CPMR Discussion Paper No. 21

Evaluation in the Irish Health Sector

Michelle Butler

First published in 2002
by the Institute of Public Administration
57-61 Lansdowne Road
Dublin 4
Ireland

www.ipa.ie

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN 1 902448 75 8
ISSN 1393-6190

Cover design by Creative Inputs
Typeset by the Institute of Public Administration
Printed by ColourBooks

Foreword

This paper is the twenty-first in a series undertaken by the Committee for Public Management Research. The Committee is developing a comprehensive programme of research designed to serve the needs of the future developments of the Irish public service. Committee members come from the Departments of Finance, the Environment and Local Government, Health and Children, the Taoiseach, and Public Enterprise, and also from Trinity College Dublin, University College Dublin and the Institute of Public Administration.

This series aims to prompt discussion and debate on topical issues of particular interest or concern. The papers may outline experience, both national and international, in dealing with a particular issue. Or they may be more conceptual in nature, prompting the development of new ideas on public management issues. They are not intended to set out any official position on the topic under scrutiny. Rather, the intention is to identify current thinking and best practice.

We would very much welcome comments on this paper and on public management research more generally. To ensure that the discussion papers and wider research programme of the Committee for Public Management Research are relevant to managers and staff, we need to hear from you. What do you think of the issues being raised? Are there other topics you would like to see researched?

Research into the problems, solutions and successes of public management processes and the way organisations can best adapt in a changing environment has much to contribute to good management, and is a vital element in the public service renewal process. The Committee for Public Management Research intends to provide a service to people working in public organisations by enhancing the knowledge base on public management issues.

Carmel Keane, Chair
Committee for Public Management Research
Department of Finance

For further information or to pass on any comments please contact:

Pat Hickson

Secretary
Committee for Public Management Research
Department of Finance
Lansdowne House, Lansdowne Road
Dublin 4

Phone: (+353) 1 676 7571; Fax: (+353) 1 668 2182
E-mail: hicksonp@cmod.finance.irlgov.ie

General information on the activities of the Committee for Public Management Research, including this paper and others in the series, can be found on its world wide web site: www.irlgov.ie/cpmr; information on Institute of Public Administration research in progress can be found at www.ipa.ie.

Acknowledgements

This project could not have been completed successfully without the active co-operation of a range of individuals involved in the development and use of evaluation in the Irish health services. Among those who participated in the study were representatives from health boards and the Eastern Regional Health Authority, the Department of Health and Children, the Office for Health Management, the Health Services Accreditation Implementation Steering Group and BUPA Ireland.

I would also like to thank Orla O'Donnell for her contribution to the section on economic evaluation, Richard Boyle for his guidance and expertise in evalution, Carolyn Gormley and Karen Isaac for their help in the preparation and presentation of the paper and the IPA's publications division for its production.

Responsibility for the content of the paper, however, rests with the author.

Michelle Butler
February 2002

Executive Summary

Evaluation has a vital role to play in enabling health service planners and managers to attain the highest standards of effectiveness, efficiency, equity, quality and value for money in the services that they provide, and to demonstrate that attainment for accountability purposes. Evaluation has the potential to provide the evidence required for effective decision making at all levels of the health system and across all areas of health care provision. The important role of evaluation and the need to enhance evaluation demand and capacity in Irish health services is underpinned in recent policy documents, such as the two most recent health strategies (2001 and 1994), the Department of Health and Children's (1998) Statement of Strategy, and the Report of the Commission on Health Funding (1989). In addition, recent legislative changes aim to enhance accountability, further endorsing the importance of monitoring and evaluation in health care management.

Public expenditure allocated for the provision of health services has doubled since 1996 to just over €8bn in 2002. Over that time there has also been increasing interest in how effectively and efficiently funding is used. This shift in thinking is reflected in the increased emphasis placed on results-based management and evidence-based decision making in Ireland and internationally, in health services and in public services in general. The health strategy Quality and Fairness: A Health System for You, launched by the Department of Health and Children in 2001, outlines four national goals for the Irish health system: 1) better health for everyone; 2) fair access; 3) responsive and appropriate care; and 4) high performance. It also outlines a range of measures aimed at improving the performance of the health system and enhancing accountability for performance. The development of monitoring and evaluation is a central part of the strategy and the range of measures outlined include: the establishment of agencies with specific responsibility for setting standards in relation to monitoring and evaluation; organisational reform to support the evaluation function and to provide a better link between evaluation and health service planning; the development of formal reporting arrangements between health service funders and providers; additional investment in health research and the development of information technology to support evidence-based decision making.

Three key roles for evaluation can be identified in the literature:

- *Accountability* – to ascertain that a programme is being implemented in accordance with the agreed implementation stages. Accountability is a key element of modern governance which aims to provide greater flexibility and autonomy for managers in return for them taking on accountability for what has, or has not, been achieved.
- *The design and management of policies and programmes* – evaluation can feature throughout the life cycles of policies and programmes. Evaluations can be used in strategic planning: to analyse and better understand the present situation; to understand the impact of previous decisions and apply them to the current situation; and to provide information necessary at most stages of the planning process. Evaluation can be used to ensure that a programme is on track to meet its objectives, to detect early problems requiring corrective action, and to monitor the effectiveness of corrective action. Evaluation findings can also be used to support decision making in the budget allocation process.
- *Learning* – as a feature of healthy organisations, evaluation provides insight into the underlying generative mechanisms of problems, their underlying causes and consequences, thus enabling the organisation to better respond to contextual changes and to identify changes required in rules and behaviour within the organisation.

Currently in the Irish health system, the most common form of evaluation is ex-post evaluation, mainly through expenditure reviews conducted with the Department of Finance, or through value for money examinations conducted by the Comptroller and Auditor General. In 2001, the Department of Health and Children also commissioned a number of one-off reviews, for example the Review of the 1994 *Health Strategy* (conducted by Miriam Wiley of the ESRI) and the review of value for money (conducted by Deloitte and Touche).

At regional level, health boards are required to produce annual service plans against which health board performance can be assessed. A range of approaches to developing evaluation demand and capacity within health boards was identified in this research, through interviews with representatives from a selection of health boards. These included the development of structures to clarify responsibility for evaluation; the development of managerial processes including needs assessment, continuous self-examination and project management; and, external evaluation, such as hospital accreditation and external inspection.

Evaluation can be defined as '… systematic analytical studies conducted occasionally or on an ad hoc basis, to answer specific management questions about performance' (DAC, 2000). Several key themes can be identified in current international thinking on evaluation, including the following:

- *Issues in the design of evaluation in health services*

  As social programmes, health services are inherently political in nature. They have multiple, diverse stakeholders, often with very different, and sometimes competing, priorities and expectations of programmes. All stakeholders will have vested interests, which will need to be accounted for in the design of the evaluation.

  As social programmes, health services involve interplays of individual and institution, of agency and structure, and of micro and macro processes. Accordingly, key issues to be considered in the design of evaluation are: the embeddedness of all human action within a wider range of social processes, social structures and social relations; the underlying social mechanisms or hidden workings involved in outputs; the pre-existing social contexts that are crucial in explaining the success or failure of a programme; the role of underlying mechanisms in the patterns or regularities found; and the unpredictability of social systems.

- *Evaluation and performance measurement*

  Evaluation and performance measurement can be viewed as two sides of the same coin in producing data to support management decisions. Performance measurement can be used to provide on-going feedback on progress against a small number of key measures. Evaluation can then be used to provide more in-depth information about a particular issue identified through the performance measurement process. Evaluation can help to address some of the shortcomings of performance measurement, such as its inability to capture qualitative aspects of performance. In addition, it can help to overcome some of the pitfalls of performance measurement. The two approaches can be viewed as interactive and interdependent.

- *Utilisation-focused evaluation*

  The utility of evaluation is a key concern in the evaluation literature, which also emphasises the need to consider how evaluation findings will be used, and by whom, at the outset of evaluation design. Maximising the utility of an evaluation will involve identifying the primary intended users of the findings and their

particular data needs. It is also recommended that primary intended users are closely and actively involved in the design and conduct of the evaluation.

- *Evaluation as a unique event*
  This suggests that each evaluation should be designed afresh focusing on the key questions that are to be asked about a policy, programme or project. As such, evaluation should represent a unique effort to meet the needs of stakeholders and primary intended users, in an effort to provide maximally useful data to decision makers.

- *Participative evaluation*
  It is suggested that active engagement of primary intended users in the evaluation process alone can influence change. In addition it can enable shared understandings to be developed between managers and staff on the organisation's missions and goals. Involvement in evaluations can give voice to individuals who otherwise might not be heard and can help to improve communications between service providers and users. However, the literature also cautions evaluators against trying to address the needs of too many different stakeholders at once, such that the evaluation questions and findings become too diluted to be of significant practical value to anyone.

The timing and focus of an evaluation are key considerations at the design stage. Decisions will need to be made about the stage(s) of the policy/programme/project cycle where evaluation is to take place. Evaluation can be conducted prospectively (ex-ante) to inform decisions that have yet to be made about the design or content of a policy/programme/project. Evaluation(s) can also be conducted throughout the life of a policy/programme/project to provide interim information about performance. Once a policy/programme/project has been implemented evaluation can (ex-post) provide a retrospective assessment of what has been achieved.

Also to be considered when deciding when a policy/programme/project should be evaluated is whether the programme is ready for an evaluation. The timing of an evaluation can influence the accuracy of the findings because sufficient time will be required for the programme to have an effect, the programme may not yet be operating at its full scope and data may not yet be readily available. Specific guidance for such decisions is outlined in section 4.2.2 of this report. In timing the evaluation, attention also needs to be given to the availability of current and timely data for decision making.

Decisions will also have to be made about which of the number of programmes that could be evaluated in any given year should have highest priority. Several criteria for such decisions are outlined in section 4.2.4.

Once the decision is made about when to conduct an evaluation, a set of relevant and pertinent questions will need to be identified, drawing on the priorities for evaluation and practical and political considerations. Five different types of evaluation questions are identified. Questions can be posed about the rationale for a particular intervention, continued relevance, effectiveness, efficiency or impact. The programme logic model is outlined in this report as a method to assist in the identification of evaluation questions and the design of evaluation studies.

Four main evaluation paradigms are described in this report: 1) post-postivism/scientific approaches – using quantitative methods such as experiments, quasi-experiments, systems analysis and causal modelling; 2) pragmatic evaluation approaches, developed in response to the perceived failure of quantitative approaches to provide timely and useful information for programme decision making. Methods include surveys, questionnaires, interviews and observations; 3) economic evaluation, focusing on issues such as cost-effectiveness analysis, cost-minimising analysis, cost-benefit analysis and cost-utility assessment; 4) interpretive/qualitative approaches, which focus mainly on the experiences that the various stakeholders have of programmes. Methods used include observation, interviews, case studies and documentary reviews.

The credibility and rigour of the evaluation findings will rely heavily on the appropriateness of the methods selected for the evaluation. The purpose of an evaluation is to provide maximally useful evidence and particular evaluation methods may be suitable in one situation but not in another. However, in selecting methods, theoretical considerations will need to be balanced with practical considerations and a trade-off may be required between methodological rigour and the utility of the evaluation.

The uses and limitations of the various approaches are explored in considerable detail in section 5.4. Scientific approaches are useful for impact and effectiveness evaluation questions. Survey methods are useful for questions relating to the continued relevance, effectiveness and impact of programmes. Economic evaluation can be used to measure the costs and consequences of programmes to varying degrees, thus providing data to

address all five evaluation questions. Qualitative methods can provide in-depth insight into programmes as socially constructed entities, exploring the experiences and perceptions of stakeholders. Qualitative methods can provide evidence across all five research questions.

Theoretical considerations will come into play when selecting the most appropriate method(s) for the evaluation question in order to ensure the quality or rigour of the evaluation, and accordingly the credibility and accuracy of the findings. Theoretical considerations will include the value of the types of data – there will be a trade off between breadth and depth, and between generalisability and specificity. Scientific rigour relates to how the methods selected add to the objectivity, accuracy, validity and reliability of the information produced. There are also philosophical distinctions between approaches, relating to understandings about the nature of knowledge and how it is best acquired through evaluation.

Practical considerations include those relating to how stakeholders are likely to view the findings and their credibility, based on their perceptions of what is a good evaluation study and their particular preferences for different methods. This implies that stakeholders could possibly reject the findings of a well-designed and executed study on the basis of their dislike or scepticism of the methods used. Practical considerations will also include the availability of knowledgeable and skilled staff, given that some methods are more easily taught or undertaken; the costs of undertaking alternative methods and the resources available; and the time constraints.

The CSF Evaluation Unit (1999) recommend that evaluations should be: analytical, systematic, reliable, issue-oriented, user-driven, transparent and objective.

This research explored current thinking on evaluation across countries. There is considerable variation, reflecting to some degree differences in thinking on the role of evaluation and how it should be organised. Efforts to develop evaluation have focused mainly on supporting accountability and decision making, with less attention being paid to learning. In a small number of countries where evaluation is well established, there is a greater emphasis now on developing performance indicators to provide on-going feedback on performance, with evaluations being undertaken when needed to inform decisions. There is a greater emphasis on collaborative evaluation, particularly in the areas of health technology assessment and the conduct of systematic reviews. This review would suggest that in the design and development of an evaluation framework

considerable thought should be given to the consistency between proposals and the principles, structures and culture of the health system; the success or failure of any previous attempts to develop evaluation; and how the various approaches would meet with the functions envisaged for evaluation in the particular system.

This review also explores the location of evaluation within systems. A clear distinction is found in the literature between evaluation conducted at the national level and that conducted at the organisational level. For example, in Australia evaluation is driven from the centre and all ministries are required to adopt the central evaluation strategy and to identify in the annual portfolio budget statement areas where evaluation will be conducted. At state level, evaluation is organised in different ways by states. In Canada, two reviews of evaluation found that evaluation was most successful when focused at the operational level. The issue about whether to locate evaluation in the legislature or the executive, or both, is significant. It would appear that the key issue as to the location of evaluation depends on the purpose of evaluation. Is it to improve programmes or to challenge them? How can evaluation deal with impact and relevance issues? The location of evaluation will influence primarily how, and how well, the findings get used to affect change, the accessibility of evidence to evaluators and objectivity. Evaluations conducted within the organisation are more likely to produce information required by managers to make decisions. Access to information is less likely to be an issue than for someone coming from outside of the organisation, but objectivity is more difficult to ensure. Evaluators coming from outside of the organisation will not have the detailed knowledge about the working of programmes and so on that someone from within the organisation would have.

The following key themes emerged in the review of approaches to establishing evaluation within systems:

- the role of the centre in promoting the need for evaluation and leadership, a key influence
- the use of a formal requirement to establish evaluation
- making evaluation a key element of on-going reforms
- linking evaluation to funding, for example in contracts between funders and providers; ear-marking funding specifically for evaluation; funding specific programmes aimed at innovation, with evaluation as a key element.

Two aspects of establishing evaluation in systems are explored in this report: the development of evaluation demand and the development of evaluation capacity. Toulemonde (1999) identifies various 'carrots, sticks and sermons' to build evaluation demand, which are explored in detail in section 7.4. Evaluation capacity demands: sound data systems that will provide good reliable data; developed social science systems that will provide the knowledge of evaluation methodology and the thinking to set the scene for evaluation; a cadre of trained analysts/evaluators; good governance including an ethics infrastructure, an effective legal framework and effective accountability mechanisms (Boyle et al, 1999).

The review in this report of the current status of evaluation in the Irish health system suggests that a framework is needed to provide better coherence to current approaches. The 2001 Health Strategy deals with several aspects of evaluation that could be brought together coherently in a system-wide and system-deep framework. Various challenges are identified to building evaluation demand and capacity within the Irish health system. The main challenges identified are: to build an adequate information infrastructure; to develop a stronger information management culture; to promote better use of evaluation findings in decision making; and to develop evaluation expertise and skills.

1
Introduction

1.1 Focus of the report

This report on evaluation in the health sector was carried out by the Committee for Public Management Research. The research seeks to establish a specific role for evaluation in the management of health services in Ireland through a review of current practice and, in the context of recent reforms, to identify ways in which evaluation can be enhanced further. The research also draws on current thinking internationally on the potential role of evaluation in enhancing management and accountability, and the essential elements of effective approaches to evaluation. The range of macro and micro approaches being adopted across countries is also considered and the research seeks to identify the lessons learned from experience thus far. In order to enhance the applicability of the research to those considering or undertaking evaluation, the research also outlines the range of possible evaluation questions and specific techniques that can be applied across evaluation situations.

1.2 Background and context

The need to improve evaluation capacity within health service management in Ireland is well stated in the literature. For example, the *Report of the Commission on Health Funding* (1989) noted that evaluation was underdeveloped and recommended that the roles of information and evaluation should be further developed in planning health services, allocating resources and decision making. Evaluation also has a vital role in enabling health service planners and managers to ascertain and demonstrate the attainment of the highest standards of effectiveness, efficiency, equity, quality and value for money in the delivery of health services. These are the principles underpinning the health strategies, *Quality and Fairness: A Health System for You* (Department of Health and Children, 2001), and *Shaping a Healthier Future* (Department of Health, 1994), and the department's Statement of Strategy, *Working for Health and Well-Being* (1998). In addition, recent changes in legislation, such as *The Health (Amendment) (No.3) Act 1996*, *The Public Service Management Act 1997*, and *The Health (Eastern Regional Health Authority) Act, 1999*, aim to enhance accountability in health and public services, further endorsing the importance of monitoring and evaluation.

A clear role for evaluation is outlined in the health strategy, *Quality and Fairness: A Health System for You*, (Department of Health and Children, 2001a), launched in November 2001, in order to ensure that resources are used to best effect in the development of a quality culture and in better planning and accountability. The development of effective decision making supported by the production of robust evidence is also a key theme. The range of measures contained in it relating to improving performance, monitoring and evaluation are outlined in further detail in Chapter Two of this paper. While the measures outlined will support the development of evaluation, equally evaluation will be an essential tool in achieving those measures. Evaluation can be used to measure performance against standards set; to better understand the factors contributing to poor and outstanding performance; and to measure the effectiveness of the measures identified and progress in the implementation of the strategy itself. Further endorsing some of the themes detailed later in this paper, the strategy draws on a range of detailed studies recently undertaken of fundamental issues in the Irish health system, and the process through which it was developed is a comprehensive example of a participative approach to planning.

The need to develop evaluation capacity in health services was identified in two recent CPMR research studies on service planning (Butler and Boyle, 2000, CPMR Discussion Paper No. 13) and performance measurement in the health sector (Butler, 2000, CPMR Discussion Paper No. 14). This study further complements that research in raising awareness and promoting discussion on issues of public management in the health sector. It is also anticipated that the findings will be useful to anyone contemplating or involved in evaluation in public services.

1.3 Terms of reference

The terms of reference for this study were to:

a) explore the importance of evaluation in health services and potential uses, drawing on current health service reforms both in Ireland and internationally;
b) review current approaches to evaluation in Irish health services in order to identify areas where evaluative capacity could be further enhanced, drawing on the findings of a review of thinking on best practice internationally;
c) outline the key features of evaluation and the range of possible approaches for evaluation in health services, with reference to particular case studies as examples of how such methods can be usefully applied; and

d) identify the key issues to be considered in undertaking an evaluation.

1.4 Research approach and methods

The approach involves a review of Irish and international literature on current thinking in evaluation, examples of approaches being developed and issues arising. This review also includes thinking on evaluation in areas outside of health services. Initially, the research involved in-depth interviews with a range of individuals involved in the promotion and development of health service management, in order to focus the research on issues that would be relevant to health service managers and professionals involved in evaluation. Interviews were also undertaken with key individuals involved in health service monitoring and evaluation in the Department of Health and Children, health boards and provider agencies, to identify the current status of evaluation, to explore the lessons to be learned for experiences thus far of evaluation in the Irish health service context; and to identify areas where evaluation capacity needs to be improved.

1.5 Structure of the report

Chapter Two sets the context for the research. The potential role for evaluation is explored in the light of changes in thinking internationally and in Ireland about public management and recent public service and health sector reforms in Ireland. Key themes are the growing emphasis on results-focused management, enhanced decision making and accountability. This chapter also provides an overview of the current status of evaluation in the Irish health sector. In Chapter Three, current thinking on evaluation is outlined. This review also explores the relationship between evaluation and performance measurement, evaluation and modes of decision making, and the need for evaluation that is focused on the needs of users. In Chapters Four and Five, issues in evaluation design – such as the key elements of a good evaluation, when to evaluate and what to focus on in evaluation – are discussed. Programme theory is introduced as a method to help to focus an evaluation. In Chapter Five, a range of possible evaluation questions are identified that can be applied at various stages of the life of a project or a programme, which are then linked to specific techniques that can be used to address different types of questions.

In Chapter Six, the development of a framework for evaluation in Irish health services is explored drawing on approaches being used in other countries. This is followed in

Chapter Seven with a review of current thinking on the establishment of evaluation and the development of evaluation capacity. This also draws on experience thus far in other countries. Chapter Eight, by way of conclusion, draws together the key points made throughout the report around the terms of reference set out in this chapter.

2
The case for and current status of evaluation in the Irish health service

> At a time when sometimes simplistic questions are asked about whether 'enough money' is being spent on healthcare, it is perhaps well to remember that the major issue for the future is what we are achieving with the available resources (Lynch, 1998, p. 107).

2.1 Introduction

Since 1997, public expenditure on health services increased by almost 100 per cent to £5.4bn (€6.9bn) (Deloitte and Touche, 2001). This was increased by an additional €255m in the December 2001 budget. Over that same period, public concern has continued and in some areas increased about Ireland s creaky health system (Ó Morain, 2001). Images of patients waiting for unacceptable periods of time for treatment for life-threatening illnesses, waiting on trolleys for emergency admission to hospital for hours, and accident and emergency units bulging at the seams, continue to abound in the media (e.g. *Irish Times,* 2000a, *Unhealthy State* series). While health outcomes, such as premature mortality rates for cancer and cardiovascular disease, have improved over recent years, Ireland still lags behind the European average, partially because health outcomes in other countries have also improved over that time. Further, data is beginning to emerge to highlight significant inequalities in health in Ireland between geographical regions and between specific groups within society (Chief Medical Officer, 1999). Yet the pace of change and reform in Irish health services is unprecedented. Given the extent of the recent attention which the health services have received, a major question in the minds of policy makers, public representatives and others is, what impact increases in spending will actually have on improving health outcomes, the performance of health services and addressing the needs of local, regional and national populations. The above quote by Lynch (1998) reflects an increasing shift in interest from traditional concerns with controlling health expenditure, to how resources are used and what is achieved in return for increased investment in health services.

In response to the notion of 'crisis' in the health services, one question that is increasingly posed in the media is: Do we spend enough on our health services? While

acknowledging that following cuts in spending in the 1980s by as much as 15 per cent we are still playing catch up, comparisons with spending in other countries shows that despite the recent increases in spending, we are only now spending at the EU average per head of population (see Table 3, page 43, of the *Health Strategy*, 2001). Of the 100 per cent increase in spending from 1996 to 2000, two thirds went on pay costs and other technical items (community drug schemes etc) rather than service development (source: Department of Health and Children). In addition, a sizable proportion of the increases will go on the provision of social services (expenditure allocated to health in Ireland is also for the provision of what would be considered in other countries to be social services). 'Thus, much of what (was) announced as increases in health spending, while going on these very necessary services, has no impact on waiting lists or on acute hospital services, the areas where the public perceives a 'health' crisis' (Wren, 2000).

The question is also asked: 'Is the problem … that we spend too little, or that we spend it badly?' (Wren, 2000). The trouble is that without evaluation feedback we are unable to answer such questions conclusively, and it is true to say that many decisions made currently in the health area are based on anecdotal, or at best, inadequate information on the performance of health services, the effectiveness of different patterns of care, or of the needs to be targeted in the provision of health services. In addition, a lack of investment in IT (less than half of one per cent of investment in health services) makes the collection, analysis and dissemination of data on performance difficult.

The concern about how (well) money is spent in health services has been emphasised by the Minister for Finance on several occasions. For example, at the meeting of Ministers in Ballymascanlon in May 2001, he reiterated that he expects value for money from funding allocated in the estimates at the end of the year and that he was not satisfied with how 'the 'enormous' amounts of money already allocated had been spent'. At the same meeting, the Minister for Health and Children, Micheal Martin TD, emphasised that 'money is not the only solution' and reform of the structure of the health services will also be required (O'Connor, 2001). This is reflected in plans in the recently launched health strategy to commission an audit of organisational structures and functions in the health system, and to 'consider the number and configuration of existing health boards and other agencies and the scope of rationalisation' (p.130).

The Irish health system represents a huge investment both in terms of public confidence and public funding. It occupies a central place in the collective consciousness in so far as all members of society, at some time or other, will be users of services, often at a very vulnerable time in their lives. It accounts for almost 20 per cent of total public expenditure (which is only second to expenditure on social welfare) and is extensive (with about 86,000 employees) and complex. In this context, questions about how the health system and health services meet their objectives, the contribution they make to enhancing health and social gain within society, and their effectiveness, efficiency, equity, quality and value for money, are entirely appropriate. In addition, feedback on performance is required to enable those who are responsible to steer health services in the right direction. This chapter considers the case for evaluation in the Irish health sector, and in that light reviews the current status of evaluation.

2.2 Towards a focus on results and evidence-based decision making

Ireland and its health sector are not alone in this increasing focus on results in management. At the heart of public service reforms across countries is the shift away from traditional command and control management towards an emphasis on governance and accountability, where clear objectives are agreed that incorporate both national and local priorities and where managers are allowed to get on with managing to achieve these objectives. The thinking in the development of results-focused management is that, rather than bind organisations in rules and compliance with processes that can divert attention from important priorities, programmes and services should be focused on what needs to be achieved. Thus, a focus on results (outcomes and impacts) rather than on processes can:

- free agencies and staff from a preoccupation with complying with regulations and enable them to redirect their efforts towards creative problem solving, resulting in innovative high quality programmes
- provide stakeholders with the opportunity for 'collective, shared deliberation' about what constitutes valued outcomes
- illuminate whether investments are adequate to achieve expected results (Greene, 1999).

As Schick (1996) suggests, current reforms are based around the development of accountability frameworks in which governments entrust spending agencies with flexibility in using resources, in return for holding them responsible for results.
Evaluation has a key role to play in results-focused management. Through evaluation, robust data is generated on what has been achieved and how effectively achievements have been managed. As the basis for effective decision making, evaluation also enables managers to move services towards achieving objectives by identifying where corrective action needs to be taken and to monitor the progress of such corrective action.

