

BENCHMARKING

A POLICY ANALYSIS

Suzanne Wait PhD



The Nuffield Trust
FOR RESEARCH AND POLICY
STUDIES IN HEALTH SERVICES

BENCHMARKING

A POLICY ANALYSIS

Suzanne Wait PhD
Judge Institute of Management
University of Cambridge

70 years

75 years

85 years

Foreword by John Wyn Owen CB

65 years

70 years

75 years

80 years

85 years

65 years

70 years

75 years

80 years

85 years

65 years

70 years

75 years

80 years

85 years

65 years

70 years

75 years

80 years

85 years

ISBN 1-902089-95-2
© The Nuffield Trust, 2004

Published by The Nuffield Trust
59 New Cavendish Street
London W1G 7LP

Telephone: 020 7631 8450
Facsimile: 020 7631 8451

E-mail: mail@nuffieldtrust.org.uk
Website: www.nuffieldtrust.org.uk

Charity Number: 209201

Designed by Nicholas Moll Design
Telephone: 020 8879 4080

TABLE OF CONTENTS

List of tables	4
List of figures	5
Foreword	6
Preface	7
Executive summary	8
PART I: INTRODUCTION	11
PART II: BENCHMARKING IN HEALTH	14
2.1. Categorisation of benchmarking	15
2.1.1. Level of comparison	16
2.1.2. Level of assessment	16
2.1.3. Focus of measurement	16
2.1.4. Intended use of benchmarking data	16
2.2. Criteria for the selection of indicators	17
2.3. Limitations of existing data	18
2.4. The need for quality health information systems	19
PART III: AN EXPLORATION OF POLICY MODELS	20
3.1. The performance measurement model	20
3.2. Rational versus political models of policy	21
3.3. The policy process	22
3.4. Policy implementation analysis	24
3.5. Top-down and bottom-up models of policy implementation	24
3.6. Evidence-based policy	25
PART IV: BENCHMARKING IN PRACTICE	26
4.1. International benchmarking	26
4.1.1. The Organisation for Economic Cooperation and Development (OECD) ...	26
4.1.2. The World Health Organisation	27
4.1.3. The Commonwealth Fund	28
4.1.4. Other international players	28
4.2. National benchmarking systems	29
4.2.1. The United Kingdom	29
4.2.1.1. England	29

4.2.1.2. Wales	30
4.2.1.3. Scotland	30
4.2.1.4. Northern Ireland	31
4.2.2. France	31
4.2.3. Canada	34
PART V: ASSESSING BENCHMARKING FROM A POLICY PERSPECTIVE	35
5.1. A framework for the assessment of benchmarking efforts	35
5.2. What is the underlying conceptual framework?	36
5.3. Are policy objectives made explicit within health care systems?	36
5.4. Do benchmarking indicators reflect policy objectives?	40
5.5. Does the collection of benchmarking data influence and guide policy?	43
5.6. Does benchmarking stimulate change?	44
5.7. Is the policy process underpinning benchmarking coherent?	45
5.8. Does benchmarking follow a top-down or a bottom-up approach?	46
PART VI: DISCUSSION	48
References	50
Appendix 1: List of persons interviewed	57

LIST OF TABLES

Table 1: Application of the “management by objectives” concept to health policy	14
Table 2: Broad categorization of indicators in health	15
Table 3: Directions and purposes of indicator development in health	17
Table 4: Criteria for selection of benchmarking indicators	18
Table 5: Common limitations of health information systems	19
Table 6: Conceptual framework for the top-down and bottom-up approaches to policy implementation	25
Table 7: Criteria used to evaluate policy recommendations	25
Table 8: Priority areas for public health action in France	33
Table 9: NHS policy objectives as stated in key policy documents	39
Table 10: Underlying principles proposed in the new “Projet de loi sur la politique de santé publique” in France	41
Table 11: Example of targets and indicators set out for injury and suicide prevention, Public Health Framework 2003	42

LIST OF FIGURES

Figure 1: A model of the performance measurement process	21
Figure 2: Linking policy objectives to indicator measurement	21
Figure 3: The Walt-Gibson model for policy analysis	23
Figure 4: An amended systems model of the policy process	23
Figure 5: List of key questions to serve as basis for assessing benchmarking efforts from a policy perspective	35
Figure 6: The OECD framework for defining health system objectives	37

FOREWORD

In commending Policy Futures for UK Health 2000 to the Prime Minister I wrote and explained that the Nuffield Trust would be pursuing its interest in benchmarking. The whole issue of health and its longer term perspectives I believe must be retained as a priority across the whole of the government's programme. Health increasingly has a European and wider international dimension. There is a growing trend towards measuring and benchmarking both health and the performance of healthcare systems over time and across nations: a trend towards seeing expenditure on health as an investment and not a cost – good health is after all good economics – and giving health priority a wider focus beyond the NHS will bring dividends. The Nuffield Trust believes that benchmarking is an important contribution to policy review and policy development.

This paper by Dr Suzanne Wait is an important contribution to improving the policy making process and is part of the Trust's development of Policy Futures Mark 2 to be published in Spring 2005. It is complementary to a recent publication by Professor Martin McKee and Dr Ellen Nolte, separately supported by the Trust, on Does Healthcare Save Lives?

John Wyn Owen, CB
Secretary, Nuffield Trust
February 2004

PREFACE

Benchmarking has infiltrated national and international health policy at a rapid pace over the past few years. Indicators are being developed to help nations benchmark the performance of their health care systems against that of other countries, and governments assess the performance of their health care systems at devolved levels of service delivery. There is a considerable literature assessing the merits and caveats of different indicator systems. What is less available is an analysis of the policy relevance of benchmarking: is benchmarking achieving its objectives? Has it contributed to the betterment of health care systems? Do the generated data help guide policy-making? These are ambitious questions to address, yet it is our hope that this report may serve as a starting point for exploring them.

This report aims to propose a framework for assessing the value and policy relevance of benchmarking initiatives in health. This work is undertaken within the context of the Policy Futures research programme supported by the Nuffield Trust, in that it aspires to offer a tool to evaluate the capability of benchmarking to guide future health policy development. The framework proposed in this document will serve as a basis for future work exploring specific areas of benchmarking and health policy.

The author is grateful to those experts, listed in **Appendix 1**, who generously gave their time and shared their thoughts to help advance this research. Sincere thanks are also due to Dr Sheila Leatherman, Pam Garside, Dr Charlotte Sausman, Dr Alan Ingram, John Wyn Owen, Professor Sandra Dawson, Dr Ellen Nolte and Professor Martin McKee for their continuing guidance throughout this project.

Suzanne Wait
Cambridge, January 2004

EXECUTIVE SUMMARY

The past decade has witnessed an explosion in the development of indicators and targets in healthcare. These data are used for comparative performance assessment within countries as well as international health system comparisons.

Several scholars have expressed fears that indicator measurement has become an end in itself and that its impact on guiding future policy has not been evaluated. Within this context, this research aims to take a policy perspective on benchmarking and propose a framework for assessing the policy relevance of benchmarking efforts.

The present report is informed by a review of the published literature, policy documents and discussions with policy experts and academics involved in national and international benchmarking initiatives. It draws on observations of international health system comparisons as well as national benchmarking initiatives within the UK and its four devolved administrations (England, Wales, Scotland and Northern Ireland), Canada and France. These health care systems are at the grips of testing the potential role that indicator systems may play to improve their health care systems and thus provide valuable practical testing ground for the concepts explored in this research.

The recent experience of international and national benchmarking initiatives has brought to light the challenges inherent in collecting, evaluating and interpreting indicator data to guide policy-making. It has illustrated the need to define policy objectives within health care systems, ensure coherence between different evaluative and regulatory processes, and move away from a culture of measurement towards a culture that embraces performance and is truly evidence-driven.

Based on the theoretical literature, a review of policy models and critical reviews of benchmarking, a framework for assessing benchmarking efforts is proposed. This consists of a list of key questions that may guide the assessment of benchmarking initiatives either retrospectively or prospectively.

These key questions are as follows:

- What is the underlying conceptual framework?
- Are policy objectives and goals explicit to guide benchmarking within health care systems?

- Do benchmarking indicators reflect policy objectives? Do policy objectives drive the selection and collection of indicators?
- Does the collection of benchmarking data influence and guide policy?
- Is the policy process underpinning benchmarking coherent?
- Does the publication of benchmarking data stimulate change (better performance of the health care system? better health care?)
- Does the articulation from objectives to targets to indicators and back to policy follow a top-down or a bottom-up approach?

This list of questions provides an initial framework for assessing benchmarking efforts and will be applied to more specific areas of health policy, namely cancer and public engagement, in subsequent phases of this research.

