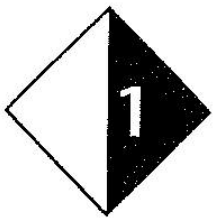

What is evaluation?

What are figures worth if they do no good to men's bodies or souls?

(Eyler 1979)

Practical life cannot proceed without evaluation, nor can intellectual life, nor can moral life, and they are not built on sand. The real question is how to do evaluation well, not how to avoid it.

(Scriven 1991)

Introduction

This book is about how to use or make an evaluation of a health treatment, service or policy. It aims to increase your ability to carry out a simple evaluation, and to understand and appraise reports of different types of evaluations. It has a practical focus: on 'evaluation for action'; on making and using evaluations to improve people's health and the health services in which we work. It is about a multidisciplinary approach to developing evidence-based healthcare practice, as well as evidence-based management and policy-making.

This book does not argue for or against different approaches to health evaluation. Rather, it shows the need to choose a design and methods which are most suited to the purpose of the evaluation and to the type of intervention which we are evaluating. It describes a variety of approaches within the four broad categories of experimental, economic, developmental and managerial evaluations. It shows the strengths and weaknesses of each, and the different criteria by which evaluations carried out within each perspective should be assessed.

> *We are in danger of replacing indifference about effectiveness with a dogmatic and narrow view of 'evidence'.*

This plural and practical approach is because our subject is evaluating health interventions of many different types. Even for an intervention of one type, such as a change to how teamwork is organized, many different evaluation designs and methods are possible. Both evaluators and users of evaluations need to know which type of evaluation could be used and the advantages and limitations of each. All need to understand the strengths and weakness of qualitative and quantitative methods and when and how to combine them.

Evaluation is not a boring subject, although some of the literature is not designed for those with attention deficit disorder: this is not a criticism of experimental evaluations, which are often well written for busy clinicians,

but a criticism of the unnecessary jargon and long-winded discussions which hide a lack of substance in some evaluation writings. This book is not without its slow moments, but I hope to convey to you some of the fascinating issues and human drama within this rapidly evolving field, and to develop your understanding in a relatively painless way, but without sacrificing necessary scientific rigour. We will encounter conflicts between different views about how to study people and how to enable change, about what counts as 'evidence', about the type of explanations which an evaluation should give and about whether evaluators should take more responsibility for ensuring that action follows from their work.

In modern healthcare it is unethical not to be concerned with evaluation, and no longer acceptable to be 'evaluation illiterate'

Health evaluation raises questions about our own and others' values and about what we consider important in life. It involves questions about how we change clinical and managerial practice and about how governments and organizations make health policies which affect us all. We will see how evaluations are used to advance and defend powerful interests, and how the evaluator has continually to choose between the ideal and the practical. We will see how simple evaluation methods are breaking the division between research and practice, and how more practitioners are applying the methods to improve the organization of care as well as the treatments which they give to patients.

As health professionals we monitor our patients' response to a treatment and have an ethical obligation to evaluate our own practice. As managers and policy-makers we need to ensure that resources are used to the best effect and that we can explain to the public why we have made certain choices on their behalf – a public of consumers and taxpayers who are increasingly using the findings from health evaluations. We can all use evaluation principles and methods to improve our practice and how we provide our service.

Who is the book for?

This book is for people who need to read and use evaluation reports and for those planning or making an evaluation. One group is health practitioners or researchers new to evaluation. They may be wanting to make better use of evaluations reported in their professional journals, or to evaluate their own or their colleagues' practice (self-evaluation). They may work for an organization which uses the treatment they are evaluating, or work for an organization which wants to evaluate its own service or a new policy. They may be 'external evaluators' planning an evaluation or already carrying one out. The main interest of those making an evaluation is in the phases of defining the item to be evaluated, deciding criteria of valuation and gathering and analysing data.

The book is also for managers, advisors and policy-makers who use or sponsor evaluations: they are evaluation 'users', as are practitioners and, increasingly, patients and patients' groups. Those using evaluations need to be able to understand an evaluation and critically appraise it in a short

period of time, and then decide whether the evaluation suggests that they should act differently. They may ask internal or external evaluators to undertake an evaluation. Their interest is mainly in the phases of defining the item and criteria, assessing proposals for an evaluation, judging value and planning and carrying out action.

The purpose of the book is to increase evaluators' ability to make evaluations which are useful, and to increase the ability of others to make better use of an evaluation. Evaluation is becoming a more common activity in health services. The reader is almost certain to be involved in an evaluation or using at least one type of evaluation. All health professionals, whatever the position in which they work, need to understand the fundamentals of evaluation. They will need even more in the future to be able to apply findings from evaluations, and to use evaluation methods to improve their own clinical or managerial practice and the surrounding organization of care.

How to use the book

Why are you looking at this book? What do you want from it? When I am new to a subject, I do not know enough about the subject to say what I need to know. For some this book will give you an introduction and help you to decide whether you need to know more or which skills you need, and guide you to other sources. The following gives suggestions for how different readers might make the best use of their time looking through this book. In Appendix 4 the book gives a number of learning exercises which readers can use individually or in a group to apply and work through the ideas that are covered. The first is a 'learning needs and objectives exercise' which you can use to clarify what you want from this book, or from a training course or conference on evaluation. The last section of this chapter gives a guide to all the chapters.

Health practitioners

Health practitioners – doctors, nurses, therapists and others – might find that starting with Chapter 3 is the best way to get a simple and rapid introduction. This chapter uses an evaluation of a treatment to give an example of different evaluation designs. To gain an overview of evaluation you could continue with this introductory chapter, which will also help you to decide which chapters you want to turn to next. Do spend a few minutes on the 'learning needs and objectives exercise' in Appendix 4, after you have read some of the introduction, so that you can clarify what you want from the book. If you want to become more skilled in critically assessing an evaluation report, and learn how to understand a complex study quickly, then test and improve your skills by analysing the case examples in Chapter 4.