Specifically in the health sector, it has been suggested (*Report of the Commission on Health Funding,* 1989) that health system structures confuse political and executive functions and that the devolution of decision making to regional/local level would enable the 'proper balance between local and national decision making' to be achieved. Further, it is suggested that devolution of accountability and responsibility would also enable the department to refocus efforts on policy (Department of Health, *Shaping a Healthier Future,* 1994). The Health (Amendment) Act 1996 and the Eastern Regional Health Authority Act (1998) set out the provisions for such separation and clarify the responsibility of the health boards/ERHA and other agencies in the new structures.

Four national goals are set out in the new health strategy, *Quality and Fairness: A Health System for You* (Department of Health and Children, 2001a):

- better health for everyone
- fair access
- responsive and appropriate care
- high performance.

The strategy aims to improve the performance of the health system and accountability for performance, by developing standardised quality systems to support best patient care and safety and by ensuring that evidence and strategic objectives underpin all planning and decision making. There is an objective related to this goal, that decisions across the health system will be based on the best available evidence from research findings, qualitative or quantitative data or other documented trends and behaviours, or on agreed standards, protocols or models of best practice (point 68).

It proposes that accountability will be strengthened through further development of performance indicators in the service planning process, to provide a stronger framework for assessment of health board performance by the Department of Health and Children on an annual basis, and by developing the monitoring function of the department (point 70). In addition, health boards will be expected to include performance indicators in service agreements with providers, including voluntary agencies. Increased investment in health research is proposed, to increase the evidence base for decision making, to support health professionals to undertake research and to foster an 'active research environment' (point 73).

Six frameworks are identified to support the four national goals, two of which concern organisational reform and health information. The framework for organisational reform aims to support effective decision making based on the best available evidence and to promote high quality services. It aims to strengthen health board accountability and service planning to ensure the best possible value for money and the pursuance of high quality standards. It will require health boards to have explicit responsibility for driving change at regional level and to focus on the actual outcomes of services for which they are responsible. It sets out the basis for the establishment of a Health Information and Quality Authority, which will:

- ensure services meet nationally agreed standards
- assess if health and personal social services are managed and delivered to ensure the best possible outcomes within the resources available
- have responsibility for:
  - developing health information systems
  - reviewing and reporting on selected services each year
  - overseeing accreditation and developing health technology assessment
- be established on an independent statutory basis.

In addition, a new division dealing with population health is to be established in the department to facilitate health impact assessment (health proofing) of both health and non-health policy and government decisions. It also states that departments of public health within health boards will be developed further as population health functions; and will work closely with the new division in the monitoring and evaluation of regional and local initiatives and the development of appropriate performance targets and indicators.

A framework is also identified for the development of health information. On the basis of the findings of the Deloitte and Touche Report (2001) that inadequate information is 'a critical weakness which limits the capacity for prioritisation, planning, evidence-based decision making, efficient service delivery and monitoring and evaluation at all levels' (p.131), the strategy highlights the need to develop information as a clear basis for identifying priorities, demonstrating performance and value for money and establishing the evidence for decisions.

Once the range of goals, objectives and targets are outlined in an action plan, the strategy devotes a section to implementation, 'making change happen'. Monitoring and evaluation are key elements identified. It states 'Monitoring and evaluation must become intrinsic to the approach taken by people at all levels of the health services' (p.179). The following arrangements will be put in place to support these functions.

- A formal organisational function for monitoring and evaluation will be established in the department and the health boards. The 'function at health board level will be to monitor progress against targets and to evaluate outcomes over the medium to long term (p.180).
- High-level standards that are challenging but attainable will be set. These will reflect the targets outlined in the strategy, and will monitor progress using robust information and will facilitate international comparisons on the basis of WHO guidelines.
- There will be further development of performance indicators for service planning, monitoring and evaluation and a national set of performance indicators, through joint work between the department and health boards, with input from the Health Information and Quality Authority and the National Hospitals Agency.
- Local indicators will be developed for local management to supplement the national set.

The strategy states that the development of good performance indicator information should enable:

- managers to judge that service delivery is effective and quickly identify difficulties arising
- policy makers to judge how well policy is being implemented
- evaluation and review of services and policy, thus informing future developments

- better communication of achievements, understanding of actions required, and participation in management across professional boundaries
- the public to be better informed (p.180).

The specific role identified for evaluation is to provide a more focused and in-depth assessment of the quality, equity and patient-centredness of particular services. It calls for a more systematic approach to the evaluation of services on a national basis, both at national and local level. At the national level, external evaluation will be conducted by the Health Information and Quality Authority. At local level, evaluation will be carried out by a monitoring and evaluation function to be developed. The department will require health boards to specify the formal evaluation to be undertaken each year as a part of the service planning process.

Further details for the development of external monitoring and evaluation and for cross-sectoral monitoring and evaluation are provided. In addition, the need to develop a robust monitoring and evaluation culture and to develop evaluation capacity to support that culture are emphasised. In particular, it identifies the need for major investment in information and communication systems and a major programme of human resource development to develop a 'supportive rather than policing approach' (p.182).

2.3 The current deficit in information and evaluation

The current deficit in information and evaluation in health service management is noted in several policy documents1. A steering group, with related working groups, was set up by the Minister for Health and Children in 2000 to develop a national Health Information Strategy. In the objectives for the strategy there is a strong emphasis on ensuring that individuals – managers, professionals, users of services – have the information that they require to be able to make informed decisions; on the availability of information required to monitor the effectiveness (for example, improvements in population health) and the quality of health services, and on accountability.

The National Health Information Strategy Steering Group's discussion document (NHIS, 2000) suggests information is a vital requirement across several aspects of health service management, including:

- *at the level of government and the Oireachta*s – for public accountability, priority setting and allocative efficiency
- *at the level of the policymaker* – for the assessment of need; for performance measurement, quality assurance and policy evaluation; and to predict and respond to emerging health issues
- *at the level of the manager* – for strategic planning and commissioning; clinical governance; and evaluation and monitoring

> Managers must be able to evaluate the services which they provide according to the objectives and targets of those services. Evaluation requires similar information to that which is required for planning and commissioning (NHIS, 2000, p. 11).

- *at the level of the public health practitioner* – to promote and protect the health of the population; monitor health status, disease patterns and health inequalities; assess health needs; plan, monitor and evaluate health services; and for evidence-based policies and effective interventions
- *at the level of the clinician* – for communications and exchange of information, e.g. GPs accessing laboratory results electronically; access to information on evidence-based medicine, clinical guidelines and protocols; clinical audit, risk management and clinical governance; and access to information about new health threats
- *at the level of the user or member of the public* – to provide information about health and illness; for them to understand and follow advice and treatment; to provide information on entitlements and locally available services; and to provide information about the quality of hospital or GP services.

Evaluation features at all six levels, from ascertaining the effectiveness of services overall to making choices and judgements at the level of the user, on the basis of information provided. Also across the six levels, elements of *ex-ante* evaluation (identifying the needs to be addressed through a particular policy or programme), *on-going* evaluation (monitoring the progress of policies or programmes in addressing the needs identified and on which the policy or programme is founded) and *ex-post* evaluation (retrospectively examining the effectiveness and impacts of policies or programmes) are identified.

The importance of information and evaluation in health services, and its current absence, are well noted in documents dating back to the mid 1980s, for example, in the Department of Health's strategy document *Health: The Wider Dimensions* (Department of Health, 1986). Further, the Report of the Commission on Health Funding (1989) suggests that allocation decisions are made without sufficient knowledge of the consequences and are based on intuitive rather than objective criteria. The importance of information and evaluation is underlined in informing the decision-making process. The report states that 'all decisions should, to the greatest possible extent, be based on information that is accurate and sufficient' (p. 180). Three broad types of research are identified which can be used to produce this type of information for decision makers:

- *epidemiology,* providing information on the determinants and distribution of disease
- *clinical or biomedical research*, developing and assessing treatments
- *health services research*, providing information on the effectiveness and efficiency of the delivery of health services.

The focus of this study is primarily on the third type of research – health services research and information obtained through the evaluation process.

Further, the Commission identifies three categories of evaluation in the management of health services:

- compiling and evaluating information as an integral part of the day-to-day management of local services
- coordination and assessment of this information, for the provision of technical support to local evaluation, as part of the central management of health services
- research and evaluation in areas such as priorities for resource allocation, technology appraisal and the development of service protocols (p. 192).

Accountability is also enforced in the following legislation. *The Comptroller and Auditor General (Amendment) Act 1993* sets out the responsibilities of the Comptroller and Auditor General (C&AG) to oversee moneys administered to or by government departments, ensuring that expenditure was applied for the purposes for which appropriations were made. It also requires the accounts of health boards to be audited by the C&AG, ensuring that expenditure was applied for the purposes for

which appropriations were made. The Act also enables the C&AG to carry out examinations in relation to whether and to what extent the resources of the department (or health board) have been used economically and efficiently, and to examine the systems, procedures and practices employed by a department to evaluate the effectiveness of its operations.

The Public Service Management Act 1997 requires each government department to produce a strategy statement outlining key objectives, outputs and related strategies, including use of resources. Departments must ensure that resources are used in accordance with the Comptroller and Auditor General (Amendment) Act 1993 and must manage and develop means to improve the provision of cost effective public services.

The Health (Amendment) (No.3) Act 1996 sought to enhance accountability at health board level. It states that, in pursuing its functions, a health board should have regard for the need to 'secure' the most beneficial, effective and efficient use of resources, 'wherever originating'; produce an annual service plan outlining the services to be provided and estimates of income and expenditure for the period; supervise the implementation of its service plan to ensure that the net expenditure for the financial year does not exceed the net expenditure determined by the minister. The chief executive officer shall implement the plan. Each board should produce an annual report in relation to the performance of its functions during the preceding year that will include a statement of services provided and financial statements.

The Health (Eastern Regional Health Authority) Act 1999 provided for the establishment of the Eastern Regional Health Authority (ERHA), the dissolution of the Eastern Health Board and the establishment of three area health boards accountable to the ERHA. It requires the newly established ERHA to put in place systems, procedures and practices to enable it to monitor and evaluate services provided, to provide in its annual report an account of measures taken to monitor and evaluate services and an account of the outcome of such measures. It also requires the ERHA to enter into written agreements with each service provider, the agreements to contain standards relating to the efficiency, effectiveness and quality of services to be provided. In addition, the Act outlines the requirement of the Regional Chief Executive to provide evidence to a Committee of Dáil Éireann, whenever required on: the economy and efficiency of the ERHA and the area health boards in their use of resources; the systems, procedures and practices employed by the ERHA and the area health boards

for the purpose of evaluating the effectiveness of their operations; and any matter referred to the ERHA by the C&AG. The area health boards are also required under the Act to put in place systems, procedures and practices to enable them to monitor and evaluate services provided by anyone with whom they make an arrangement to provide a service. The new health strategy requires all health boards to put similar arrangements in place.

Butler (2000) argues that, along with deficits in information and evaluation, there are deficits in relation to the appropriate use of data in decision making and that current data is under-utilised and is not seen as a management tool. The need to support the development of skills and competencies required to analyse and interpret data at all levels of management is also noteworthy. These issues are also raised in the new health strategy.

Although a formal framework for evaluation has yet to be established in the management of health programmes and initiatives, there is evidence of increasing activity in the area of evaluation, some of which is explored for the purposes of this study.

2.4 The potential role of evaluation and modes of decision making

The CSF Evaluation Unit (1999) suggests that at a general level the role of evaluation should be to promote a culture of critical analysis of state expenditures. At a more specific level, the impact of evaluation on programmes will include adjustment, discontinuance of obsolete programmes and enhancement of the watchdog role. Three key roles for evaluation can be identified in the literature – to enhance accountability, to improve the design and management of programmes and to facilitate organisational learning.

2.4.1 Accountability

During programme implementation, evaluation can be used to ascertain that a programme is being implemented in accordance with the agreed implementation stages. It can also examine whether the best use of resources is being made in implementing the project – efficiency. Highlighting the importance of accountability, Patton (1997) suggests that lack of systematic accountability is the reason why government programmes fail.

Mayne and Ulrich (1998) explore the concept of accountability in the context of modern governance and management in public service. In modern governance and management, many alternative delivery approaches are being tried that replace traditional hierarchical relationships and responsibilities. There is also a greater focus on results-based and performance-based management, paralleled with greater flexibility and autonomy for organisations and managers to achieve results. Transparency is also an essential feature of public sector accountability, and an awareness of the need to address concerns about the integrity of government. In the light of these four developments, they redefine accountability as:

> …a relationship based on the obligation to demonstrate and take responsibility for performance in light of agreed expectations (p.4).

Further, they define accountability for results as follows:

> Accountability for results asks if you have done everything possible with your authorities and resources towards affecting the achievement of intended results and if you have learned from past experience what works and doesn't work. Accounting for results of this kind means demonstrating that you have made a difference, that through your actions and efforts you have contributed to the results achieved. It means you are accountable for what you can influence as well as what you can directly control. And the greater management flexibility which has often accompanied a greater focus on results provides the needed means to better manage your ability to influence outcomes. Demonstrating the results you have achieved, including what you have influenced, provides the evidence of effective stewardship of the greater flexibilities made available (Mayne and Ulrich, 1998, p.7).

This shift, from command and control to devolution and accountability, clearly places the onus on managers to demonstrate what has been achieved, in return for being given the flexibility to manage services in whatever way they deem necessary to achieve results. Within such arrangements, performance measurement and evaluation are important management tools.

2.4.2 The design and management of programmes

Rist (1990) suggests evaluation can feature throughout the 'life-cycle' of policies and programmes, with front-end (*ex-ante*) analysis to identify needs and to help in the

design of the policy, programme or project. Work is now going on within some health boards, and with the small areas assessment unit at Trinity College, to develop needs assessment. Another recent example of a form of ex-ante evaluation in Ireland is the review of the drugs strategy to inform the development of a new National Drugs Strategy.

During the life of the policy, programme or project, evaluation can be used to ensure that the programme is on track to meet its objectives, to detect problems early on and to monitor the effectiveness of corrective action. Evaluation can 'throw light' on the key issues affecting 'the means and the end' (McKeown, 1999). Evaluation findings can also be used to support budgetary decision making and in the allocation process – to make sure resources are allocated to those activities that contribute most effectively to achieving the objectives of the organisation (Baird, 1998).

Bastoe (1999) explores the link between evaluation and strategic planning, budgeting, monitoring and auditing in public administration in general. He suggests evaluation can have an important role both at governmental and organisational level strategic planning, in an analysis of the current situation, to get an understanding of what works well and what needs to be changed, and to examine the costs and benefits of existing policies, which 'is a key to informed, tough-minded, policy analysis and formulation (World Bank, 1994b)'. The real value of evaluation is in providing a good basis for planning by providing data about previous and ongoing programmes and policies; without a good basis for planning, strategic plans are likely to unrealistic and vague. Evaluation can be used to close the 'planning-implementation-feedback policy loop' (Bastoe, 1999, p.97, citing Boyle, 1996). Bastoe suggests that in strategic planning, evaluations can be used: to analyse and understand the present situation; to 'systematise' past experiences and clarify possible decisions; to analyse plans based on previous experiences; and, to provide necessary information in most stages of the planning process. However, plans need to have clear objectives that can be evaluated.

Retrospective or *ex-post* evaluation can be used to assess outcomes or impact once policies, programmes or projects have been established long enough to effect change. Evaluations carried out at the local level can also provide useful information to be fed into decision making at the national level about national programmes and in identifying national priorities. Regardless of the focus of evaluation and whether it informs decisions at the national, regional or local level, if evaluation is to be useful in decision making, it must provide information that is timely, relevant, reliable and available in the

appropriate format. The 'utility' of evaluation is an issue that is discussed later in this report.

2.4.3 Learning

Experience-based learning is identified by Bastoe (1999, citing Olsen, 1993) as a general characteristic of healthy organisations. This involves looking back and building on the transfer of information and evaluation of earlier experiences.

> A learning organisation is also characterised by a management and staff that are constantly concerned with internal and external development. This presupposes something more than methods and techniques. An 'inner understanding' in the form of insight, values, attitudes, and personal development must be included to generate change and growth (Bastoe, 1999, p.108).

Gray et al (1993) suggest that evaluations close the learning loop in the policy process, whereas in the past incrementalism failed to contribute to effective learning. Evaluation can be used to determine what worked well in the past and could be worth trying again, what did not work well, to identify innovative ideas, and to identify good practice to be shared. McKeown (1999) suggests that organisations are now more strategically aware of the impacts that they have on society and stakeholders and use experience to learn and to grow. Evaluation can provide the sort of feedback that organisations need to learn from experience.

Hummelbrunner (2000) identifies three types of organisational learning:

- *single loop learning* (learning to adapt). This type of learning results in a change in strategy or action but is only likely to be successful in the short term as it focuses on apparent symptoms of larger problems and does not address the underlying values or assumptions.
- *double loop learning* (learning to change). This type of learning involves reflecting on values and assumptions to understand the generative mechanisms of problems, their underlying causes and consequences, and leads to better mid-term and long-term reactions to contextual changes.
- *Deutero learning* (learning to learn). Learning occurs by reflecting on the learning mechanisms so that it is possible to change rules and behaviour, such as different ways of recognising and handling problems.

This suggests that evaluation can support more successful forms of organisational learning by providing insight into underlying generative mechanisms, the effectiveness of previous responses to contextual changes, and the consequences of alternatives in future responses to change. Hummelbrunner suggests that if evaluation is to lead to learning in an organisation, the actual functioning of a programme and key context factors must be taken into account in evaluation. That is, it is not sufficient to focus on questions of achievement (e.g. results, impacts) alone.

2.4.4 Evaluation for accountability, decision making and learning at various levels of the health system

Based on the previous discussion, the potential role for evaluation, at various levels of the Irish health system, is outlined in Figure 2.1. At the national level the focus of evaluation would be on accountability for the achievement of national health service objectives and to demonstrate progressive improvement in the provision of health services, especially in the light of dramatically increased spending on health. This would also support constructive debate on the design, organisation and performance of health services and enable priority areas for attention to be identified. Secondly at the national level, findings both from evaluations conducted at the national level and from those conducted at the local level, provide the evidence base required for effective decision making on national health issues and for strategic planning. Thirdly, collaborative evaluation, across similar service areas and regions, can facilitate learning and help to share some of the costs and expertise required to conduct evaluation.

Figure 2.1 The role/potential role of evalution in the Irish ealth system

Level	**Accountability**	**Effective decision making**	**Learning**
National	• Achievement of: – health strategy objectives – objectives in national strategies /action plans falling out of health strategy (e.g. cancer, cardiovascular strategies, dental health action plan) – specific national initiatives • Demonstrate progressive improvement of health services and effective governance of health system	• Strategic development of health services – what works, what does not • Evidence-based provision of effective and appropriate services/programmes • Allocation of resources to maximise return and to meet greatest needs	• National and sectoral collaborative learning, e.g.: patient satisfaction survey, hospital accreditation • Piloting programmes / innovative approaches • Learning from other sectors and jurisdictions • Identifying and sharing good practice • Meta-evaluation
Regional	• Evaluation of achievements against service plan	• Programme management /project management	• Piloting / innovation • Identifying and sharing

	objectives • Value for money for new and existing services	• Decisions about establishing and retaining services • Evidence-based practice • Allocation of resources	good practice • Meta-evaluation
Organisa-tional	• Provider plans • Service agreements	• Project management • Services appropriate and meeting users' needs? • Evidence-based practice	• Piloting / innovation • Identifying and sharing good practice

At the regional level, evaluation can be used for accountability, for health boards to demonstrate both that they are meeting the objectives agreed in service plans, and that they are holding providers accountable for the objectives agreed between them and the health board. In terms of decision making, evaluation findings can enable health boards to know if their programmes are on track to meet objectives and time scales agreed and to identify where corrective action is required. Evaluation can also enable health boards to identify local population needs and particular areas where additional efforts or resources are required to address assessed inequalities in health or access to services. Evaluation can also be used to distinguish the most effective treatments and patterns of care from those that are less so, thus ensuring cost-effective use of resources. Similarly, evaluation can be used to assess the impact of existing programmes, identifying those that work well and should possibly be extended, and those that are not very effective and that may need to be reformed or replaced.

Evaluation can also support accountability, decision making and learning at the local level in a similar way to evaluation conducted at regional and national level. One particular benefit at the local level can be to improve the responsiveness of services by assessing the degree to which services meet the direct needs of users. In addition, the focus of evaluation at the local level is more likely to be on process (delivery of care) and outputs, rather than high level outcomes. Evaluation at the local level is also useful in developing innovative approaches to service delivery, which if proved successful, can be extended later to other areas. It is also a useful element of project management at the local level.

While the three roles for evaluation outlined above (accountability, decision making, and learning) would appear to be vital elements of effective management, the concepts of accountability and management improvement may not sit well together. Accountability requirements can shift the focus to outputs, rather than the more difficult-to-measure outcomes or impacts, and without understanding the cause and effect mechanisms involved in them. Accountability is concerned with measuring what

has been achieved against pre-defined targets rather than analysing how or why it was achieved, whereas management improvement 'is concerned with analysing the context and factors influencing performance, and with drawing lessons for improving performance' (Development Assistance Committee (DAC), 2000, p120). DAC also suggest that the two approaches imply different data collection and analysis approaches, and may influence management's behaviour differently – a managing-for-results focus would tend to encourage risk-taking, experimentation and learning, whereas an emphasis on accountability-for-results may encourage a more conservative, risk-averse approach (examples given include avoiding potentially risky projects, focusing on lower-level results, and setting easily attainable targets).

2.5 The current status of evaluation in Irish health services

At national policy level and in relation to national programmes, the most common format for evaluation is *ex-post* evaluation. Two key mechanisms are used to conduct evaluation of national programmes. Expenditure reviews are conducted with the Department of Finance as a part of its review of all government departments. Three such reviews have been conducted dealing with dental services (further information is provided in Appendix One), nursing home subvention and intellectual disabilities. Value for money examinations are conducted by the Comptroller and Auditor General. Three such reviews were conducted in relation to: energy management (1995), prescribing practices and the development of GP services (1997) and the emergency ambulance services (1997).

Leading up to the launch of the recent health strategy, the Department of Health and Children commissioned several national reviews. The department commissioned Professor Miriam Wiley, from the Economic and Social Research Institute, in 2001 to undertake a critique of the 1994 *Health Strategy,* as part of the preparations being undertaken towards the development of a new health strategy. The department also commissioned Deloitte and Touche to conduct a value for money (VFM) review of health services in 2000. The report (Deloitte and Touche, 2001) identifies the lack of a culture of performance measurement and management in the health system and recommends that 'structures to improve the planning, implementation and monitoring and evaluation of both core service delivery and national strategies need to be established' (p.162). The report recommends steps that might be taken towards the development of a performance measurement culture. The development of 'a prescriptive approach to VFM, with the establishment of the appropriate processes for

monitoring and evaluation' is recommended, along with the development of systems and processes for performance measurement and evaluation in relation to service delivery and health outcomes, internal audit, audit of management practices, clinical audit and governance. The report also highlights the need for a major investment in IT to support these recommendations.

At regional level, the *Health (Amendment) (No.3) Act, 1996* clearly makes the service plan the key accountability document between the department and health boards. The service plan outlines the health board's commitment to provide a specific range of services/service developments over the coming year, against which performance can be evaluated. Although there is a requirement for health boards to produce an annual report, previous research by the CPMR (see Butler and Boyle, 2000) found that there was considerable variation between health board annual reports, particularly in terms of how they reported board performance and the results of evaluations carried out. Although the ERHA Act explicitly requires boards in the Eastern region to have monitoring and evaluation arrangements in place, these are not yet formally required in other regions. However, significant work has been undertaken in some boards to enhance evaluation. Some of these measures are outlined in the following section of this report.

As discussed in the two previous CPMR reports on service planning and performance measurement, accountability until very recently was essentially focused on boards staying within budgets. More recently, there has been a greater focus on providing a more balanced view of performance and this is reflected in the development of performance indicators in the service planning process (first introduced in 2000). These will be developed further year on year. As is outlined later in this paper, evaluation can be a very useful tool to further examine problems identified through assessment against performance indicators.

This review also explored the development of evaluation at health board level and below through interviews with a small number of representatives. The aim of the review was to identify the range of approaches being developed rather than to conduct a comprehensive audit of all health boards. Therefore, it is likely that there are good examples that exist but that have not been included in the following overview. Three types of efforts to enhance evaluation were identified: the development of structures, processes and external evaluation.

2.5.1 Structures

Examples were found within health boards of structural changes to formalise the evaluation requirement and to establish specific responsibility for evaluation. For example, in the Eastern Regional Health Authority, a planning and evaluation directorate was established. Evaluators work across a number of service areas, and once the planning and evaluation team was established, a series of priority areas for evaluation were identified and a programme of evaluation outlined for 2001. In some service areas, strategic reviews were undertaken to identify a framework for the strategic development of services. For instance, the focus in one service review was on assessing the relevance of services, assessing services in relation to recommendations made in key documents relating to services, and mapping out the full range of services provided and how they relate to assessed needs.

In the Midland Health Board, the remit of its Department of Public Health has been extended to include an explicit link between evaluation and planning. An additional and complementary Directorate of Corporate Fitness has also been established to bring risk management, clinical audit, service audit, health and safety, and occupational health together in one function, with a formal role in relation to how performance complies with policies, governance and value for money.

In the Southern Health Board, the remit of the recently established Strategy and Planning Directorate is business evaluation, linking evaluation into the business planning process and also the development of performance indicators using the balanced scorecard.

Plans to formalise evaluation through the development of organisational structures in health boards and within the department are outlined in the new health strategy (Department of Health and Children, 2001a, see section 2.2).

2.5.2 Processes

Several approaches to managerial reform featuring evaluation as a key element can also be identified. Evaluation is a key element of the continuous quality improvement (CQI) approach being pursued within some health boards, and elsewhere. For example, the Midland Health Board has a formal quality strategy, from which quality initiatives are identified for each care group. These initiatives are outlined in the board's annual service plan. The CQI approach is also being adopted by several of the major academic teaching hospitals that have applied to take part in a national quality accreditation programme (outlined later in this report). The CQI approach involves

continuous self-examination of performance to identify areas where improvement is required, and to ascertain the effectiveness of corrective action. The accreditation process supports this approach by providing the basis for comparisons with other similar hospitals and peer review of services.

Evaluation as a key element of the service planning and business planning process has been developed well in some health boards. For example, in the Southern Health Board, its Department of Public Health provides information on population needs for specific services, based on needs assessment, to inform the strategic plan. The strategic plan informs the annual service plan, which informs the business planning process. As a part of the business planning process, a detailed work plan is agreed which sets out what is to be done, by when and by whom. The progress made against the work plan is tracked using MS Project and is monitored quarterly. In addition, each service is being reviewed in turn to develop 'roadmaps for the future', involving the development of service strategies and assessment of the impact of programmes. At BUPA, evaluation is a key element of the project management process which is based around a standard project management framework. The process includes on-going monitoring of progress against specific timetables and debriefing to share experiences, to identify what went well and to identify what went wrong.

