QAP INFORMATION SHEET

Number 1

EVALUATION FOR BEGINNERS

by Barbara Stocking

'Evaluation helps us learn about what we do so that we can provide better health care'

These information sheets aim to provide some basic information and reading about commonly requested issues in Quality Assurance. The views expressed by authors and their selection of items for reading are their own.

KING'S FUND CENTRE

Quality Assurance Programme

Series editor: Brian Glasser

A mystique has grown up around evaluation but in essence the concept is very simple: it means looking at the value of something, assessing how worthwhile it is. An action, event, development will have a different value to different people. An evaluation implies value judgments, so it is important to note who is making the evaluation.

In this general sense, all of us evaluate actions and events daily, but in people's work in the health service, evaluation requires more scientific rigour. It is important because we cannot assume that the treatment and care we provide necessarily improves health or is the most effective way of achieving some health care objective. Evaluation helps us learn about what we do so that we can provide better health care.

Scientific evaluation

Individuals or groups providing a service, or more commonly changing a service, may consider evaluation. What they want to know is: what are the effects of doing things this way and can we be sure that the changes we have made are the causes of these effects. To be really sure about cause and effect requires very strong evidence, and in its most pure form the randomised controlled trial is the nearest we get to such a 'gold standard'.

In such a trial you try to exclude the effects of all the other influences by randomly assigning people [or even organisations, though this is rarely done] between one treatment and another or a treatment and no treatment. All the other factors which might be causing the effect should then on average appear in both groups and the only difference between them is the particular intervention made. This 'gold standard' trial requires that people are assigned prospectively, i.e. in advance, though there are other less rigorous methods, such as case control studies which look at data retrospectively. Again, although less satisfactory, evaluation can be used to look at what happens before and after an intervention, but then other factors may be influencing a change, not just the intervention being evaluated.

Randomised controlled trials are commonly used to evaluate clinical practice but are very difficult to do in evaluating service delivery. To give a simple example: it may be that you would like to know whether giving particular literature to pregnant mothers affects their satisfaction with the service provided. You could randomly allocate mothers to groups getting or not getting the information. If you tried to do this within a hospital your results would be likely to be confused by the mothers passing the literature to each other, or just informing others that they had received such literature. If you took a number of hospitals at a considerable distance apart you would no longer be randomising the patients but in effect trying to randomise the hospitals. This is not likely to give clear-cut results either because it would be impossible to have enough hospitals in your sample to control all the other variables that might influence the outcome. For example, one hospital may already be working to a policy of giving similar information to mothers verbally, while another had never provided much information at all. Giving written information would be expected to have

different effects in the two hospitals. Thus if hospitals were assigned in the trial to be written information givers or not the results would be completely confused by the previous policies.

The randomised controlled trial, case control studies, etc. focus on inputs and outcomes. There are other complementary approaches to evaluation which focus more on the processes involved. These are called illuminative evaluation methods. Illuminative evaluation stresses the importance of qualitative [i.e. non-numeric] information, of information from different sources and people, and of information about all the processes related to the setting up of a given innovation, not just the outcome of the innovation itself. Reports can therefore be expected to contain a great deal of qualitative material, even stories and accounts of processes. This may look less 'scientific' to those who are used to controlled trials, or who are used to 'science' being about numbers. But it is equally scientific, provided the information is obtained in a systematic and unbiased way. It is widely recognised in the social sciences as often the most appropriate way of understanding interventions in the social world. It results in information which can be extremely useful for managers who may wish to understand how outcomes are achieved and enable them to change parts of their programme. It also provides useful insight for others wishing to duplicate a particular programme.

Because service delivery and organisations are so complex and include so many interacting variables, it is difficult to do rigorous scientific trials and therefore difficult to prove cause and effect. It will probably be impossible to prove scientifically that some change you made really did have a particular effect and that the effects of this change could be predicted to occur in other similar circumstances. Nevertheless, having explained just how difficult a scientific evaluation can be, that is no reason for giving up entirely. You need to be clear and explicit about the limitations of your evaluation, but even within limits evaluation can be very useful in helping you know whether what you are doing is any good or whether or not a change you are making is better. Evaluation doesn't have to be boring either, it can stimulate people to think a lot more about their work and find it more enjoyable.
So where do you start?

Carrying out an evaluation

The very first thing to be clear about is what you are evaluating; what it is this treatment, service, or innovation is meant to do. The purpose of evaluation is to investigate the gap between intention and what actually results in practice. These intentions may, of course, be complex— there may be a variety of different intentions and people may have different views, but if that is the case it is important to be clear about this from the start. Define clearly what you intend to evaluate — sometimes harder than it sounds!

Then you need to ask yourself a series of questions about your evaluation starting with the most important one about the objectives or purpose of your evaluation.