Service managers

Health managers should go straight to Chapter 9 if they come to this book with a view to managing an evaluation. Chapter 9 follows a colleague who decided to buy in an evaluation of a service which she managed in order to

help her to decide whether to set up a similar service in another area. Then you could finish reading this introduction, and next turn to Chapter 8 on managerial evaluation, which will be more familiar territory, and will help you decide where to go next.

Policy makers and purchasers

'Policy makers' are politicians and their advisors, and managers with authority to make policies; for example, about allocating resources, changes to how health organizations work and measures to promote health. Purchasers and others working for organizations which pay for health services are increasingly using evaluation for a number of purposes. This chapter shows the range of approaches, and Chapter 14 considers how to make better use of evaluations of different types. You may not need to know some of the details of different designs and data gathering methods, but the discussion of economic evaluation (Chapter 6) and of quality and outcome evaluation (Chapter 13) will be of use and will help you to decide which other parts of the book to look at.

Teachers

If you are reading this because you already run courses or sessions on evaluation, then you will be able to decide without my suggestions which chapters might be of most use. Some of this book has already been used for international distance learning courses in health evaluation. Have a look at the ten learning exercises in Appendix 4: you can use or adapt some of these for individual or group work. Teachers new to the subject might like to think about which type of knowledge and skills their students most need, and the time their students have to gain these. How many need to be able to carry out an evaluation, and of which type? How many only need to be able quickly to understand and critically assess an evaluation report?

This first chapter gives some basic definitions which help students to get orientated. I have found the models of evaluation design in Chapter 3 a quick and effective way of building students' ability quickly to read, understand and critically assess most types of evaluation report, and Chapter 4 will test and deepen their ability to do this. (Appendix 2 gives a framework which students can use to appraise any evaluation critically, and Appendix 3 gives six 'empty' design formats for them to write notes summarizing an evaluation.) After this, there are some chapters which apply to all types of evaluation, such as the one on data-gathering methods, and some which are more relevant to specific subjects. I have found debates on 'what counts as evidence' some of the most interesting teaching events: the last section of Chapter 14 is a good basis for such stimualting debates.

Researchers

The main value of this book to researchers is to give a simple introduction to types of evaluation which they have not encountered, or to help those making the transition from pure research to decide how best to deal with the practical issues involved in making an evaluation. One question is what, if

any, is the difference between evaluation and research? Chapter 9, which describes the phases of an evaluation, will help you to answer this question and to understand how a sponsor and user are thinking when they ask you to carry out an evaluation.

Newcomers may find Chapter 2, on theory, history and perspectives in evaluation, a good starting point. Then you can judge whether you need to study the description of designs in Chapter 3 by reading a case example in Chapter 4 and testing how quickly you can analyse it. You will find the discussion of data gathering methods which you have already used a bit basic (Chapter 11), but Chapter 14 will challenge you to decide what your responsibilities for implementation are, and Chapter 10 will alert you to the practical issues and politics of evaluation.

More about the contents of each chapter is given at the end of this chapter. Each chapter finishes with a list of key points, which can help those who have not read the chapter to decide if they need to. Next, we look at some basic terms and which types of 'intervention' need to be evaluated in the health sector.

Evaluation: some basic terms

Evaluation is simple – it is something which we do all the time without thinking. We are always judging the value or importance of things, or of what we and others do. In everyday language, when we evaluate something we usually mean we judge its value. Sometimes we also mean that we look more closely before judging value: for example, we look at how we have spent our time, before then judging the value of how we spent our time. But many of us are very quick to make judgements.

More thoughtful evaluation is more complex. We are more careful to get the right information and to use it in the right way in order to reach our judgement of value. We are clearer about what is important to us: our criteria of valuation. We are also more careful about how we make a link between our judgement of value and how we could act. In systematic evaluation the usual approach is to separate the collection of information (the task of the 'evaluator') from judging value and acting (the task of the 'user' of the evaluation). Other differences from everyday evaluation are the evaluator's careful definition of what is to be evaluated, and of the information needed, as well as his or her careful selection and use of methods for collecting and analysing the information.

Systematic evaluation is also complex in that the evaluator has to know about a range of methods and approaches so as to be able to choose the best methods for the purpose of the evaluation. We do not use methods like participant observation and grounded theory to get information about the cost of home care in comparison to hospital care, but we may use these methods if the purpose is to find out how patients and carers judge or make use of these types of care. Ideally the evaluator has to be able to use a variety of methods, or at least to know which method is best for the purpose of the evaluation. Ideally those sponsoring and managing an evaluation should be able to make an independent judgement about whether other methods

Basic terms

Evaluator: the person making the evaluation.

Sponsors: those who initiate or pay for the evaluation.

Users: those who make use of or act on the evaluation.

Intervention: an action on, or attempt to change, a person, population or organization which is the subject of an evaluation.

Target: the part or whole of a person, population or organization which the intervention aims to affect.

Outcome: the consequences of the intervention; that which 'comes out' of it.

Target outcome: the change effected by the intervention on the target (the difference which the intervention makes to the target, whether intended or not).

would be more cost-effective and what can and cannot be expected from the evaluation.

In the pages to come we will continually return to the purpose of an evaluation and to the criteria for valuing something. Next come some basic terms, then a discussion of the term which we use to describe the subject of an evaluation ('an intervention'), and then we introduce the different types of intervention which are evaluated.

As a test of your understanding of basic terms, what is the 'target' of an evaluation of a primary care programme for expectant mothers? This is a trick question, because normally we talk of the 'subject' of an evaluation rather than the 'target' of an evaluation. The 'subject' of the evaluation is a primary care programme (which is the intervention) and the 'targets' of the primary care programme are expectant mothers. The point is not to confuse you, but to alert you to the fact that some reports use the word 'target' to refer to the target of the evaluation (e.g. Breakwell and Millward 1995). This book uses 'target' to describe the person or object which the intervention

Who makes evaluations?