Also within health boards and service provider organisations, there is an element of evaluation that is initiated by professionals within their own disciplines focusing on the effectiveness of particular treatments or patterns of care.

2.5.3 External evaluation

The Social Services Inspectorate (SSI) and the Mental Hospitals Inspectorate (MHI) are examples of bodies with explicit responsibility for external assessment of the quality of services provided.

The SSI (for further information visit the SSI website: www.issi.ie) was established by the Minister for Health and Children in 1999. In its initial period it is an independent professional body administered by the Department of Health and Children but the *Health Strategy* (2001) states that it will, by 2003, become a statutory body. The current role of the SSI is to carry out inspections of services provided for children in care, but it is intended to extend this cover later to a wider range of social services. The SSI works in consultation with health boards and other agencies in the conduct of inspections and the Chief Inspector reports to the department on the work of the inspectorate and provides advice on the quality of residential child care, social work

and social care services. Inspections focus on service delivery at the point of impact on users and carers and approaches include: interviews with users of services; interviews with staff involved in the delivery and management of services; site visits; practice observation; the systematic collection and analysis of data to make professional judgements; and the evaluation of services against previously agreed standards and criteria. Findings and recommendations are reported to the agency inspected and reports are made available to the public.

The role of the MHI (under the *Mental Treatment Act* 1945) is to conduct annual inspections of all psychiatric hospitals. Draft reports are presented to health board chief executive officers and to medical and administrative directors of private and voluntary hospitals for observation, before being presented to the Minister for Health and Children. The *Mental Health Act 2001* makes provision for the establishment of a Mental Health Commission as an independent agency to raise standards and practices in the delivery of mental health services and to take all reasonable steps to protect the interests of detained persons. The Act also provides for the Mental Health Commission to employ an Inspector or Mental Hospitals, whose principal functions include annual inspections of every approved mental hospital and to furnish a report in writing to the commission on the quality of care and treatment provided.

In 1999, thirteen acute hospitals took part in the Irish Society for Quality in Healthcare National Survey of Patient Satisfaction. Participation in this collaborative effort enabled organisations to receive comparative feedback on patient satisfaction with the services that they provide, along with recommendations for improvements required and advice on how they could be achieved. The new *Health Strategy* (Department of Health and Children, 2001a) outlines plans to establish a national standardised approach to measure patient satisfaction by the end of 2002, and the systematic collection and analysis of complaints following the publishing of appropriate legislation by the end of 2002.

The third example of external evaluation is the acute hospital accreditation scheme currently being developed with major academic teaching hospitals in Dublin, Cork and Galway. The development of the accreditation scheme will provide a framework within which evaluation will have a significant role. Initially, the scheme is aimed at the major academic teaching hospitals, but once established, the scheme will be extended to include other hospitals. Hospitals apply for accreditation on a voluntary basis rather than being required to. At the time of writing the scheme is well advanced and five

hospitals have applied for accreditation. An independent statutory body is being established to accredit organisations. Once organisations make the application for accreditation, the accreditation process will involve the establishment of self-assessment teams, self-assessment against the standards, specific education programmes provided to the organisation by the accrediting body, and a peer review survey visit. An organisation may be awarded accreditation where it is predominantly compliant with standards; where the organisation complies with all standards, which, if not complied with, would represent a danger to patients/ clients or staff; and, where there is organisation-wide commitment to and application of the principles of quality improvement. Once an organisation has been accredited, it will be required to submit a progress report to the accrediting body twelve months later and the organisation receives a high-level visit eighteen months post survey. The organisation will also receive focused visits according to the organisation's level of compliance with the standards at the time of the last survey and the organisation is resurveyed every three years by the accreditation body.

Assessment is made against standards relating to five aspects: care/service; environmental management; human resource management; information management; and leadership and partnerships. Standards are described in statements that outline what organisations set out to achieve. For each standard, there are criteria relating to the processes required to achieve the standards. This will include planning, implementation and evaluation. Although organisations are required to have provisions in place for evaluation, there are no specific requirements in terms of the type of evaluation to be conducted. However, part of the peer review process includes providing advice on evaluation. In addition, organisations are required to provide evidence of compliance, for example the implementation of policy or performance ascertained through evaluation or against performance indicators.

The standards were developed through a series of working groups by the central project team and with the support of the Canadian Council for Health Services Accreditation. The standards were tested and further refined in pilot self-assessments undertaken in the major academic teaching hospitals from May 2000. The refined standards were then evaluated and received international validation from the International Society for Quality in Healthcare's (ISQua) ALPHA (Agenda for Leadership in Programs for Healthcare Accreditation) Programme. Assessment of an organisation against the standards can provide the basis for identifying improvements that are required.

Evaluation is also a key feature of the development of the accreditation scheme itself and a joint Irish/Canadian study is planned to evaluate its effectiveness. The evaluation will focus on:

- 'The extent to which introduction of accreditation meets its objectives of:
  –establishing and promulgating Continuous Quality Improvement in Health System settings
  –promoting development and use of indicators for quality measurement and improvement
- The study of different approaches to implementation of the accreditation process' (Health Services Accreditation Steering Group (HSASG), 2001).

Work is also ongoing to develop accreditation in private hospitals through the Independent Hospitals Association of Ireland (IHAI) Accreditation Group. Although this work has gone on along-side that of the HSASG, it is planned that, once established, the statutory body will be the single body responsible for accreditation for the health system as a whole.

An explicit role for external evaluation is established in the new *Health Strategy* (Department of Health and Children, 2001a). Responsibility for external evaluation at the national level will be assigned to the newly established Health Information and Quality Authority, and at health board level to the monitoring and evaluation functions to be established within each health board.

2.6 Conclusions

This exploration of evaluation in the Irish health context suggests there are considerable benefits to be derived from the development of evaluation, in terms of enhancing accountability for effective decision making and to support organisational learning. The review also suggests that there are considerable differences in practice between boards, and in terms of how evaluation is organised. Although several different examples of evaluation are identified within boards, when compared to the potential role of evaluation outlined, it seems that current approaches fall short of their potential and efforts tend to be ad-hoc and poorly co-ordinated. The aim of the following chapters is to promote and support the development of evaluation in Irish health services. In Chapter Three, current thinking on evaluation is outlined. This is

followed in subsequent chapters with a review of evaluation methods and issues to be considered when undertaking an evaluation.

3
Current thinking on evaluation

3.1 Introduction

Having made the case in the previous chapter for the need to enhance evaluation in health service management in Ireland, this chapter seeks to explore some of the key themes in current thinking internationally on evaluation. The chapter begins by attempting to define evaluation and how it relates to management decision making and performance management. Other themes are also identified in the literature. These relate to designing evaluations of health services, building usefulness into evaluations to ensure that the findings get used, the role of evaluation in the learning organisation, the role of participation in designing and implementing evaluation, and other issues in the effectiveness of evaluations.

3.2 Definition of evaluation

The international forum – the Development Assistance Committee Working Party on Evaluation (DAC) (2000) – provides a very comprehensive definition of evaluation:

> *Evaluations* can be defined as systematic analytical studies conducted occasionally or on an ad hoc basis, to answer specific management questions about performance. Evaluations may assess and explain any of a variety of project or program performance issues, but are particularly well suited for dealing with more complex issues such as impact/attribution, sustainability, and relevance. They are often conducted by experts external to the project/program being evaluated, either from the inside or outside the agency. But some may be self-evaluations conducted by project/program managers and may have participation by stakeholders or beneficiary groups. They not only present evidence about results achieved (often obtained from performance measurement systems), but they interpret, explain, and make judgements about the performance in the light of the conditions that influence the outcomes/impacts. Moreover, evaluations typically provide recommendations for actions to be taken that flow from their analysis. In other words, evaluations may draw their findings from performance monitoring results data, but go well beyond simple presentations of results, by drawing conclusions,

> interpretations or judgements based on an understanding of the broader context, and then making recommendations. Without an understanding of the underlying causes of performance shortfalls, which evaluations can provide, management may take inappropriate actions. Moreover, evaluations can often draw broader lessons for future project designs and/or for formulation of agency policies and program strategies (DAC, 2000, p. 107).

As this definition suggests, evaluation can be as broad or as narrow as deemed appropriate and any number of approaches can be used to reflect the aims of the evaluation and the rationale for doing it. Evaluations can be comprehensive (all-embracing) or focused on specific concepts and will employ social research methods to gather valid, reliable evidence (Rossi and Freeman, 1993).

3.3 Issues in the design of evaluations of health services

3.3.1 The political nature of evaluation

A key issue in the evaluation of social programmes such as health services is that the programmes themselves are inherently political in nature, and as such will have multiple, diverse stakeholders, often with very different, and sometimes competing, priorities and expectations of programmes. Weiss and Greene suggest that this is to be expected:

> Social programs are manifest responses to priority individual and community needs and are themselves 'the creatures of political decisions. They [are] proposed, defined, debated, enacted, and funded through political processes, and in implementation they remain subject to [political] pressures – both supportive and hostile' (Weiss, 1987, p. 47). So program evaluation is integrally intertwined with political decision making about societal priorities, resource allocation, and power. 'By its very nature [evaluation] makes implicit political statements about such issues as the problematic nature of some programs and the unchallengeability of others' (Weiss, 1987, p. 48), (Greene, 1994, p. 531).

Further, Shaw (1997) suggests that the parameters for evaluation are themselves often political, as many goals set for organisations are set outside of the organisation, for example by government ministers. This, he suggests, influences their acceptability to

members of the organisation and accordingly their chances of being adopted and pursued. In addition, he suggests that, often

- goals imposed externally are too ambitious or too vague to be measured
- goals set are often replaced by conceptions of goals that arise empirically or from 'impinging reality'
- and targets set are not the real goals of the organisation but the result of a process of negotiation and conflict between groups within and outside the organisation.

Greene also highlights the difficulties that arise in evaluation from the fact that there are multiple, possibly competing, potential audiences – all stakeholders with vested interests in the programme being evaluated. So in constructing an evaluation, she suggests evaluators must 'negotiate whose questions will be addressed and whose interests will be served by their work'. In addition, she suggests that there is no one right way to do an evaluation because no one method can meet the range of diverse criteria that stakeholders will have or their evaluation questions. Further, she suggests that the choice of methods is more a political decision than a paradigmatic one and should 'devolve substantially' from the information needs of stakeholders.

Shaw (1997) argues that evaluations should be focused on the realisation of the real goals of the organisation rather than those set through negotiation, and that somehow evaluation needs to be decontextualised and de-politicised. He suggests that the current tendency in evaluation is to focus on criteria such as efficiency and economy and that this focus can hide quality loss in a service due to resource constraints, where the organisation may be performing well on what is assessed, but at the same time not manage to further any of the organisation's core aims. However, although this view points out that evaluation could be better focused on core aims, it ignores the fact that services do not operate in a political vacuum, and that the issue of context is very real. For example, many inequalities in health outcomes that governments are currently striving to address can be attributed to local factors (see for example, Fox and Benzeval, 1995).

3.3.2 Health services as social programmes

Pawson and Tilley (1997) argue that 'social programs are undeniably, unequivocally, unexceptionally *social systems* … (which comprise) … interplays of individual and institution, of agency and structure, and of micro and macro processes' (p. 57). On

that basis they describe five features of social programmes, which will need to be considered in evaluation:

- *Embeddedness* – the embeddedness of all human action within a wider range of social processes, social structures and social relations. As such, human action needs to be understood 'in terms of its location within different layers of social reality'.
- *Mechanisms* – that there are a range of underlying social mechanisms or hidden workings involved in outputs, such as people's choices and capacities, and the combination of the agency and structures. Thus the evaluator needs to go beneath the surface to look for explanations; and programme mechanisms need to be understood in terms that 1) reflect embeddedness; 2) account for how both macro and micro processes constitute a programme; and 3) that demonstrate 'how programme outputs follow from stakeholders' choices (reasoning) and their capacity (resources) to put these into practice'.
- *Contexts* – 'the relationship between causal mechanisms and their effects is not fixed but contingent' (p. 69). Social programmes are introduced into pre-existing social contexts, which are crucial when it comes to explaining the success or failure of a programme. Along with spatial or geographical contexts, the evaluator needs to consider the social rules, norms, values and interrelationships that impact on the efficacy of programme mechanisms. Pawson and Tilley claim that context is one of the greatest omissions in evaluation research.
- *Regularities* – the goal of Pawson and Tilley's 'realist explanation' is to explain social regularities, but in terms of how they relate to mechanisms and context. So, regularities such as rates, outcomes or patterns are related to the underlying mechanisms that generated them, and to 'how the workings of such mechanisms are contingent and conditional, and thus, are only fired in particular local, historical or institutional contexts' (p. 71).
- *Change* – relating to the unpredictability of social systems. This is so because people may desire to change patterns which they can or cannot effect or, because of imperfect knowledge of contextual conditions, they may take actions which have unanticipated consequences. Programmes or policies are essentially about change and trying to shift patterns towards more acceptable levels. Evaluators need to look at the potential for change '… it is not programs that 'work' but their ability to break into the existing chains of resources and reasoning which led to the 'problem' (p. 75).

3.4 Evaluation and performance measurement

Evaluation and performance measurement can be seen as two separate but complementary approaches to producing data to support management decisions. Performance measures are usually used to provide on-going feedback on the performance of government programmes. Ideally, this on-going feedback is based on a small number of key measures reflecting the purpose or objective of the programme, or to signal if the programme is worth continuing. Data produced can be evaluated against specific standards, against performance in similar organisations, or over time as a time series analysis. But there are limitations to performance measurement, '… measures do not normally speak for themselves; they have to be analysed and interpreted' (Bastoe, 1999, p. 102). Evaluation can be used to get beyond the measures, to assess performance in greater depth, and to put measures and indicators into a wider context (Bastoe, 1999). Evaluation can address some of the limitations of performance measurement by rendering judgement of programme quality, whereas programme quality is not well-captured or well represented in performance measurement (Greene, 1994). The use of evaluation in conjunction with performance measurement can help to overcome some of the 'pitfalls' of performance measurement alone, for example, by providing the know-how to determine causality in the measurement of outcomes (Davies, 1999).

The DAC (2000) also suggest that the two approaches should be viewed as inter-active and inter-dependent, where for example performance monitoring provides early warning of possible gaps in performance, which can be further explored through evaluation. In addition, evaluations can draw on the findings of performance measurement, and performance measurement can be developed further to address gaps identified during evaluations. Although the two approaches can be viewed as complementary, it is also useful to clarify the different situations in which either performance measurement or evaluation is more appropriate. Such a distinction is also provided in the DAC report:

> Both performance measurement and evaluation involve analysis and reporting on project/program performance and results. Evaluations are increasingly seen as more substantive, in-depth analytical efforts that can supplement the simpler forms of performance measurement analysis and reporting. Evaluations are undertaken when there is a need to better understand or explain

> project/program performance in its broader context or to generate recommendations for appropriate actions in light of that understanding. Moreover, whereas performance measurement analysis generally focuses on effectiveness (that is, whether results are being achieved as planned/targeted) and other simple performance measures, evaluation is better suited to address performance issues requiring more sophisticated methodologies (e.g. attribution) (DAC, 2000, p. 105).

All of this emphasises the complementary relationship between evaluation and performance measurement. Further, it is suggested that there may be an element of convergence between the two approaches in the latest approaches to management, and a redefinition of their respective roles. The DAC (2000) explores the relationship between evaluation and performance measurement in results-based management and how approaches have developed over recent years in donor agencies and in OECD countries in general. They suggest that before the rise of result-based management, traditional distinctions between monitoring (performance measurement) and evaluation were fairly clear cut but latterly there is considerable effort to clarify their relationships, both in terms of distinctivenes and complementarity.

3.5 Utilisation-focused evaluation

One important theme in the recent CPMR paper on performance measure in the health sector (see Butler, 2000, CPMR Discussion Paper No. 14) is how to ensure that information produced through performance measurement gets used by decision makers and is viewed by them as a key management tool. Issues identified in that research related to:

- The 'decision-usefulness' of data (Hyndman and Anderson, 1997) – the relevance and timeliness of data to those making decisions, the confidence that decision makers have in the quality of the data, and the comparability of data.
- Management style – that management values and managerial culture are receptive to using performance data as the basis for decision-making, that managers feel empowered to use the data and have the skills and competencies to analyse and interpret data effectively.

Along similar lines, a concern in the evaluation literature is that evaluation findings get used. Patton (1997) suggests that evaluations should be judged by their utility and

actual use and emphasises that the eventual impact of a study will be determined by what happens from the very beginning of the study (long before the final report is published). As such, utility should be a key consideration from the design stage to the end of an evaluation. Further, he suggests that in utilisation-focused evaluation, an explicit focus is required on actual primary intended users and their specific uses, from the range of possible audiences and potential uses. This makes utilisation-focused evaluation highly personal and situational and requires the evaluator to frame the evaluation in terms of the values of those who have responsibility to apply evaluation findings and to implement recommendations, rather than acting as an independent judge. The evaluator will be required to work closely with intended users in the design and conduct of the evaluation, while ensuring that the approach adopted meets with the requirements of good evaluation practice. Patton also suggests that the utility of the evaluation can also be enhanced through the active involvement of primary intended users in the evaluation process, through which users are more likely to develop understanding and ownership of the evaluation process and findings.

3.6 Evaluation as an idiosyncratic effort

Pollitt and O'Neill (1999) suggest that there are different basic purposes for evaluation and that these different purposes frequently demand different approaches to evaluation, different types of evaluation product and different relationships between evaluators, their audiences and their paymasters. In addition, Cronbach (1987) suggests that it is not appropriate to set out to produce a standard evaluation framework, or to cast evaluations in a 'single mould':

> Designing an evaluative investigation is an art. The design must be chosen afresh in each undertaking, and the choices to be made are almost innumerable. Each feature of a design offers particular advantages and entails particular sacrifices. Further merits and limitations come from the way various features combine (P. 5).

Several authors emphasise that evaluations should be designed around the key questions that are to be asked about a programme or a project. As such, each evaluation should represent an idiosyncratic effort to meet the needs of the stakeholder, or 'primary intended users' as proposed by Patton (1990). Evaluators need to be responsive to the context in which they are working, including the evaluation's 'policy sphere' (Rossi and Freeman, 1993):

> … evaluations need to be designed and implemented in ways that recognise the policy and program interests of the sponsors and stakeholders, and that will yield maximally useful information for decision makers given the political circumstances, program constraints, and available resources (Rossi and Freeman, 1993, p. 30).

Thus evaluations, and their purpose and intent, are different to scientific investigation, which strives to meet a set of research standards set by investigative peers. Nonetheless, there are several principles of good practice to be found in the literature that can be used for guidance in the design and implementation of an evaluation study.

3.7 Participative evaluation

Patton suggests that a list of specific primary intended users of the evaluation should be drawn up from the range of possible users and stakeholders, and the evaluation should be focused on their intended uses. Thus, a key element of the planning stage will be the identification of primary intended users, and working with them to agree on what is meant by evaluation and to engender commitment to both evaluation and use. It will involve the generation of meaningful evaluation questions with them – their questions. It will involve getting a commitment to what Patton refers to as 'reality testing' – where evaluation acts as a mechanism for finding out if what is supposed to be going on within an organisation is in fact going on. As Patton states:

> Some people would just as soon not be bothered dealing with programmatic or organizational reality. They've constructed their own comfortable worlds built on untested assumptions and unexamined beliefs. Evaluation is a threat to such people. Evaluators who ignore the threatening nature of reality testing and plow ahead with their data collection in the hope that knowledge will prevail are engaged in their own form of reality distortion. Utilization-focused evaluators, in contrast, work with intended evaluation users to help them to understand the value of reality testing and buy into the process, thereby reducing the threat of evaluation and resistance (conscious or unconscious) to evaluation use (Patton, 1997, p. 28).

Patton emphasises the importance of identifying the intended uses of the evaluation from the outset, and that will then guide the design of the evaluation, in much the same

way as the design of a programme is guided by the goals it wishes to achieve. Patton also suggests that the evaluation process itself can have a lot to do with how evaluation findings get used, and further, that being engaged in the evaluation process itself can influence change. For example, gaining insight into the logic of evaluation and 'thinking in terms of what's clear, specific, concrete, and observable' can be of value to those involved.

> Learning to see the world as an evaluator sees it often has a lasting impact on those who participate in an evaluation – an impact that can be greater and last longer than the findings that result from that same evaluation (Patton, 1997, p88).

Patton also highlights several other benefits from adopting a participative approach. The process of identifying explicit criteria for evaluation that are communicated to all can be used to clarify thinking and to enhance shared understandings on the organisation's mission and goals, and to increase shared understandings between managers and line staff. The logic and principles of evaluation can be used in negotiations between different parties with different perspectives and to focus on the attainment of results. Evaluation can also be used to 'give voice' to those who might normally be excluded and to improve dialogue between providers and service users.

However, Pollitt and O'Neill (1999) warn against trying to meet the needs of too many stakeholders at once:

> One way of avoiding the question of who an evaluation is for is to claim that it is for *all* stakeholders … An evaluation addressed to everyone tends to be heard – and owned – by no-one (p. 31).

In this light, they suggest joint evaluations are to be avoided, and although multi-stakeholder evaluations may work well in some circumstances, they may be inappropriate in others.

3.8 Conclusions

The review of current thinking and definitions of evaluation identifies some of the challenges that health services, by their very nature, pose when designing or undertaking an evaluation. Also identified is the need to treat each evaluation as an

idiosyncratic effort. It is suggested that the identification of research questions and selection of appropriate methods will need to be based on the particular purpose of the evaluation and also around meeting the needs of those who are most likely to use the findings or to require them for decision making. The range of stakeholders that is likely to be involved in health programmes/projects and the political nature of programmes can present particular challenges when it comes to identifying the most relevant evaluation questions to be addressed. In the following chapter, some of the more practical considerations in the design of an evaluation are outlined.

4
Timing and focusing the evaluation

4.1 Introduction

Following the review in the previous chapter of definitions and current themes in evaluation, this chapter focuses specifically on issues in the design of evaluation studies. While doing so, it aims to guide would-be evaluators, in a practical way, through the key decisions to be made in designing a sound evaluation study. It explores the principles of robust evaluation, the appropriate purposes for evaluation, different approaches to evaluation and the timing of evaluation. Programme theory is then explored as a framework for the identification of the key elements of an evaluation.

4.2 Timing the evaluation: When to evaluate?

4.2.1 At what stage of the programme management cycle?

Rist (1990) distinguishes programme evaluation – retrospective assessment of policies or programmes – from policy analysis, which prospectively seeks to inform decisions yet to be made. He makes the point that whether the approach is prospective or retrospective will influence the kinds of questions addressed. The emphasis of policy analysis is on the likely effects, whereas the emphasis in evaluation is on the actual effects. Rist (1990) suggests that rather than see the two approaches as very distinct entities, it is more useful to view them as 'two stages of an interactive process: decisions are made, information is gathered about the effects of those decisions, further decisions are made with data available on the results of previous decisions, etc. etc. etc.' (p. 4). Further, he suggests that current thinking is that programme evaluation can encompass all of the various stages of the life cycle of the programme or policy.

To illustrate the point, Rist (1990) identifies at least three distinct phases in government policy life cycles where retrospective assessment can take place: 1) policy creation or formulation, 2) policy implementation and 3) policy outcome or impact (citing Chelimsky, 1985). So, the focus of evaluation at the first stage will be front-end analysis and evaluability assessment. Process evaluation will feature at the implementation stage, and the focus of evaluation in the third stage will be on effectiveness or impact evaluation. In addition, programme and problem monitoring is on-going during the life of the programme and findings from evaluations may be used

along with those from other evaluations for meta-evaluations, aimed at generating knowledge and learning in relation to policy issues in general.

Evaluation can also be described as *summative* or *formative*. The focus of summative evaluation is on final or once-off assessment of outcomes or impact. Summative evaluations:

> … are often carried out when the programme has been in place for some time (*ex post* evaluation) to study its effectiveness and judge its overall value. These evaluations are typically used to assist in allocating resources or enhancing accountability. The clients are usually external, such as politicians and other decision makers. The objectivity and overall reliability of findings is considered important, and external evaluators are therefore often commissioned to conduct an evaluation. Questions of outcomes and overall relevance of the programme are expected to be addressed (PUMA/PAC, 1999, p. 12).

Formative evaluation is concerned with providing interim feedback on performance, or further insight, on the basis of which changes are made to programmes – 'the *formative* study … regards the programme as fluid and seeks ways to better it' (Cronbach, 1987, p. 11). In this way it contributes to the learning process, its purpose being to 'support and improve the management, implementation and development of the programme' (PUMA/PAC, 1999, p12). Formative evaluation may be undertaken periodically or on an on-going basis over the life of the programme, and typically focuses on processes and intermediate outcomes. Evaluators and clients are usually internal, and the applicability of results is of more concern than the objectivity of findings (PUMA/PAC, 1999).

However, Cronbach (1987) suggests the distinction between summative and formative evaluation is a false division, and that an evaluation that seeks to establish whether a programme was successful or not can also be used formatively, for example, to identify the factors that contributed to the programme's success or failure. These findings can then be drawn on to weigh up alternative approaches where a programme has failed, or to plan similar programmes in the future.

4.2.2 The evaluability of the programme

Further emphasising the importance of the timing of an evaluation, Cooksy et al (2001) highlight the influence that timing can have on the accuracy of evaluation results. They suggest doing an 'evaluability assessment … to discover whether a program is ready to be evaluated so that costly summative evaluations are not conducted prematurely; that is, before there are clear and logical links between a program's resources, activities and outcomes' (p. 120) and 'to avoid costly evaluations in situations where it is unlikely that program effects will be observed' (p. 121). Burt et al (1997) suggest evaluation planners should consider if the project or programme is yet operating at its full scope and if it is stable, before embarking on an evaluation.

Harrell et al (1996, p. 6) identify a series of questions to be considered when assessing whether a programme is ready for evaluation and whether programme evaluation is justified, feasible and likely to provide useful information:

- Are the programme goals realistic and the intervention strategies well-grounded in theory and/or prior evidence?
- What kinds of data will be needed, from what number of subjects, and what data are likely to be readily available? Are resources available to collect new data if crucial data needs are not met with existing data?
- Are there adequate financial resources, time, expertise and community and government support? Are there any factors that constrain access to these resources?
- Can the evaluation be achieved in a timeframe that will allow the findings to be useful in making programme or policy decisions?
- Does evaluation information already exist on the same or closely related intervention and, if so, what is the value of a new evaluation?
- To what extent could the findings be used to assess whether the programme should be extended to other settings or areas, that is, how generalisable are the findings likely to be?

