AN APPROACH TO EVALUATING HEALTH OUTCOMES

Les Irwig, Associate Professor, Department of Public Health, University of Sydney

Health services often seem to expend enormous resources without a clear picture of the outcomes achieved. With rapid improvements in computer hardware and software, it is enticing to use health service data to generate information about the outcomes of health care. Evaluating health outcomes, which also encompasses measuring the process of care, has enormous potential for improving health. To achieve maximum gains, we need a clear framework to guide this work and its interpretation. In this article I propose a framework based on four major questions. By assessing which of the questions is being addressed and whether the methods are appropriate for that question, you may discover ways in which your own work or your interpretation of others’ findings could be improved.

**QUESTION 1: WHAT IS THE RIGHT THING TO DO? (EFFECTIVENESS)**

This is the most critical of the questions. It is concerned with deciding policy whereas the remaining questions are concerned with quality assurance of its implementation. Without reasonable certainty about the effectiveness of the policy, addressing later questions may do a disservice by encouraging uniform application of a policy which may be harmful. Which methods should be used to ensure Question 1 is answered adequately?

1. **Choose the right study design: randomised trials**

Health policies or practice guidelines should be based on randomised trials and not on observational studies or opinion. This is contrary to much of the health outcomes literature which suggests the efficacy of interventions can be ascertained from sophisticated analysis of outcome variability between different sources of care, often using routinely collected data. I do not think such data can be used to test interventions. The effects of therapy are often small and easily overshadowed by the selection biases inherent in observational studies. Selection biases are likely to be much greater in studies of interventions than in studies of environmental, lifestyle or occupational exposures. This occurs because both physicians and patients are deliberately choosing interventions on the basis of patient characteristics, not all of which will be explicit. Moreover, it will not be possible to adjust adequately for important predictors of outcome, such as case-mix, because they are imperfectly measured.

2. **Measure outcomes relevant to patients: quality of life and survival**

Quality of life has been ignored for too long because it is difficult to measure. However, rapid advances are being made in the methodology which should enable its wider application.

Apart from their quality of life, patients will be interested in their survival. We therefore should examine survival, or its proxy, all-cause mortality. Mortality from the disease at which the intervention is aimed may be reduced at the expense of an increase in mortality from other causes. For decision making, we need to shift away from measures of relative improvement, such as relative risk, to measures of absolute improvement, such as risk difference, or its inverse the number of people who need to be detected/treated to prevent one death or illness. For example, one needs to screen 23,000 women aged 35-44 for high blood cholesterol and treat 4,600 of them for five years to prevent one death. Moreover, this calculation is based on the assumption of a 20 per cent reduction in mortality from cholesterol intervention in all age-sex groups, which seems far too optimistic an estimate.

3. **Meta-analyse all the trials**

Meta-analysing all randomised trials will provide an estimate of the efficacy of an intervention with narrower confidence intervals than any of the individual trials and may give sufficient power to examine efficacy in different subgroups of patients. Continuously updating meta-analyses and making results readily accessible in computer format will help health practitioners make rational decisions and should form the basis for clinical practice guidelines.

This approach assumes the existence of randomised trials. I expect our hands will be full with addressing policy formulation and quality assurance for common conditions for which trials have been done. Doing randomised trials for important conditions where none has been done will be made easier if clinicians and patients become more accepting of the idea that there is uncertainty about the best treatment and more willing to enter trials. Trials should be made easier to incorporate into clinical practice by simplifying inclusion criteria and consent procedures, reducing the data requirements and changing funding arrangements.

**QUESTION 2: DID WE DO THE RIGHT THING? (APPROPRIATENESS OF CARE)**

Assessing appropriateness of care should be based on measuring the process of care (practice) against the ‘gold standard’ of the evidence-based policy generated in response to Question 1. It will have the greatest impact if the condition is common, actual practice varies considerably from the policy and clinical opinion leaders are keen to collaborate. Commonly, appropriateness is measured by examining what proportion of people on whom a procedure was done should have had it. This approach does not indicate how many patients got the procedure when they should have. Knowing how many people who needed the procedure did not get it is clearly as important as knowing how many got the procedure inappropriately. Methods for monitoring appropriateness of care should include both components and take steps to deal with error in the measurement of appropriateness.

**QUESTION 3: DID WE DO THE RIGHT THING RIGHT? (PERFORMANCE)**

The assessment of performance includes three main components: technical performance, patient satisfaction with the process of care, and efficiency. Technical performance often requires assessments which are more detailed than can be ascertained from systems set up to monitor appropriateness. Examples include the proportion of ‘lumpectomies’ in which the breast cancer has been removed with an adequate margin. Patient satisfaction includes several dimensions concerned with the process of care. Efficiency is concerned with whether care was carried out in the most streamlined way, at least cost and without delays.

**QUESTION 4: DID IT HAVE THE RIGHT RESULT? (OUTCOME)**

A flow diagram of the expected sequence of events of an intervention helps in deciding which, if any, outcomes to measure. Long-term outcomes such as five-year survival are often not worth measuring as it may not be possible to disentangle the effects of intervention from other sources of variability in outcome. Furthermore, distant events will not provide timely feedback for monitoring quality. Surrogates (indicators) of outcome should be used only if there is good evidence that they are an important intermediate step.
Evaluating health outcomes

between the intervention and long-term outcome. Surrogates may suggest mortality will be reduced when the opposite occurs. Outcomes or their surrogates need to be unambiguously interpretable as an effect of intervention. This may occur if the effects are large, immediate, or rarely occur in the absence of the intervention, for example post-operative morbidity and mortality. Measuring immediate adverse events such as post-operative mortality may be useful for weighing up long-term benefits, estimated from randomised trials, against the risks in your patients.

In instances where outcomes will not be unambiguously interpretable as an effect of the intervention, quality assurance should be based on measuring the appropriateness and performance of the intervention, and we will need to assume it will have the benefit demonstrated in randomised trials.