External evaluators: researchers or consultancy units not directly managed by and independent of the sponsor and user of the evaluation.

Internal evaluators: evaluation or development units or researchers that are internal to the organization, and that evaluate treatments services or policies carried out by the organization or one of its divisions.

Self-evaluation: practitioners or teams who evaluate their own practice so as to improve it.

aims to change. It uses the term 'subject' to describe the intervention which is evaluated.

Note also that 'outcome' is the end result of the intervention and is a broad concept which encompasses a variety of consequences of the intervention, some of which are intended and some of which may not be. Many evaluations study the outcome for the targets of the intervention. In this case the outcome of the primary care programme is the effect of this intervention on expectant mothers (the targets) – whether they are changed in any way by the programme. Evaluations often measure some aspect of a patient or population before and after the intervention. Many things apart from the intervention may account for any before–after difference which the evaluation detects. 'Target outcome' here refers to the difference produced which can be attributed to the intervention. It does not refer to the before–after difference in general, just that which can be attributed to the intervention.

What do we evaluate?

The concept of 'intervention' (otherwise termed 'the evaluated')

Evaluations gather data for the purpose of valuing an intervention. An intervention is an action which results in a change: for example, a nurse gives a mother an information leaflet about breast feeding and the mother reads the leaflet. An intervention is something which someone does, which 'comes between' (*inter venire*) what would otherwise happen. The aim of an intervention is to produce a change, and to make a difference to people's lives.

Most evaluations examine an intervention which aims to alter the course of events so that people gain a health benefit from the intervention – people or populations are the targets of the intervention. There are also interventions to health organizations, such as a training programme, or a change to how people delegate work, or a change to how primary care personnel cooperate with a hospital and social services – in this case health personnel or organizations are the immediate targets of the intervention.

What is wrong with calling the things (or phenomena) which we need to evaluate 'interventions'? After all, most of what we do in the health sector aims to make a difference – to intervene in people's lives or in how we organize. This is true, but what about assessment or diagnostic methods which we need to evaluate: for example, a new type of brain scanning technique, or a care management assessment system? Are these 'interventions'? The purpose of many new diagnostic techniques is to help to assess a patient, but without making an intervention in the sense of directly causing a change in the patient. And what about an evaluation of whether patients from different social groups get equal access to and use of a service? Are all mothers given a leaflet about breast feeding, and one which is in a language which they can understand?

There are thus two problems with using 'intervention' to describe the different subjects of evaluations: first, not all interventions aim to change people; second, not all evaluations aim to find out if the intervention did change people. We may take for granted the value of breast feeding or the effectiveness of the programme, or not be interested in effectiveness, but be more concerned with how many mothers are reached by the programme. The purpose of some evaluations is to assess not the effect, but whether, for example, the service is provided to all equally. Using the term intervention implies that our main criterion of valuation is whether 'the evaluated' produces a change, when we may want to concentrate on other criteria of valuation such as costs or equity. We may know that the intervention has an effect but we may want to judge its value according to other criteria.

There is no real need to 'split hairs' at this stage about whether 'intervention' is the best general term to use to describe the subject of an evaluation – you will need all your energy for more important hair-splitting discussions later. The point is that by calling the subject of an evaluation an 'intervention' we imply that the thing has already made a change or had an effect. Yet it is precisely this which many evaluations try to find out. Strictly speaking, intervention means come between what would otherwise have happened and does not necessarily mean that a change was produced. To avoid the implication that a change was produced, the book sometimes uses the even less elegant term 'the evaluated' instead of 'intervention' to describe the subject of the evaluation. This term describes only the subject of the evaluation without implying any characteristics of the subject, such as whether it does intervene in someone's life and produces a change.

Defining the subject of the evaluation and assessing effects

There are two further points to be made before we look at different types of interventions. The first is that how we define and specify the thing or phenomenon to be evaluated is of the utmost importance if we are to make a useful evaluation. By define I mean 'draw a boundary around' (*de-finire*) which includes and excludes certain other things or phenomena. In the example of the nurse and mother, what was the intervention? Do we define it as the nurse giving the leaflet? Or was it this action and the mother reading the leaflet? What do we do if the mother does not read the leaflet? Do we say that there was no intervention, or that the intervention had no effect?

In the example it was implied that the intervention was the nurse giving and the mother reading the leaflet. A plan for an evaluation and a report must precisely describe the intervention, and also specify the key elements 'inside' the boundary of the intervention: for example, specify the details of the leaflet. Some evaluations aim only to get a good description of the evaluated – to specify possibly important elements and to define what should and should not be considered part of the evaluated. Thus there may be a difference between the intended intervention and what was actually done, and we need to know exactly what was done and evaluated to be able to interpret and apply the findings.

The second point concerns which changes we look for and when we look for them. Discovering that an intended intervention does or does not produce a change in the target makes it easy to judge its value, according to one common criterion of valuation. Note, however, that there may be other types of change which may be of value – for example, the nurse may feel that she or he is doing something to encourage breast feeding and this may raise morale. Further, there is such a thing as an unintended intervention, an unplanned event or action which produces a change: evaluations looking back into the past can often discover that something a practitioner or service did had an important effect, even though it was not intended.

In summary, many evaluations try to describe or measure changes which the intervention produces in the target. They aim to collect data about how big the valued changes are, and whether there are any changes which are not valued or which are harmful. These data help others to judge the value of the intervention.

Evaluations differ in the scope of the possible changes which they investigate: they may look at one type of effect, or at many effects and at short- or long-term effects. In the example, a limited investigation is of any change in the mother's attitude. A broader one would also look for changes in her behaviour: did she breast feed, or persuade others to? Broader still would be to include studying any change to the mother's or baby's health. Evaluations also differ in the time period over which possible change is investigated. The scope and time period depends on the evaluation purpose and questions or hypothesis to be tested.