Weidman et al (1975) suggest that an evaluation should not go ahead unless it can satisfy three conditions:

- 'Those who will use the evaluation results must agree on definitions of the program's or project's activities, the conditions it is supposed to change and the kinds of outcomes expected.

- The key assumptions on which the program is based must be stated in forms that can be tested objectively.
- Program or project managers must spell out clearly defined use for evaluation information in making a decision or in initiating administrative action' (p. 12).

4.2.3 Providing timely information for the decision- making process

As discussed in Chapter Three, evaluation needs to be focused on meeting the needs of the relevant decision makers. This will include providing information that is relevant, reliable and available in time to be used as the basis for decision making – 'evaluation is most useful when it is timely' (Coldren et al, 1989, p. 17). In planning the timing of an evaluation, the evaluator will need to consider the time-lag involved in the conduct of the study and how this will enable the study to dove-tail with key steps in the decision-making process. However, the evaluator must also ensure that the time elapsed will not compromise the currency of the data and findings produced.

4.2.4 Which programmes to evaluate, and when

On a similar theme, Coldren et al (1989), suggesting that in any one year not all programmes will or can be evaluated, identify a series of considerations in deciding which programmes to evaluate. The evaluator will need to consider which of the range of programmes are accessible to evaluation. For example, geographical location or having to cover multiple jurisdictions may make evaluation too costly or difficult to organise. Secondly, which programmes will be in operation for some time into the future? – evaluation takes time. In addition, a programme needs to be in place long enough for the impact to be felt. Thirdly, how expensive are programmes? Coldren et al suggest more expensive programmes usually justify or warrant evaluation and there will be more demand for information about them. Fourthly, which programmes are controversial? It is suggested that programmes may be controversial in a negative or positive sense and in some cases this will necessitate evaluation whereas in other cases it will work against it. However, Coldren et al also suggest that this criterion should not exclude programmes that have been in operation for years under the assumption that they work. They suggest such assumptions should be questioned and evaluated. The fifth point is that evaluators should consider which programmes are identified as priorities by their office or other significant officials.

4.3 Focusing the evaluation: what to evaluate?

A key element of evaluation design is to identify a set of relevant and pertinent questions from the range of possible questions and to decide on priorities and where the emphases should be in the evaluation. As well as substantive considerations, the selection of questions will need to be guided by practical and political considerations (Cronbach, 1987). The evaluator must recognise that, within the likely time and resource constraints of an evaluation, focusing in on a particular question is at the expense of attention to other possible questions.

> He must recognize the opportunities what will be forgone … and he must determine whether, on balance, the credibility of the study will be increased or decreased. The cost of answering one question well must be weighed against the cost of leaving other questions unanswered (Cronbach, 1987, p. 7).

Harrell et al (1996, p. 2) suggest that 'the design of any evaluation begins by defining the audience for the evaluation findings, what they need to know and when'. They suggest decision makers want to know what programmes accomplish, what they cost and how they should be operated to achieve maximum cost-effectiveness. This may require the evaluation to focus on impact, performance monitoring, process or cost. A comprehensive evaluation would include all four activities.

Rist (1990) outlines the types of questions to be asked in evaluation for each of the three stages of policy and programme evaluation identified by Chelimsky (1985).

At the *policy formulation* stage, the key questions to be addressed include:

- *the nature of the problem* – the exact contours of the issue and if it is larger or smaller than before, or about the same, and what is known about how it has changed; and how well can the condition be defined and measured
- *what has taken place previously in response to the condition or problem*? – projects initiated, how long they lasted and their success; what level of funding and staff were required? the receptiveness of populations or organisations to the initiatives; did previous initiatives address the same problem or issue, and what additional efforts are required in this case?
- *what is known about the impacts of previous effort* – time-frames before one sees evidence of impact.

At the policy or programme implementation stage, questions might relate to:

- *the implementation process* – costs; the degree to which the programme is reaching its targeted audience; whether the services set out to be delivered were, in fact, delivered; aspects of the programme that are or are not operational; the similarity of programmes across various sites
- *on-going monitoring of the problem or situation that prompted the policy or programme in the first place* – whether it has improved, worsened or remained static; whether it still applies to the same population as before, or has spread or contracted; whether the aims of the programme still match the condition
- *efforts made organisationally to respond* – conceptualisation of the situation and the organisation's response; the expertise and interest shown by staff; whether the organisational structure reflects organisational goals; 'what means exist to decide among competing demands'; what kinds of interactive information or feedback loops are in place to assist managers in their efforts to move the programme towards stated objectives.

The third stage is concerned with the *evaluation of outcomes* and the *impact* of the policy or programme and *accountability*. The timing of this stage is very important as programmes and policies need time to 'mature' before they can be expected to yield results. Key questions at this stage of evaluation include:

- *what the programme did or did not accomplish* – objectives met; the effectiveness of implementation strategies in moving the programme or policy in the desired direction; mid-course corrections made to keep the programme on track; confidence in measures used to determine programme influence
- *changes in the programme/policy or condition* – whether the condition has changed or not; proportion of change that can be attributed to the programme interventions; whether further action is necessary to bring about change
- *accountability* – management supervision, attention and procedure; did the programme stay within budget? were personnel matters handled appropriately and legally? are records complete and available? are attrition levels reasonable? is all equipment accounted for and working?

More specifically, relating to evaluation in the CSF context, the CSF Evaluation Unit (1999) suggests evaluation should address five key questions:

- the *rationale for* a particular *intervention*

- the *continued relevance* of an intervention
- *effectiveness* – the extent to which objectives have been achieved, by comparing outputs and/or expenditure with associated targets
- *efficiency* – could the same outputs or benefits of the intervention be achieved at a lower cost?
- the *impact* of the intervention – net effects/changes in socio-economic situations that can be attributed to the programme; causality between the intervention and changes – could it have occurred anyway?

These five questions can also be linked to the criteria identified by Rist (1990) in the previous section (see Figure 4.1).

Figure 4.1Key questions in evaluation at each stage of the programme life cycle

	Rationale for intervention	*Continued relevance*	*Effectiveness*	*Efficiency*	*Impact*
Policy/programme formulation	3				
Implementation		3	3	3	
Outcome/ impact			3	3	3

4.4 Programme theory and the programme logic model

Programme theory can be used to identify key elements of a programme for evaluation, by relating a programme to the intended outputs and outcomes, potential impacts and the activities and inputs required to achieve these outputs and outcomes. The approach is based on the premise that programmes or projects are based on explicit or implicit theory about how and why a project will work. The evaluation is then designed around collecting and analysing data to track the unfolding of theory-related assumptions at several 'microsteps' over the course of the programme. The thinking is that in this way, if events do not work out as expected, there can be a certain confidence about where, how and why the breakdown occurred (Baker, 2000).

Cooksy et al (2001) suggest that the theory-driven approach to evaluation (such as programme theory) is more likely to focus on programme effectiveness than traditional method-driven approaches. They also outline the two major types of theory-driven approaches identified by Chen (1990). The first is normative evaluation, which explores inconsistencies between a prescriptive theory of what the programme should be against the data on the programme in operation. The second approach is causative evaluation, which explores programme impact by focusing on the causal relationships underlying a programme and the causal mechanisms associated with programme effects.

Baker (2000) identifies two key benefits to theory-based approaches. First, early indicators of programme effectiveness are provided during project implementation. This can then be used as the basis for corrective action. Second, it is possible to explain how and why effects occurred by following the sequence of stages and tracking the microsteps from inputs to outcomes. Drawing on the work of Weiss (1998), Baker identifies four potential shortcomings: difficulty identifying assumptions and theories because they can be inherently complex; problems in measuring each step when the right instruments or data are not available; difficulty testing effects because theory statements are too broad or loosely constructed; and, problems of interpretation that make results difficult to generalise.

The programme logic model is one approach to identifying the theory of a programme and is an integrative framework for the design and analysis of evaluations using qualitative and quantitative methods. Cooksy et al (2001) define logic models generally as 'flow charts that display a sequence of logical steps in program implementation and the achievement of desired outcomes'. They also suggest that logic models are not rigid and can be easily adapted to different programme theories. Millar et al (2000) describe logic models as:

> … word or pictorial depictions of real-life events/processes that depict graphically the underlying assumptions or bases upon which the undertaking of one activity is expected to lead to the occurrence of another activity or event. It involves 'modelling or simulating' real-life in such a way that the fundamental 'logic' becomes apparent. Logic models show causal relationships as they relate to one another – a systems approach to portraying the path towards a desired reality (p. 73).

Cooksy et al (2001) identify the key elements of the programme logic approach:

- programme theory guides an evaluation by identifying key programme elements and articulating how these elements are expected to relate to each other
- within the framework, data collection is designed to measure the extent and nature of each element's occurrence
- data is collected and then analysed within the framework
- data collected from different sources is triangulated
- the patterns of relationships found in the data are compared with the patterns of relationships articulated in the programme theory.

As an integrative framework, Crooksy et al (2001) suggest that the logic model enables data from different sources and collection methods to be organised by programme element and examined for consistency. Further, they suggest that the model forces the evaluator to look at patterns in the data across sources and in relation to the intended sequence of events. In this way, the process may help to prevent conclusions being drawn on evidence that is incomplete and enables conflicting data to be examined. In addition, the model enables antecedents and consequences for each element to be examined so that evidence about each element is interpreted in the light of its expected relationships, rather than in isolation, and so that the success of implementation can be examined in terms of initial, intermediate and longer-term outcomes. An example of a programme logic model for the Dental Treatment Services Scheme (DTSS) is provided in Figure 4.2. Although the author has developed this model solely as a worked example, it draws on some very real service-based information, thinking and planning.

Figure 4.2 A programme logic model for the DTSS

Inputs	Activities	Short-term outcomes	Intermediate outcomes	Long-term outcomes

Planning and strategies
 	
Health Strategy:
• Principles – equity, accountability, quality
 	
DTSS –
• Eligibility criteria
  o 65 years or over
  o 16-34 year olds
  o full dentures
• Funding mechanisms
• Access to private practitioners
• Objectives
• Performance indicators (post-1999)
 	
Dental health action plan –
• Objectives and goals for dental services as a whole
 	
Resources
• £13.4m
• 858 private dental practitioners (1998)
• Health board dentists
	DTSS schemes
	• Emergency treatment
	• Routine treatment
	• Full denture scheme
	
	Extension of scheme to 35-64 year olds
		• Reduced demand for dentures
		• Increased ratio of restorations to extractions
		• Increased number of private dental practitioners in scheme
		• Allocated funding is spent as intended
		• Increased patient uptake rates as a proportion of those eligible
		• Increased number of dentists enrolling in scheme
		
		By the year 2000*:
		• Average number of missing teeth for 16-24 year olds is 1 (2) or less
		• No more than 2% of 35 to 44 year olds with no natural teeth
		• No more than 42% of persons 65 years or older with no natural teeth
			• Improved oral health for adult medical card holders
			• Reduced discrepancies between health boards, and urban /rural areas in number of contracting dentists
			• Increases in funding targeted at those who need it most
			• Equitable distribution of funding
				• Reduced equity gap between adult medical card holders and general population
				• Increased level of oral health in general population*
				• Equity of funding, provision and access

	• Remuneration of dentists to ensure quality work • Quality initiatives	• Scheme is popular with those who use it • Acceptability to those who use it		• The provision of a quality service
		• Reduced average cost per patient • Reduce ratio of expenditure on routine to emergency treatment • Reduce variations in costs between health boards		• Cost-effective provision of services
	• Examining dentists • Reform DTSS contact • Probity and investigations officer GMS(P)B • New forms and validations, • Restructuring public dental services • Research into developing accountability mechanisms	• Effective arrangements for managing DTSS scheme • Effective mechanisms for monitoring dentists and taking action where required		• Enhanced accountability

** From the dental health action plan – for further details see Appendix One.*

A summary of the DTSS review is provided in Appendix One. The DTSS was established in 1994 and through it three particular types of treatment are provided to eligible medical card holders – emergency treatment, routine treatment and the provision of full dentures. The DTSS review report provides considerable background information to support the review, including: material from the *Health Strategy* and the related dental health action; the objectives of the DTSS; the Dental Treatment Benefits Scheme and how it relates to the DTSS; national, regional and international trends in oral health; and the resources allocated to the DTSS. For the purposes of developing a programme logic model for dental services, these are seen as inputs to the DTSS. Also included in inputs is the selection of those eligible or targeted by the DTSS – eligible medical card holders. At the time of the review they were adult medical card holders between sixteen and thirty-four years old and those aged sixty-five and over. It is also stated in the review that there are plans to extend eligibility to include also persons aged between thirty-four and sixty-four years old. The long-term outcomes identified relate to reducing inequalities in oral health, equity, accountability, quality and cost-effectiveness. From these, several related intermediate and short-term outcomes are identified and these are linked to specific activities to achieve them. This fulfils steps one and two in the model for developing a programme logic model identified by Burt et al (1997) (see Figure 4.3). Steps three and four then involve: identifying factors, relating to the backgrounds of those involved, that would influence outcomes independently; identifying factors that would affect participation in the programme; and identifying events or factors happening during or after the project that could influence how or whether the project accomplishes its objectives. For example, in the DTSS review a model was developed to predict uptake rates. In addition, considerable background information is provided on the medical card population, including epidemiological information. The extension of the scheme to sixteen to sixty-four year olds is an example of a factor that could well influence the outcomes of the DTSS, both in terms of outcomes and in terms of resources required.

Burt et al (1997) suggest the programme logic model can be used in evaluation in the following ways:

- *in impact evaluations* – the model can be used to identify how, and for whom, project activities are expected to attain specific goals.
- *in process evaluation* – the model can be used to identify expectations about how the project should work. This can be used to assess deviations in practice, why they have occurred and how they might affect achievement of goals.

- *in on-going evaluation/performance monitoring* – the model can be used to focus in on the kind of indicators that are appropriate for target populations, communities or time periods. This can help to identify gaps in planning or differences between staff and other stakeholders in assumptions about how the project will operate and the various responsibilities of participants.

The University of Ottawa (1997) suggests that for evaluation purposes a programme logic model will:

- summarise the key elements of a programme
- explain the rationale behind programme activities
- clarify the difference between activities and intended outcomes
- show cause and effect relationships between activities and outcomes
- help to identify the critical questions for the evaluation
- provide the opportunity for programme stakeholders to discuss the programme and agree upon its description.

Burt et al (1997) identify a series of steps to be taken in constructing a programme logic model. These are outlined in Figure 4.3.

Figure 4.3 Key steps towards developing a programme logic model

Step 1.
Make a list of project goals
These will include both immediate outcomes and longer-term outcomes

Step 2.
Make a list of project services and activities
These will include all services and activities to be provided by the programme

Step 3.
Make a list of all the background characteristics (factors) of the people involved
The problems and resources available to individuals served by the programme may influence the outcomes independently of the programme and may also influence what services they receive, e.g. language, employment, education, income.

Step 4.
Make a list of all the events or factors happening during or after the project activities that could influence how or whether the project accomplishes its objectives
This might include things that may affect the desired outcomes either in combination with the programme services, or in addition to them, e.g. support from family/ friends, availability of needed services, court/police response.

4.5 Conclusions

The focus of this chapter has been on the key issues to be considered in the design of an evaluation and the selection of the evaluation questions. The programme logic model is presented as one approach to enable this to be achieved. Chapter Five builds on this by outlining the range of methods that can be used to enable different evaluation questions to be addressed.

5
Evaluation techniques

5.1 Introduction

In Chapter Three it was suggested that each evaluation should be designed anew in order to identify and answer questions that are relevant and timely and that will provide useful information to decision makers. However, the selection of appropriate methods and techniques in evaluation is also an essential consideration in the design of evaluation, and this is a key theme in this chapter. The chapter begins by introducing the notion of methodological appropriateness in evaluation, before moving on to outline a range of evaluation techniques and the types of questions that they can help to answer.

5.2 Evaluation paradigms

Four main evaluation approaches (or paradigms) can be identified in the literature including:

- *Post-positivism/quantitative approaches* – using quantitative methods such as experiments, quasi-experiments, systems analysis and causal modelling. Greene (1994) suggests that post-positivism remains dominant amongst evaluators and evaluation audiences, with a particular emphasis on the use of control groups.
- *Pragmatic evaluation* – Greene (1994) claims pragmatism, which is particularly orientated to decision making and management, arose largely in response to the 'failure of experimental science to provide timely and useful information for program decision making' (p.52). Methods used include structured and unstructured surveys, questionnaires, interviews, observations, focused on which parts of the programme work well and which parts need improvement, how effective the programme is in relation to the organisation's goals and in relation to the needs of beneficiaries.
- *Economic evaluation* – often conducted 'slightly apart from evaluation in general' and focusing on issues such as cost-benefit analysis and cost-effectiveness assessment.
- *Interpretive/qualitative approaches* – this approach is based on the epistemological view that experimental and economic evaluation are not appropriate for evaluation

in a society that is socially constructed. Instead the use of qualitative methods, such as case studies, interviews, observations and documentary review are advocated. Interpretivism is concerned with how the programme is experienced by various stakeholders and is 'part of the responsive tradition in program evaluation' which 'seek to enhance contextualized program understanding for stakeholders closest to the program … and thereby promote values of pluralism as well as forge direct channels to program improvement' (Greene, 1994, p.533).

5.3 Methodological appropriateness in evaluation

The soundness of the methods selected for an evaluation will have a major effect on the credibility of the evaluation, and accordingly, the extent to which the findings are likely to be used in decision making (PUMA/PAC, 1999). The emergence of distinct paradigms for evaluation underlines differences in perceptions about what constitutes a credible evaluation. Cronbach (1987) refers to the perceived 'conflicting ideals' between 'scientific' (quantitative) and 'humanistic' (qualitative) approaches. Advocates of the scientific ideal would equate a 'quality' evaluation with randomised experiment and propose that non-experimental approaches should only be considered in situations where political infeasibility, costs and practical difficulties of random assignment rule out experimental approaches. He suggests that humanists, on the other hand, would find experiments unacceptable and propose that programmes already in place should be studied, that observations should be opportunistic and responsive to the local scene rather than being pre-structured, and that assignment to groups (intervention and control groups) should occur naturally rather than being made for the sake of research. Humanists also believe that different questions should be asked of different programmes and that the benefits of a programme should be described rather than being reduced to a quantity.

Cronbach suggests that more recently the two ideals have been reconciled and the value of multiple methods that are deliberately complementary has become accepted. Nonetheless, he emphasises that 'eclectic tolerance' is not enough to guide the evaluator and careful decisions must still be made about the most appropriate methods to be used in an evaluation. He suggests that evaluations are different to social research in so far as evaluations fit into a different institutional and political context and key considerations in selecting methods are the 'context, purpose and expected payoff' of the evaluation.

As advocates for mixed method approaches to evaluation, the US National Science Foundation (NSF) (1997) suggests that the reliance on quantitative techniques alone can mean that evaluators miss 'important parts of a story' and that 'experienced evaluators have found that often the best results are achieved through the use of mixed method evaluations, which combine both quantitative and qualitative techniques' (p.1). They suggest that quantitative and qualitative methods have different strengths, weaknesses and requirements which should be considered when deciding which methods to use.

Current thinking in the evaluation community is that *'methodological appropriateness'* (should be) the primary criterion for judging methodological quality' rather than allegiance to any particular philosophical paradigm (Patton, 1990, pp. 38-39). The purpose of evaluation must be to produce 'maximally useful' evidence and evaluation methods might be suitable in one situation but not appropriate in the next (Cronbach, 1987). Thus, evaluation methods must be targeted at meeting the information needs of the identified evaluation audiences (Greene, 1994). Critical decisions need to be made at the planning stage that are informed by the relevance of the criteria selected for evaluation, the capacity of the evaluation to produce adequate evidence, and the extent to which the evaluation produces reliable data and results in clear findings. Pragmatic considerations have to be balanced with epistemological ones and there may be a trade-off between methodological rigour and the utility of evaluation (PUMA/PAC, 1999).

5.4 Selecting evaluation methods

In the previous chapter, five key types of evaluation questions were identified for a policy, programme, or project: 1) the rationale for intervention 2) the continued relevance, 3) effectiveness, 4) efficiency and 5) impact. Once agreement has been reached on the purpose of the evaluation and the questions to be addressed, one or more evaluation methods will need to be selected to provide the data required for the evaluation, in order to provide 'maximally useful data' for decision makers, as emphasised by Cronbach (1987). As previously suggested, the context in which a policy, programme or project exists will also be an important consideration in the selection of methods.

In advance of a discussion of the range of methods that can be used in evaluation, possible choices of methods to address the five evaluation questions identified by the

CSF (1999) are provided in Figure 5.1. This suggests that for each evaluation question, one or more methods may be appropriately selected.

Figure 5.1 Evaluation questions and techniques

Evaluation techniques		Evaluation questions				
		Rationale for intervention	**Continued relevance**	**Effectiveness**	**Efficiency**	**Impact**
Scientific / quantitative	Experimental			3		3
	Quasi-experimental			3		3
	Non-experimental			3		3
Survey / Questionnaire			3	3		3
Economic analysis		3	3	3	3	3
Humanistic /naturalistic /qualitative	Observation		3	3	3	
	Documentary analysis	3	3	3	3	
	Interviews	3	3	3	3	3
	Focus groups	3	3	3	3	3
	Case studies	3	3	3	3	3

5.4.1 Scientific approaches

Harrell et al (1996) suggest that experimental, quasi-experimental, and non-experimental designs, all of which compare programme outcomes with some measure of what would have happened without the programme, can be used for impact evaluation. They suggest experimental approaches are the most powerful and produce the strongest evidence, but in cases where they are not possible, one of the other two methods could be used.

They describe each of the approaches. The key elements of experimental designs are that individuals are assigned at random to one or more groups prior to the beginning of the intervention. Individuals within each group may receive an intervention (intervention or treatment groups) or not (control group). After a period of time the outcomes of the groups are compared. This may involve comparing the outcomes of groups receiving various interventions or comparing those receiving interventions with a control group. In impact evaluation, the assessment of outcomes of an intervention, or interventions, are compared with those of a control group. There are various approaches to randomly assigning individuals to groups. Harrell et al (1996) identify a number of limitations to experimental designs, for example the difficulty assigning

individuals randomly to groups in real life situations, ethical considerations and difficulties collecting comparable data from control groups where contact with programme staff is not as continuous as with those receiving treatment. In addition, experimental evaluation approaches can be very costly.

The critical difference in quasi-experimental evaluations is that assignment to groups is not random. This makes them relatively easier and less expensive to conduct than experimental designs. Comparison groups are made up of members of the target population that are as similar as possible to treatment recipients, based on factors identified to affect selected outcomes. In some cases, individuals in different groups may be 'matched' for certain characteristics. In addition, pre-intervention data may be recorded as a baseline to be used in the analysis of within-individual changes occurring over the programme period. Multi-variate analysis techniques are used to control for remaining differences between the groups.

In terms of the limitations of quasi-experimental approaches, Harrell et al (1996) identify three threats to validity:

1. *Maturation* – the possibility that age-related processes will contribute to outcomes independently of the programme intervention. Matching groups for age may help to overcome this.
2. *History* – the risk that unrelated events may affect outcomes. To avoid this, such events could be identified and controlled in the analysis.
3. *Selection* – the factors that determine who receives services, for example, 'program participants may receive services because they are more motivated, skilful, or socially well connected than non-participants' (p. 10).

Harrell et al suggest that controlling for such threats to validity can be a particular challenge.

In non-experimental designs, impact is assessed on the basis of aggregate data or data on each individual within a group. Aggregate data can be used in *before and after comparisons* of programme participants, where the assessment of programme impact is inferred from the differences in the average score of a group before and after the services (Burt et al, 1997). Or, aggregate data can be used for repeated measures of outcomes before and after the programme – *time series* designs. The time series study is an extension of the before and after study. Several measures are taken before and

after the intervention. The impact of the programme is assessed based on whether a statistically significant change in the outcome occurs at or shortly after the time of the intervention (Burt et al, 1997).

Data at the individual level can be used for *panel studies*, where repeated measurements of outcomes are taken of individuals within the same group of participants in a programme. Data is collected from individuals at the same intervals from the time they enter the programme. In *cross-sectional comparisons,* data at the individual level is collected through surveys of groups of participants after programme completion. Data collected is used to estimate correlations between the outcomes experienced by individuals and the duration, type, and intensity of the services they received.

The limitations identified by Harrell et al (1996) to non-experimental designs are that, because non-participants are not included, the full impact of the programme cannot be estimated, i.e. compared to no service at all. In addition, the design cannot control for the effects of other factors outside of the programme. Furthermore, generalisability of the findings to other settings or groups is limited, because not enough is known about how the outcomes of participants compare to those of non-participants.

5.4.2 Survey methods

The purpose of a survey is to collect a structured or systematic set of data about the same variables from a set of cases. Questionnaires are used widely, but other methods such as observation and in-depth interviews can also be used to collect the structured data. Data for each case, for each of the variables, is arranged in a data matrix (a variable by case matrix). Analysis of the data can be used to describe the characteristics of the set of cases, to compare cases, or to search for causal links between variables. In contrast to the experimental approach, the survey approach seeks naturally occurring variation within a sample as the basis for analysis. De Vaus (1996) provides a comprehensive source on the survey approach, including how to develop indicators for concepts, find a sample, construct questionnaires and conduct analysis.

Surveys are used widely in health service evaluation to obtain information about patient satisfaction. A well-designed patient satisfaction survey can provide useful feedback on the general sense of satisfaction, particular service areas or issues to be addressed and ideas about possible solutions to issues raised. However, they are

limited in the extent to which they adequately establish causal connections, provide insight into the meanings of social action and explore individual beliefs. In addition, they are often perceived to assume that human action is determined by external forces, and to be a 'sterile, ritualistic and rigid model of science centred around hypothesis testing and significance tests' (de Vaus, p. 8).

5.4.3 Economic analysis

Drummond et al (1997) define economic evaluation as the comparative analysis of alternative courses of action in terms of both their costs and consequences. Economic evaluation can be particularly useful where projects are 'non-routine and therefore relatively expensive; dependant on specialist expertise; driven by intellectual curiosity; and quite rightly, having to justify itself continually in an increasingly sceptical world' (Cairns, 1998, p.52, quoting Hutton, 1994). Four types of economic evaluation can be identified: cost-effectiveness analysis, cost-minimisation analysis, cost-benefit analysis and cost-utility analysis.