The concepts and practical experience of benchmarking explored in this report allow to draw some tentative conclusions about the policy relevance of benchmarking:

- Problems of data availability remain a significant challenge to benchmarking within and across countries. More focus needs to be given to developing rigorous health information systems capable of ensuring systematic, high-quality data collection.
- The pursuit of meaningful indicators to evaluate and compare health system performance within and across health care systems is essential in a health policy environment that aims to foster accountability and stewardship for health.
- Further research is still needed to identify the most reliable and valid indicators to measure different dimensions of health system performance. The public, healthcare managers and clinicians, policy-makers and the media need to be made aware of the limitations of existing indicator data to avoid misinterpretation.
- The rational policy model suggesting that clear policies and objectives should drive indicator measurement and that evidence collected should guide policy-making is not to be entirely dismissed. Clearly stated objectives and a solid conceptual framework underlying benchmarking lend credibility to the indicator and target data being collected and may contribute to the permeation of a true performance culture throughout the health care system.
- Frameworks for evaluating the performance of health care systems should be policy-driven, as opposed to data-driven. The starting point for any benchmarking framework should be the policies and objectives that one is wishing to evaluate. The collection of data that may appropriately and reliably reflect progress against these goals should be prioritised.
- The interpretation of indicator data should not lose sight of the policy context within which they are measured, the players involved in formulating and implementing policy, the time lag needed to assess the impact of different policies, and aspects of health care that remain unmeasured by available data.
- The policy relevance of benchmarking hinges on the overall cohesiveness of the policy process. Evaluating indicator data across countries without any understanding for the regulatory and evaluative policies underpinning them may hamper data interpretation.

- The move to further decentralisation within health care systems is likely to have a significant impact on the future of benchmarking. For true devolution to be achieved, local players responsible for data collection will need to accept greater ownership for performance data, whilst policy-makers will need to shift policy priorities to reflect local needs.
- Policy-makers have fully embraced the notion that public engagement in all policy-making will act as a key driver to improving health system performance. Whether policies may truly be effective in engaging the public and how they may translate into improved performance and health outcomes has yet to be determined.

In summary, reliable comparisons of indicator data within and across countries may provide a powerful tool to guide policy and stimulate learning both within and between health care systems. Further research is encouraged to help policy-makers use comparative data appropriately to help understand discrepancies in health care and, ultimately, improve the quality of care.

PART I: INTRODUCTION

Benchmarking: “a detailed analysis of comparative performance to help identify what underlies differences between two similar bodies.”

(UK Parliament Select Committee, 2003)

“The ultimate goal is to manage quality. But you cannot manage it until you have a way to measure it, and you cannot measure it until you can monitor it.”

(attributed to Florence Nightingale, c.1875)

“Targets can be used to set standards to achieve greater equity. Targets provide better accountability.”

(Public Services for the Future, 1998)

“The problem with measurement is that it can be a loaded gun: dangerous if misused and at least threatening if pointed in the wrong direction.”

(O’Leary 1995, p. 354)

The past decade has witnessed an explosion in the development of indicators and targets in healthcare. Concepts of performance assessment and management, target-setting, indicator measurement and benchmarking have become integral features of most health system reforms over the past decade. International health system comparisons suddenly came to the forefront of high-level political discussions with the publication by the World Health Organisation (WHO) of its controversial report “*Health Systems: Improving Performance*” in 2000 (WHO 2000). Since then, many countries have used international benchmarking as a basis for guiding national health policy priorities. The most illustrative example is that of the UK, where evidence of poor relative cancer survival outcomes compared to other European countries (Berrino *et al.*, 1995), coupled with the fact that Britain spends less of its GDP on health than most other Western nations, prompted the Labour government to

inject significant funds into the National Health Service (NHS) (Wanless 2002; Appleby and Boyle, 2000).

The adoption of an indicator-based culture in health policy raises several issues. There is a growing literature highlighting the measurement limitations of existing indicator systems in terms of the validity and reliability of measures collected (Musgrove, 2003; Hurst and Jee-Hughes, 2000). The various measures adopted are not necessarily evidence-driven, nor are they articulated within a cohesive framework. Indicators are very rarely evaluated over time, thus preventing incremental changes to the health care system from being observed. Several scholars have expressed fears that indicator measurement has become an end in itself and have urged for the evaluation of benchmarking indicator systems (Blalock 1999; Goddard et al, 2000; Smith 2002; Walshe 2003). Others have suggested that the role that the adoption of indicators has played in improving the quality of the health policy process and the performance of health care systems has yet to be ascertained (Sheldon, 1998). Similarly, the impact, intended or unintended, that benchmarking may have on guiding future policy is too rarely acknowledged or evaluated (Smith, 1995).

Within this context, the aim of this report is to propose a framework for assessing the value and policy relevance of benchmarking initiatives within and across different countries. This work is based on a review of the theoretical literature, policy documents, critical reviews of benchmarking initiatives and unstructured interviews with a number of policy-makers and academic experts. Drawing from these combined sources, this work proposes a list of key questions that may be applied to the analysis of benchmarking efforts from a policy perspective.

This report forms the initial part of a broader programme of research on benchmarking in health. It aims to provide the background to more focused analyses of benchmarking, which will be carried out in selected policy areas. Next steps in this research include the empirical analysis of existing indicator frameworks in two priority areas of policy: cancer and public engagement. Together, it is hoped that these complementary pieces of research may help further the methodology of benchmarking and inform future health policy in the UK and elsewhere. As such, the objectives of this work fit into the overall objectives of the Policy Futures work carried out under the auspices of the Nuffield Trust (Dargie *et al.* 1999; 2000).

Two important caveats deserve mention in introducing this report. First, this work examines benchmarking systems from a general policy perspective, thus no attempt is made to give a critical appraisal of the technical merits of existing indicators as such. Similarly, the term 'benchmarking' is applied loosely to all benchmarking efforts that aim to guide national health policy, thus it denotes both cross-country comparisons and performance assessment frameworks used within countries to benchmark health system performance.

The present report is informed by a review of the published literature, policy documents and discussions with policy experts and academics involved in national and international benchmarking initiatives. It draws on observations of international health system comparisons as well as national benchmarking initiatives within the UK and its four devolved administrations (England, Wales, Scotland and Northern Ireland), Canada and

France. These health care systems are at the grips of testing the potential role that indicator systems may play to improve their health care systems and thus provide valuable practical testing ground for the concepts explored in this research.

This document is divided into six parts, the first being this introduction. **Part II** presents an overview of benchmarking as applied to health, whilst **Part III** explores policy models that may help inform the conceptual framework underlying benchmarking initiatives. **Part IV** then provides a brief overview of ongoing benchmarking initiatives in the UK, Canada, France and international comparisons. In **Part V**, a framework for assessing benchmarking efforts is proposed and applied to the experience of the benchmarking initiatives described previously. Implications for future research and preliminary conclusions on the policy relevance of benchmarking are presented in the **Discussion**.

PART II: BENCHMARKING IN HEALTH

Benchmarking was translated to health policy from the field of management. McGinnis first applied the ‘Management by Objectives’ culture to health policy in the United States in his “Objectives for the Nation” work in the 1980s (McGinnis, 1982). During this same period, increased recognition of epidemiology as a science promulgated the use of health targets (van Herten and Gunning-Scheppers, 2002). The “New Public Management” ethos that affected public services in the 1990s also played a role by shifting expectations for public accountability and encouraging the release and reliance on league tables in the UK and elsewhere (Nutley and Smith, 1998). McGinnis’s ‘Management by Objectives’ framework is presented in Table 1 below.

Table 1: Application of the “management by objectives” concept to health policy

Objective bases	Business application	Health application
Outcome	Profits	Morbidity and mortality reductions
Strategy	Product type and mix	Risk factors
Productivity	Labour/capital mix	Scope and quality of services
Marketing	Client attitudes and awareness	Public/professional attitudes and awareness
Innovation	Product improvement	Surveillance, evaluation, and research

Adapted from McGinnis, 1982.

Evidence of gross medical errors and of severe inadequacies in the quality of care also contributed to the rise of benchmarking in health. In the UK, the Harold Shipman tragedy in 2000 and the Bristol Royal Infirmary Inquiry in 2001 highlighted the failings of existing internal quality assurance and professional regulation systems and pointed to the need for external audit procedures to ensure safe and high-quality standards of care (O’Neill, 2000; <http://www.bristol-inquiry.org.uk>). In France, the publication of comparative hospital outcomes data in 1997 shocked citizens and government into acknowledging that huge

discrepancies in hospital care existed and since then, emphasis on accreditation and evaluation have been a constant thread in French health policy (*La liste noire des hôpitaux*, 1997; Matillon *et al.*, 2001). In the US, the Institute of Medicine report on the quality of care, “Crossing the Quality Chasm”, highlighted the chasm that exists between average care and the best quality care within the United States and led to a series of initiatives aimed at improving the appropriateness of care (Institute of Medicine, 2001; 2001b).

Finally, the increased focus on indicators and benchmarking reflects current health policy trends in developed countries. In efforts to contain spiralling health care costs, governments are devolving budgetary control over health to regional and local levels (Saltman and Figueras, 1998), thus rendering them accountable to central government for the flow of funds to deliver health services. Meanwhile, the rise of consumerism and increasing media attention to health care issues have forced central governments to be accountable to the public for the overall performance of the health care system and to acknowledge their stewardship over the health care system (Blalock, 1999). On an international level, globalisation has affected health policy like other disciplines, adding visibility to cross-country comparisons and removing some of the parochialism formerly typical of health policy.

