifications of health workers; (2) health care and administrative process; the appropriateness of case finding and care of patients with the target disorder; and (3) patient or citizen health outcomes: changes in symptoms, disability, and mortality. The selection of markers needs to take into account the representativeness of the spectrum of disease and age groups, accuracy, and feasibility [90].

The use of “sentinel sites” and community diagnosis systems as monitoring strategies has received attention recently [91]. Sentinel sites are chosen according to the extent to which they are judged as representative of the surveillance targets of interest. Hence, problems identified at these selected sites can be used as indicators of problems in the overall setting where a program is being implemented. At these sites, a small amount
of additional manpower may be provided to ensure careful reporting of the data of interest.

Community diagnosis techniques consist of population-based monitoring of the health events of interest rather than the selection of specific sites and can include indicators relevant to monitoring specific programs. This approach involves an annual census plus regular reporting of events by community health workers or appropriate others generated through regular surveillance of all homes in a predefined area.

Finally, case-control and cohort studies have been proposed as cost-effective, time-efficient methods for monitoring the impact of vaccination programs [92–94]. The impact of BCG mass immunization programs in Sri Lanka has been assessed using the case-control approach with the results suggesting an unacceptable level of impact [94].

As is evident from the review of community effectiveness presented above, monitoring is crucial in the successful implementation of community-based programs of measles immunization and ORS. The selection of indicators relevant to provider and patient compliance and coverage (e.g., potency of measles vaccine, number of children effectively immunized, inappropriate use of antibiotics in the treatment of acute diarrhea, willingness of mothers to administer fluids to their sick child) will provide the information needed to identify problems and institute the changes necessary to keep the program on target with regard to the expected decrease in measles incidence and diarrhea case fatality rates.

CLOSING THE LOOP: REASSESSING THE BURDEN OF ILLNESS

Reassessment of health needs and the burden of illness is essential to assess the overall success of interventions—the loop is "closed," returning again to the residual burden of illness. With regard to our two examples, this entails periodic surveys of the incidence of measles and diarrhea case fatality rates.

CAN THE LOOP BE USEFULLY APPLIED TO CLINICAL AND POLICY DECISIONS?

The usefulness of the loop approach in clinical and policy decisions can be illustrated in two ways. First, the loop can be applied to the available evidence on measles immunization and ORS to make recommendations for needed research. Measles immunization is one example where adequate information exists for each step in the loop to the extent reasonable and possible. Currently available evidence should provide a sound foundation in most settings for clinical and policy conclusions. All children are potentially susceptible to measles without vaccination. Efficacy has been studied in rigorous randomized trials that examined the impact of vaccination on the actual incidence of measles. Even though conducted in an industrialized setting, there is no evidence to suggest that the efficacy estimate of 85 percent is not appropriate for developing country settings. Estimates of the other components of community effectiveness, however, are certainly subject to variation between different settings, but it should be possible to approximate these or carry out quick surveys to obtain the required information. Improved methods for cost-effective, accurate monitoring of priority health problems continue to be called for in the literature [95]. The standardized cluster sampling technique developed by WHO's EPI has proved to be extremely useful and is applied regularly to assess immunization uptake [97]. Finally, new and potentially better vaccines, such as Edmonston-Zagreb, suggest even higher levels of efficacy may be possible [99].

ORS, on the other hand, is an intervention where, by applying the loop to assemble the available information, we can identify gaps in our knowledge and make recommendations for needed research. Although there is no question about the potential of ORS to reduce deaths due to the dehydration caused by diarrhea, significant gaps remain regarding optimal strategies for its delivery including (1) methods to enhance compliance by the mothers and coverage and (2) the role and appropriate use of home-based fluids; further information on the cost-effectiveness and cost-utility of alternative strategies is also needed.