1. In *cost-effectiveness analysis*, costs are related to a single, common effect, which may differ in magnitude between the alternative programmes. Results of cost-effectiveness analysis may be written in terms of cost per unit of effect (cost per life-year gained) or in terms of effects per unit of cost (life-years gained per euro spent). Effects per unit of cost is useful when dealing with budget constraints, as long as alternatives being compared are of similar scale. For example, the costs and consequences of hospital dialysis can be compared with kidney transplant in terms of prolongation of life after renal failure. The alternatives are compared by calculating the prolongation and comparing cost per unit of effect (cost per life-year gained) (Drummond et al, 1997). Cost-effectiveness analysis can be used on any alternatives that have a common effect (e.g. kidney transplant compared to heart surgery if the common effect of interest is life-years saved).
2. *Cost-minimisation analysis* aims to identify the least cost. For example, a comparison of the effectiveness of two minor surgery programmes shows that data differ in no other respect except that one involves hospital admission for at least one night while the other, a day surgery programme, does not. Focusing on the common outcome of interest – that the operations were successfully completed – it is identified that this could be achieved to the same degree in either programme, but at different costs (Drummond et al, 1997). Drummond et al suggest that cost-minimisation analysis is really a special form of cost-effectiveness analysis where the consequences of alternative treatments being compared turn out to be the same.

3. An economic evaluation which measures both the costs and consequences of alternatives is called a *cost-benefit analysis*. The results may be stated as a ratio of euro costs to euro benefits, or as a sum (perhaps negative) stating the net benefit (loss) of one programme over another. Drummond et al (1997) suggest that one of the advantages of cost-benefit analysis is that it provides information on the absolute benefit of programmes and additional information on their relative performance.
4. An economic evaluation where utility is used as a measure of the value of programme effects is called *cost-utility analysis*. Drummond et al (1997) suggest that utility is used as a measure of value by analysts who are reticent about assigning values to benefits in monetary terms (euros, dollars). They suggest the term utility is used in a general sense to refer to the preferences individuals or society may have for any particular set of health outcomes. For example, the use of quality-adjusted life-years (QALYs) may be used as a generic outcome measure for comparison of costs and outcomes in different programmes.

Each category of economic evaluation measures the costs and consequences of programmes or services to varying degrees. In relation to which form of analysis is most appropriate, Drummond et al (1997) suggest that, in addition to the question of which is the most appropriate approach for the problem being tackled, the institutional framework, the practical measurement challenges and the perspective the analyst takes on the role of economic evaluation will also be relevant considerations. They suggest that at the outset of a study an analyst may not be able to predict what form the final analysis may take, as this may depend on the results of an associated clinical evaluation (e.g. it may not be known in advance that a clinical evaluation will show two treatments to be medically the same, resulting in a cost-effectiveness analysis being transformed into a cost-minimisation analysis). Further, they suggest that the different forms of economic evaluation are used together on occasions.

5.4.4 Qualitative methods

Observation

Observational methods can be used to gather first-hand data on programmes, processes or behaviours, to develop a holistic perspective of how activities and operations fit into or are impacted by a sequence of events, or to learn about things that staff may be unaware of or unable to discuss in an interview or focus group (NSF, 1997). The NSF also suggest that observations can be a good way to find information about:

- the setting or environment in which a project takes place
- the ways in which all actors interact and behave towards each other
- what goes on in the life of a project, what the various actors do and how funds are allocated
- how staff and participants understand their experiences, described in their own language or jargon
- non-verbal cues about what is happening on a project
- what is not occurring (although the expectation is that it should be occurring as planned by the project team) or noting the absence of some other noteworthy activity/factor.

Information is recorded in field notes and the process of recording events can be enhanced by the use of a tape recorder, camera, video camera or laptop computer.

Observation can be described as *participant* observation (where the observer will also take part in the activities observed) or *non-participant* (where the observer is a 'fly-on-the-wall'). Observation can also be *overt* (where participants are aware that observation is taking place) or *covert* (where they are not).

One of the biggest criticisms of observation is its potential for distortion. What we choose to observe in the research situation will be determined by our understanding of what we are studying and will be influenced by the baggage that we bring with us to the research situation (our philosophical perceptions of research and our perceptions of life in general). Endacott (1994) identifies potential sources of distortion a) in the planning stage – poorly defined terms, unclear research role or role inappropriate, biased sample selection, and a poorly designed research instrument; b) in the observation stage – the 'Hawthorne effect'2, poorly planned explanations, and inaccurate interpretation of the research situation; c) in the analysis stage – subjectivity, exclusion of contextual information and bias in the selection of data. Endacott goes on to discuss how such distortions can be avoided.

Documentary analysis

Secondary data can also provide valuable information to be used in an evaluation. This can be in the form of records kept or documents produced relating to the project. Documentary analysis can provide information relating to the aims or purposes of a project, target populations to be served by the project, project objectives, outputs and outcomes targeted, the proposed activities to achieve these and the inputs (i.e. financial

resources allocated, staff numbers). These can be used in identifying criteria for the assessment of impact or effectiveness, to assess the rationale for the particular intervention(s) targeted by the project, and the continued relevance of the project in the light of changes occurring over the time of the project.

Documents relating to evaluations previously carried out are also useful sources of secondary data, but as with all data, the evaluator will need to make judgements about the credibility of the data provided before using them as the basis for assessment of the project. In addition, documents relating to strategic planning and on-going monitoring of the project will be useful.

Interviews

Interviews are purposeful conversations that are initiated and guided by the researcher. Interviews can be unstructured, semi-structured, or structured. In the unstructured interview, questions are not scripted ahead of time and the researcher asks pertinent questions as opportunities arise and listens closely to the responses for clues about what to ask next. Unstructured interviews are very useful for in-depth exploration of issues, particularly when little is known about the subject area. The structured interview is based on a carefully worded interview schedule and the format of the schedule may be very similar to a questionnaire – requiring only short answers or ticking boxes. Structured interviews are useful when there are a lot of questions to ask that are not likely to be particularly contentious or thought-provoking. In the semi-structured interview, an interview schedule is used but more latitude is permitted than in the structured interview. Respondents are allowed to express themselves at length but the researcher is provided with enough shape to be able to close down the discussion when it moves away from the particular questions of interest. This format is useful where some deliberation may be required or where the researcher may need to probe deeper to gain further insight into the responses given.

Interviews are particularly useful in obtaining the views and understandings of respondents of their experiences of aspects of a project. This is an advantage over observation. In the semi-structured or unstructured approaches, the researcher can probe further to gain additional information (to expand information), to obtain more detail about something that was already said (to focus information) or to seek clarification on particular points raised. This can be used to enable the evaluator to better understand or get behind data obtained. These options would not be available in a questionnaire survey. In addition, interviews permit face-to-face contact with

respondents and enable the researcher to experience the affective as well as the cognitive aspects of responses (NSF, 1997).

However, there are certain issues that the evaluator should be aware of in the interview situation such as: the influence that the researcher has on the interview; how the interviewee's perception of the interviewer may limit the amount and quality of the information provided; the difficulty attributing what people say to what actually happened; and difficulties in inference and interpretation relating to differences in knowledge and language usage between the interviewer and the respondent. The NSF (1997) also caution that interviews can be time-consuming and expensive. They require well-qualified, highly trained interviewers, whose flexibility can result in inconsistencies across interviews and, furthermore, the volume of information produced may be too large and difficult to transcribe and reduce.

Focus groups

Focus groups are group discussions used to explore a specific set of issues. Group interaction is used explicitly as research data to explore not only what people think, but also why. Differences between group participants are discussed with them and participants are encouraged to reflect on each other's ideas. Conflicting views amongst participants are explored with them to get them to clarify what they believe and why. Marczak and Sewell (1991) define a focus group as:

> … a group of interacting individuals having some common interest or characteristics, brought together by a moderator, who uses the group and its interaction as a way to gain information about a specific or focused issue.

Marczak and Sewell identify a number of advantages to using focus groups:

- the approach takes advantage of the fact that people naturally interact and are influenced by each other
- a focus group may be one of the few research methods that can be used to obtain information from children or from individuals who are not particularly literate
- data can be provided more quickly and at a lower cost than conducting individual interviews and groups can be assembled at shorter notice than doing a systematic survey

- the researcher can interact directly with respondents, for example, to follow-up on issues raised, to seek clarification, to probe, and can use information gained from non-verbal responses to supplement verbal responses
- a deeper level of meaning can be obtained by using the respondent's own words, making important connections and identifying subtle nuances
- the approach is very flexible and can be used with a wide range of topics, individuals and settings, and it is comparatively easy to prepare and conduct
- the results are easy to understand, and are more accessible to lay audiences or decision makers, than for example complex statistical analyses of survey data.

They also identify a number of limitations of the focus group approach, including lack of control over the group and information produced, the relatively 'chaotic' nature of data produced making analysis difficult, the lack of generalisability of findings due to the small numbers of participants and convenience sampling, the need for a trained interviewer who knows about group dynamics to conduct the group, and uncertainty about the accuracy of what participants say within the group situation.

Case studies

Stake (1994) describes a case in terms of a pre-existing bounded entity. In the case study approach the case is treated as the focus for an in-depth, holistic study of the subject in the context in which it exists – described by the NSF (1997) as the 'real-world setting'. This is usually on the understanding that context is inherently linked to the phenomenon being studied. The study may be based on a single case, or multiple cases may be used to enable comparisons to be made between the cases or to add to the generalisability of the findings. Usually, more than one method will be used to study the case, for example a mix of documentary analysis, observation and in-depth interviews. Case studies can be useful in comparative analysis and to examine the ways in which differences within cases (e.g. the structure of a programme, differences in the resources allocated, socio-economic differences in different populations) contribute to the outcomes of policies, programmes or projects. The disadvantages of the case study method relate mainly to its complexity, sometimes requiring considerable resources. In addition, findings would not be generalisable to a wider population in the same way that survey findings would, but in-depth study of the particular may be considered to be more appropriate in some evaluations than would generalisability.

5.5 Issues in the selection of evaluation methods and the quality of evaluation

As outlined in Figure 5.1, it is likely that there will be more than one method that could be used to answer each of the five evaluation questions identified earlier. This will require a judgement to be made about which method(s) is (are) most appropriate given the evaluation questions to be addressed, given the constraints (e.g. tight deadline, sensitivity of the subject) within which the evaluation must take place and the resources (e.g. time, funds, expertise) that will be available.

Weidman et al (1975) suggest that seven factors should be considered in the development of an evaluation design:

- the measurements required
- the data sources for those measurements
- the comparisons to be made
- the analytical techniques for making the comparisons
- the level of confidence provided by the design
- the cost of conducting the necessary work
- the type and volume of findings that the design can be expected to produce (p.18).

5.5.1 Theoretical considerations

Decisions about which methods are most appropriate will need to be guided by judgements about how each of the range of methods will add to or ensure the quality or rigor of the evaluation, and accordingly the credibility and accuracy of the findings. The NSF (1997) identify three theoretical issues to be considered in whether to use a quantitative or qualitative method:

- *the value of the types of data* – there is a trade-off between breadth (quantitative) and depth (qualitative) and between generalisability and targeting the very specific
- s*cientific rigour* – standardised methods are used in quantitative approaches to yield what is believed to be more objective and accurate information which can be replicated and analysed using sophisticated statistical techniques. On this basis, qualitative methods have traditionally been used for formative evaluation and quantitative for summative evaluation, where 'hard' measures are used to provide an ultimate assessment of the value of a project. However, they suggest that this distinction is too simplistic and that quantitative researchers are becoming more aware that their data may not be so accurate and valid, whereas qualitative

researchers have developed better techniques for classifying and analysing large bodies of descriptive data

- *philosophical distinctions* – both approaches are founded on different philosophical perspectives about the nature of knowledge and how it is best acquired.

Those designing an evaluation will need to consider the validity of the methods selected and their ability to capture or measure that which they set out to capture or measure. The Joint Committee on Standards for Educational Evaluation (JCSEE) (1994) has done considerable work in the US to develop a comprehensive set of programme evaluation standards. They define validity in terms of 'the soundness or trustworthiness of the inferences that are made from the results of the information gathering process'.

Reliability is also an important consideration. The JCSEE (1994) describe reliability in terms of 'the degree of consistency of the information obtained from the information gathering process'.

> Most important when assessing the reliability of the information obtained is to distinguish unwanted variability that is attributable to noise or random error in the information collection procedure from variability due to systematic, explainable sources. Systematic sources include differences attributable to different instructional effects, characteristics of program participants, and conditions in the environment or context of the program being evaluated. Reliability will be called into question whenever the evaluation procedures yield information and results that cannot be explained by these systematic sources (JCSEE, 1994, p. 153).

Some methods (e.g. case studies) may be criticised for their lack of generalisability, and this can have implications for an evaluation. Judgement will need to be made about whether generalisation is sought or, whether instead, a more appropriate focus would be on a particular case. An in-depth focus on a particular case will provide very rich data and in-depth insight, but at the expense of knowing how the findings can be applied to a wider population. This may be appropriate, for example, to explain the relationships between particular events involved in outcomes. Additional cases can be used later to gain wider resonance for the findings or to test their applicability beyond the original case.

The JCSEE (1994) also include 'defensible information sources', 'justified conclusions' and 'impartial reporting' in their standards relating to evaluation design. In terms of defensible information sources, they state that many different sources of information can be used in evaluation that is obtained from a variety of sources and 'tapped' in different ways. Some degree of sampling will be involved and the evaluator can only hope to obtain a portion of the information that is potentially available. Accordingly, 'evaluators should document, justify, and report their sources of information, the criteria and methods used to select them, the means used to obtain information from them, and any unique and biasing features of the obtained information. These descriptions should be sufficient to permit others to determine the adequacy of the information for the evaluative questions to be answered' (pp. 141-142). They suggest that the conclusions of the evaluation should be explicitly justified/defended and defensible, otherwise they may be disregarded by stakeholders who do not receive sufficient information to determine that the conclusions are warranted. They further suggest that where possible, the conclusions should be accompanied by alternative plausible explanations of the findings and an explanation of why these were rejected.

The JCSEE (1994) also emphasise the importance of findings being reported impartially. Sources of distortion identified by them include not reflecting all of the perspectives that should have been taken into account or favouring one interpretation over others, carelessness, or pressure from a particular client/sponsor/audience. They also identify that bias can occur where there is continuous reporting of findings and where earlier reports constrain subsequent reports. In addition, they suggest that where findings from earlier reports are used to refine a programme, evaluators may bias reports because they are evaluating their 'own' programme.

5.5.2 Practical considerations

The NSF identify four practical issues to be considered in the selection of quantitative and qualitative methods:

- *the credibility of findings* – acknowledging the range of audiences/stakeholders that an evaluation will have, it is suggested that evaluators should be cognisant of possible sceptical audiences or stakeholders who may discredit findings that are too critical or uncritical of a project's outcomes. They may seek to reject the findings on the basis of the methods used, or because the findings are weak for a specific case. The authors provide a useful example:

The major stakeholders for EHR projects are policymakers within NSF and the federal government, state and local officials, and decision makers in the educational community where the project is located. In most cases, decision makers at the national level tend to favor quantitative information because these policymakers are accustomed to basing funding decisions on numbers and statistical indicators. On the other hand, many stakeholders in the educational community are often sceptical about statistics and 'number crunching' and consider the richer data obtained through qualitative research to be more trustworthy and informative. A particular case in point is the use of traditional test results, a favorite outcome criterion for policymakers, school boards, and parents, but one that teachers and school administrators tend to discount as a poor tool for assessing true student learning (NSF, 1997, p.5).

- *staff skills* – they suggest that qualitative methods such as in-depth interviewing, observations or interviews require good staff skills and considerable supervision to provide trustworthy data. In contrast, they suggest that some quantitative methods, such as small-scale, self-administered questionnaires where most questions are answered using yes/no checkmarks, can be mastered easily 'with the help of simple training manuals'. However, the design, management, and analysis of large-scale, complex surveys will require more skilled personnel.
- *costs* – different methods will have different cost implications, which need to be balanced with the robustness of the data provided.
- *time constraints* – within the time allowed for the evaluation, choices will have to be made about the amount of data and the complexity of the data to be collected, and the time required to develop and use particular techniques. For quantitative methods it takes considerable time at the outset to design a good survey and to pre-test questions, although technological advances have shortened the time required for analysis. However, they suggest that qualitative methods may be even more time consuming because data collection and analysis overlap and new evaluation questions may be explored as they arise. Having to cut short an evaluation to meet severe time constraints, by curtailing the amount of data to be collected or reducing the analysis process, may reduce the value of the findings.

5.5.3 Recommended evaluation standards

Based on the evaluation standards identified by the Community Support Framework (CSF) Evaluation Unit (1999)/European Commission (1997), and the European Commission (2000), it is suggested that evaluations should be:

- *analytical* – based on recognised research techniques
- *systematic* – involving careful planning and consistent use of chosen techniques; in addition, the evaluator should have access to the data sources that are relevant for the evaluation questions to be addressed
- *reliable* – a different evaluator with access to the same data and similar techniques of data analysis would arrive at similar findings; also, the final report should present findings based on carefully described assumptions, data analysis and rationale
- *issue-oriented* – should address important issues such as relevance, efficiency and effectiveness of a programme
- *user-driven* – should be executed in ways that provide useful information to decision makers and evaluation questions and recommendations should be realistic and relevant to the intended audience; stakeholders in the programme or action to be evaluated should be identified, and to the extent possible, their involvement should be clearly stated
- *transparent and objective* – the final report should present the findings and conclusions determined by the evaluator and these should not be amended without the evaluator's consent; also where there are significant dissenting views these should be indicated.

5.6 Conclusions

Building on the thinking presented in Chapters Three and Four relating to issues to be considered in the design of evaluations, this chapter emphasises the importance of methodological appropriateness in evaluation and those issues to be considered in the selection of evaluation methods. The key features of a range of methods are outlined in order to assist in the selection of methods that are linked to five types of evaluation question. There are several very useful sources of further information on each method identified here and listed in the reference section. In the following chapter, approaches being adopted in other countries to develop evaluation frameworks in health services are reviewed.

6
Towards the development of an evaluation framework for health services

6.1 Introduction

This chapter explores the development of evaluation, including evaluation of health services, across countries in public services. The aim of the review is to identify issues to be considered in the design of an evaluation framework for health services in Ireland. In considering the relative merits of approaches used in other countries, caution is advised by Saltman (1997):

> It is simultaneously seductive and deceptive to seek to compare health care systems across different countries. Seductive in that clever comparisons ought to make it possible to identify best practice about health system structure and policy. Deceptive in that health systems are deeply embedded with the social and cultural fabric of that society, and thereby defy simple economic or financial characterization. If a cross-national exercise is to be reasonably valid, comparisons must reflect national social contexts, rather than assuming that health system arrangements exist in splendid social and political isolation (p.S10).

It is within such an understanding that this review attempts to identify different ways in which governments have sought to develop evaluation frameworks and to attempt to consider how they might usefully be applied in the Irish health service context.

6.2 The development of evaluation in other countries

Considerable variations are to be found across countries in approaches to organising evaluation in health and public services. These variations reflect differences in structures (for example, whether government is organised through federal, centralised or decentralised systems); funding arrangements; understanding of the role of government; and the maturity of evaluation systems. It also suggests that in the development of an evaluation framework, considerable thought will need to be given to the principles, structures, and culture of a health system, the success or failure of

any previous attempts to develop evaluation, and how various approaches would meet with the functions envisaged for evaluation in a particular system.

Across OECD countries, three stages of development of evaluation are outlined by Derlien (1990, cited in PUMA/PAC, 1999).

- In the 1960s and 1970s, where public programmes were launched by 'social-liberal' governments to solve social problems, evaluation was focused on improving programmes and providing feedback to programme managers.
- In the 1980s, the concern of mainly conservative governments was to curb public programmes given fiscal constraints, with evaluation being used to reconsider the justification of policies and to rationalise resource allocation within budget.
- Since then, evaluation has been used to strengthen the accountability of government due to concerns about the legitimacy of the public sector.

More recently the emergence of meta-evaluation is reported in countries such as the US, where evaluation has been firmly established for some time and where because of the amount of evaluation that has taken place, general conclusions can be drawn from a series of evaluations. Particularly in the health sector, there is the recent emergence of more collaborative approaches to evaluation based on supporting evidence-based practice; for example health technology assessment and systematic reviews of the effectiveness of treatments and models of care. Several themes can be identified across countries, and these are explored in more detail in the following sections.

6.3 Evaluation, accountability, decision making and learning

In this comparative review of evaluation frameworks, two functions for evaluation can be identified clearly: accountability and decision making. As can be seen in countries that have adopted a purchaser/provider split, there is a clear emphasis on accountability between the purchaser and the provider built into the contracting process. The emphasis in the British model is on performance indicators and standards, with evaluation being used for closer examination where issues are identified. In the emerging model in New Zealand, explicit auditing and monitoring requirements are built into funding agreements between the ministry and district health boards, and between district health boards and providers. An emphasis on accountability is also apparent in other countries. In some countries, there is a strong emphasis on both accountability and decision making. In Australia, although recent efforts are focused

on the development of performance indicators, performance measurement and benchmarking for accountability purposes, there is also a strong emphasis on evaluation, and there is evidence to suggest that evaluation results are used to inform decision making at all levels of the system. There is a strong emphasis on evaluation in the Netherlands and Canada. In Canada, it is suggested that, to date, evaluation findings have been used most effectively for decision making at the programme management level but that, more recently, there is a greater emphasis on building evaluation into decision making at the national level.

Despite current thinking in evaluation about the value of evaluation in organisational learning, this concept does not feature strongly in the literature relating to evaluation in the health and public services. The exception is Canada where recent reforms aimed at building a culture of results-based management emphasise learning – making decisions on the basis of empirical information obtained through evaluation, audit, performance measures and scans (Auditor General of Canada, 2000).

6.4 Evaluation and the development of performance indicators

It seems that, in some administrations, thinking has moved on from the establishment of evaluation to developing suites of performance indicators that will provide continuous feedback on performance. This shift in thinking was identified in Chapter Three. In Australia, the Department of Health and Family Services (now called the Department of Health and Aged Care) (HFS, 1997) reports a shift towards performance assessment with evaluations being undertaken when needed to inform decisions rather than on a cyclical basis as was the practice before. PUMA/PAC (1999) suggest it is likely that mandatory requirements for evaluation in Australia will be replaced by an integrated performance management framework, which will encompass performance monitoring and evaluation. Mackay (1998b) suggests that the increased emphasis on performance measurement in the 1990s in Australia is not to suggest that evaluation has been inadequate, but that the emphasis and success has been in the area of the development of evaluation. Less attention has been paid to the setting of programme objectives and the collection of frequent performance information for on-going monitoring of programmes.

The National Health Performance Committee (NHPC) was formed to develop and maintain a national performance measurement framework for the Australian health system, and has been working since February 2000 (NHPC, 2001). Performance

indicators are being developed to monitor three dimensions ('tiers') of health service performance: health outcomes, the determinants of health and health system performance. There are nine dimensions to health system performance: effective, appropriate, efficient, responsive, accessible, safe, continuous, capable and sustainable. In addition, the NHPC state that equity and quality are integral to the framework and overarching themes in the health system performance tier.

The NHPC promote performance information as a tool and emphasise the need for caution when using performance indicators, to ensure that a balanced view of performance is sought and also emphasise that further work will be required to make sense of performance information:

> In considering the selection or development of relevant health system performance indicators it is important to keep in mind that indicators are just that: an indicator of organisational achievement. They are not an exact measure and individual indicators should not be taken to provide a conclusive picture of an agency or system's achievement. A suite of relevant indicators is usually required and then an interpretation of their results is needed to make sense of the indicators. Performance information does not exist in isolation and is not an end in itself, rather it provides a tool that allows opinions to be formed and decisions made (p. 20).

This national set of performance indicators will complement performance measurement and benchmarking that occurs at state/territory level (for further information on performance measurement at the state/territory level, see NHPC, 2001). Performance measures are also built into agreements between the Commonwealth and states/territories, and between states/territories and health service providers (Health Care Agreements and Health Service Agreements respectively).

In Britain, most of the effort in recent reforms has been on the development of performance indicators and the use of information technology to provide accurate, up-to-date information to support the drive for quality and efficiency in the NHS (Department of Health, 1997). The National Health Performance Assessment framework, comprising a set of high level performance indicators, was road-tested in 1998 and following some amendments, introduced into the system in 1999. This emphasis on performance measurement is complemented by the development of clinical governance arrangements which make health service providers responsible for

the quality of services they provide. In addition, health authorities and trusts are subject to mandatory audit by the Audit Commission. Each year specific service areas are identified nationally that are subject to external 'value for money' audit by the Commission within each health authority and trust. In recent reforms, the Commission for Health improvement was established with an explicit role to oversee the quality of services at local level, to tackle shortcomings, and to intervene where necessary. The development of performance measurement and evidence-based practice are key elements of recent reforms towards a 'new NHS'. This includes the development of national service frameworks setting out standards in relation to access to and quality of services to be expected across the country. In addition, the recently established National Institute for Clinical Excellence has a strong lead role in the development of national guidelines on clinical and cost-effectiveness. Explicit quality standards reflecting these and other national standards and guidelines are included in service agreements between the NHSE and health authorities, and between health authorities and service providers.

6.5 Evaluation and evidence-based practice

More recently, there has been a greater use of collaborative evaluation, in particular in the assessment of health technologies and in systematic reviews of the effectiveness of treatments and approaches to care. There is a strong emphasis in both approaches on the wide dissemination of findings. Findings are also expected to feature strongly in decision making at all levels of health systems.

6.5.1 Health technology assessment

Health technology assessment (HTA) is being carried out in several countries in the light of the burgeoning range of new technologies emerging in health care. The approach is aimed at providing an independent review of both new and existing technologies. For example, in Australia the National Health Technology Advisory Council is responsible for the assessment of both emerging and existing technologies. In addition it is responsible for the development of guidelines for planning and delivery of specialised health services. Abel-Smith et al (1995) report a variety of approaches to HTA across EU countries. They report that approaches include the establishment of centres for the evaluation of health technologies, such as the National Commission for Hospital Planning in Belgium, or the development of assessment methodologies such as the Agence Nationale pour le Developpement de l'Evaluation Medicale in France. Saltman and Figueras (1997) report that governments tend to play a major direct or

indirect role in the development of health technology assessment, mainly to ensure independence of commercial influence and out of concern about some research undertaken in the past by the pharmaceutical and medical equipment industries. Thus, the agencies that have been established in France, Germany and Spain (although established at regional level in Spain) are very closely linked to government. In other countries, the approach is decentralised, with research undertaken by universities but as part of a national framework of commissioned research. A third approach identified by Saltman and Figueras is for agencies responsible for paying for health care, including sickness funds and insurance agencies, to undertake the research. The general approach is to identify existing and emerging technologies and to evaluate them on the basis of existing evidence which is collected and analysed, or, where appropriate, to conduct evaluative research. A key element of the approach is the dissemination of findings. Despite these developments, Abel-Smith et al (1995) suggest that HTA in Europe is largely an informal process and could be improved through better circulation of information produced through assessment activities and by clinicians having more timely access to findings, better resourcing, and co-ordination of HTA activities. Saltman and Figueras (1997) suggest that much clinical practice is still inadequately evaluated despite the considerable volume of activity being undertaken, largely because of the growth of health care technology. Furthermore, the indications and contexts within which technologies are used are also changing frequently.