2.1. Categorisation of benchmarking

Like many evolving areas of research, the study of benchmarking in health is characterised by considerable confusion over terminology. A proposed categorisation for health indicators is proposed in **Table 2**.

Table 2: Broad categorisation of indicators in health

Feature	Alternatives
Level of comparison	International comparisons National assessment
Level of assessment	Individual practitioner Hospital/trust Primary care group/managed care organisation Locality/district/region National
Measurement focus	Process Outcomes Quality Performance
Use of data	Public reporting, accountability Internal reporting only, self-learning and improvement

2.1.1. Level of comparison

As mentioned previously, indicators have been developed for the purposes of international comparisons as well as to enable national governments to evaluate performance within their health care system. This report is limited in scope to the Canadian, UK and French benchmarking initiatives as well as international comparisons. For a thorough review of other benchmarking frameworks, the reader is referred to Shaw and Kalo (2002) and Hurst and Jee-Hughes (2000).

2.1.2. Level of assessment

Most indicator data measure hospital activity, as ambulatory settings often lack the health information infrastructure necessary for reliable data collection. Notable exceptions exist, however, such as the National Healthcare Quality Reports in the United States (Institute of Medicine, 2001) and the extension of performance ratings to Primary Care Trusts in the UK (Commission for Health Improvement, 2003).

2.1.3. Focus of measurement

The focus of indicators may be on the processes, quality, outcomes or performance of care. Most often, process measures are used as a substitute for outcome measures due to lack of data availability (Goddard *et al.*, 2000). A widely accepted definition of quality is “the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge” (Institute of Medicine, 2001b). Several countries have initiated comprehensive initiatives to monitor the quality of health care (Leatherman and Sutherland, 1998; Shaw 2001). Performance is a multidimensional concept, which incorporates dimensions of safety, effectiveness, appropriateness and timeliness and responsiveness of services (Girard and Minville, 2002). Typically, assessments of the performance of health systems are based on rather crude measures such as the level of expenditure on health care, resources used in the system or summary measures of population health such as mortality or life expectancy (Hurst and Jee-Hughes, 2000). The WHO made a bold attempt to develop a more sophisticated measure of performance in its 2000 Report, however their approach is far from universally accepted (McKee, 2002).

2.1.4. Intended use of benchmarking data

The publication of indicators may have several objectives:

- To secure accountability to funders and other stakeholders
- To identify areas of poor performance and centres of excellence
- To help patients and purchasers of health care choose a provider
- To enable providers to focus on areas requiring improvement
- To provide epidemiological and other public health data.

(Nutley and Smith, 1998)

Freeman (2002) and Soldberg *et al.* (1997) suggest that indicators may be developed for two main purposes: first, internal and external control and accountability, and secondly, quality improvement. Whilst most of the indicator systems developed in health simultaneously strive to achieve both these objectives, the rationale and culture underlying both streams of indicators are, or should be, different (Freeman, 2002). These differences are outlined in **Table 3**.

Table 3: Directions and purposes of indicator development in health

Approach	Indicators for assurance/accountability	Indicators to stimulate improvement
Purpose	Verification and assurance	Learning to promote continual improvement
Emphasis	Measurement oriented	Change oriented
Rationale	Provide external accountability and renew legitimacy	Promote change and improvement in care quality
Culture	Comparisons in order to make summative judgements on care quality. League tables. Blame and shame	Emphasis is to learn from differences between providers and encourage improvement. Informal benchmarking to promote discussion and change

Adapted from: Freeman, 2002.

2.2. Criteria for the selection of indicators

The value of benchmarking initiatives can be greatly enhanced if target indicators are selected for their relevance and usefulness as evaluation tools rather than merely on data availability. Pringle *et al.* recommended that indicators should be valid, communicable, effective, reliable, objective, available, contextual, attributable, interpretable, comparable, remediable and repeatable (Pringle *et al.*, 2002). Another helpful framework for selecting indicators was proposed by the Institute of Medicine and has been adopted by the OECD Healthcare Quality Indicator project (described later). This framework is presented in **Table 4** overleaf.

Table 4: Criteria for selection of benchmarking indicators

<p><i>Importance of what is being measured</i></p> <ul style="list-style-type: none"> – What is the impact on health associated with this problem? – Are policy makers and consumers concerned about this area? – Can the health care system meaningfully address this aspect of problem? <p><i>Scientific soundness of the measure</i></p> <ul style="list-style-type: none"> – Does the measure actually measure what it is intended to measure? (core validity) – Does the measure provide stable results across various populations and circumstances? (reproducibility) – Is there scientific evidence available to support the measure? <p><i>Feasibility of using the measure</i></p> <ul style="list-style-type: none"> – Is the measure in use? – Can the information needed for the measure be collected in the scale and timeframe required? – How much will it cost to collect the data needed for the measure? – Can the measure be used to compare different groups of the population?
--

Source: Institute of Medicine 2001b.

2.3. Limitations of existing data

Although this report will not elaborate on the technical merits of existing benchmarking indicators, it is important to note that limited data availability and lack of uniformity of data across different settings plague most benchmarking initiatives. Musgrove revealed that only 39% of the indicator values included in the WHO 2000 report were based on real data, the rest being estimates imputed using regression analyses and other means (Musgrove, 2003). Discrepancies in the quality of available information are also prominent in within-country comparisons, sometimes even despite the existence of common minimal datasets and standardised protocols for data retrieval (Wait, 2001). Dixon *et al.* found that it was impossible to compare outcomes data on waiting times, waiting lists and patient satisfaction data across Scotland, Northern Ireland, Wales and England as no policy initiative had ever encouraged consistent data recording across the four devolved administrations of the UK. The authors raised the question of what the ‘currency of assessment’ should be in comparative research (Dixon *et al.*, 1999).

Other difficulties in capturing valid and reliable indicator data are that performance is multidimensional, that different users have different information needs, that many outputs are qualitative, that the assessment of performance will be affected by the timing of measurement and that many measures are susceptible to manipulation (Nutley and Smith,

1998). Who analyses the data will have a significant impact on results, as each observer's value system and perspectives will inevitably influence interpretation of data. Moreover, establishing causality between observed results and possible causative factors must also allow for consideration of factors external to the health care system that are likely to be confounding factors (Goldstein and Spiegelhalter, 1996).

2.4. The need for quality health information systems

A necessary, yet oftentimes unmet, condition to ensuring accurate collection of indicator data is the existence of reliable and well-established health information systems. Indicators for international comparisons are particularly susceptible to failings in consistency in reporting, as they require case-mix and other adjustments that may sometimes be avoided in internal data comparisons within a single health care system (Shaw and Kalo, 2002). Some of the common problems with using existing health data systems for the purposes of evaluation are described in **Table 5** below.

Table 5: Common limitations of health information systems

Design and primary focus is financial management, not evaluation
Focus on hospitals and resources rather than populations and performance
Lack of agreed minimum data sets and definitions
Absence, inaccuracy and misinterpretation of aggregated data for indicators
Failure to integrate population and patient-based data
Lack of incentives to collect accurate data
Legal and ethical conflicts between freedom of information and data protection
Lack of linkage between diagnostic data and outcomes of care.
Partiality of data – often institution-specific and not representative of entire trajectory of care.

Adapted from Shaw and Kalo, 2002

PART III:

AN EXPLORATION OF POLICY MODELS

“What is needed is a model which fits reality while being directed towards its improvement and which can in fact be applied to policy making while motivating a maximum effort to arrive at better policies.”

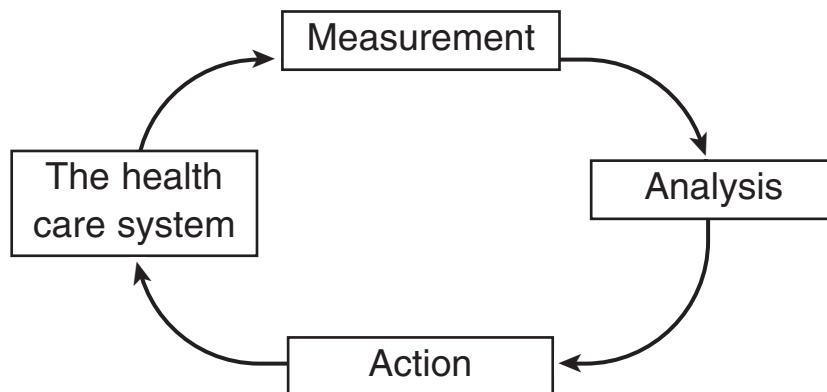
(Dror 1964)

The machinery of indicator development, benchmarking and performance assessment is so fully engaged that it is timely to pause to assess the value of such an exercise. A first step in any policy analysis is to determine the appropriate conceptual framework for analysis. What follows is a brief exploration of some possible policy models that may be pertinent to the evaluation of benchmarking.