The main points are that the subject of an evaluation is an intervention, and evaluations will do one or more of the following: describe the intended and actual intervention, describe or measure its consequences, provide information for judging the value of the intervention and explain why what was described did occur.

Evaluation defined

Evaluation is making a comparative assessment of the value of the evaluated or intervention, using systematically collected and analysed data, in order to decide how to act.

A slightly longer definition highlights the important features of evaluation:

Evaluation is attributing value to an intervention by gathering reliable and valid information about it in a systematic way, and by making comparisons, for the purposes of making more informed decisions or understanding causal mechanisms or general principles.

Let us unpack this definition and look at each feature.

. . . attributing value to an intervention . . .

The intervention is the thing or process of which we judge the value. The 'evaluated' that we will look at are health treatments, services, policies and interventions into, or changes to, health organizations. These subjects are very different in their nature and complexity. Some are easier to define than

others. Defining or specifying what we are evaluating is important because we want to be sure exactly what it is that we are judging the value of. Even treatments can be difficult to define exactly, as we will see when we look at 'alternative' treatments such as aromatherapy and homeopathy.

Exact definition often gets more difficult as we move from biophysical systems to social systems such as services, and then to 'higher level' social systems when we are looking at a health policy such as a ban on smoking which is implemented across health and other sectors. We will see that all evaluators spend some time defining the evaluated before they start, and that the only purpose of some evaluations is to get a clearer description of an intervention.

If we are able to define the evaluated, how do we then judge its value, or 'attribute' value to it? The first point is that it is the 'users' of the evaluation who judge the value of the evaluated; that is, politicians, managers, citizens, health practitioners and others. Strictly speaking, evaluators do not judge value. They simply collect, analyse and present information. The information which they decide to collect and the way they present it is designed to concentrate on certain things which some, or all, users think are important for them to be able to judge value.

Why does the definition of evaluation include 'attributing value'? Surely this suggests that the evaluators make a judgement of value rather than just the users? We will return to the questions of the evaluator's values and whether or not the evaluator judges value in Chapters 9 and 14 – note at this point that there are different views on these questions. The definition includes 'attributing value' because it aims to encompass the whole process of evaluation and to draw attention to the part which users play in evaluations, rather than keeping this separate. I think that there is something lacking in a definition and understanding of evaluation which does not include valuation. This definition recognizes the process of valuation and this book shows how values enter into most parts of the evaluation process.

. . . gathering reliable and valid information about it in a systematic way . . .

This second part of the definition shows one way in which an evaluation helps users to attribute value: the evaluation presents to them information about the evaluated which is reliable and valid and has been gathered by evaluators in a systematic way for the purposes of judging value. This distinguishes evaluation from journalism, where the journalist does not use the same rigorous methods. What then is the difference between evaluation and other types of research, which also give valid and reliable information about the evaluated?

. . . by making comparisons . . .

Comparison is the main way in which evaluation helps users to attribute value, and, together with valuation, is what distinguishes evaluation from some other types of research. Different types of evaluation carry out different types of comparison. One type we are familiar with is where the evaluation compares one group which gets an intervention with another group which does not, to see if an intervention has any effect on the first group (the experimental controlled trial). This is a comparison of the measured

state of two groups, and evaluations of this type will use different ways to try to be sure that any difference in the measured state can be attributed reliably to the intervention. Other types of comparison are to:

- compare the state of one or more people, populations or organizations before an intervention to their state after the intervention;
- compare the needs of people or populations to their needs after an intervention;
- compare the objectives of the intervention to the actual achievements;
- compare what is done to a set of standards or guidelines (e.g. audit).

We can only judge value if we make a comparison. Many evaluations use 'value criteria' to help to make the comparison, or only one criterion, such as symptom relief, cost or the number of eligible people using a service. In the two-group evaluation model the criterion is the measure used to measure the state of the two groups. In 'before and after' evaluations the criterion is the before and after measure. In the other two types above the criterion is the objectives, standards or guidelines.

Where do criteria come from? This question is important because if we can only judge value by making a comparison, and if the criterion is the way we compare, then the type of criterion we choose will shape the data we collect and how we judge the value of the evaluated. The criteria come from one or more users of the evaluation, or are predefined (e.g. established standards), and users agree to adopt them to help them to judge the value of the evaluated. Recognize, though, that people judge the value of the evaluated using many criteria, most of which are below awareness or half-formed criteria, and are different for different people and interest groups. Evaluators may help people to clarify the criteria of valuation which they want to use. This point, that an evaluation makes a limited comparison using limited criteria, takes us to the next part of the definition:

. . . for the purposes of making more informed decisions . . .

Examples of decisions are if or when to use a treatment, how to improve a service or whether to reduce or discontinue it. Decisions to be made may be about how to set up other similar services or whether or not to set them up, or whether to abandon, extend or modify a policy.

We need to add to the phrase *'more informed decisions'* the phrase *'than they would otherwise do'*. People do not need an evaluation to judge the value of something and to decide how to act. Very few treatments, services or other interventions have been or ever will be evaluated. Generally, mangers, practitioners and citizens go about their business as if evaluation and reports do not exist. The purpose of evaluation is to add to the information which people have and to improve the way that they judge value, so that they can make more informed decisions than they would otherwise do.

The way evaluation does this is to make clear the criteria used in the evaluation for judging value, so that people can decide whether or not to accept these criteria or how much weight to give them in making their judgement of value. It also helps decisions by giving information about the intervention and its performance in relation to these criteria. But people will decide how to act by using criteria and values not included in the

evaluation, and also by making judgements about the feasibility and consequences of different actions.

. . . or understanding causal mechanisms or general principles.