Why are you doing it? What differences are the results going to make? If you are not able or willing to change anything in the light of your results, what is the use of doing it? Be wary of undertaking evaluations just to try to persuade others. Service evaluation by its nature is often flawed and people, especially doctors, who know what scientific evaluation involves are likely to be able to pick holes in what you have done. Evaluation is a political football, in part because no evaluation is entirely 'pure', the methods you choose and the data you collect will be influenced by your own values and others may challenge what you have done. But a good evaluation will give you useful information about the services you are providing and encourage you and others to look at your work critically.
Who is the evaluation for? Who will be the audience for your results? Is it for your own department only, are you going to use it in discussions with particular professionals, the Health Authority, consumer groups, etc.? This should be borne in mind in designing your study and the questions to be addressed, and in considering any ethical implications in the analysis.

Remember that in the real world information, even good data from evaluation, will be only one of the factors taken into account in making a change or reaching a decision. You should be prepared to 'sell' your evaluation but recognising that people have many motivations for doing something. If you do want to influence events though it will be important that your data is as timely and informative as possible.
How are you going to do it? Weigh up methods in relation to your objectives and the nature of the project you are evaluating. For example, if the evaluation is of a major new form of service provision, it may be important to find the resources for a thorough evaluation carried out by external investigators. In other cases you may simply wish to have some before and after evidence relating to a change to check out your perceptions, and some fairly simple measures will suffice.

Evaluation costs time and money, so the type of evaluation you do needs to be in proportion to the importance of the innovation itself and the purposes for which you are carrying out the evaluation.
When are you going to do it? If you are evaluating an innovation, are you going to do a before and after study, or are you going to wait until the innovation has been implemented? Are there times to be avoided because they would be 'atypical'? Weigh up the pros and cons of different timings.
Where are you going to do it? If you are going to interview people or give them questionnaires, where is the best place to do this?

Finally, don't be afraid to seek advice. If it is the first time you have carried out an evaluation you are not likely to know the number of observations which make sense, or the size of the survey necessary. There are people in universities, polytechnics and in the health service itself, e.g. community physicians, who are trained in these areas and should be able to help you.

This section has set out how to start thinking about

an evaluation, it does not give enough information for you to carry out an evaluation, and for further help readers should refer to Reference 1.

The next sections give some factors which should be helpful in working through the questions set out here.

Spin-offs, side effects and the Hawthorne effect

It is one thing to know that your intervention did or did not succeed in meeting your objectives, but once you intervene in a system there may be all sorts of other impacts. You may want to find out about these effects too. For example, if you plan to change an outpatient booking system your objective might be to reduce the length of time patients have to wait. You could measure the lengths of wait before and after that change to see if your new approach has been effective. However, it could just be [even though this may seem unlikely] that despite shorter lengths of wait patients are more dissatisfied. Perhaps a group of people were used to waiting together and had enjoyed the companionship. You might want to find out then what the patients' reactions are and not just assume that if you meet your objectives they will be happier. If you had done a pilot study in your evaluation you should not be too surprised about this since a pilot will help you find out what aspects you might want to evaluate more thoroughly.

This example also illustrates the difference between subjective and objective evidence. Recording the length of waiting time is an objective measure, the fact of how long a patient waited can be recorded by a clock. What the patients feel about the length of time they have to wait is an entirely different matter, you can ask them and they will give you their subjective views — a different but equally important sort of evidence.

You also need to be aware of something called the Hawthorne effect. A researcher in the US was measuring productivity in a factory and looked at such environmental factors as lighting levels. If he turned them up he found the workers produced more compared to the control group, but then he turned them down only to find that this had a similar effect. In other words, it was the stimulation of the change itself which caused the effect, not any particular change. The Hawthorne effect can be very difficult to exclude in research studies. In the example above,a method would be to measure lengths of wait at quite long intervals after the change. Then you could be sure that it was not the interest in the change process which reduced lengths of wait but the way the work is being organised [which is what you hope is the reason for the change].

The Hawthorne effect isn't just a nuisance though, you can use it very positively: making a change and taking an interest in people's work by helping them evaluate it can be positive stimuli, motivating them to do better.

While service providers may want to know about the benefits or disadvantages of particular forms of service provision and organisation, they are also likely to want to know something about costs, and about cost benefit and cost-effectiveness. Cost benefit analysis requires costing the particular service and procedures and then costing [usually in monetary terms] all the benefits to see if the costs equal the benefits. Putting everything into monetary terms is often inappropriate in health care so cost-effectiveness is more likely to be used. Here, you need to know, for example, whether for two different approaches with the same effectiveness, one costs more or less than the other, or for the same costs different levels of effectiveness are achieved. This is difficult to do properly. It is often difficult to obtain costs of treatments or service provision.

Even if this is available different approaches may mean that the costs are borne by different people and often patients and carers are forgotten.

For example, if hospitals versus home care is being evaluated, the costs may be much less in the community if only direct service costs are considered. If the costs to patients and carers in the two circumstances are included then the reverse may be the case.

There are other pitfalls too. It is often assumed that if one method of service provision is more cost-effective than another it will reduce expenditure. This will not be the case if more patients are treated or if a higher quality of care results for the same costs. For example, shorter lengths of stay where the costs per patient treated are reduced will not save the hospital any money if those beds are then filled with more patients.

Cost-effectiveness is something of a buzz word, but for these various reasons it should be used with considerable caution. Again though evaluators may still need to get some measure of the costs of alternative approaches but should be careful not to imply more than has really been measured.