The HTA programme website (http://www.hta.nhsweb.nhs.uk/abouthta.htm) outlines the framework for Health Technology Assessment developed over recent years in Britain. The key elements are:

- a national programme of research established and funded by the Department of Health's research and development programme
- a national co-ordinating centre for HTA (NCCHTA) which co-ordinates the HTA programme on behalf of the R&D programme

  advisory panels, who once the HTA programme has actively consulted users and solicited suggestions, helps to decide which of the many suggestions received from the NHS and its users should become research priorities
- the call for proposals and the commissioning of research to answer the questions identified
- the publication of the results of research.

The NCCHTA monitors the progress, timeliness, relevance and value for money of all projects. The final report of every project is peer-reviewed, and if accepted for publication, is published as a report in the HTA monograph series (available from the HTA website).

While not wishing to be drawn on which structure is most effective, Saltman and Figueras (1997, p.198) identify several key tasks that should be undertaken in health technology assessment:

- a systematic identification of priorities based on national circumstances and a review of existing information
- commissioning reviews of existing evidence and basic research
- a means of ensuring that the results are collated in accessible form
- mechanisms for disseminating and implementing them (such as regulation, financial incentives and education).

The new health strategy (Department of Health and Children, 2001) recognises the importance of developing health technology assessment in the Irish health system to maximise the potential that advances in science and technology bring, while guarding against the associated threats. Accordingly, 'The strategy must establish the mechanisms and structures to support the health system in monitoring and evaluating the benefits and risks which technology can bring so that the system can take advantage of the benefits and respond quickly to challenges that may arise' (p.55). It also notes the findings of the Deloitte and Touche Report (2001) that there is currently no coherent structure for carrying out evidence-based HTA. It identifies an explicit role for the Health Information and Quality Authority in the development of HTA.

6.5.2 Systematic reviews and the dissemination of evidence-based practice

A similar approach to health technology assessment has been adopted in Britain for the assessment of the effectiveness of specific treatments and the organisation of care. The NHS centre for reviews and dissemination (CRD), based at the University of York, conducts systematic reviews of research carried out on a particular topic, and on the basis of what is found and the quality of each piece of research, makes an assessment of the effectiveness of the treatment or approach to care. The findings are disseminated to health professionals and managers and are recorded in on-line databases (DARE – containing abstracts of quality assessed systematic reviews, and NHSEED – economic evaluations of health care interventions). The centre also

provides access to details of publications and projects conducted by health technology assessment organisations.

Also involved in reviews and dissemination is the Cochrane Collaboration, which evolved from the Cochrane Centre, which was established initially to conduct systematic, up-to-date reviews of all relevant randomised control trials of pregnancy and childbirth. This was later extended to cover all areas of health care and to collaborate with others, in the UK and elsewhere. Funds were provided by the NHS Research and Development Programme. In October 1993 the first in a series of annual Cochrane Colloquia (seventy-seven people from eleven countries) co-founded 'The Cochrane Collaboration'.

The Cochrane Collaboration aims to help people make well-informed decisions about health care by preparing, maintaining and ensuring the accessibility of systematic reviews of the effects of health care interventions. The Collaboration is built on ten principles: collaboration; building on the enthusiasm of individuals; avoiding duplication; minimising bias; keeping up to date; striving for relevance; promoting access; ensuring quality; continuity; and enabling wide participation. Further information can be found on the Cochrane Collaboration's website: www.cochrane.org/cochrane/general.htm.

6.6 Evaluation as a national or local function, or both

A clear distinction can be found in the international literature between evaluation at the macro level, evaluation of national or federal systems or programmes, and evaluation at the micro level, evaluation at the level of the organisation. In some countries, such as Australia, there is a central evaluation strategy which all ministries must adopt and evaluation is a key element of national public health strategies and national initiatives such as the National Demonstration Hospitals Program and the National Health Priority Areas initiative. Performance indicators are identified in the Department's (the Commonwealth Department of Health and Aged Care) Portfolio Budget Statement (budget allocation to the Department), which also identifies any evaluations to be conducted over the year. Below the national level, performance indicators are used in agreements (Health Care Agreements) between the Commonwealth Department of Health and Aged Care and states/territories, which also require programmes being introduced to be evidence-based. Within states/territories, evaluation is organised in different ways. For example, the Health Department of Western Australia has its own

performance evaluation unit which undertakes programme evaluations and is involved in performance review of the health system. In the Northern Territory, the Attorney General's Department monitors the effectiveness of the Territory Health Services Executive's role in programme evaluation.

In Canada, it would appear that although federal policy requires all federal departments to evaluate their key policies and programmes on a cyclical basis, the most effective evaluation occurs at the programme management level. Divorski (1998) reports on the findings of two audits of evaluation carried out in 1993 and 1996, which found that evaluation was most successful (in producing changes) where it was focused at the operational level (to meet the needs of programme managers). The audits found few examples of evaluations resulting in significant changes to programmes, or relating to the overall effectiveness, of larger programmes. While the 1996 audit identified 'important progress' in linking evaluation to the budget process, Divorski also reports that a systematic presentation of government evaluation priorities and clear links between overall government evaluation plans is still lacking.

In Britain, the development of standards and performance indicators have been the key mechanisms identified towards improving performance and quality in health services. Although health care provider organisations have been given greater autonomy over recent years, national frameworks (as outlined in section 6.4) exist to enforce performance measurement and evaluation. The establishment of clinical governance brings a greater balance between national and local evaluation practice. Clinical governance places an implicit responsibility for evaluation at the organisation level and requires organisations: to accept responsibility for developing and maintaining standards within the organisation; to put in place quality improvement processes; to ensure that evidence-based practice is in day-to-day use and that there is the infrastructure to support it; and to ensure that good practice, ideas and innovations are evaluated and systematically disseminated within and outside the organisation.

Johnson (1996) reports on the integrated framework for evaluation of the US public health service (PHS) programmes. All federal programs (including PHS programmes) are subject to systematic evaluation, the results of which are fed into decision making by Congress and Federal Government. The Assistant Secretary for Planning and Evaluation (Department of Health and Human Services (DHHS)) co-ordinates evaluations overall and provides guidance on priority PHS policy areas for evaluation. Each PHS agency is then required to submit a plan of its evaluation strategy and proposed projects for the immediate fiscal year, or subsequent financial years for multi-

year projects. Most evaluation projects are developed at the programme level and a committee of policy and planning staff members at agency level conducts the initial review. The project will also be reviewed for technical quality before it is implemented by a second committee, of staff members who are skilled in evaluation methodology, and following a set of criteria for quality evaluation practice established by the agency. The General Accounting Office (GAO), universities or large research firms, then usually undertake the evaluations. Larger projects are conducted in two stages – a design contract and a government procurement to implement the design. The DHHS policy information centre monitors the progress of projects and a departmental evaluation database and library is maintained. The results of evaluations are disseminated through targeted distribution of final reports, articles in refereed journals and presentations at professional meetings and conferences. It is reported that the principal stakeholders of PHS evaluations have been government programme planners and managers and the larger public health community as consumers. In addition, the PHS agencies and the Office of the Secretary for Planning and Evaluation have striven to take a leadership role in producing new knowledge about the effectiveness of public health programs and interventions and in developing evaluation tools for use in the larger public health community. Over recent years there has been an increasing emphasis on the need for more comprehensive program outcome and impact evaluations, a view supported also by the GAO and reinforced in the National Performance Review. In addition, the Government Performance and Results Act (1993) links performance measures to federal agency strategic planning systems and potentially towards performance budgeting.

Despite the decentralised approach to the delivery of health services in the Netherlands, a strongly co-ordinated approach to evaluation is being pursued. Responsibility for the provision of health services is shared between the Ministry of Health, Welfare and Sport and local authorities in a decentralised system of care delivery. At central government level, the Court of Audit is responsible for auditing ministries and boards connected with central government (statutory tasks or funding). The Court of Audit has a constitutional brief to carry out performance audits (similar to programme evaluations, operational audits, or value for money audits (Court of Audit, 2001)). Performance audits are focused on the efficiency of management, organisation and policies of central government and the Court of Auditors has freedom to select specific audit areas. Certain evaluation measures and responsibilities are required by government, including the following (PUMA/PAC, 1999):

- The heads of ministries are ultimately responsible for evaluation.
- Policy directorates are responsible for evaluation themselves by adopting either a centralised or decentralised approach in managing activities.
- Internal budget directorates are responsible for co-ordinating, promoting and ascertaining the use of evaluation and ensuring that necessary advice, guidance and research expertise are provided. They draw up evaluation programmes for individual projects, encourage the periodic evaluation of policies and monitor the quality of evaluation and its use. Other directorates provide help with their special expertise (on personnel and organisation, auditing, legislation).
- The Ministry of Finance has the same responsibilities for promoting and co-ordinating internal budget directorates.
- The Court of Audit reviews the efficiency of evaluation management and organisation and publishes reports on them (p.18).

In 2001, the local authorities representative organisation (VNG) and the organisation representing healthcare providers (GGD) entered into a National Contract for Public Health with the ministry and other organisations, to agree to work together towards improving public health, including the development of local monitoring of the effects of health services. In a separate arrangement, health care providers are required under the Care Institutions Quality Act (1996) to develop quality systems to systematically monitor and improve care, including patient feedback. In addition, a hospital accreditation scheme has been developed, drawing on the work of the Canadian Council on Healthcare Accreditation.

6.7 Issues in locating evaluation

Mayne et al (1999) discuss the issues to be considered about where evaluation is to be located in the development of evaluation frameworks, in particular whether evaluation should be anchored in the legislature, the executive, or both. They suggest that the responsibility for planning and managing evaluations can be assigned to a variety of locations within a jurisdiction. Legislative 'anchors' include: audit offices that undertake performance audits or evaluations to assess how well government programmes are working; and legislative bodies undertaking or commissioning evaluations to examine what the public is getting for the taxes it pays. Executive anchors include: programme managers planning and undertaking evaluations of their respective programmes; organisational corporate staff groups with responsibility for planning and undertaking evaluation studies or programmes within the organisation;

and central corporate staff within the executive evaluating the performance of major programmes, programmes that cut across several organisations, or programmes that might need special attention. They also identify 'Outside Government Anchors' such as universities and research groups who can be a source of evaluation expertise, undertaking studies of the effectiveness of government programmes, or non-governmental organisations, for example, community groups or consumer groups that are affected by government programmes, evaluating the benefits that they or their client members receive.

Mayne et al (1999) highlight the wide range of kinds of evaluation undertaken in the public service, along with the wide range of individuals or groups who use evaluation findings. The upshot of this diversity is that no one single evaluation or evaluation function can handle all of the data collection and analysis required. In addition, they claim that the location of evaluation significantly influences the focus it takes, and evaluation needs to be reasonably well-focused and specific if it is to be useful. In deciding where to anchor evaluation, they suggest two strategic questions need to be addressed: 'Which evaluation issues are to be addressed? Who is going to use institutionalized3 evaluation?' (p.27).

Mayne et al (1999) make two key points in relation to deciding which evaluation issues are to be addressed.

1 *Whether the purpose of evaluation is to improve programmes or to challenge them.* It is suggested that different locations are better suited to dealing with different types of issues to be addressed.

> … as we move from operational issues to impact issues, to rationale issues, evaluations become more challenging in terms of organizational resistance. The move from operational to impact issues also entails challenging technical difficulties. Assessing success is more difficult than questions regarding program operations. Determining the continuing need for a program is the most challenging question of all. Evaluation regimes quite often will address operational performance issues well. They are usually easier to measure and analyse, are of immediate interest to those close to the program who may control access to the program and its information, and speak directly to questions of improving the program rather than challenging its existence (p.29).

2 *How evaluation can deal with impact and relevance (continued necessity) issues.* These issues, it is suggested, pose more fundamental and threatening questions about the well-being of organisations, usually require more sophisticated measurement practices to get credible evidence, and are usually conducted from outside the organisation. In contrast, operational issues usually meet much less resistance in organisations:

Managers need information on operational performance to get on with their job and, hence, sometimes with help from an internal evaluation group, tend to have or want this type of information available and to use it … A key conclusion is that the closer the evaluation anchor is to the program, the less likely is it able to deal adequately with continued rationale and impact issues. … There is also the very real practical difficulty of not being able to step back and look at the program from 'outside the box'. Independence does often make an objective, outside and more questioning look easier to undertake (p.31).

In terms of the second strategic question – who is going to use institutionalised evaluations? – it is suggested that no matter where evaluation is located, there will be others who will make use of the evaluation findings and conclusions. Accordingly, in deciding where to anchor evaluation, thought needs to be given to the likely use of the findings by other key players in the jurisdiction. The usefulness of the evaluation as perceived to other key players will relate to: the scope of the evaluation – whether the evaluation addresses the 'right' issues; the evaluation entity – 'just what is being evaluated'; evaluation timing – if findings are available when needed; and, the perceived reliability and credibility of the evaluation.

Mayne et al (1999) discuss the benefits and drawbacks of the different locations for evaluation. To summarise, they suggest that the most common way to locate evaluation is within the organisation, either with programme managers or with a corporate group within the organisation. The benefits include the contribution to day-to-day management and the improvement of programmes because of the strong focus on operational issues. Further, the evaluation can draw on the knowledge of programme management and staff, making evaluation easier to perform and enhancing the buy-in and use of findings. It may also be less costly to perform. Access to programme data and information can be less of an issue because the user is closer to the programme being evaluated. Genuine interest by programme personnel will ensure a focus on critical programme areas and programme areas where there are known problems. Evaluations can help managers to find organisational support or resources

needed to improve critical organisations and to find and implement solutions. In terms of drawbacks, it is suggested that there is a danger of evaluation becoming internal management consulting with a focus on operational issues, at the expense of impact issues and continued relevance; and such evaluations are unlikely to challenge the existence of a programme. Where issues of interest to corporate government and legislative users are addressed, it is unlikely that the findings will be of use unless the results of a number of individual evaluations are rolled up, and this is not likely to occur without strong corporate government direction and co-ordination.

PUMA/PAC (1999) also explore the benefits and drawbacks of locating evaluation within and outside of the organisation. The benefits outlined relate to learning within the organisation and access to data and knowledge of programmes. The drawbacks identified relate to the time and skills of staff required and available, and the possible lack of objectivity.

Evaluation located in corporate government groups, such as finance, treasury and budget offices, which are usually responsible for ensuring the good management of government as a whole and for developing and assisting to implement government-wide policies, can result in an emphasis on larger government issues relating to the improvement of management and or assisting in resource allocation decisions. The potential drawbacks relate to access to information, where the group would not have direct access to programme knowledge or information, or where they would be expected to produce findings that could be critical of programme managers, threaten programmes, or lead to restrictions on the discretionary authority of managers. In addition, programme managers may be reluctant to accept and use findings, and evaluations would not be used in the same way in programme management as those conducted by programme managers. The use of findings by legislative bodies may be limited and access by legislative bodies to findings may be restricted.

Evaluations located in the legislative branch are more likely to address issues that are relevant to legislators, and to address issues that the executive branch may wish to avoid. Drawbacks include lack of access to detailed knowledge about the working of programmes and, because legislative evaluators in their role of providing objective comment on performance must avoid partisanship, the evaluators may be restricted from addressing certain issues such as the merits of government policies.

Mayne et al also explore supply and demand for evaluation. Wherever evaluation is located, it will produce or supply evaluation to meet information needs. In real terms, this will result in multiple 'markets' for evaluation, with users with different information requirements. Each market will have its own demands with its specific characteristics, and accordingly its own supply requirements. Thus, they suggest that evaluation may need to be anchored in several locations to meet the multiple market demands in a jurisdiction.

6.8 Towards a coherent approach to performance assessment

This review of the development of evaluation across countries highlights the range of different approaches being taken, which to some extent also reflect differences in thinking about how system and organisational performance is best controlled and enabled. The review of evaluation in the Irish health system also suggests that different activities are occurring at the national, regional and organisational level to improve performance, which draw on evaluation in different ways. Bastoe (1999) explores the possibility of developing a coherent approach to performance management in public systems, which would bring all of the elements present together into an integrated framework, or some sort of 'schematic order' (p.107). This would involve setting out the functions and their institutions in an interlocking structure of responsibilities, with networks of processes which contribute to and draw on each other to produce management synergy. It assumes that there is both a deliberate strategy of organisational design and a link between administrative and political forms in the policy process. The key elements of such a 'strict rational model of integration' are:

- performance objectives and targets identified for programmes
- managers responsible for programmes having the freedom to implement processes to achieve the objectives and targets identified
- the actual level of performance against targets being monitored and reported
- the performance level achieved feeding into decisions about future programme funding, changes to programme content or design and the provision of personal rewards or penalties
- the information being provided to ex-post review bodies, whose views may also feed into decisions about programmes.

Bastoe comments that while there are 'obvious' problems to such a strict model, it can be argued that a rational model is required for the necessary coherence and structure. However, implementation approaches need to take account of some of the features of organisations that affect implementation and how learning actually takes place. Several pre-conditions required in order to integrate the different functions, based on experiences from different countries, are identified. These include: availability of data and personnel; political authority to change and integrate systems in order to overcome organisational inertia at the lower levels; and a firm policy to develop and link systems and the design of strategies to implement policy.

6.9 Conclusions: towards an integrated performance management framework in Irish health services

In Figure 6.1 the key elements of a possible integrated framework for performance management, incorporating evaluation, for the Irish health system is outlined. Performance indicators provide the overarching framework for on-going monitoring, perhaps based on changes in performance over time. Where negative trends over time are identified, or where an organisation performs poorly in comparison with similar organisations, evaluation may be instigated to understand more about factors to be addressed. The second element – external evaluation – would involve the undertaking of evaluation by agencies or units outside of the organisation. External evaluation could be required: a) for all new national programmes, b) through a programme of work for existing programmes, and c) where particular areas or problems are identified for further examination. Evaluation (undertaken as an internal function) would also become an integral part of the local management of health services.

Figure 6.1 An integrated performance management framework

Approach	Source	Key elements of approach
Performance indicators	• Service plans and provider agreements • Falling out of strategies	• On-going monitoring • Benchmarking
External evaluations incorporating current expenditure review approach	• Possibly undertaken by unit within DoHC or independent agency	• Cyclical for all new programmes ⇒ action planning and review • Programme of work for existing programmes ⇒ continued relevance – recommendations to retain, amend or disband • Priority areas – hot spots identified through PIs/ benchmarking / specific criteria (e.g. high cost/ controversial programmes)
Integral to programme management – e.g. CQI approach	• Undertaken by programme /project managers	• Strategy and business planning as drivers • Clear objectives set out in strategies • Evaluation plans/criteria set out in project proposals

Collaboration and support to enhance evaluation	• DoHC or independent agency taking lead	• Establishment of networks • Joint programme evaluations /piloting of new programmes • Dissemination of findings • Health technology assessment, systematic reviews and meta-evaluations • Support for evaluation – advice, training on skills and competencies, resources, exchange programmes

The framework also includes a strand concerned with promoting collaboration and supporting evaluation through investment in evaluation infrastructure, including training.

Building on the discussion in this chapter on the development of a framework for evaluation in the Irish health system, the following chapter focuses on the development of evaluation capacity to support and be supported by such a framework.

7
Building evaluation capacity

7.1 Introduction

Building on the exploration of methods and approaches to evaluation in health services and a possible framework for evaluation, the focus of this chapter is on how evaluation capacity can be enhanced. First, themes identified in a review of approaches to establishing evaluation across countries are outlined. This is followed by an exploration of what evaluation capacity is and what are the factors influencing supply and demand. Finally, the chapter concludes by identifying some of the challenges identified to evaluation supply and demand in the Irish system.

7.2 Approaches to establishing evaluation

PUMA/PAC (1999) report considerable differences between OECD countries in terms of the degree to which evaluation has been established. In countries such as the United States, Sweden and Canada, evaluation is firmly institutionalised and Australia and the Netherlands have also developed sophisticated evaluation systems. Other countries 'are just beginning to reflect on the possibility of systematising evaluation activities' (p. 15). Variations are also reported between countries and between policy sectors in terms of the maturity of evaluation systems, with areas such as developmental aid, labour market policies, health and education having a long tradition of evaluation across countries.

7.2.1 The role of the centre and leadership in building evaluation capacity

The literature (PUMA/PAC, 1999, Mackay, 1998a, 1998b, Guthrie and English, 1997) suggests that evaluation has been well established across the Australian Public Service. For example, there is clear evidence of a high level of evaluative activity, with evaluations being used both to assist cabinet decision making (including budget prioritisation) and to support internal programme management (Mackay, 1998a). Evaluation has been systematically integrated into corporate and programme management and planning, with all public programmes or significant parts of them being reviewed every three to five years. In addition, all new policy proposals are required to have an evaluation component and the results of all major evaluations are expected to be made public and ministries are required to produce an annual evaluation

plan (PUMA/ PAC (1999). 'Evaluation in the Australian Government – as measured by the extent of evaluation planning, conduct and use – had achieved a healthy and vigorous state by the mid-1990s' (Mackay, 1998b, p.27).

The successful establishment of evaluation and its institutionalisation in Australia is largely attributed by Mackay (1998a, 1998b) to the role of the centre and leadership. He states that there has been sustained commitment and support for evaluation for more than a decade, and the process has been supported by an explicit government strategy and the ongoing development of public service infrastructure. Key success factors have included a combination of formal requirements for evaluation along with strong advocacy from the Department of Finance (a powerful central department), which have enabled evaluation to be linked to budget decision making and the on-going management of government programmes. The Department of Finance is also identified as a committed champion of evaluation, and has continually identified and created new opportunities for influence and development. For example, in 1987, it undertook a diagnostic study of departments' evaluation progress which identified: the lack of integration of evaluation into corporate and financial decision making; that rather than focusing on overall programme effectiveness, evaluations tended to focus on efficiency and process; a poor level of evaluation skills and analytical capacity; and that the role of central departments was unclear. This study led to the establishment of the requirement for a formal, ongoing evaluation strategy for all departments. In 1988, the Commonwealth Government launched an evaluation strategy. The objectives of the strategy were to: provide better information to assist managers in improving programme performance; assist government in decision making and prioritisation; and contribute to improved accountability to parliament and the public. The Department of Finance is the co-ordinator of the evaluation strategy and provides training, encouragement and guidance to support evaluations. However, the responsibility for planning and conducting evaluations and using evaluation findings is with individual departments and agencies.

In the Netherlands, in the early 1990s, the Court of Audit recommended that ministries pay more attention to programme evaluation. Accordingly a three-track strategy was established by government. This included: 1) setting out responsibilities of ministries for evaluation in regulations; 2) requiring ministries to provide information on current and future evaluation activities, findings and utilisation in annual reports accompanying ministerial budgets; and 3) setting out who is responsible for the management and co-ordination of evaluation at ministerial level. It was also emphasised that evaluation

should not be only an external activity but also internal and not only ex post but also ex ante and intermediate (PUMA/PAC, 1999).

In Canada, the centre has a specific role in overseeing evaluation across all government departments. Specific structures are in place to reinforce the responsibility for programme evaluation with key actors: the Office of the Comptroller General (OCG), who has system-wide responsibility for programme evaluation; the deputy heads of departments and agencies; evaluation units within departments; the Treasury Board; cabinet policy makers; and the Auditor General. In the 1970s and 1980s, the Treasury Board developed policy to ensure that all departments and agencies of the federal government periodically evaluated their programmes and this was supported with guidance on the evaluation function and the principles of evaluation. Evaluation is seen more as a management tool, 'one element in the management cycle', rather than an external 'scientific' study (Segsworth, 1990). The OCG provides oversight of programme evaluation within departments and agencies, ensuring that appropriate corrections and actions are taken when necessary and providing advice and assistance to departments on programme evaluation. Legislation also allows the Auditor General's staff to include value for money issues and the quality of evaluation capability and studies in assessments. The Treasury Board Secretariat requires departments and agencies to submit evaluation plans and summaries of evaluations as a part of the expenditure management process, and departments are expected to include indications of how and when new or enhanced activities are to be evaluated when making requests for new money.

7.2.2 From evaluation as a formal requirement to a voluntary approach

Mackay (1998a) identifies a recent shift in Australia from a system based on tight, formal controls and the mandatory requirement for evaluation, to 'more of a voluntary, principles-based approach' and explores the possible implications of this shift in thinking. He suggests that formal requirements for evaluation since the 1980s have helped to establish evaluation as a management practice and the acceptance of performance measurement as an integral part of performance management. He suggests that if this is reflected in greater incentives to achieve a stronger focus on performance and outcomes, rather than processes to achieve them, this will provide greater support for devolutionary processes that provide departments and managers with greater autonomy and flexibility to achieve this performance. In terms of evaluation, he suggests that the thinking is that 'if the environment of public sector governance is strongly conducive to evaluation being conducted and used, then it will

happen' (p.23). However, he suggests that there are risks attached and the success of these reforms will depend on the extent to which an evaluation culture has been achieved to date.

7.2.3 Building evaluation into reforms

Evaluation is a vital element of the results-based approach to management launched by the Canadian Government in 1995. This approach is based around ministers, senior officials and managers making 'decisions based on what a programme has achieved for Canadians – the results that citizens value – and at what cost'. The thinking is that 'Holding managers accountable for results encourages them to focus more on results' (Auditor General of Canada, 2000, pp. 20-8). The Auditor General emphasises 'evaluation as a vital element' and states that evaluation should be an integral part of programme management and managing for results. The specific role for evaluation is to provide important information on programme performance that is not gathered by ongoing monitoring systems, and to help managers understand why programmes are working or why they are not.

In Australia, reforms have given considerable autonomy to line managers and emphasise 'bottom-line results', thereby providing incentives to conduct and use evaluation findings. Mackay (1998a, 1998b) traces the development of evaluation capacity in the Australian federal government, where he suggests high priority has been given to ensuring that programmes are evaluated and the findings used. The reforms in the late 1980s and early 1990s emphasised the need to make managers manage, the importance of programme objectives being realistic to help guide managers and staff, and the need for a focus on programme performance through the collection of performance information and regular programme evaluation (Mackay, 1998b). The ongoing evaluation strategy for all departments introduced in 1988 had three main objectives: to encourage programme managers to use evaluation to improve programme performance; to provide fundamental information on programme performance to cabinet to aid decision making and prioritisation; and to strengthen accountability within a devolved environment by providing formal evidence of programme managers' oversight and management of programme resources (Mackay, 1998b).