3.1. The performance measurement model

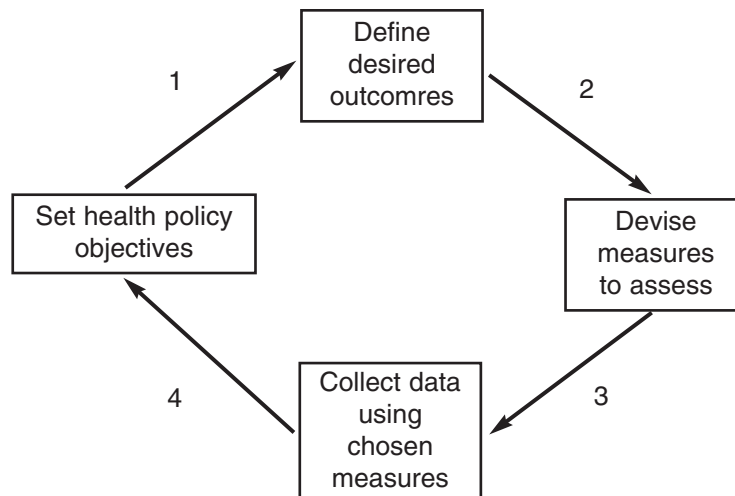
Policy has been defined as “the translation of government’s political priorities and principles into programmes and courses of action to deliver desired changes” (National Audit Office for England, 2001). Applied to benchmarking, the assumption is that the systematic measurement and monitoring of performance across the health care system using selected indicators will lead to the achievement of policy objectives. This is the basis for the *Performance Measurement Model* proposed by Nutley and Smith (1998) and inspired from the Continuous Quality Improvement literature (Berwick 1989, 1990; Brennan 1998). This model, presented in **Figure 1** opposite, underpins the OECD international comparison framework and sees data collection, measurement and analysis as a stimulus to change policy.

Figure 1: A model of the performance measurement process
(Adapted from Nutley and Smith, 1998)



A variant on this model is presented in **Figure 2**. This model would suggest that policy objectives drive the selection and collection of indicators (arrows 1-2) and that the collection of benchmarking data (arrow 3) influences and guides policy (arrow 4).

Figure 2: Linking policy objectives to indicator measurement.
Adapted from Nutley and Smith, 1998; Hurst 2001.



3.2. Rational versus political models of policy

The assumption that benchmarking indicators reflect health system objectives is consistent with a *rational model* of policy-making, which defines policy-making as a normative process that draws a linear connection between problem formulation, evaluation of alternatives, selection of actions, and implementation. Several policy analysts from a wide range of disciplines see this model as a “dignified myth” which oversimplifies the understanding of policy-making (Gordon *et al*, 1977; Smith and May, 1980). Two of its most important deficiencies are that it does not allow for the intended

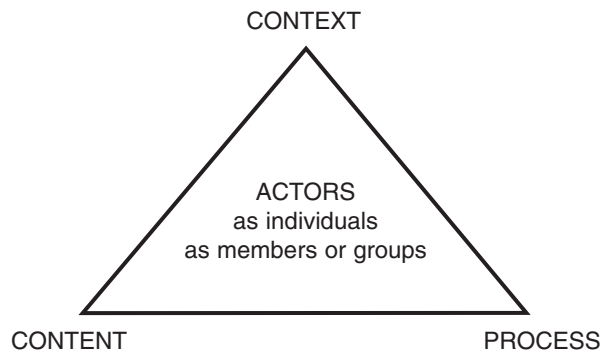
or unintended consequences of policies, and that it fails to acknowledge that the means and ends of policy-making are often derived in parallel, yet unrelated processes. Thus post-hoc analyses of policies using this model may run the risk of trying to establish artificial causality between two parts of the policy process that were, in fact, never jointly planned. This caveat may, in some cases, apply to the search for a relationship between a given health care system's policies and objectives and the benchmarking indicators used to assess performance.

An alternative to the rational model is the *political model* of policy, where policy is considered to be a 'bargained outcome', the result of an essentially political activity (policy-making) in which the values, judgments and preferences of individual actors enter into all stages of play. Whereas the rational model suggests that policy must invariably precede, and lead to, action (in this case indicator measurement), the political model suggests that policy is a dynamic set of actions and values reflecting a dynamic course of action in a given area. Lindblom has proposed an *incremental model* of decision-making, which he refers to as "disjointed incrementalism" (Lindblom, 1959). He sees the decision-maker starting not with a clean slate, but with existing policies from which he can make only incremental changes. Supporting this notion, Gordon states that it is not necessarily possible to identify policies as distinct, clearly articulated items which emerge or become apparent at a given time in the decision-making process (Gordon *et al*, 1977). Instead, the most relevant unit of analysis is the policy process, where policy objectives, goals, players, and policy content interplay in a dynamic manner. Moreover, a nation's health policy is not a single entity, but rather a composite of individual policies that address different goals (eg. policy on cancer, policy on ambulatory surgery, policy on patient information).

3.3. The policy process

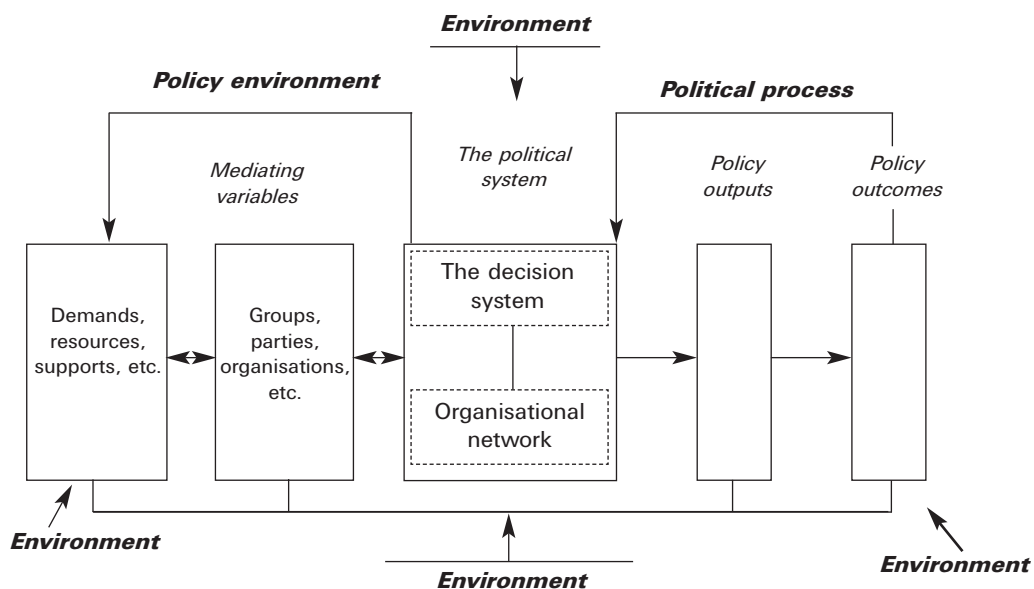
Wildavsky has suggested that policy is as much a product as a process (Wildavsky, 1979), that policy is adaptive to its environment and that information is constantly fed into the policy process to inform policy-making. Walt and Gibson (1994) stress that policies cannot be divorced from the context within which they are formulated and implemented. Their policy model describes policy content in terms of processes, boundaries, underpinning values and objectives. Their essential premise is that policy analysis typically focuses too much on the content of policies and neglects the importance of other essential components, such as the actors involved, the processes needed to implement change and the context or processes (social, political, cultural, economic, historical) which may explain why policy outcomes were not achieved. This model can be used both prospectively (for policy planning) and retrospectively (for policy impact assessment). It is presented in **Figure 3**.

Figure 3: The Walt-Gibson model for policy analysis
(Adapted from Walt and Gibson, 1994)



Jenkins¹ proposes a framework which may serve as a starting point for analyzing a given policy process (Jenkins, 1978).² In his “extended systems model” (**Figure 4**), he proposes a conceptual model that draws attention to the importance of the reciprocal control between the environment and policy inputs, and the intended and unintended consequences of policies in terms of outputs and outcomes. Although allowing for more external influences on policies and their outcomes, this model still adopts the rational model of decision-making as its core, as goals are translated into outputs through a fairly linear process.

Figure 4: An amended systems model of the policy process
(Adapted from Jenkins, 1978)



1. Jenkins has borrowed from the systems framework proposed by David Easton in D. Easton. A Framework for Political Analysis, Prentice Hall, 1965; or D. Easton. The Political System. A. Knopf, 1953.
2. Jenkins refers to a number of authors who have adopted the systems model to the study of policy, including I. Sharkansky. Public Administration: Policy making in government agencies (Markham, 1970) or T. Dye, Understanding Public Policy, Prentice Hall, 1972.

3.4. Policy implementation analysis

The measurement of system performance may also be seen as part of the implementation of given policies. Implementation may be defined as the “bringing about, by means of outputs, of outcomes that are congruent with the original intentions.” (Lane, 1987, 1997). Lane (1997) sums it up in the following equation:

Implementation = F (Policy, outcome, formulator, implementer, initiator, time).

Applied to benchmarking, Lane’s model would suggest that to assess whether a given policy has been implemented, one needs:

- Information about the extent of congruence between policy objectives and outcomes (measured through benchmarking indicators)
- A decision about the time span needed to see the outcomes arising from policies
- Understand the processes for implementation, including the actors engaged, trust and accountability.

(Lane, 1997).