This last part of the definition draws attention to the fact that the primary purpose of some evaluations is only scientific knowledge, and here evaluation clearly overlaps with scientific research. Medical researchers and other scientists use certain evaluation designs because they are useful for investigating cause–effect mechanisms in different phenomena and for explaining and predicting. The 'general principles' refers to the aim of most evaluations being to produce findings which can be generalized beyond the specific intervention studied – this makes evaluation more than consultancy or development work.

Other definitions

The definition above is just one of the many definitions of evaluation. More are listed at the end of Appendix 1. For example:

> Program evaluations aim to provide convincing evidence that a program is effective. The standards are the specific criteria by which effectiveness is measured.
>
> (Fink 1993)

This defines evaluation in terms of finding evidence of effectiveness. It is true that many evaluations examine effectiveness, but there are other ways to judge value. We also tend to find what we are looking for. Should we not have a more sceptical approach than to aim to provide convincing evidence that a programme is effective? Should not the evaluator instead assume that the programme has no effects?

Another definition of evaluation is in terms of assessing whether the evaluated achieves its goals:

> The critical assessment, on as objective a basis as possible, of the degree to which entire services or their component parts (e.g. diagnostic tests, treatments, caring procedures) fulfil stated goals.
>
> (St Leger *et al.* 1992)

Achieving goals is also a criterion we could use to judge the evaluated, but whose goals? Those of management? Suppose meeting needs is not a goal? Both achieving goals and effectiveness are criteria we could use to judge the evaluated, but there are other possible criteria. The broader definition given at the start of this section allows us to include other types of descriptive or developmental evaluations and 'goal-free' evaluation (Scriven 1973), but also distinguishes evaluation from other activities.

What is the difference between evaluation and 'fundamental research'?

How does the earlier definition distinguish evaluation from other activities? Evaluation is a form of research in the sense that it is a systematic

investigation of the evaluated and aims to discover new knowledge. Evaluators use surveys, interviews, measuring instruments and other data gathering methods which other researchers also use. Some evaluations also use experimental design – for example, a controlled trial of a surgical technique – and these evaluations are also called scientific or medical research. The definition above gives room for traditional research within the category of evaluation by defining one purpose of evaluation to be discovering causal mechanisms, where such types of conceptualization are appropriate.

In my view it is not data gathering or study design which is distinctive of evaluation, but gathering data for the purposes of judging value, the element of comparison and the practical focus. That is, the methods are used within an overall 'process of evaluation' which clarifies the criteria which are to be used to judge value (for example, a criterion of effectiveness) and then gathers information about the evaluated against this one criterion or more criteria. Thus, it is how the methods are chosen and combined in a process called 'evaluation' that has a practical purpose which distinguishes evaluation from other types of pure research, investigative journalism and other activities. This process is termed here 'evaluation for action' and described in the following pages.

This book emphasizes the practical purpose of evaluation and makes a distinction between pure or fundamental research (the aim of which is only scientific knowledge) and evaluation (the aim of which is to help people to make better informed practical decisions). Some scientific research does not involve comparison, or does not enable people to judge the value of something or to make better informed decisions. There is certainly an overlap between evaluation and research but while most evaluation could probably be termed research of some type, only some types of research are evaluation. Because of the practical emphasis of the book, it makes a greater distinction between evaluation and some types of research than that made by other books and writings, and some do not make this distinction at all.

What is the difference between evaluation and clinical audit or quality assurance? In my view none, and later chapters consider clinical and other types of audit as one type of evaluation. Quality assurance is a catch-all term, and most quality assurance is also a type of evaluation. There are also different uses of these terms in different countries (Øvretveit 1997a). Some writers distinguish audit from research. For example, Black (1992) points out that audit uses 'evaluative research' as a basis for defining what is good quality care, and also that the data gathered in some audits are useful for some research, giving as an example the UK Intensive Care Society's database from 28 intensive care units.

I think it is useful to distinguish between audit and some types of research because there has been a tendency to redefine some research as audit in order to secure finance, and because some practitioners doing audit try to use research and outcome methods which are too sophisticated and expensive and not necessary for answering the practical questions with which they are concerned. As Black (1992) puts it, 'there is a danger of audit masquerading as research but without the necessary scientific rigour, or research pretending to be audit, but without any attempts to improve the quality of care being studied.'

My own view is that some types of audit are research and are also evaluations, but most audit and quality assurance does not and should not meet the same rigorous scientific criteria which are used in basic or pure research. The purpose is different. Audit is one type of evaluation which asks 'Are we (or they) doing things right?' Experimental evaluation (the activity which Black and some others refers to as 'evaluative research') asks, 'What is the right thing to do?' when there is no established knowledge to guide us.

Definitions: different activities for different purposes

Basic or pure research: using scientific methods which are appropriate for discovering valid and generalizable knowledge of a phenomenon for the purpose of contributing to scientific knowledge about the subject.

Audit: an investigation into whether an activity meets explicit standards, as defined by an auditing document, for the purpose of checking and improving the activity audited. The auditing process can be carried by external auditors, or internally for self-review, and the knowledge produced is specific to the service and cannot be generalized. The standards can be external and already made, or can be developed by the service providers for self-audit – in clinical audit, ideally by using scientific research.

Monitoring: continuous supervision of an activity for the purpose of checking whether plans and procedures are being followed (audit is a sub-type of the wider activity of monitoring).

Review: a single or regular assessment of an activity, which may or may not compare the activity to an explicit plan, criteria or standards. (Most audits or monitoring are also types of review. Many 'managerial evaluations' are reviews or monitoring.)

Evaluation: a comparative judgement of the value of an intervention in relation to criteria, for the purpose of making better informed decisions about how to act.

Action research: a systematic investigation which aims to contribute to knowledge as well as solve a practical problem. (Some action research is a type of evaluation. Much 'developmental evaluation' is action research.)

Summary: What evaluation can and cannot do

- Evaluation cannot itself carry out changes, but it can give a more informed basis for others to carry out changes.
- An evaluation study cannot attribute value to the evaluated, but it can give information which helps others to attribute value. In selecting which information to collect the evaluator decides which information is relevant to attributing value, and excludes other possible information.