Second, several examples exist where substantial problems could have been avoided through the use of the loop in formulating clinical and policy decisions. For example, the widespread implementation of nutrition programs in the developing world in the 1960s and 1970s could have benefitted from consideration of the issues summarized by the loop. Gwatkin, in his review, highlights the role of the various factors that determine community effectiveness and their relation to the poor performance of the programs undertaken [96].

Finally, through the application of the loop to summarize the evidence on measles immunization and ORS, the wide range in the available evidence becomes evident, as is shown in Table 27-10 and 27-11. With such wide ranges in the estimates, it is extremely difficult to establish rankings that allow the analyst or health care planner to determine the relative impact and costs of the various alternatives under consideration. For example, the top three interventions in terms of maximum health impact per dollar spent relevant to diseases of the under-5 population of LDCs.

An important, associated point concerns the validity of rankings based on efficacy estimates, as opposed to community effectiveness. The two examples used here illustrate the potential impact that the other factors that influence community effectiveness can have. It seems reasonable to conclude that high levels of efficacy may not necessarily translate into high levels of community effectiveness, and, accordingly, rankings based on efficacy estimates should be viewed with caution.

CONCLUSIONS

These guides will not and should not be expected to provide the answer to the question of which programs should be recommended for funding. It is fully appreciated that only rarely
(or never) will ideal information be available. Rather, they are intended to provide a set of commonsense guidelines that can be useful as a framework for identifying some of the important factors and assessing the quality of information used in decision making in health care.

On the other hand, the amount of evidence available to decision-makers can be substantial, as our two examples illustrate. Nevertheless, because there are almost always gaps in the evidence, it is important to systematically organize the available information and explicitly identify the gaps so that an informed decision may be made, whether it be a recommendation for funding or for further study. It is recognized that policymakers at all levels have to make decisions now on the basis of the best available information, but it is crucial to recognize a weak information base when it exists. All involved in the planning and provision of health care have a responsibility to identify the strengths and weaknesses of the evidence upon which decisions are made. Researchers have a responsibility to generate evidence that can be used by policymakers. Policymakers need to capitalize on the expertise and insights that health researchers can bring to the decision-making process and setting of priorities. Thus, we can iterate toward a higher standard of information on which decisions are based.

Finally, it cannot be overemphasized that the issues discussed above are not unique to LDC settings. However, given the extreme scarcity of resources for health, the severity of many of the problems of LDCs, and the explosion of expensive new medical technologies, the need to set priorities and to identify strategies that maximize the impact of the health resources that are available is of extreme importance to LDCs. What we have presented raises issues common to all involved in the planning and provision of health care worldwide. As has already begun to evolve, the challenges and opportunities offered by this common need set the stage for international partnership and a worldwide sharing of ideas and strategies for solution.

REFERENCES

22. Department of Clinical Epidemiology and Biostatistics: How to read clinical journals. V. To distinguish useful from useless or even harmful therapy, Can Med Assoc J 124:985–990, 1981
43 Minot GR, Murphy WP: Treatment of pernicious anemia by a special diet. JAMA 87:470, 1926
44 Goldsmith GA: Experimental niacin deficiency, J Am Dietetic Assoc 32:312, 1956
45 Bennett KJ, Tugwell P: Measles immunization—A critical review of the evidence for community effectiveness. Submitted for publication
47 Bennett KJ, Tugwell P: Oral rehydration—A critical review of the evidence for community effectiveness. Submitted for publication
49 Levine M: Case study A—Oral rehydration therapy for diarrheal diseases, in Status of Biomedical Research and Related Technology for Tropical Diseases. US Congress, Office of Technology Assessment, OTH-225, September, 1985
58 Veterans Administration Co-operative Study Group on Antihyper- tension Agents: Effects of treatment on morbidity in hypertension II—results in patients with diastolic blood pressure averaging 90 through 114 mmHg. JAMA 213:1143–1152, 1970
60 Himshaw HC, Feldman WH, Plouetze KH: Treatment of tuberculosis with streptomycin: Summary of observations on 100 cases. JAMA 132:778–782, 1946