According to the above equation, the capability for policy implementation is contingent upon the relationships between the policy makers (formulators) and those expected to implement them (implementers) – and the power of the latter to adapt policies to meet their own needs.

3.5. Top-down and bottom-up models of policy implementation

Lane’s distinction between policy formulators and implementers brings to light the tensions endemic within all devolved health care systems where policies, targets and objectives are determined centrally, yet policy implementation is devolved to local levels. This dichotomy is illustrated in the top-down and the bottom-up models of policy. The top-down model is similar to the rational model, in that it assumes a linear articulation from problem identification to policy formulation, implementation and evaluation. In contrast, the bottom-up model rejects this linear and centralized view of the policy process and sees it as iterative – i.e. results will drive policies at all levels as they emerge. Which model is most suitable to a particular health policy context depends on the level of effective decentralisation, the role of centre, the ability of top policy makers to force lower-level authorities to conform with guidelines, regulatory and evaluation mechanisms and incentives and sanctions (Sabatier, 1997; Sabatier *et al.*, 1979). Sabatier compares these two approaches (**Table 6**) and cautions that whilst the bottom-up model may seem more intuitive, it fails to acknowledge the role that values or overall goals decreed at the centre may have on shaping actions and policy implementation ‘in the field’.

Table 6: Conceptual framework for the top-down and bottom-up approaches to policy implementation

	Top down approach (Sabatier and Mazmanian, 1979)	Bottom-up approach (Hjern <i>et al</i>, 1982)
Initial focus	Central government decision	Local networks and activists
Accountability between major actors in the process	From top policy-makers to local decision-makers and service providers. Incentive structure key in understanding relationships	From local decision-makers to top decision-makers
Evaluative criteria	Extent of attainment of formal objectives, which is carefully collected and analysed.	Locally-determined outcomes, evaluated based on available data
Overall focus	How does one steer system to achieve top policy-makers' intended policy results?	Local engagement and ownership for results

Adapted from Sabatier, 1997.

3.6. Evidence-based policy

A final consideration is that given health policies may be ineffective not only because of poor implementation, but because they are poor policies. Hogwood and Gunn (1984) state that 'perfect' policy implementation is only feasible if there is a perfect agreement and understanding of policy objectives and that the policy to be implemented is based on a valid theory of cause and effect. Thus, attention must be given to the quality of policies before assessing their implementation. A series of criteria against which policy recommendations may be evaluated is proposed in **Table 7**.

Table 7: Criteria used to evaluate policy recommendations

<ul style="list-style-type: none"> ● Supported by systematic, empirical evidence ● Supported by cogent argument ● Scale of likely health benefit ● Likelihood that the policy will bring benefits beyond health ● Fit with existing or proposed government policy ● Possibility that the policy might do harm ● Ease of implementation ● Cost of implementation.
--

(McIntyre et al, 2001)

PART IV:

BENCHMARKING IN PRACTICE

This section provides a brief description of the overall benchmarking frameworks adopted in three countries (Canada, France and the UK, including its devolved administrations) as well as international benchmarking initiatives. Given that continual changes are being brought to these benchmarking frameworks, this description will limit itself to a brief overview of each initiative, and refer to weblinks for further information and updated lists of indicators.

4.1. International benchmarking:

4.1.1. The Organisation for Economic Cooperation and Development (OECD)

International benchmarking was traditionally the domain of health economists interested in comparing different countries' health expenditures and the share of GDP devoted to health (Kavanos and Mossialos, 1996; Hurst and Jee-Hughes, 2000). One of the pioneers of international comparisons was the Organisation for Economic Cooperation and Development (OECD). The OECD promotes cross-national comparisons as a means of capturing a range of different approaches in health care, thereby allowing the experience of each country to provide "an experimental laboratory for others" (OECD, 1990). Its 1985 publication, "Measuring health care, 1960-1983 expenditure, costs and performance", was the first in a series of international studies designed to provide an empirical basis for a comparative understanding of the differences and similarities between OECD countries' health systems (OECD 1985; Schieber 1987). In 1993, it published the OECD Health Data Set, which has since evolved into a comprehensive dataset that includes data from 30 countries covering over 1200 indicators (www.oecd.org/health/healthdata). Whilst the main focus of these data is on health care resources, financing and expenditure, other key aspects of health care systems are also covered (OECD, 2003). The organization also publishes summary indicators for OECD countries in its 'Health at a Glance' publications series (http://www.oecd.org/statisticsdata/0,2643,en_2649_37407_1_119656_1_1_37407,00.html).

In 2001, the OECD Health Policy Unit initiated the OECD Health Project, which focuses on 'measuring and analysing the performance of health care systems in Member countries and factors affecting performance'. The goal of this project is to help decision-makers formulate evidence-based policies to improve their health systems' performance. This project is described in full on the OECD website. (Weblink: http://www.oecd.org/document/28/0,2340,en_2649_37407_2536540_1_1_1_37407,00.html).

As part of its Health Project, the OECD initiated the Healthcare Quality Indicator Project (HCQI) in an effort to develop quality indicators for target setting and policy evaluations. Quality is defined as the technical quality with which medical care is provided, i.e. changes in health status attributable to preventive or curative care. The HCQI combines the work of two pre-existing initiatives, the Commonwealth Fund International Working Group on Quality Indicators and the Nordic Council of Ministers project, whose focus is on developing indicators for benchmarking purposes within Scandinavian countries. The goal of the HCQI Project is to propose a standard minimal dataset of indicators to be used across all OECD countries to evaluate the quality of health care. The starting point for this dataset is a list of 31 indicators initially proposed by the Commonwealth Fund International Working Group on Quality Indicators (see below). A second phase will involve the development of further indicators in policy areas considered to be priorities by the participating countries (S. Mattke, personal communication). The strength of this initiative lies in the direct involvement of top-level health representatives from the 19 participating countries.

4.1.2. The World Health Organisation

The WHO made its first entry into benchmarking in 1977 with the launch of its 'Health for all by the year 2000' strategy, which proposed 28 targets to be achieved by member states by the year 2000. This platform was subsequently modified into "Health 21" for Europe, outlining 21 targets to be achieved within Europe in the 21st century (WHO Regional Office for Europe, 1998). The health targets were born from a population health perspective, thus their focus was mainly on public health achievements and bore no links to management and financing mechanisms of health care. In Europe, the twenty-one proposed targets include two targets on equity, three on 'better health in Europe', four on prevention of disease and injury; five on multisectoral strategies for creating sustainable health; four on outcome-oriented health; three on mechanisms for managing change (van Herten and Gunning-Schepers, 2002). Member states were also expected to develop their own indicators in order to help them prioritise their resources towards specific goals (van Herten *et al*, 1999; (Ritsatakis, 2000). A core set of 200 indicators was subsequently developed into the extensive Health for all (HFA) database, allowing benchmarking of health trends between the 51 countries of the European region over time (WHO Regional Office for Europe, 2003). The European Health for All database is updated periodically and is available online (<http://hfadb.who.dk/hfa/>). The WHO Statistical Information System (WHOSIS) also provides extensive comparative statistics on a global basis (www3.who.int/whosis/menu.cfm).

In 2000, the WHO issued its highly publicised report, *Health Systems: Improving Performance*, which aimed to set out “new concepts and measures which lay the empirical basis for assessing health system performance” (World Health Organisation, 2000; p. 144). The report presented international comparisons based on the WHO Performance Assessment Framework, which builds on the concepts of health system boundaries, objectives, goals, health system efficiency, health system functions, and enhancing policy relevance of health system performance frameworks (Arah *et al*, 2003; Murray and Frenk, 2000). Overall health system performance is assessed as a composite measure that includes the level and distribution of health attainment, level and distribution of health system responsiveness, and degree of fairness of financing. This aggregate measure is then compared with what might be expected given the country’s level of economic and educational development, thus producing two measures of performance, or efficiency, one at the health level and one on overall performance. In the 2000 report, the 191 WHO member states were then ranked based on these performance measures, producing the highly controversial league table of the world’s health systems.

The WHO approach has been the subject of intensive debate, ranging from the implied values underlying the approach taken (Navarro, 2000; Williams, 2001; Musgrove, 2003; Richardson *et al.*, 2003) to technical considerations of specific aspects (Almeida *et al.*, 2001). These debates played an indisputable role in raising awareness of the methodological challenges inherent in conducting and interpreting international comparisons (Richardson *et al*, 2003). The WHO is continuing its work in international benchmarking and an update to its 2000 report is expected during the course of 2003-2004. All publications and reports relevant to its Performance Assessment work are available on the WHO website (<http://www.who.int/health-systems-performance/>).

4.1.3. The Commonwealth Fund

The Commonwealth Fund has been a key player in providing international comparisons of healthcare systems. They commissioned the seminal “Quality Chartbook”, which laid out different measures to assess the quality of health care in the United States and has since been used as a template for application to other countries (Leatherman and McCarthy, 2002). The International Working Group on Quality Indicators, which has reported quality indicators for five Commonwealth countries (New Zealand, Canada, United Kingdom, United States, Australia) based on 31 criteria, provided the starting point for the OECD HCQI project described above. The Fund is currently engaged in a number of international collaborations focused on benchmarking (<http://www.cmwf.org/programs/policy.asp?link=2>).