While the evaluation does not itself attribute value, the way in which it is done shapes how others attribute value.

- Evaluation makes clear certain criteria which others can use to judge the value of the evaluated.
- An evaluation study cannot include all the criteria which people will use to judge the value of the evaluated, or all the things which people need to consider in deciding how to act.
- An evaluation does not have to be an expensive three-year randomized controlled trial – it can simply be a description of something such as a new policy and how it is implemented.
- Not all health evaluations look at effectiveness – there are other criteria of valuation – but most do look for the effects of the evaluated.

Evaluation for action

As a result of an evaluation, someone should be better able to act or make a decision – an evaluation user should be more informed.

As well as giving a general overview of evaluation, this book describes an approach which emphasizes the practical aims of evaluation, summarized in the phrase 'evaluation for action'[1] (elsewhere as 'action evaluation'; Øvretveit 1987a). Evaluation for action is an objective and systematic approach for making a comparative judgement of the value of an item against criteria, in order to decide how to act. The 'item' may be a defined treatment, a service or a policy (termed the 'evaluated' or the 'intervention'). Evaluation for action:

- recognizes the different perspectives of different groups ('stakeholders') who have an interest in the results of the evaluation and in the item being evaluated (it is politically aware);
- uses criteria agreed by one or more stakeholder groups (the primary users of the evaluation) to decide which data to gather and how to judge the value of the item (criterion-based);
- considers, at each step of the evaluation, the practical actions which the findings imply and which others might take if they were to act on the findings (an action orientation is an integral part of the approach);
- uses the most relevant theory and methods from different disciplines for the purposes of the evaluation, to help to decide the evaluation criteria, to gather and analyse data, and to clarify the implications for judging value and for carrying out action (multidisciplinary);
- covers the phases of defining the item to be evaluated, clarifying evaluation criteria, gathering and analysing data, and judging value and planning and carrying out action.

Does this mean that the purpose of evaluation is change? Certainly change is the aim of evaluation – the aim is to enable practitioners,

1 I am indebted to Steve Harrison for suggesting that evaluation for action was a better term than action evaluation, which could be confused with action research.

managers and others to do things differently and better as a result of the evaluation. The aim of evaluation, like the aim of a health intervention, is to make a difference, even if the difference is only that people continue to do what they did before, but with more confidence that they are doing the right thing or doing things right. If a change is made at all – and we will see in the last chapter that it takes more than a conclusive evaluation to produce a change – then usually people make a change after the evaluation is carried out. But it is also true that some approaches to evaluation – many 'developmental evaluations' – use methods such as action research methods to change the people or organization being evaluated while doing the evaluation.

How can you evaluate something while you are changing it at the same time? Here we touch on one of the differences between different evaluation perspectives: usually evaluations using an experimental perspective do everything which they can to ensure that the intervention does not change during the evaluation. Developmental evaluations often feed back findings to people in a service during the evaluation so that they can make immediate changes.

The main point here is that the purpose of evaluation is practical action. 'Evaluation for action' is thus a broad umbrella term for a variety of different approaches which can be used for the purpose of practical improvement. It extends beyond the phases of an evaluation study which are usually described in research-oriented evaluation texts because evaluation for action pays attention to how to assist the practical actions which could follow from the data gathering.

What can you evaluate?

One reason for using the broad definition of evaluation discussed above is that this book is for health personnel and researchers. Most people working in the health field now need to be able to use or make evaluations of four categories of things or phenomena: treatments, services, policies and changes to health organizations. In this section we note differences in the nature of these items which, as we will see in later chapters, have implications for how we evaluate them and collect data about them. Evaluation theory and techniques for one type of intervention can draw on and contribute to evaluation theory and techniques for evaluations of interventions in other categories. This is another reason why this book considers all four categories – because there is scope for more cross-fertilization between sub-fields of evaluation in health.

Health treatments, programmes or services and policies are all 'interventions' which are used to intervene in people's lives in order to improve their health. They are all different in nature, and also have different 'targets'. There are three different 'levels' of evaluation in the sense that the target of a treatment (a part or whole of a person) is a different level from the target of a service or programme (a population), which in turn is a different level of target from that of a health policy (a large population). The breadth and nature of each of these targets are different. The targets can all be viewed as

'systems', but the 'level' of a biochemical system is different from the level of a single human being as a system, which in turn is different from a population or a service as a system.

A fourth category of phenomena which are evaluated in health is changes to how health services and health personnel work. These changes include training programmes, a new system for devolving responsibility or a personnel appraisal system. These changes are also interventions but they do not have patient or population health as their primary aim – the aim may be to save money.

Table 1.1 helps to clarify which type of design and method is best for an evaluation, and it also introduces the way in which certain terms are to be used in this book. Note that some interventions directly improve health (D) and some may aim to improve health, but act indirectly (I) – better health is the ultimate aim. They will both be termed 'health interventions' (strictly speaking, 'intended' ones). There are also interventions which do not aim to improve health directly or even indirectly, but to increase efficiency or save money (O for operational interventions).