Improving the system by which departments and agencies make decisions, manage and evaluate achievements, and enhancing public accountability and scrutiny, were key elements of the Financial Management and Improvement Plan (FMIP) introduced in

1984, along with budgetary and financial management reforms. Performance measurement and evaluation were also key elements of the growing 'managerialism' philosophy within the Australian Public Service (Guthrie and English, 1997).

In New Zealand, a purchaser/provider split has been pursued in the organisation of health services for some time, but more recently the emphasis has switched to co-operation. There have been a number of structural changes in the health system, including the Health Funding Authority being subsumed into the Ministry of Health and the establishment of twenty-three district health boards (DHBs) in 2001, with responsibility for the planning of health services to meet the needs of their populations. DHBs will take over assigned responsibilities for audit and monitoring from the ministry, but they may choose to contract some or all of the work out instead of using their own employees. This is aimed at enabling DHBs to address any initial concerns about audit and monitoring capacity. All organisations are responsible for their own financial management and actively managing risks that could jeopardise achieving their objectives (Ministry of Health, 2001b, 2001c).

The purpose of audit and monitoring identified by the Ministry (MoH, 2001c) is to: 'ascertain if appropriate service coverage is being achieved; measure progress towards the achievement of goals and identify whether any action is required to enhance the potential for meeting those goals (for example, the goals set out in the New Zealand Health Strategy) … ; ascertain if the provider is delivering the service being provided, and level of adherence to best practice service guidelines where applicable; identify problems in a timely fashion so that action can be taken to prevent further occurrences; check that any other data required under the service agreement … is being provided' (p.2).

The guidance explicitly states that a comprehensive programme of audit should be in place, which should be prioritised based upon a risk management approach in order to make the best use of limited resources. This would involve identifying each auditable area within the DHB's responsibility, for example, providers or service lines, and ranking them according to risk. In addition, a number of ad-hoc or issues-based audits may be necessary as problems are identified. Ministry of Health service directorates will provide guidance and information to DHBs to support them in audit and monitoring.

7.2.4 Linking evaluation to funding

Requiring evaluation in contracts

In Britain and New Zealand, where the purchaser/provider split has been pursued in recent reforms, performance measurement, quality assurance and evaluation are included in contractual requirements between funders and providers. The key accountability arrangements in New Zealand are the Crown funding agreements between the ministry and the DHBs and service agreements between DHBs and providers. DHBs are required to produce district strategic and annual plans that contain explicit objectives and performance targets against which progress can be measured. Under the New Zealand Public Health and Disability Act (2000), DHBs have an obligation to monitor the delivery and performance of services provided by the DHB and by those it engages to provide services, including both the quality of service provision and whether contractual obligations have been met. This will require DHBs to put in place processes for monitoring (Ministry of Health, 2001a). Under the Crown funding agreements, DHBs are required to take part in national reviews or audits and to collect particular sets of data on behalf of the Crown. Service providers will have primary responsibility for ensuring the safety and quality of services provided but the need for a collaborative effort between providers, funders and policy makers in achieving safer and better quality services is stressed.

Ear-marking funding for evaluation

Under US legislation, the Secretary of the Department of Health and Human Services is permitted to use up to 1 per cent of funding allocated to PHS programs for evaluation. Amounts available from the 1 per cent are determined on an annual basis and distributed among the PHS agencies for use, under guidance developed by the Assistant Secretary for Planning and Evaluation (ASPE), who also co-ordinates evaluations overall. Johnson (1996) reports a diverse range of approaches that have been employed on PHS evaluations, including outcome evaluations, impact evaluations, implementation/process evaluations, policy assessments, cost-benefit analysis/cost-effectiveness analysis, survey data analysis, management studies, evaluation syntheses, evaluation feasibility studies, evaluation designs and instrument development. PHS evaluation funds also support other activities, such as providing assistance to PHS agencies on any aspect of evaluation.

Funding evaluation to encourage innovation

Two examples are to be found of funding evaluation in such a way as to encourage innovative ways to deliver services. In Canada, the Health Transition Fund has been established, worth $150m. Based on the thinking that 'Continued evolution of the health system is essential if it is to be sustainable and responsive to the health needs of

Canadians' (Health Canada, 2001), the fund is to support projects across Canada in order to test and evaluate innovative ways to deliver health care services. Thirty million dollars have been allocated to national projects and the remaining $120m to projects at provincial and territorial levels. Four priority areas have been identified: home care, pharmacare, primary care reform and integrated service delivery.

The second example is the National Demonstration Hospitals Program in Australia. This is a collaborative benchmarking model enabling participating hospitals to compare their practices with those identified in lead hospitals and to identify innovative practices to be used in their own hospitals. A commonwealth grant was provided through contractual arrangements with consortia, and in turn to lead hospitals and collaborating hospitals, to cover management of consortium activities, project management at each collaborating hospital and at the lead hospital – to allow service enhancement and to expand, develop and conduct additional evaluation of these service models. The project is run in phases (currently in phase three), and clear benefits have been identified during the two phases completed so far. Phase one concentrated on three aspects of the management of elective surgery: patient pre-admission and admission processes; operating suite management; and discharge planning and post-acute care. 'Substantial gains' were reported from phase one, such as reduced length of patient stay, reduced unused operating theatre sessions, and increases in the number of patients admitted on the day of surgery (CDHAC, 1999). The focus of phase two was on bed management, because despite the gains made in phase one, elective surgery was still frequently cancelled because of the unavailability of beds. For further information on the project see: http://www.health.gov.au/hsdd/acc/ndhp/overview.htm.

7.3 Enhancing evaluation capacity

Boyle et al (1999) look at several of the issues around building evaluation capacity in public management, based on thirty years experience across countries attempting to build evaluation capacity in public policy making and to integrate evaluation and decision making. Linking evaluation capacity to the practice of evaluation, they describe evaluation capacity as the hardware – the human capital (skills, knowledge, experience etc.) and financial/material resources required to enable the 'doing' of evaluation'. Thus, the development of evaluation capacity is central to developing a results-based approach to management. They suggest that in any system, four elements are required for evaluation regimes to be successful and the development of evaluation capacity should be focused on these four elements. The four elements are:

- sound data systems that will provide good reliable data
- social science disciplines (such as political science, economics, sociology, management studies, public administration studies), providing the knowledge of evaluation methodologies and the thinking to set the scene for evaluation
- the presence of a trained cadre of trained analysts /evaluators
- the presence of good governance – such as, an ethics infrastructure, an effective legal framework and effective accountability mechanisms.

Boyle et al also suggest that the institutional context for evaluation is very important, and institutional anchorage influences the supply of and demand for evaluation. The development of evaluation capacity (the capacity to supply evaluation) alone does not ensure that the findings will be used, or indeed that evaluations will be conducted. Thus, 'this demand and supply link is crucial for evaluation use' (p.12). They also warn that this supply-demand model should not be interpreted in a simplistic manner. For example, there are different types of demand for evaluation as a management tool (such as the accountability/decision-making differences identified above), which will have implications for where supply is located, how it impacts on organisational learning and where and by whom it is used. 'Achieving a suitable balance between the demand for and the supply of evaluation becomes a key issue in evaluation institutionalisation' (p.12). Where demand for evaluation is high but supply (capacity) is weak, evaluation will not be done well. Where capacity exists but there is no demand for its use, the evaluation function remains weak (quoting the World Bank, 1994a). Evaluation can be said to be fully institutionalised where evaluation supply and demand are both strong and supporting each other.

Boyle et al (1999) identify seven key issues to be considered in attempts to institutionalise evaluation and to enhance the demand for and the supply of evaluation:

- *anchoring evaluation* – whether evaluation should be anchored in the legislative or executive branch, or both
- *anchoring evaluation within organisations* – the advantages and disadvantages of: a) a centralised approach and b) using internal versus external evaluators
- *evaluation coverage* – the kind of activities to be covered by evaluation
- *linking evaluation with other functions and institutions* – how and to what extent evaluation should interact with budgeting, auditing, strategic planning and monitoring

- *the use of evaluation in decision making* – 'what are the key domains of use, and how can evaluation utilisation be improved?'
- *professionalising the evaluation function* – to what extent evaluation should be a discipline in its own right, and what are the skills and competencies required of evaluators
- *fostering demand* – what are the most effective ways to foster demand for evaluation activity?

7.4 Generating evaluation demand

Toulemonde (1999), in order to identify how evaluation demand has been promoted, examines thirteen cases where evaluation systems have been created successfully. He describes his findings using the metaphor of 'carrots, sticks and sermons', previously used by Bemelmans-Videc et al (1998). The lesson emerging from the review is that all three instruments need to be applied together. However, this may be difficult to achieve in some cases resulting in a slower development of evaluation demand. Carrots refer to the creation of evaluation demand using incentives. For example, budgetary incentives, such as channelling money specifically for evaluation; and, career and turnover incentives, such as developing persons as full-time evaluators. In terms of career and turnover incentives, an example is given of how a programme to develop a pool of full-time evaluators in Canada, although initially aimed at the supply side, a few years later began to contribute to the demand side, as evaluators, as a relatively independent group of evaluation professionals, helped to export evaluation demand from one department to another. Toulemonde suggests that training professionals is not enough to generate evaluation demand, incentives are required to generate the interest in evaluation.

Various 'sticks' are identified by Toulemonde. One example is compulsory evaluation – the example given is the European Community Structural Funds, which require systematic evaluation. However, he also points out that if evaluation capacity is not fully developed, a rule of systematic compulsory evaluation is probably not workable. Another stick identified is to grant stakeholders, outside of the usual network of programme supporters, the right to ask evaluation questions. This can help to avoid the tendency whereby decision makers steer evaluation away from sensitive areas and censor critical questions. However, there is also the tendency for decision makers to ignore answers when they have not asked the question. The right to ask questions can be given to members of powerful and well respected groups such as the General

Accounting Office in the United States. Another stick identified is to require compulsory access to field data on programmes, so that evaluators can understand how impacts are made and make judgements about programmes. Toulemonde questions if it is possible to develop evaluation demand by making utilisation compulsory. Such an approach would require some serious threat (such as a budget cut) for those who do not comply. He suggests that based on experience, this approach is not likely to succeed and could result in decision makers fulfilling their legal obligations without any real commitment to evaluation. Commenting generally on the use of sticks in evaluation, he suggests that the best use of constraints is to use them sparingly, as a deterrent. 'Smart evaluation promoters … build up evaluation systems that are strongly based on power but whose superficial appearance is one of partnership and discussion' (p.159).

'Sermons' relate to the use of culture-building to enhance evaluation demand.

> Once this culture (evaluation culture) is well established, evaluation is deeply rooted in the administrative values, is seen as an undisputed duty and becomes one of the fundamentals of the governing system. The culture provides the collective pressure that makes decision makers overcome their reluctance, even when evaluation deeply contradicts their self-interest' (Toulemonde, 1999, p.167).

This process of developing an evaluation culture is described by Toulemonde as preaching faith, with the sermon as the relevant instrument. The process requires a sustained communication effort but, equally important is the content of the message. Sermons take the form of conferences, workshops, training courses, newsletters and journals. In addition, extensive use is made of demonstration projects, success stories, visits to good practitioners, prizes and awards. 'Those who listen to the sermon should be convinced that they belong to a community of people who trust that evaluation is part of sound public management' (p.167).

7.5 Building evaluation capacity in the Irish health system

This review of evaluation in the Irish health system identified a range of approaches to developing evaluation at national, regional and organisational level. However, the following observations suggest that evaluation is underdeveloped, opportunistic rather than systematic, and efforts lack coherence.

Until recently efforts at the national level were based on a small number of reviews (only six major reviews) initiated outside of the health system (Department of Finance, Comptroller and Auditor General). The commissioning by the Department of Health and Children of the Critique of the Health Strategy, the VFM review, and the review of beds capacity, all conducted in 2001, may be indicative of a sea-change in this regard.

At health board and provider levels, considerable variation is noted in terms of the approaches being taken to develop evaluation capacity, and the extent to which evaluation is currently used for accountability and to inform management decisions.

Most of those interviewed for the purposes of this research noted the lack of leadership and co-ordination from the Department of Health and Children in terms of evaluation and the absence of a national strategy on evaluation in the health services. It was suggested that evaluation priorities need to be cascaded down from the national level rather than conducted on the current ad-hoc basis, with every team and manager having a role in evaluation. It was also suggested that the approach taken should be collaborative rather than autocratic, if it is to support effective evaluation. While boards were reported to be working together in some areas, for example in the area of immunisation and the development of performance indicators, it was suggested that boards are generally doing their own thing and that they would benefit from national guidance and direction. It was also suggested that the department could develop a very useful role in identifying and sharing good practice, developing networks and the maintenance of central databases on studies undertaken. Such an activity could presumably be extended to include meta-evaluation in areas where significant evaluation is taking place.

The review of evaluation being undertaken abroad suggests that a central body such as the Department of Health and Children could also play a very important role in co-ordinating research in the area of health technology assessment and systematic reviews and disseminating results.

Co-ordination is also important to ensure that the findings from evaluations, regardless of the level of the health system at which they are conducted, find their way into decisions made at national level, and where appropriate, at local level. One interviewee also noted the absence of a national ‘think-tank’ on health policy, including evaluation. Perhaps the role of such a body could include building evaluation findings into policy.

7.5.1 Challenges to building evaluation capacity and evaluation demand

Several issues/challenges relating to building evaluation capacity in the Irish health system were identified by those interviewed and through the comparative review. The issue most mentioned was lack of good information on performance, relating both to an inadequate information infrastructure and the lack of an information management culture. For example, the absence of base-line data is a problem for someone evaluating an intervention, and some areas of the health services are not networked and only hard copies of information are kept. Under-investment in IT, incompatibility between data and systems, and the lack of an ICT strategy were highlighted as particular issues. Less than half of a per cent of current investment in health services is on information technology.

Similar issues identified in the new *Health Strategy* (Department of Health and Children, 2001a) include:

> ... inadequate and poorly integrated information systems to support the measurement of inputs and outcomes on a quantitative or qualitative basis in the health system; insufficient investment in the development of intellectual and organisational capacity to carry out comprehensive research and analysis of policy options; lack of an overriding national structure for the development, dissemination and evaluation of the impact of agreed national quality protocols and standards; … concerns about a 'blame' culture in which quality audits and evaluations make individual practitioners feel isolated and vulnerable. High-standard, well-integrated and reliable information systems are central to quality. While a number of good information systems exist or are being developed, the ability to identify health needs or to evaluate equity, efficiency, effectiveness and overall quality of health services is limited. This is due, in part, to inadequacies in the availability, quality and integration of health information systems (p.50).

However, it was also suggested in this study that where information is available, it might still not be used in decision making. One interviewee identified two particular examples where similar recommendations were made in subsequent reports, suggesting that change had not occurred in the intervening period. It was suggested that allocations should be made on the basis of sound practice and needs assessment, but such information is either not available or does not find its way into allocation

decisions. On a related theme, there can be pressures for services instigated on a pilot basis to be extended in advance of being evaluated. The example given was 'Breastcheck', which was extended nationwide before the model of care and quality of the service provided in the pilot schemes could be evaluated. It was suggested that often the role of evaluation is seen as a mechanism to justify the introduction of a service rather than a tool to develop models of care based on evaluation of pilot schemes.

The need to develop evaluation expertise to ensure credible and rigorous evaluation was also identified and it was suggested that this would require significant investment in training and developing learning opportunities. It was suggested that organisations such as the Office for Health Management and the Institute of Public Administration could provide valuable input into developing evaluation expertise. In addition, it was suggested that there could be a role for the Health Research Board where significant evaluation expertise already exists. Details of an interesting approach taken to developing evaluation expertise by the American Cancer Society (ACS) are reported by Compton et al (2001). Their 'Collaborative Evaluation Fellows Project' (CEFP) is described as 'an innovative program designed to develop evaluation capacity within the ACS and to study the outcomes and effects of its programs, while providing program evaluation and training for graduate students in public health'. The CEFP is a national programme whose purpose is to establish stronger links between the ACS and university-based professional training in programme evaluation. Graduate students are selected to become CEFP evaluation fellows for one year and during this time gain 'practical, real-world cancer-focused program evaluation experience' on grant-funded projects. Over the three years that the programme has been running, it has resulted in the development of partnerships and it is anticipated that programme evaluation will be demystified for ACS staff and that the programme will assist in the development of evaluation courses in schools of public health. Similar exchange schemes could be developed between Irish health service managers and professionals and organisations such as the Health Research Board, universities and consultancies.

7.6 Conclusions

This review suggests that the establishment of evaluation and the development of evaluation capacity are critical elements in the development of evaluation in Irish health services, along with the development of evaluation demand. Several key challenges

are also identified that will need to be addressed towards the development of an evaluation framework.

8
Conclusions

8.1 Introduction

This study set out to explore the importance of evaluation in the Irish health services and to review current approaches to evaluation. The intent was to identify areas where evaluation capacity could be further enhanced, and by way of practical support, to explore the range of approaches to conducting evaluation and the circumstances in which each would be useful in health service management. In this chapter, conclusions are drawn on each of these topics.

8.2 The importance of evaluation in the Irish health context

The case is made very early in this report for the importance of evaluation in the Irish health context in terms of accountability, providing information to support effective decision making, and to enable organisations to grow as learning organisations. The perception of a health service crisis promulgated in the media, as explored in Chapter Two, goes some way towards describing current views on the state of Irish health services and, by inference, the quality of health service management. One problem that has existed for some time is that the information is not available for health service managers and policy makers to demonstrate that progress has been made over recent years to improve health services, or, to show that patients receive a good standard of care. In addition, the criteria used by the media in their judgement of the overall performance of health services are limited to a small number of emotive issues, such as the length of waiting lists, waiting times and facilities in accident and emergency units and staff shortages. While these are important issues, especially to those who are affected by them, a greater emphasis is required on all of the other aspects of health care delivery and on in-depth analysis of the factors contributing to problems identified, in order to accurately assess the performance of health services and identify issues to be addressed. For instance, how are waiting lists managed and how does this contribute to the length of time people have to wait? What do we know about the management of outpatient services, where patients often wait for a considerable period of time to be seen, and the quality of the interaction and care that they receive there?

What do we know about the quality of care that patients receive once they are admitted to hospital via a waiting list? What do we know about the availability and use of hospital beds that contribute to some degree to bottlenecks in accident and emergency? Can we be sure that the most effective use is being made of the limited funding available? Can we be sure that patients are receiving the most effective treatments available? These are all questions that can be answered through evaluation.

While it is important to be able to demonstrate that health services are effectively managed and that resources are used effectively, evaluation (and its twin, performance measurement) also has tremendous potential as a tool for effective management. The information and insight produced through evaluation and performance measurement can enable managers to know that projects are progressing as they need to; to know that services are of a good quality, meeting the needs of users, and that they are effective; to identify what works well and what does not; and, to understand the factors contributing to good or poor performance. The process of introspection facilitated by evaluation also contributes to learning within an organisation and among individuals. In addition, examples are given of how evaluation can be used to develop innovative approaches to care and service delivery that if successful can be applied more widely.

8.3 Key issues to be considered when undertaking an evaluation

Several pointers on issues to be considered when undertaking an evaluation are spread throughout this report, drawing on current thinking on good practice and based on experiences in other countries. These are drawn together in the following section.

The first point to consider is who is the audience for the evaluation and what information needs they are likely to have. This point is seen as critical to ensuring evaluation findings get used. While this sounds straightforward, the reality in health service evaluation is that there will be many stakeholders, with sometimes very different needs, and these needs may be competing or even conflicting. While the evaluation literature promotes the importance of participative evaluation, it also cautions the evaluator against trying to please too many stakeholders at once, such that the focus of the evaluation becomes so diluted that the evaluation fails to meet the needs of any stakeholder, and that the evaluation becomes impossible to conduct within the time and resources available. Also, different stakeholders will have different

interests in evaluation, asking very different questions, and requiring different evaluation methods.

The utility of evaluation is another major theme in the literature on evaluation. In addition to identifying the primary intended users of evaluation and their information needs, it is suggested that the active involvement of these primary intended users in the design and conduct of the evaluation will also enhance utility by developing understanding and ownership of the process and findings. The profile of the evaluation and its implications for stakeholders can be raised through promotional events, through seminars and workshops and through local networks and newsletters. Some of these activities can also serve as information gathering events, perhaps helping to inform the identification of evaluation questions and the design of the evaluation study. Networking can also be used to identify other studies or events that are likely to impact on the evaluation, or to influence the relevance of findings. The time taken to conduct an evaluation can also influence the utility of the findings to stakeholders. Interviewees participating in this study suggested that for evaluation studies to be useful in the decision-making process, findings need to be available at key decision-making times and also that they need to be conducted within a reasonably short time frame to ensure findings are both current and timely. Once the evaluation is conducted, wide dissemination of findings, particularly amongst stakeholders, is vital to promoting the use of findings.

The evaluability of programmes is also a major consideration in deciding when to conduct an evaluation, and further, if an evaluation is appropriate at all. There is no point evaluating a programme or project before there has been sufficient time for it to have an effect. The evaluability of programmes/projects can be enhanced also by ensuring that clear aims, objectives and timescales are set out in plans. Evaluation can also be enhanced by outlining the requirement to evaluate the programme at specific times over the course of the programme and by outlining the methods to be used.

In deciding which evaluation methods are to be used, consideration should be given to which methods are most likely to answer the evaluation questions within the time and resources available. Thus, once the primary intended users are identified, a set of evaluation questions will need to be drawn up. Five key types of questions are identified in Chapter Four: rationale for an intervention, continued relevance, effectiveness, efficiency, and impact. Also, as outlined in Chapter Four, different methods will provide different types of information and often in evaluation more than

one method is used. As emphasised in Chapter Five, the credibility, and as a result the extent to which findings are likely to be used subsequently, will depend to a large degree on judgements about the appropriateness of the methods chosen. The purpose of the evaluation must be to produce maximally useful data. Decisions on which methods to be used will need to be based on the capacity of the evaluation to produce adequate evidence, reliable data, and clear findings. Decisions will also need to relate to the expertise, funding and time available. The criteria identified by the European Commission for good evaluation, outlined in Chapter Five, are useful when considering the value of various evaluation approaches. They state that evaluation should be analytical, reliable, issue-oriented, user-driven, and transparent and objective.

8.4 Where to now? Building a framework for effective evaluation

One dominant theme in this review of current evaluation practice in Irish health services is the need to develop a more coherent and co-ordinated approach to evaluation across the health system. This review suggests evaluation has yet to become firmly established in health policy and planning, and in health care management in general. The current demand for evaluation is patchy and largely based around accountability. Different approaches are being used within the health boards to build evaluation capacity and to support evaluation.

In moving towards the development of a coherent and co-ordinated approach to evaluation, the basis for a framework for evaluation is presented in Chapter Six in Figure 6.1. Several of the elements identified are referred to in the new health strategy, *Quality and Fairness: A Health System for You*, (Department of Health and Children, 2001a), in particular:

Performance indicators

- strengthening the accountability of health boards in the service planning process and the monitoring of service plan objectives by the Department of Health and Children
- the development of service agreements containing performance indicators between health boards and providers

External evaluations

- establishing the Health Information and Quality Authority (HIQA) with an explicit role in external evaluation

- developing hospital accreditation further
- establishing the social services inspectorate on a statutory basis

Developing a CQI approach

- establishment of a formal monitoring and evaluation function within each health board to monitor progress against targets and to evaluate outcomes over the medium term
- developing hospital accreditation and a quality assurance programme

Collaboration and support to enhance evaluation

- establishing/expanding a role for HIQA, the Health Board Executive (HeBE) and the Office for Health Management as change agents
- recognising and promulgating the view that decision making must be evidence-based and quality must be central to all aspects of health care management and delivery
- promoting the need for a co-ordinated national approach to quality management
- a quality programme, to include staff training
- development of national guidelines and protocols, evidence-based practice and health technology assessment
- development of health information and integrated health information systems.

In addition, a role for ex-ante evaluation is established in the development of needs assessment and policy proofing. This is to be undertaken by a new division of population health to be established within the Department of Health and Children, working closely with population health functions to be developed within health boards.

The establishment of evaluation is discussed in considerable depth in Chapter Seven. This considers both how an emerging framework for evaluation helps to promote evaluation, and how evaluation supply and demand can be developed. PUMA/PAC (1999) outlines Boyle's (1997) assessment of what needs to be done to systematise evaluation in public services in general in Ireland. In the light of the discussion on the various aspects of evaluation in this study, Boyle's assessment seems applicable to developing effective evaluation in the health services also.

> According to Boyle (1997), lessons from other countries suggest that a range of evaluators is needed (semi-independent evaluation bodies in the administration may be a good model). Evaluation priorities have to be set:

> comprehensive reviews conducted every three years may be too challenging in the early stages. User perspective has to be involved when possible. Training and development support is needed for evaluators and users, both 'hard' and 'soft' skills are important. Stimulation of effective demand – sticks, carrots and sermons, earmarking resources, specification of questions by those commissioning evaluations – is necessary. Finally, linkages between budgeting, strategic management and evaluation have to be established (PUMA/PAC, 1999, p.41).

These views are largely consistent with those outlined in Chapter Seven in relation to the establishment of evaluation in the Irish health services. PUMA/PAC (1999, p.7) also identify the need for top-level support for evaluation – 'Support for evaluations is demonstrated through willingness of politicians, policy managers and central management agencies (e.g. Ministry of Finance), to make effective use of policy advice generated in evaluations'. Emphasising the importance of consultation, they also suggest that evaluation without ownership is unlikely to have an effect and that consultation can help to overcome internal resistance to evaluation. They identify a role for government in supporting an evaluation culture that encourages innovation and adaptation to a changing environment – 'The basic message should be that to stay relevant, organisations need to continue learning from feedback about results … training and professional dialogue, competent evaluators, well-informed commissioners and enlightened and enthusiastic users all contribute to an evaluation culture'.

8.5 Conclusions

This study focuses on two key themes in evaluation – evaluation practice (the doing of evaluation) and the development of a framework for health service evaluation (generating the demand and supply of evaluation). A key function of the study was to identify the main issues to be considered by anyone interested in conducting an evaluation of health services. However, most of what is discussed could apply to any public service evaluation. The review of different approaches is intentionally light, aimed at providing an overview of the key features of the range of approaches with an indication of the types of situations where their use would be appropriate. Further information can be obtained by accessing the references listed.

To focus on the practice of evaluation alone would be to suggest that evaluation exists in isolation of the structures, incentives and culture required to support effective

evaluation. It is those factors that ensure that evaluation findings result in the progressive improvement of health service management and the delivery of health care. Thus, a substantial focus of this research was on the developing framework for evaluation in the Irish health service context, and identifying the key challenges to building evaluation capacity.