4.1.4. Other international players

A new policy group involved in international benchmarking is the *Bertelsmann Foundation International Health Policy Network*, whose aim is to help identify best practice in health policy. The network consists of health policy experts from 15 countries who provide information on a 6-monthly basis of policy developments within their respective countries. Information is provided by means of a semi-structured questionnaire, which asks to

describe the policy issues and how they are being implemented. A more detailed questionnaire then captures information about the process for policy development, stakeholder groups involved, and how implementation and evaluation were carried out. Data from the survey are posted on the Network's website (<http://www.healthpolicymonitor.org/>) and have been published in a comprehensive report (Busse and Schlette, 2003).

At a European level, the *European Observatory on Health Care Systems* “supports and promotes evidence-based health policy-making through comprehensive and rigorous analysis of the dynamics of health care systems in Europe.” (<http://www.euro.who.int/observatory/>). The Observatory is a partnership between WHO Regional Office for Europe, the European Investment Bank, the Open Society Institute, the World Bank, the London School of Economics and London School of Hygiene & Tropical Medicine and some national European governments. It produces a number of individual country reports and international comparisons on selected themes. Most of these reports are available on their website (<http://www.euro.who.int/observatory/>).

A new, and somewhat elusive, player in the international benchmarking arena is the *European Commission*. The Commission has made available significant grant money to fund benchmarking and indicator development efforts in Europe as of 2003 (M. McKee, personal communication). It is unclear at this stage what the scope of these initiatives will be, or how they may overlap with ongoing efforts described above. The Commission has sponsored the publication of the Eurobarometer data for several years, which focuses on European-wide opinion surveys on a broad range of topics, which can include health (http://europa.eu.int/comm/public_opinion/standard_en.htm).

It should be added that many academic centres of excellence are engaged in international comparisons. It is beyond the scope of this report to describe all of these efforts; however the work of the London School of Economics and the London School of Hygiene and Tropical Medicine merit particular mention.

4.2. National benchmarking systems

4.2.1. *The United Kingdom*

Performance management is a prominent feature of all public services in the UK. Devolution within the UK has meant that, despite some common policies, England, Scotland, Wales and Northern Ireland have each developed their own health strategies as well as benchmarking frameworks.

4.2.1.1. *England*

A pervasive culture of benchmarking dominates the NHS in England (Department of Health 2000; 2001a; 2001b; 2001c). In 1997, the Labour government introduced the Performance Assessment Framework in England following publication of the 1997 White Paper “*The New NHS: modern, dependable*” (Department of Health, 1997). The PAF aims to compare

NHS providers using specific indicators across 6 areas: health improvement, fair access to services, effective delivery of care, efficiency, patient and carer experience and health outcomes of care. The number of indicators was originally 60 but it has since been reduced. Targets are built into accountability agreements between different levels of the NHS (Goddard *et al*, 2000).

The PAF has been somewhat superseded in England by the NHS Performance Rating System, introduced in September 2001 (http://www.doh.gov.uk/performance_ratings/). The star-rating system now evaluates the performance of all NHS acute and specialist hospital trusts, ambulance trusts, mental health trusts and primary care trusts in England and in Wales. Star ratings from 0 to 3 stars are awarded annually depending on how an organisation has performed against a set of Performance Indicators (PIs) and a Clinical Governance Review by the Commission for Health Improvement (CHI). The intention is to reward three-star organizations with increased financial and 'earned autonomy'. Zero-star hospitals are investigated and senior management changed where appropriate. Further information about the star-rating system is available on the CHI website (<http://www.chi.nhs.uk/eng/ratings/index.shtml>). CHI is being replaced by the Commission for Health Audit and Inspection as of 2004 (<http://www.doh.gov.uk/statementofpurpose/>).

These initiatives fit into a wide range of reforms aimed at improving performance within the NHS. The creation of the National Institute for Clinical Excellence in 1999 is a notable example, as is the development of National Service Frameworks (NSFs), which define the strategic plans, clinical governance structures and treatment standards in priority areas of care (<http://www.doh.gov.uk/nsf/index.htm>).

4.2.1.2. Wales

Wales is subject to many of the same regulatory and evaluative frameworks as England, such as star-ratings, other auditing functions of CHI and NICE guidance. One important development in Wales has been its commitment to focus investment on the building of a new health information infrastructure, namely through the creation of electronic patient records (Health Information in Wales, 2003). Wales also adopted a public health strategy, which follows from the WHO European 'Health 21' framework and emphasises the socio-economic context for health (Welsh Office, 1998). Some interesting initiatives are also underway at the local level to define care delivery pathways and monitor the quality and outcomes of care (J. Bullivant, personal communication).

4.2.1.3. Scotland

Scotland has a long history of benchmarking, made possible by the existence of linked datasets throughout the health care system since the 1960s. In 1992, Scotland published its first comprehensive set of 45 Clinical Outcomes Indicators (Scottish Office, 1998). These indicators are still being collected throughout Scottish Trusts under the auspices of the newly-formed NHS Quality Improvement Scotland (www.nhshealthquality.org).

Scotland set out an ambitious public health strategy that, like its Welsh equivalent, focuses on improving health and reducing health inequalities (Scottish Office, 1999). The 2003 White Paper identified four priority areas that required particular attention: early years; teenage transition; workplace and communities (Scottish Executive, 2003).

Performance targets for NHSScotland are set out in a Performance Assessment Framework that comprises 60 quantitative performance indicators and 30 qualitative measures. The framework is intended to reflect national priorities and allow the public to judge the performance of NHSScotland (www.paf.scot.nhs.uk/paf/index.html). Whilst standards are defined nationally, accountability for achievement falls on each of the NHS Boards, which must outline their objectives and achievements in a cohesive Local Health Plan.

Other important developments include the creation of the new Scottish Health Council, which monitors the performance and effectiveness of Boards in relation to public involvement. Scotland has also developed its own National Service Frameworks for coronary artery disease, cancer, and mental health.

4.2.1.4. Northern Ireland

In Northern Ireland, the Executive sets out its Programme for Government every 3 years to signal priorities and plans to improve the work of public services. (Department of Health and Social Care in Northern Ireland, 2002; 2003). The implementation of strategic plans falls upon the four Health and Social Services Boards and the HSS Trusts, who are responsible for organising their service delivery to achieve the goals set out by government.

As a complement to the Programme for Government, a public health strategy document 'sets direction, identifies priorities and sets objectives and targets to help measure success.' (Department of Health, Social Services and Public Safety for Northern Ireland, 2002b). Like its Welsh and Scottish equivalents, the proposed strategy follows the WHO European 'Health 21' framework. It is rooted in two overarching goals (to improve the health status of all people and to reduce inequalities in health) and seven objectives (mental health, poverty reduction, neighbourhood and environmental health, affordable housing, and reduction in workplace and road accidents). Specific targets are defined against the two overarching goals. Targets and objectives are reviewed every three years in order to be in line with the Programme for Government timeframe. Monitoring and accountability schemes bind locally-established Investment for Health partnerships and central government.

4.2.2. France

The French health care system has traditionally lacked a strong evaluation culture and its health insurance system, involving multiple sickness funds, has never encouraged the creation of a single health information infrastructure. This situation promises to change significantly, however, with the emergence of a proposed new bill on public health (*Projet de loi de politique de santé publique*, 2003). This bill proposes the creation of a new evaluative framework that will help build up the capability within the health care system for collecting the necessary data to evaluate progress against national objectives over time.

The proposed public health bill has two main components: the development of strategic plans in designated priority areas, and the establishment of a framework of 100 targets/objectives against which specific indicators are to be collected over a five-year period. The strategic plans are similar to the National Service Frameworks in England and cover five topics: cancer, environmental health, violence and behavioural risks, rare disorders and the quality of life of persons affected by chronic illnesses.

Priority areas for action were determined through a broad public consultation in September 2002. On the basis of these, an expert committee set out an assessment framework of 100 objectives or targets for public health to be achieved over a five-year period. Achievement against these targets is to be assessed using specific indicators at the national and regional level. Achievements against these targets are to be evaluated and debated within Parliament every 5 years.

The priority areas for indicator collection were selected on the basis of

- Their significance in terms of burden of illness
- Their fit with societal values and priorities
- Evidence of inequalities in health outcomes for the condition/problem within the country, or poor outcomes in France compared to other countries.
- Current state of knowledge about the condition/health problem's etiology, determinants or risk factors, treatment options and the effectiveness of actions to impact upon it.

(Projet de loi de santé publique, 2003)

The retained priority areas for target and indicator development are listed in **Table 8** below.