Quantitative vs. qualitative data

Quantitative data is the sort which can be put in numerical terms. One of the worries of researchers in health and social services is that the emphasis on numerical data means that what can be measured easily is emphasised and sometimes the more qualitative effects are ignored. You can measure reduced waiting times in outpatients but how do you find out whether the whole quality of the environment in outpatients has changed? The nearest you can get may be to obtain the subjective judgments of many different groups of people and see if they agree. It may not be possible to put the results in straight numerical terms but the qualitative data may be very important, for example giving you insight into many other issues which perhaps need resolution too. Your end product might then be a descriptive report rather than statistical data.

Qualitative data is very important for patients and clients. What they want out of a service is often not expressed in numerical terms but in descriptive, qualitative feelings.

For example, it may be that someone takes a drug which successfully lowers their blood pressure but if

asked they may say they actually feel worse, perhaps because of side effects, perhaps because of the fact that they now feel like a patient and worry all the time about what they should or should not do. An issue in evaluation is how to measure quality of life, as judged by patients and professionals, who may differ, but this is dealt with in a separate review.

Reducing bias

If you are involved in both the intervention and in measuring its effect then you are open to accusations of bias. Because you want the intervention to work you are more likely to evaluate it positively. Even with objective measures bias can creep in. For example, if you are recording lengths of wait at weekly intervals, did you happen to choose a day when the staff were most likely to make the change work? For the more subjective data which probably involves interviewing or questionnaires, the possibility of bias is stronger. The questions may be biased towards giving a particular answer, or the interviewer's tone and "body language" may give clear signals about what you would like to hear. The way round this, of course, is to use external evaluators who have no stake in what you are doing. However, most people do not have access to external researchers or the resources to fund such evaluations.

Also, it can often be difficult for an organisation to act on the results of an external evaluation if there is no "ownership" by the organisation. One way round this is by the use of "collaborative evaluation" methods, where researchers and practitioners work closely together to identify issues and decide how to evaluate and interpret the results.

What other approaches can you use to try to avoid bias? One method is to use the principle of triangulation as far as possible. A distinction is often made between method and person triangulation. For method triangulation the principle entails using more than one way of obtaining information, for example, get your information from direct observation, from interviews/questionnaires, and from reading documents. Person triangulation means building up a picture by obtaining your information from as many different people as possible, for example, looking at waiting times: it may be relevant to talk to patients, doctors, nurses, receptionists, and anyone else who may have a view to contribute.

If you are using interviews or questionnaires it is almost certainly better [unless you are experienced in survey and questionnaire design] to use existing well validated instruments rather than writing your own. If modifications are necessary to meet your particular circumstances check the changes you make with an experienced researcher.

Thirdly, can you find people not committed to your change to help in the evaluation? The CHC may be willing to survey patients for you, or a student or trainee from your own or another department may be interested in doing a small research study, or enlist your department of public health [community medicine].

Ethical issues

There are ethical considerations involved in undertaking research which evaluators should be aware of. If patients/clients are directly involved even through completion of questionnaires then the protocol should go through an Ethics Committee — or at minimum the Chairman should be consulted, and consideration be given to how informed consent can be obtained.

Whatever data is collected, the following questions should beasked:

- Is the information you collect confidential? Can you guarantee this?
- Do people [or organisations] expect to be kept anonymous?
- Where is the report going— who is going to see the information?

Issues like these should be cleared up from the outset.

Key References

1. A. Selwyn St Leger, Harold Schneider & Joanna P Walsworth- Bell, *Evaluating Service Effectiveness: A Handbook for Community Physicians and other Health Service Managers* North Western Regional Health Authority, 1987. (gives concepts, definitions and is a very practical guide to how to do service evaluation - the next thing to read).
2. J M Abramson, *.Survey methods in community medicine* Churchill Livingstone, Edinburgh, 2nd Edition, 1979. (A systematic introduction to epidemiological and evaluative studies)
3. W W Holland, *The Evaluation of Health Care* Oxford University Press, Oxford, 1983. (a classic text describing scientific evaluation)
4. Drummond M F, Stoddart G L, and Torrance G W, *Methods for the Economic Evaluation of Health Care Programmes* Oxford University Press, Oxford, 1987. (a good introduction to economic evaluation, cost effectiveness,etc.)
5. Ann Cartwright, *Health Surveys: In practice and in potential* King's Fund, London, 1983. (a thorough guide to health surveys)
6. Patton, MQ, *Practical Evaluation* Sage Publications, London, 1982. (describes collaborative evaluation methods)
7. Humphreys, S, and Blunden, R, *A collaborative evaluation of an individual plan system.* In: British Journal of Mental Subnormality, **33**, 19-30. (a practical example of collaborative evaluation)
8. Parlett, M, and Hamilton D, *Evaluation as Illumination: A New Approach to the Study of Innovatory Programs* In: Glass GV, [ed] Evaluation Studies Review Annual Volume 1 [Beverly Hills: Sage], 1976 (the classic reference on illuminative evaluation).