Treatments

The category of 'treatments' includes different therapies, such as drugs, surgical, physical, psychological and social therapies. Therapies differ in their nature: the active agent in most drug therapies is a chemical agent; in

Table 1.1 Different direct and indirect health interventions

Focus or target of the intervention	*Examples of interventions*
An individual: one patient or person	*Treatment:* hernia surgery (D) *Care:* voluntary worker bringing shopping for an older person just out hospital (D) *Assessment or test:* an X-ray or care management needs-assessment (I) *Health promotion:* advice about reducing risk factors for heart disease
A population: a group of patients	*Service:* a surgical service (D); a care management service (D) *Programme:* a health promotion programme to encourage 'healthy living' (D) *Project:* increasing community participation in the management of a health service (I)
A large population	*Policy:* people over 65 will be offered influenza vaccination (D); smokers will not be given certain types of heart surgery.
A system of care (the way a system is organized, and elements within it, such as health practitioners)	*Reorganization:* all units will have one general manager, who will report to the director; a fax communication system will be installed between the hospital discharge unit and all primary care centres (I) *Payment system:* performance-related pay (O) *Policy:* there will be no overtime working for the next six months (O) *Training:* training in how to treat and prevent bed sores (I)

physical therapy it is physical manipulation, usually by a physical therapist. Strictly speaking we should say 'the active agent is thought to be', because many therapy evaluations aim to find out exactly what is the active agent, as well as to test the hypothesized effectiveness of what is defined as the treatment. In the past most evaluations have been carried out on medical treatments which aim to cure an underlying pathology, and by using a controlled trial design. The purpose of some treatment evaluations is to investigate the nature of the treatment: for example, to describe a new type of treatment such as a plant treatment used by people in Lapland for treating frostbite.

However, all health treatments have the same type of target – a person, or part of a person. The definition of a treatment does not usually include the organizational context within which the treatment is used. For example, we can evaluate a new drug for hypertension or a new dressing for leg ulcers, regardless of the setting. To some extent the boundary between what is a treatment and what is a service is arbitrary. Very few 'treatments' are actually single interventions like a drug treatment; indeed, many drug treatments are not single interventions because the efficacy and effectiveness of many drugs depend on them being combined with other interventions, even if it is only information to the patient about when to take the drug. Cancer treatments are one example of treatments which are usually multiple, and which, at a certain point, merge into what some would call a service. There are different views about whether some disease-prevention interventions or health education should be considered a treatment, a service, a 'programme' or a policy. As we will see in later chapters, the way in which we conceptualize and define the intervention has implications for how we evaluate it.

Services and programmes

When we evaluate a service or programme the item we are evaluating is larger in scope and more complex than a treatment. The effects and costs of some treatments may depend on how and where they are applied, in which case we may define the thing to be evaluated as the treatment plus the organizational context. A service is one or more treatments as well as the way the treatment is given to the patient – it includes the organization and environment of care.

Is a primary health care centre a programme, a service or both? The word 'service' is sometimes used to mean the same thing as 'programme', but a service usually means an organization which provides a range of programmes. In the USA programme is more common a term than service, and 'programme evaluation' has grown into a large industry, mostly paid by government to make independent evaluations of public welfare programmes. Generally speaking, a service is an ongoing organization or institution, whereas a programme is time-limited or renewable with specific objectives (like a project). But further to confuse things, a service can sometimes also be a policy: for example, 'case or care management' can be a service for coordinating care or it can be a policy of an organization which marks out a general intention of that organization.

'Health care project' is sometimes used to mean a time-limited international health care programme: for example, a three-year overseas aid funded child health project in a developing country. This book uses the terms 'service' and 'programme' interchangeably, with one exception: when referring to an institution which provides education it uses the term 'educational service', but the term 'educational programme' is used to describe a specific type of education and training for health personnel or patients. Evaluations of services may involve an assessment of outcomes, of processes, of inputs or of the needs of patients, populations or health personnel, and often of more than one of these dimensions of a service or of the targets of a service (shown in Figure 2.1).

Policies, reforms and interventions to organization

Policies are directives or rules which aim to change or regulate how people behave. In this booklet we distinguish two types. The first type is health policies which have people's health as their primary and direct target: for example, health policies to increase immunization, reduce smoking or prevent infections.

The second type is organizational policies, such as non-discrimination, a policy to reduce waiting times, decentralization or a policy to reduce personnel overtime. These are interventions whose immediate purpose is to change how a health organization operates. Many organizational policies have improving health as their ultimate purpose, and it is sometimes difficult to distinguish a 'health policy' from an 'organizational policy'. In such cases the evaluation distinguishes between the immediate effects (change to organizational functioning) and longer-term effects on people's health. We use different methods to evaluate different types of organizational policies: the policies may be local or specific to one organization, or may be regional or national and applied to many organizations. Local managers often have to interpret national policies and they interpret them differently in different settings. The fact that the policy is being evaluated will influence how managers and practitioners implement the policy.

Sometimes large-scale policies are called reforms, typically when a government reforms how services are administered, such as introducing patient charges or new ways for patients to change their doctor. A reform is 'a significant set of changes to the method of financing, organisation, or running of health services, or to patients' rights' (Øvretveit 1996c). Examples are the transfer of responsibility for care for the elderly from counties to communes in Sweden, the decentralization programmes in a number of African countries or the British NHS 'market reforms' of 1991. Evaluation has an important part to play in planning reforms: for example, by local 'pilot' testing, by comparative analysis of similar reforms elsewhere, by theoretical policy analysis or by simulation modelling. However, to date most health reforms have been driven by ideology rather than by evidence of effectiveness or other findings from evaluations.

There are also 'interventions to organization'. Examples are a training programme for health personnel, a quality assurance system or a new

weekly meeting to decide collectively how to allocate patients who are referred to a service. Each of these interventions could be evaluated for the immediate effect on health personnel or for the effect on patients. These interventions to organization are sometimes called 'new policies' or 'health reforms': for example, the law requiring a quality assurance system in Norwegian hospitals is one type of intervention to organization, which could also be termed a policy or a reform.

Health promotion

We need more evaluations of health promotion: how do we best evaluate health promotion?

This is an assertion and a question which is increasingly being made, but what type of intervention are we referring to? Is health promotion a treatment, a service, a policy or an intervention to organization? An intervention to promote health can be any of these. Treatments to help someone stop smoking, such as counselling, hypnosis or a nicotine patch, are all health promoting interventions. A service which gives education in schools about the dangers of smoking is another type of health promotion, as is a policy to ban smoking in public buildings, as is a training programme to train general practitioners about the best way to help their patients stop smoking. We will see later how important it is to define precisely the intervention which we are evaluating.