Table 8: Priority areas for public health action in France
(Projet de loi de politique de santé publique, 2003)

<p>Behavioural risk factors:</p> <p>Alcohol</p> <p>Tobacco</p> <p>Nutrition and physical activity</p> <p>Occupational health</p> <p>Environmental health</p> <p>Iatrogenic infections</p> <p>Pain</p> <p>Poverty and inequalities</p> <p>Handicaps and disabilities</p> <p>Infectious disease</p> <p>Maternal and perinatal health</p> <p>Cancer</p> <p>Endocrine disorders</p> <p>Neuro-psychiatric disorders</p> <p>Sensory organ disorders</p> <p>Cardiovascular disease</p> <p>Respiratory disease</p> <p>Chronic inflammatory disease of the intestine</p> <p>Chronic renal insufficiency</p> <p>Gynecological disorders</p> <p>Musculoskeletal disorders</p> <p>Antenatal care</p> <p>Rare disorders</p> <p>Oral health</p> <p>Violence</p> <p>Vulnerable populations:</p> <p>Learning disabilities</p> <p>Reproductive health, fertility, IVF</p> <p>Health of the elderly</p>

The new bill has been accepted in principle by the French government, however it is unclear what the precise timeframe may be for its implementation.

4.2.3. Canada

Canada has a highly devolved healthcare system, with each of its 10 provinces having full stewardship and accountability for its own health care system and setting its own objectives and strategic goals (Lomas *et al*, 1997; Williamson *et al*, 2003). Nonetheless, the federal government plays a pivotal role in the evaluation and regulation of the health care system and has made the establishment of a nation-wide health information infrastructure a priority. Health system performance is part of a Canadian Health Information Roadmap Initiative Indicators Framework (Canadian Institute for Health Information, 2000). This framework is based on the population health model and aims to provide high-quality comparative information on four dimensions: health status, non-medical determinants of health, health system performance and community and health system characteristics. Both 'conceptually and operationally', the framework is strongly supported by the strength of the underpinning data (Arah *et al*, 2003). Moreover, pan-Canadian surveys of patients and the public periodically provide patient and public perspectives on the health care system (Canadian Institute for Health Information, 2001). The latest list of indicators and an illustration of the framework used to develop them are available on http://secure.cihi.ca/cihiweb/products/indicators2003_e.pdf.

The Performance Indicators are seen as the guide to monitor, evaluate and improve the quality and outcomes of care and service delivery and are well integrated with other quality-improvement strategies and evaluation policies in Canada (Arah *et al*, 2003). For example, the Canadian Council on Health Services Accreditation (CCHSA) introduced the 'Achieving Improving Measurement' (AIM) strategy that uses benchmarking between organizations to boost quality improvement. Performance Indicators are linked to quality measures captured in the accreditation framework (Canadian Council on Health Services Accreditation, 1996). Moreover, a new Health Council of Canada has been proposed in order to provide "a clear, comprehensive and consistent analysis of the outcomes the system achieves and the progress that is being made in improving quality...[which will] guide decision makers, identify areas where action is needed, compare Canada's outcomes with other countries around the world, and perhaps most importantly, provide Canadians with solid information about the performance of their health care system". (Romanow report, p. 154).

PART V: ASSESSING BENCHMARKING FROM A POLICY PERSPECTIVE

5.1. A framework for the assessment of benchmarking efforts

Whilst they may differ in their scope and approach, the different policy models discussed above raise a common set of questions that form a helpful framework to assess the value and potential impact of benchmarking initiatives. These questions are outlined in **Figure 5** below.

Figure 5: List of key questions to serve as basis for assessing benchmarking efforts from a policy perspective

1. What is the underlying conceptual framework?
2. Are policy objectives and goals explicit to guide benchmarking within health care systems?
3. Do benchmarking indicators reflect policy objectives? Do policy objectives drive the selection and collection of indicators?
4. Does the collection of benchmarking data influence and guide policy?
5. Is the policy process underpinning benchmarking coherent?
6. Does the publication of benchmarking data stimulate change (better performance of the health care system? better health care?)
7. Does the articulation from objectives to targets to indicators and back to policy follow a top-down or a bottom-up approach?

In the subsequent **Sections 5.2-5.8**, each of these questions is discussed within the context of the national and international benchmarking initiatives introduced previously in this

report (**Section 4**). The implications of these questions for future analyses of benchmarking efforts are then discussed in **Section 6**.

5.2. What is the underlying conceptual framework of existing benchmarking systems?

The underlying notion driving international benchmarking is that cross-country comparisons may provide ‘a laboratory for others’ and contribute towards improvement and standardisation across different health care systems (OECD, 1990). As mentioned previously, this is the dominant conceptual framework of the OECD international comparisons framework (Arah *et al.*, 2003). The WHO goes one step further, in that they aim, through the presentation of comparative health system performance data, to foster stewardship for health within national governments (Arah *et al.*, 2003). Even the harshest critics of the WHO 2000 report recognise that its publication played a significant role in raising the profile of accountability for health on national political agendas.

National benchmarking systems are defined by the managerial goals that drive their implementation and the overall performance objectives that the health care system is being set to achieve (Arah *et al.*, 2003). Arah *et al.* undertook a thorough review of the conceptual framework underlying different performance assessment frameworks and found significant divergence (Arah *et al.*, 2003). For example, in the UK the star-rating system and the PAF were implemented as part of the government’s drive to ensure corporate governance through benchmarking. The indicators proposed for comparing performance reflect NHS priorities, namely to ensure access to effective, prompt and high quality care. By contrast, the Canadian performance management system was devised in order to stimulate integrated management and facilitate decision-making through the provision of data. System performance is defined in terms of improved health promotion, access, effectiveness and quality of care (Arah *et al.*, 2003). In France, benchmarking efforts are rooted in a public health approach to health policy, a recognised need for better epidemiological information and a desire to reduce within-country health inequalities (*Projet de loi de santé publique*, 2003). Within the UK, Wales, Scotland and Northern Ireland also favour a public health approach to evaluating system performance, borrowing from the WHO European ‘Health 21’ framework that emphasizes the socio-economic context for health (Welsh Office, 1998; Scottish Office, 1999; Department of Health, Social Services and Public Safety for Northern Ireland, 2002b).

5.3. Are policy objectives made explicit within health care systems?

A fundamental premise in the adoption of benchmarking is that comparative analysis of indicators will contribute towards the achievement of health system objectives. This premise underlies all of the rational theory-based policy models explored in **Part IV** of this report. It assumes that health system objectives are explicit and that they govern benchmarking and policy-making.

Both the WHO Health System Performance Framework and the OECD international comparisons suggest that health care systems should be measured against set objectives. The WHO framework identifies the three fundamental goals of a health care system as improving health, ensuring responsiveness to population expectations, and assuring fairness of financial contribution and overall equity in health expenditure. Considering the distribution dimension to each of these goals, the five components of performance become:

- Maximising population health
- Reducing inequalities in population health
- Maximising health system responsiveness
- Reducing inequalities in responsiveness
- Ensuring health care equitably.

(Murray and Frenk, 2000)

The OECD framework for international benchmarking of health care systems adopts a similar approach. It too hinges on three system goals: health improvement and outcomes, responsiveness and access, and financial contribution and health expenditure. However, whereas the WHO framework considers access as a determinant of responsiveness, the OECD framework sees it as a core component of responsiveness. By considering the distribution and average levels of each of these health system goals, one arrives at a system assessment framework hinging on four dimensions of performance: health improvement/outcomes, responsiveness, equity (of health outcomes, access and finance, respectively), and efficiency (macroeconomic and microeconomic) (Arah *et al.*, 2003). This framework is illustrated in **Figure 6** below.

Figure 6: The OECD framework for defining health system objectives

	Average level	Distribution
Health improvement/outcomes	✓	✓
Responsiveness and access	✓	✓
Financial contribution/health expenditure	✓	✓
	_____	_____
	Efficiency	Equity

Adapted from Murray and Frenk, 2000.

If these international comparisons stipulate what the objectives of health care systems should be, identifying the actual objectives that national policy-makers set for their respective health care systems is a much more elusive task. All the national benchmarking initiatives studied stipulate that they aim to measure health system performance against nationally-agreed objectives and policy goals. Yet in policy documents, the term ‘policy objectives’ is used to denote at once the values underpinning policy statements (eg. “a patient-centered service” in the NHS Plan), specific targets to be met (e.g. reducing waiting

lists) and priority areas for action and injection of funds (e.g. cancer, chronic illness, safety). Often, the term ‘objectives’ is confused with ‘targets’. Targets were defined by the WHO as “specific, quantifiable and measurable objectives designed to improve the health of individuals and families, of communities and of regional and national populations” (Marinker, 2002). Objectives may be defined as the overarching goals that guide a health care system. For example, the Romanow Commission identified objectives through an extensive public consultation process and recommended that these objectives guide all future health care reforms, evaluative frameworks and decision-making in Canada. Benchmarking is based on a core indicator set of objectives that is meant to “mirror approved national goals, strategies, guidelines, standards and benchmarks”. (Arah *et al*, 2003).

The distinction between objectives and targets is not purely semantic. Without being derived from a goal-driven conceptual framework, targets may appear to be disjointed and ‘plucked from thin air’, and indicator data may be incapable of providing a reliable and relevant picture of how well the health care system is achieving its overall policy goals.