References

Abel-Smith, B., Figueras, J., Holland, W., McKee, M., and Mossialos, E. (1995), Choices in Health Policy: An Agenda for the European Union, Luxembourg: Office for Official Publications of the European Community

Auditor General of Canada (2000), *Report of the Auditor General of Canada*, February 2000

Baird, M. (1998), 'The Role of Evaluation', In Mackay, K. (ed) *Public Sector Performance – The Critical Role of Evaluation*, Washington, D.C: World Bank

Baker, J.L. (2000), E*valuating the Impacts of Development Projects on Poverty: A Handbook for Practitioners,* Washington, DC: The World Bank

Bastoe, P.O. (1999), '*Linking evaluation with strategic planning, budgeting, monitoring, and auditing*, In Boyle, R. and Lemaire, D. (1999), (eds) *Building Effective Evaluation Capacity: Lessons from Practice*, New Brunswick: Transaction Publishers

Bemelmans-Videc, M.L., Ris, R.C., Vedung, E. (1998), (eds) *Carrots, Sticks and Sermons: Policy Instruments and Their Evaluation*, New Brunswick: Transaction Publishers

Boyle, R. (1997), *Evaluating Public Expenditure Programmes. A Role for Programme Review,* CPMR Discussion Paper No. 1, Dublin: Institute of Public Administration

Boyle, R. (1996), 'Implementation of the Government Performance and Results Act of 1993, Commentary', Article in OECD, *Performance Management in Government,* Public Management Occasional Paper No. 9, Paris: OECD

Boyle, R., Lemaire, D., Rist, R.C. (1999), 'Introduction: Building Evaluation Capacity, In Boyle, R. and Lemaire, D. (1999), (eds) *Building Effective Evaluation Capacity: Lessons from Practice,* New Brunswick: Transaction Publishers

Building on Experience: National Drugs Strategy 2001-2008, Dublin: Stationery Office

Burt, M., Harrell, A., Newmark, L., Aron, L., Jacobs, L. (1997), *Evaluation Guidebook: Projects funded by S.T.O.P. formula grants under the Violence Against Women Act,* Urban Institute: Washington

Butler, M. (2000), *Performance measurement in the health sector,* CPMR Discussion Paper No.14, Dublin: IPA

Butler, M. and Boyle, R. (2000), S*ervice planning in the health sector*, CPMR Discussion Paper No.13, Dublin: IPA

Cairns, J. (1998), E*conomic Evaluation and Health Care*, Nuffield Occasional Papers Health Economic Series: Paper No.7, London: Nuffield Trust.

CDHAC (1999a), A*n Overview of Health Status, Health Care and Public Health in Australia*, Occasional Papers Series No. 5, Commonwealth Department of Health and Aged Care: Canberra

CDHAC (1999b), *A Qualitative Review of NDHP Phase 2,* Commonwealth Department of Health and Aged Care: Canberra

Chelimsky, E. (1985), 'Old patterns and new directions in program evaluation', In Chelimsky, E. (ed) *Program Evaluation: patterns and directions,* Washington, DC: American Society for Public Administration

Chen, H-T. (1990), *Theory-driven evaluations*, Thousand Oaks: Sage

Chief Medical Officer (1999), *Annual Report of the Chief Medical Officer,* Dublin: Department of Health and Children

Coldren, J.R., Bynum, T., Thome, J. (1989), *Evaluating Juvenile Justice Programs: A Design Monograph for State Planners,* Champaign, IL: Community Research Associates

Compton, D., Baizerman, M., Preskill, H., Reiker, P., Miner, K. (2001), 'Developing evaluation capacity while improving evaluation training in public health: the American Cancer Society's Collaborative Evaluation Fellows Project', *Evaluation and Program Planning*, 24, pp.33-40

Comptroller and Auditor General (Amendment) Act, 1993, Dublin: Stationery Office

Cooksy, L.J., Paige, G., Kelly, P.A. (2001), 'The Program Logic Model as an integrative framework for a multimethod evaluation', *Evaluation and Program Planning*, 24: pp. 119-128

Court of Audit (1996) *Manual Performance Audit*, The Policy Department, Court of Audit: Netherlands. www.rekenkamer.nl

Cronbach, L. (1987), 'Issues in planning evaluations', In Murphy ,R. and Torrance, H. (eds) *Evaluating Education: Issues and Methods,* Harper and Row

CSF Evaluation Unit (1999), *Review of ongoing evaluation function on the Community Support Framework (CSF) for Ireland 1994-1999,* CSF Evaluation Unit: Dublin

DAC (Development Assistance Committee) (2000), *Results Based Management in the Development Co-operation Agencies: A Review of Experience*, A Background Paper, DAC Working Party on Aid Evaluation, Paris: OECD Development Cooperative Directorate

Davies, I.C. (1999), 'Evaluation and performance management in government' *Evaluation* 5(2): pp. 150-159

De Vaus, D.A. (1996), *Surveys in social research*, (Fourth edition), UCL Press: London

Deloitte and Touche (2001), *Audit of the Irish Health System for Value for Money,* Dublin: Deloitte and Touche, in conjunction with the York Health Economics Consortium

Department of Health (1986) *Health – The Wider Dimensions*, Dublin: Stationery Office

Department of Health (1997), *The new NHS: Modern. Dependable*, London: HMSO

Department of Health and Children (1998), *Working for Health and Well-Being: Strategy Statement 1998-2001*, Dublin, Department of Health and Children

Department of Health and Children (2000) *Expenditure Review Dental Treatment Services Scheme*, Dublin, Department of Health and Children

Department of Health and Children (2001a) *Quality and Fairness: A Health System for You*, Dublin: Stationery Office

Department of Health and Children (2001b) *Progress report on the Statement of Strategy.* Department of Health and Children: Dublin

Derlein, H.-U. (1990), 'Genesis and structure of evaluation efforts in comparative perspective', In Rist, R. (ed) *Program Evaluation and the Management of Government,* New Brunswick: Transaction Publishers

Divorski, S. (1998), 'Evaluation in the Federal Government of Canada', In Mackay, K. (ed) *Public Sector Performance – the Critical Role of Evaluation: Selected Proceedings from a World Bank Seminar,* Washington, DC: The World Bank

Doig, B. and Littlewood, J. (1992), *Policy Evaluation: The Role of Social Research,* HMSO: London

Drummond, M.F., O'Brien, B., Stoddart, G.L. and G.W. Torrance (1997), *Methods for the Economic Evaluation of Health Care Programmes*, Oxford Medical Publications, Second Edition, Oxford: Oxford University Press.

Endacott, R. (1994), 'Objectivity in Observation', *Nurse Researcher*, Vol. 2, No.2, pp.30-40

European Commission (1997), *Evaluating EU Expenditure Programmes, A Guide, Ex-post and Intermediate Evaluation*

European Commission (2000), *Good Practice Guidelines for the Management of the Evaluation Function*, Brussels: Network of Evaluators of the European Commission

Expenditure Review Dental Treatment Services Scheme (1999), Dublin: Department of Finance

Fox, J. and Benzeval, M. (1995), 'Perspectives on social variations in health', In Benezval, M., Judge, K. and Whitehead, M., *Tackling Inequalities in Health: An Agenda for Action*, London: King's Fund

Gray, A., Jenkins, B. and Segsworth, B. (1993), *Budgeting, auditing and evaluation: functions and integration in seven governments,* Transaction Publishers: New Brunswick

Greene, J.C. (1994), 'Qualitative program evaluation: practice and promise', In Denzin, N. and Lincoln, Y. (eds) *Handbook of qualitative research*, London: Sage

Greene, J.C. (1999), 'The inequality of performance measurements', *Evaluation*, Vol.5, No.2, pp.160-172

Guthrie, J. and English, L. (1997), 'Performance information and programme evaluation in the Australian Public Sector', *International Journal of Public Sector Management,* Vol.10, No.13, pp. 154-164

Harrell, A., Burt, M., Harty, H., Rossman, S., Roth, J., Sabol, W. (1996), *Evaluation strategies for human service programmes: A guide for policymakers and providers*, The Urban Institute:Washington.

Health (Amendment) Act (No.3) 1996, Dublin: Stationery Office

Health (Eastern Regional Health Authority) Act, 1999, Dublin: Stationery Office

Health Canada (2001), *The Health Transition Fund,* www.hc-sc.gc.ca/htf-fass/english/whatwedo_e.htm

HFS (1997), *National Leadership Through Performance Assessment*, HFS Occasional Papers Series, No.1, Canberra: Commonwealth Department of Health and Family Service

HSASG (2001) Newsletter, May 2001, Dublin: *Health Services Accreditation Implementation Steering Group*

Hummelbrunner, R. (2000), *A systems approach to evaluation: Applications of systems theory and systems thinking in evaluation,* Paper presented for the 4th European Evaluation Society: Lausanne

Hutton, J. (1994), 'Economic evaluation of healthcare: a half-way technology', *Health Economics*, No.3, pp.1-4

Hyndman, N.S. and Anderson, R. (1997), 'A Study of the Use of Targets in the Planning Documents of Executive Agencies', *Financial Accountability and Management*, Vol.13, No.2, pp.4267-4424

Irish Times, (2000a) 'An Unhealthy State', October 2-6, 2000

Irish Times (2000b), 'Services suffer despite spending', October,3, 2000

JCSEE (1994), *The Program Evaluation Standards*, (Second Edition), Thousand Oaks: Sage

Johnson, P.L. (1996), 'Evaluation of US Public Health Service Programs: Organisation and Management', *Evaluation and the Health Professions,* Vol.19, No.3, pp.311-324

Lynch, F. (1998), 'Health Funding and Expenditure in Ireland', In McAuliffe, E. and Joyce, L., *A Healthier Future? Managing Healthcare in Ireland,* Dublin: Institute of Public Administration

Mackay, K. (1998a), 'The Development of Australia's Evaluation System', In Mackay, K (ed), *Public Sector Performance – the Critical Role of Evaluation: Selected Proceedings from a World Bank Seminar,* Washington, DC: The World Bank

Mackay, K. (1998b), *The Development of Evaluation Capacity in the Australian Government*, Washington, DC: The World Bank

Marczak, M. and Sewell, M. (1991), U*sing Focus Groups for Evaluation,* CYFERNet Evaluation: University of Arizona

Mayne, J., Divorski, S., Lemaire, D. (1999), 'Locating evaluation: Anchoring evaluation in the executive or the legislature, or both or elsewhere?' In Boyle, R., and Lemaire, D., (1999), (eds) *Building Effective Evaluation Capacity: Lessons from Practice,* New Brunswick: Transaction Publishers

Mayne, J., Ulrich, M. (1998), *Modernizing Accountability Practices in the Public Sector*: Discussion Draft, Ottawa: Office of the Auditor General/Treasury Board Secretariat

McKeown, K. (1999), 'Evaluation: A guide to its language and its logic', *Administration* Vol.47, No.1, pp.50-77

Millar, A., Simeone, R., Carnevale, J. (2000), 'Logic models: a systems toll for performance management', *Evaluation and program planning*, Vol.24, pp.73-81

Ministry of Health (2001a), *DHB Guidance Information: New Zealand Public Health and Disability Act 2000*, Ministry of Health: Wellington

Ministry of Health (2001b), *Audit and Monitoring Responsibilities,* http://moh.govt.nz/moh.nsf/

Ministry of Health (2001c), *Audit and Monitoring: Guidance for DHBs,* http://moh.govt.nz/moh.nsf/

NHIS (2001) *Health Information*, Early Draft Discussion Document, 1 June 2001, www.doh.ie/hstrat/nhis/
wihi.html

NHPC (2001), *National Health Performance Framework Report*, Brisbane: Queensland Health

NSF (1997), *User-friendly handbook for mixed method evaluations,* Virginia, USA: The National Science Foundation, Directorate for education and human resources

O'Connor, A. (2001), 'McCreevy takes tough line at health meeting', *Irish Times,* May 15, 2001

Olsen, J. P. (1993), 'Et statsvitenskapelig perspektiv pa offentlig sektor', In Laegreid, P. and Olsen, J. P., *Organisering av offentlig forvattning*, Oslo: Tano

O'Morain, P. (2001), 'Martin to outline vision for creaky health system', *Irish Times,* March 6, 2001

Patton, M.Q. (1990), *Qualitative Evaluation and Research Methods,* (Second Edition), Newbury Park: Sage

Patton, M.Q. (1997), U*tilization-focused evaluation*, (Third Edition), Sage: Thousand Oaks

Pawson, R., Tilley, N. (1997), *Realistic evaluation,* Sage: London

Pollitt, C., O'Neill, H. (1999), A*n Evaluation of the process of evaluation of EC external aid programmes*, Bruxelles: Collectif d'échanges pour la technologie appropriée – COTA

Public Service Management Act, 1997, Stationery Office: Dublin

PUMA/PAC (1999), *Improving evaluation practices: best practice guidelines for evaluation and background paper,* Paris: OECD

Report of the Commission on Health Funding (1989), Dublin: Stationery Office

Rist, R.C. (1990), 'Managing of evaluations or managing by evaluations: choices and consequences', In Rist, R.C. (ed) *Program evaluation and the management of government: Patterns and prospects across eight nations,* New Brunswick: Transaction Publishers

Rossi, P.H., Freeman, H.E. (1993), *Evaluation: A systematic approach,* Newbury Park: Sage

Saltman, R. and Figueras, J. (1997), *European Health Care Reform. Analysis of Current Strategies,* WHO Regional Publications, European Series, No.72, Copenhagen: World Health Organisation

Saltman, R.B. (1997), 'The context for health reform in the United Kingdon, Sweden, Germany and the United States', *Health Policy*, Vol. 41, Suppl., pp.S9-S26

Schick (1996), *The Spirit of Reform: Managing the New Zealand State Sector in a Time of Change*, Wellington: New Zealand State Services Commission

Segsworth, R.V. (1990), 'Policy and programme evaluation in the government of Canada', In Rist R. C. (ed) P*rogram evaluation and the management of government: Patterns and prospects across eight nations,* Transaction Publishers: New Brunswick

Shaping a Healthier Future: A Strategy for Effective Healthcare in the 1990s, Department of Health (1994), Dublin: Stationery Office

Shaw, I.(1997), 'Evaluation in health and social care: Exploring lost dimensions', *Evaluation*, Vol. 3, No.4, pp.469-480

Stake, R.E. (1994), 'Case studies', In Denzin, N. and Lincoln, Y. (ed) *Handbook of Qualitative Research*, London: Sage

Toulemonde, J. (1999), 'Incentives, constraints, and culture-building as instruments for the development of evaluation demand', In Boyle, R. and Lemaire, D. (1999), (eds) *Building Effective Evaluation Capacity: Lessons from Practice,* New Brunswick: Transaction Publishers

University of Ottawa (1997), *Program evaluation tool kit,* Ottawa: Community Health Research Unit, www.uottawa.ca/academic/med/epid/what.htm

Weidman, D., Walker, J., MacNeil, D., Tolson, F., Wholey, J. (1975), *Intensive evaluation for criminal justice planning agencies,* US Department of Justice, Law Enforcement Assistance Administration, National Institute of Law Enforcement and Criminal Justice

Weiss, C.H. (1987), 'Where politics and evaluation research meet', In Palumbo, D.J. (ed) *The Politics of Program Evaluation,* Newbury Park: Sage

Weiss, C. H. (1998), *Evaluation*, Upper Saddle River, New Jersey: Prentice Hall

WHO (1999), *Health Impact Assessment*, Gothenburg consensus paper, Brussels: WHO Regional Office

World Bank (1994a), *Report of the Evaluation Capacity Development Task Force*, Washington, DC: World Bank

World Bank (1994b), *Governance: The World Bank's Experience,* Washington, D.C: World Bank

Wren, M.A. (2000), 'Health system needs urgent funding', *Irish Times*, October 3, 2000

Notes

1 Health Research Report (Health Research Board, 2001); *Quality and Fairness: A Health System for You* (Department of Health and Children, 2001); Deloitte and Touche (2001), the *Review of Acute Bed Capacity*; and the *Health Information Strategy* (forthcoming).

2 The Hawthorne effect describes a situation where observed improvmeents in performance are the result of being observed rather than the results of an intervention.

3 '... a process of legitimisation by which evaluation formally becomes part of the decision-making process ...' (Boyle et al, 1999, p. 5).

Appendix One

Summary of the expenditure review of the dental treatment services scheme

The Dental Treatment Services Scheme (DTSS) was introduced by the Department of Health in 1994 aimed at providing more effective dental services to adult (over sixteen years of age) medical card holders. The scheme is primarily delivered by private dental practitioners, following the transfer of responsibility for the provision of dental treatment from health board dentists to the private dental sector, and is administered by health boards and the General Medical Services (Payments) Board. The review of the dental services, with a particular focus on the DTSS, began in 1999 under the Department of Finance's programme of expenditure reviews for all government departments. The DTSS was targeted for review because the service was seen as being well defined and over recent years had been evolving, become more complex and demanding increased resources. The aim of the evaluation was to 'determine the value of the DTSS to see if it has been carried out as prescribed and to discover whether the required performance and objectives have been achieved' (DoHC, 2000).

The DTSS was introduced on a phased basis, initially aimed at the provision of routine dental services for persons over sixty-five years and older, and later extended in 1996 to provide routine dental services for medical card holders aged sixteen to thirty-four, and the provision of full dentures to all medical card holders without teeth. In addition, priority for routine treatment may be given to cases where there is a serious medical condition which could be aggravated by poor dental health. Currently, it is envisaged that the scheme will be extended to the remaining age groups in accordance with the level of funding available.

The review begins by outlining the genesis of the DTSS and previous attempts to provide dental services to adults with low incomes, comparing dental services to those in a small number of other European countries and reviewing epidemiological trends worldwide. Differences in oral health in Ireland between medical cardholders and those who do not have medical cards, between health boards and between persons in different social classes are also explored.

The review then outlines the health strategy, the dental health action plan and the oral health goals set. The Dental Health Action Plan was one of the first major initiatives to fall out of the 1994 *Health Strategy* (*Shaping a Healthier Future*). It sets out, for

the first time, specific objectives for the dental services, a timescale within which they are to be achieved and methods to achieve them. In addition, a number of goals are identified to be achieved by the year 2000 (see Figure One).

Figure One The Dental Health Action Plan – objectives, methods and goals

Objectives	Methods	Oral Health Goals
• Reduce level of dental disease in children • Improve level of oral health in population overall • Provide adequate treatment services to children and all medical card holders • Equity, accountability and quality applies to all objectives	• *Investigative strategy* – oral health database for monitoring changes in oral health • *Preventative strategy* – improve water fluoridation; school-based fluoride mouth rinsing in low fluoride areas; promote use of fluoride toothpaste for adults and children; increased application of fissure sealants to children; oral health education in media, healthcare and educational settings • *Treatment strategy* – phased extension of eligibility; systematic screening and treatment of children; development of specialised services in orthodontic, oral surgery and paediatric dentistry; phased introduction of new treatment services for 1million medical card holders • *Training strategy* – dental schools to produce appropriate mix and numbers of dentists, consultants and auxiliary dental workers to meet needs of services; provision of opportunities for post-graduate and continuing education including management training	• At least 85% of 5 year olds in optimally fluoridated areas (at least 60% in less optimally fluoridated areas) will be free of dental caries (baby teeth only) • 12 year old children will have on average no more than 1 decayed, missing or filled permanent tooth in optimally fluoridated areas (no more than 2 in less than optimally fluoridated areas) • The average number of natural teeth present in 16-24 year olds will be 27.7 (compares to current average of 27.2) • No more than 2 % of 35-44 year olds will have no natural teeth • No more than 42% of people aged 65 years and over will have no natural teeth

In the chapter following in the report, the objectives, structure, operation, role and key performance indicators of the DTSS are outlined. The objectives of the DTSS are to:

> ... improve the oral health of adult medical card holders and thereby reduce the equity gap between this population and the population as a whole by providing a high quality dental service, provide dental services to medical card holders in a cost-effective and equitable manner.

Under the DTSS, services are provided to eligible adults by private dental practitioners under contract arrangements with health boards, and some health board dentists. Services are currently provided under three distinct schemes: emergency treatment – available to all eligible adults; routine treatment – available to sixteen to thirty-four

year olds and over sixty-five year olds; and the full denture scheme – available to all persons with no natural teeth. Emergency treatment is available on demand but routine and full denture treatment is subject to prior approval by the health board.

The review also outlines the nature of the contract agreement between health boards and dental practitioners, the role of the GMS Payments Board, the statutory basis of the relationship between the GMS and the health boards and the operation of the contract. The DTSS is then compared to the Dental Treatment Benefits Scheme (DTBS), which is a similar scheme run by the Department of Social Welfare (now the Department of Social, Community and Family Affairs) and through which persons are eligible for some dental services through Pay Related Social Insurance payments. It suggests that the operation of two separate schemes is not cost-effective and has given rise to inequalities in the dental health of the populations who are eligible for the two schemes. In addition, it is suggested that the average cost per patient using the DTBS (excluding patient charges) is almost half that per patient using the DTSS.

The changing role of the health board dental service is reviewed following the implementation of the DTSS. The new remit identified includes monitoring and evaluating the oral health status of health board populations, implementing and evaluating preventative programmes for the whole community, providing all the necessary dental care for children up to sixteen years of age and to special needs groups in the community, and monitoring the DTSS at the local level. Health board dentists will have a role in these areas and in addition, the provision of out of hours services, the provision of dentures to persons over sixty-five years of age, the provision of emergency treatment and the provision of routine treatment to special needs groups during night sessions. The monitoring/accountability framework for the DTSS is also outlined, and a set of key performance indicators to be assessed in the next review of the DTSS. The section concludes that immense progress was made since the introduction of the scheme and a large section of the population to whom little or no treatment was available in the past now receive relatively easy access to care. There are 'some hurdles to be crossed in the future and much negotiating to be done but a strong structure is in place to allow the scheme to develop' (DoHC, 2000, p.42).

The review examines the financial outcomes of the DTSS. The DTSS cost £14.3m in 1998 and this expenditure is broken down by health board for the percentage of medical card holders, the total cost of the DTSS, and the percentage of GMS

payments for the DTSS. Expenditure is also examined for the three schemes separately from 1995 to 1998, with a further breakdown to show the types of activities in each scheme. The analysis shows that the proportion of expenditure used for emergency services overall fell from 62 per cent in 1995 to 34 per cent in 1998, while the proportion used for routine treatment rose from 17 per cent in 1995 to 48 per cent in 1998. However, when analysed across health boards there is considerable variation between health boards. The review also estimates the future costs of the DTSS based on: the factors influencing uptake rates using a model developed by the DTBS to predict uptake rates; and factors affecting the average cost per case in each scheme. The report states that the model used is realistic and predicts incremental increases in uptake level that are inevitable, based on information collected by the GMS from participating dentists. Weaknesses identified relate to the uncertainty in predicting cohort numbers and uptake, future average costs and the number of providers in the scheme. 'Lack of certainty about when and at what level the patient uptake rates will stabilise for each cohort, is the primary limiting factor to precise forecasts' (p.54).

The review also explores health gain and the treatment profile of the DTSS. This includes allocation versus total spend on treatment in the private sector, patterns for different types of treatments/activities, trends in terms of the number of dentists in the DTSS and the number of single and group practices, the number of eligible medical card holders per contracted dentist, per health board, average payment per dentist and overall uptake rates per health board (proportion treated of those eligible). Further analysis is outlined for each health board region, of the type of treatment undertaken (whether emergency, routine or dentures) and the costs of each treatment scheme over the four years from 1995 to 1998. Average cost per patient for each of the three schemes is also calculated. The role of health board dental surgeons and expenditure on DTSS outside of the GMS are also discussed. The section concludes that the analysis shows that the introduction of the DTSS has increased oral health gain in the target population. Evidence referred to relates to the decline in the demand for dentures, indicating that this 'reservoir' of treatment is being met; and a steady decrease in the ratio of restorations to extractions, indicating that late intervention has given way to a more proactive phase of appropriate dental treatments. However, medical card patients are less likely to have their teeth cleaned than patients on the DTBS. It is stated that this trend will be monitored in the expectation that as the DTSS matures the amount of preventative care will increase. It is also reported that the stable increase in the number of private practitioners is indicative of a growing confidence in the scheme among dental providers. This trend is expected to continue,

which could have implications for the public dental sector. New re-structuring plans for health board dentists are being adopted to address this problem.

Drawing on the principles set out in the *Health Strategy* (1994), the review examines the equity, accountability and quality of the DTSS. The management of the DTSS is outlined to identify accountability structures. Health boards are responsible for monitoring the calculation of payments to be made for dental services, the making of such payments, the verification of the accuracy and reasonableness of claims in relation to such services, and the compilation of statistics and other information in relation to such services and the communication of such information to persons concerned with the operation of the services. The arrangements for the monitoring of contracting dentists is outlined, including those to deal with treatment patterns that appear to vary from what might be expected, as is the role of DTSS project officers appointed by health boards. The roles of the GMS, the national Monitoring Committee, and the DoHC's Operational Group in monitoring the DTSS are also outlined. A number of initiatives developed to strengthen accountability are identified: examining dentists; reform of the DTSS contract; the introduction of a Probity and Investigations Officer in the GMS; new forms and validations; restructuring of the public dental services; and the awarding of a national research contract to engage consultants with expertise in the development and implementation of accountability structures in dental schemes.

Equity is examined in terms of equity of funding, equity of provision and equity of access. It is reported that equity in funding is achieved by the tax-based funding of the DTSS and that recent increases in funding are highly equitable as they are targeted at those in the community who need them most. DTSS funds are distributed according to the number of medical card holders in the region, ensuring that resources are distributed equitably to those who have been targeted for treatment. However, it is stated that the scheme will need to be extended to include the thirty-five to sixty-four year age group and the rationalisation of services to a standardised routine scheme will be essential before the scheme can become truly equitable. It is noted that the contribution made by the dental services to health gain for medical card holders helps to 'close the equity gap in Irish oral health by improving the position of those at the lower end of the spectrum' (p.83).

In terms of equity of access, it is reported that geographical equity cannot be guaranteed currently, because of the large discrepancies in the number of contracting dentists in each health board area, and between urban and rural areas. The overall

uptake for DTSS patients varies between health boards, and access and uptake rates are lower for DTSS patients than those using the DTBS. It is reported that the DTSS offers medical card holders a basic package of essential oral health therapies and that this facilitates wider access to care for lower income individuals. However, restricted access to advanced restoration procedures may mean that medical card holders lose more teeth due to advanced disease. The report recommends ways in which local health board managers can facilitate more equitable delivery of the DTSS and outlines how the DoHC plans to strengthen the equity of the DTSS.

Quality is examined in terms of the popularity of the scheme among those who use it; its acceptability to them on the basis of access to good quality dental care in the private sector; and the remuneration of dentists and how it ensures quality work. It identifies a number of quality initiatives planned, which will 'greatly enhance our understanding of how efficient and effective the scheme is and also indicate how the service should evolve and develop in the future' (p.85).

The report concludes with a chapter outlining recommendations and future options.