Outline of the chapters

What are evaluations for? And what is the difference between experimental, economic, developmental and managerial evaluations? These are the two main questions addressed by Chapter 2. We look at the reasons for making or using an evaluation and why evaluation is becoming an increasingly important activity in health services. The chapter introduces the four main perspectives in evaluation, and discusses what we mean by perspective and how the perspective taken by the evaluator influences what the evaluation looks at, and how. It also gives a short history of evaluation and considers what we mean by 'evaluation theory' 'evidence' and explanation.

How would you evaluate a radical new treatment for sleeplessness? This question was put to me by a friend recently who was convinced that she had discovered such a treatment. It worked on her children too. Chapter 3 puts this question to readers, and takes them through six evaluation designs which they could use. The example introduces the newcomer to evaluation to some of the principles of evaluation and to some of the main concepts and terms, which are also listed in the glossary in Appendix 1.

How quickly can you make sense of an evaluation report? Test yourself and develop your skills in critical analysis by reading the summaries of evaluation studies in Chapter 4, and looking at the diagrams there which sum up the key features of the evaluations. You will learn how to draw a diagram of an evaluation, which is a powerful and quick way of getting to the

heart of an evaluation report. It is also very helpful for clarifying different types of design you might use in an evaluation which you are planning.

Chapters 5, 6, 7 and 8 describe different approaches to evaluation: the experimental, the economic, the developmental and the managerial. Then Susan makes an appearance in Chapter 9: she manages a service and we follow her experience as she contracts and manages an evaluation. The point of this chapter is to describe some of the common issues and considerations in the eight phases of planning, designing and carrying out an evaluation. By taking a manager's perspective, who is also the financial sponsor, we illustrate for researchers how their proposal and work could be assessed and used.

What should you look out for when you are carrying out an evaluation, and, if you are a manager, how can you best sabotage an evaluation which might make you or your service look bad? Chapter 10 discusses some of the trade secrets, and aims to help the reader to become more 'street wise' by understanding more of the politics of evaluation. It discusses roles, responsibilities and practical issues in carrying out an evaluation, such as confidentiality, access, reporting and communication.

Chapter 11 is more sober and scientific and considers a key part of any evaluation: data gathering and analysis. It summarizes different methods, such as observation, interviewing, questionnaires, measurement and existing data sources. Chapter 12 looks in more detail at data gathering theory and concepts – it considers what we mean by 'valid evidence' for an evaluation and the different types of data which an evaluator may need to gather. It also considers how to ensure the methods are valid, reliable and sensitive, as well as questions of sampling. It gives references to further literature and guidance which are of particular use to evaluators who may not be familiar with the many methods which can be used in health evaluations. Chapter 13 turns to a type of evaluation which is increasingly common and has special significance in health services – evaluating quality. It describes concepts and methods for quality evaluations, gives examples and shows the numerous ways in which evaluation methods contribute to quality improvement.

Is it evaluators' fault that their work is often not used? Chapter 14 returns fully to the main theme of evaluation for action and looks at the weak link in the chain from initiation of an evaluation to action: the link between the evaluation findings and implementation. It considers how to increase the use value of evaluations, how to maximize utilization and the shared and distinct responsibilities of evaluators and users.

If you have not done so already, look at the learning exercises in Appendix 4 and try at least one: this book is based on and tries to encourage 'learning by doing'. While you are there, look at the other appendices, which are intended as a resource that you can use at any time.

Conclusions

- Health personnel need an understanding of how to use and carry out evaluations of treatments, services, policies and interventions to

organization. Using and carrying out evaluations is part of everyday work in the health sector and we all now need to be 'evaluation literate'.

- Evaluation is attributing value to an intervention by gathering reliable and valid information about it in a systematic way, and by making comparisons. The purpose of an evaluation is to help users to make more informed decisions, or to understand causal mechanisms or general principles.
- Evaluations gather data for the purpose of valuing an intervention. The data to be gathered depend on the purpose of the evaluation (e.g. to describe, to explain, to judge any effects), the criteria of valuation (e.g. effectiveness, equity, cost, autonomy), the nature of the intervention (e.g. treatment, service, policy or intervention to an organization) and the perspective taken by the evaluator (e.g. experimental, economic, developmental, managerial).
- 'Evaluation for action' describes a process for defining valuation criteria and for collecting information about the evaluated for the purpose of helping users to make more informed decisions. It aims to include in the evaluation the criteria of all the interest groups which are important for implementing changes. This process includes, but extends beyond, what is often reported in a scientific evaluation study.
- Evaluation is different from consultancy, management review and investigative journalism in having a greater emphasis on scientific rigour, in using methods which are used in medical and social scientific research, in aiming to produce findings which can be generalized and in contributing to and drawing on published theory and research. In these respects evaluation is like scientific research.
- Evaluation is different from pure research, although some types of research, audit, monitoring and quality assurance are also types of evaluation.
- Evaluation studies differ according to the following.
  - The *subject* of the evaluation: a treatment, service, policy or organizational intervention.
  - The *target* of the intervention: a part or whole of a person, a population, service personnel or organization.
  - The *purpose*: to decide whether it works, how and why it works, whether it is worth the money, how to make it better, how well it performs (Chapters 2 and 3).
  - The *user* of the evaluation: managers, clinicians, patients, policy makers, the public as payer, other scientists/researchers (Chapter 2).
  - The *evaluation perspective*: experimental, economic, developmental or managerial (Chapters 2, 5, 6, 7 and 8).
  - The *design*: descriptive (type 1), audit (type 2), before–after (type 3), comparative-experimentalist (type 4), randomized controlled trial (type 5) and intervention to organization (type 6) (Chapters 3 and 4).
  - The *methods* for gathering and analysing the data within the broad categories of quantitative or qualitative methods (Chapters 11 and 12).