The Performance Assessment Framework and Star-rating system adopted in England are a sobering illustration of this situation and confirm the ‘disjointed incrementalism’ model of policy presented earlier. The PAF stipulates that nationally-stated objectives are translated into locally-set targets that are in turn negotiated based on the current performance and the perceived scope for improvement at the local, or individual institution, level (Smith 2002). However, it is somewhat ambiguous where the objectives are defined and how the overall objectives of the NHS are reflected in the chosen benchmarking indicators. The *Public Service Agreement* lists eight objectives against which the Treasury will evaluate the performance of the Department of Health (Her Majesty’s Treasury, 2000). Yet there is no apparent link between these objectives, the strategic goals set out in the *NHS Plan for England and Wales*, or the headings of the *Performance Assessment Framework* (**Table 9**). Which of these different “objectives” are intended to govern the evaluation of performance within the health care system remains unclear.

Table 9: NHS policy objectives as stated in key policy documents

Public Service Agreement (*Her Majesty's Treasury, 2000*):

- reduce mortality rates from major killers
- narrow health discrepancies
- treat patients at a time that suits them in accordance with their medical need
- reduce the maximum waits for outpatient appointment and inpatient treatment
- secure year-on-year improvements in patient satisfaction
- provide high-quality preadmission and rehabilitation care to older persons
- guarantee rapid access to primary care
- improve value for money

NHS Plan (*Department of Health, 2000; NHS Modernisation Agency, 2003*):

- improvements in clinical quality
- faster access, more choice
- expanding our workforce
- better ways of working
- a patient-centred service
- meeting the needs of older people and children

Performance Assessment Framework (*Freeman 2002*):

- health improvement
- fair access
- effective delivery
- efficiency
- patient/carer experience
- health outcomes

The House of Commons Public Administration Select Committee found evidence that this lack of articulation between policy objectives, targets and indicators undermined the credibility of benchmarking initiatives in the eyes of those responsible for delivering public services in the UK. They stated lack of clarity about what the government is trying to achieve and failure to provide a clear sense of direction as major failings of the current target-based evaluative frameworks (House of Commons Public Administration Select Committee, 2003).

The authors of the report state:

“Targets should never be accepted as a substitute for a clearly expressed strategy and set of priorities... The target setting process has subverted this relationship with targets becoming

almost an end in themselves, rather than providing an accurate measure of progress towards the organisation's goals and objectives."

"Targets can be good servants, but they are poor masters."

(House of Commons Select Committee, 2003)

5.4. Do benchmarking indicators reflect policy objectives? Do policy objectives drive the selection and collection of indicators?

"A growing challenge is how the Performance Assessment Framework can strike a balance between providing a broad overview of the general performance of local health systems/organisations and an assessment of their progress in relation to current strategic priorities." (Smee 2000, p. 8).

The criticisms raised above are a reminder what authors have termed the 'unintended consequences of benchmarking' (Smith, 1995). Lack of available data may force health care systems to focus measurement on what is readily available, leading to 'cream skimming', or the selective attention drawn to what can be measured at the expense of what may be important but remains unmeasured (Smith, 1995). It follows that indicators may not necessarily reflect areas that need improvement, require prioritization and match system goals and values (McKee and James, 1997; Smith, 1995; Smith, 2002; Walshe, 2003; Walshe and Freeman, 2002). Moreover, the culture of benchmarking has inadvertently led to a near-exclusive emphasis on the results of health care systems, with too little attention devoted to understanding why and how the observed results were obtained (Blalock 1999). Instead of encouraging a sense of ownership for performance data, these combined elements may contribute to a culture of 'box-ticking' amongst health care managers under pressure to comply with sometimes arbitrary performance standards. Musgrove, in his criticism of the WHO performance data, proposed that it is only when health system attainment measures (outcomes) can be linked to the causal factor (e.g. what outcomes are attributable to hospital care) that data derived can be used to focus efforts on aspects of the health care system where a difference needs to be made (Musgrove 2003).

If benchmarking is to achieve its goal of evaluating the achievement of a health system's policy objectives, then the starting point for evaluation should be the policies and objectives it is trying to achieve, not the data available to measure its outputs. The OECD Healthcare Quality Indicators project has made an important contribution towards focusing indicator selection on countries' health policy priorities. They conducted a survey of participating countries to determine their health policy priorities and then drew up a list of relevant indicators that were capable of informing on these priority areas. These indicators are to be added to the core set of 31 indicators used to compare health system quality across selected countries (S. Mattke, personal communication).³

3. At the time of writing this report, the future status of this project was uncertain given organisational changes within the OECD. To the author's knowledge, responsibility for this project has shifted to the Commonwealth Fund.

It must be underlined, however, that the articulation from system objectives to reliable targets and meaningful indicators is by no means an easy task. Marinker suggests that targets stem from the interplay between epidemiological evidence, moral values, and political will. He suggests “it is precisely these desired characteristics of the processes and outcomes of targeting (specific, quantifiable, and measurable) [that do] not fit easily with the spread of motives that are necessary to give them impetus (epidemiological evidence, moral values, and political will)”. (Marinker, 2002). Arah *et al.* tested the fit between the Performance Indicators adopted in the Canadian Health Information Roadmap Initiative Indicators Framework against its eight performance objectives. Whilst indicators were clearly available to measure appropriateness, competence, effectiveness, efficiency, and safety of health care across Canada, they were lacking to assess how well the system was meeting its objectives of continuity, acceptability and competence (Arah *et al.*, 2003). Similar exploration of the fit between system objectives and available indicators may help provide guidance on where future indicator development should be prioritised.

There are hopeful signs that several benchmarking frameworks are meeting these challenges. One notable example is the recently proposed French public health indicator framework, which sets out a five-year framework for indicator collection rooted in explicitly stated system objectives (**Table 10**), policy priority areas (**Table 8**) and specific targets⁴ and represents a clear shift from an indicator-led to a policy-led culture. This approach is innovative in several respects. First, it makes the distinction between targets that are quantifiable from those that are experimental and may serve to provide information only. Secondly, it defines for each target the policy process and structural changes that need to take place to ensure proper data collection. Thirdly, it allows for data gaps where indicators may not yet exist and uses the framework to stimulate indicator development in these areas. In other words, benchmarking is set out as an evolving process that helps build the health care system’s capability for evaluation of policy outcomes over time. An illustration of the proposed framework as applied to one specific objective (injury and suicide prevention) is presented in **Table 11** overleaf.

Table 10: Underlying principles proposed in the new “Projet de loi sur la politique de santé publique” in France

- | |
|--|
| <ul style="list-style-type: none"> ● Evidence-based decision-making ● Reducing inequalities ● Parity (equal consideration of male and female issues) ● Protection of youth ● Early intervention (“precocite”) – focus on prevention ● Economic efficiency ● Multidisciplinary/intersectoral approach ● Broad consultation ● Evaluation. |
|--|

4. It should be noted that in the original French description of this framework, the term ‘*objectifs*’ (objectives) is used to denote the 100 specific targets described above. However, the appropriate translation of this term would be ‘target’ in this context, as the ‘*objectifs*’ are specific and meet the definition proposed by the WHO, described above.

Table 11: Example of targets and indicators set out for Injury and suicide prevention, Public Health Framework 2003

Type of target	Target	Requirements	Indicators
Quantifiable target	Reduce suicide occurrence by 20% by 2008 (from 12,000 deaths by suicide per year to 10,000 deaths by suicide per year)		Incidence of suicide deaths by age group within the general population
Targets aimed at producing epidemiological data	Reduce by 50% the rate of non-intentional injuries in children aged <14 years by 2008	Set up registry of incidence and severity data on incidents by cause and age range	30-day mortality rate after an incident in children aged 1-14 years, by cause and gender
		Known: mortality rates: 8.9 deaths per 100,000 boys and 5.8 deaths per 100,000 girls aged 1-4 years	
	Reduce by 50% the number of deaths and serious injuries resulting from road traffic accidents by 2008	Need to set up an epidemiological surveillance system of road traffic accidents and their sequelae	Rate of severe injuries due to road traffic accidents (by age and gender) Rate of deaths due to road traffic accidents (by age and gender)
Target aimed at strengthening scientific knowledge	Intentional childhood traumas: defining of effective public health measures	Gather all the necessary scientific data	

Adapted from Project de loi de santé publique, 2003

5.5. Does the collection of benchmarking data influence and guide policy?

“The problem with measurement is that it can be a loaded gun: dangerous if misused and at least threatening if pointed in the wrong direction.” (O’Leary 1995, p. 354).

Perhaps the most important goal of benchmarking is to influence and guide policy (Marshall *et al*, 2003). Yet to achieve this, not only must benchmarking data provide a valid and reliable picture of health care system performance, but they must lend themselves to correct interpretation on the part of policy-makers, who in turn must have the appropriate skills and political will to make the appropriate policy changes based on results. Several authors have suggested that most policies are far from evidence-based and that very little research actually gets translated into policy decisions (McIntyre *et al*, 2001; Black, 2001). Others have warned of the risks of misinterpretation of benchmarking data brought on by political climates dominated by short-term political imperatives and high media attention (McKee, 2002; Walshe *et al*, 2001; Goddard *et al* 2000). Decision-makers armed with performance results may exhibit myopia (losing sight of the actual policy objective), tunnel vision (enacting legislation without thinking about its implementation), and blindness (turning a blind eye to the failures or limitations of policies in order to achieve results). Finally, if benchmarking is to guide policy-making, allowance for sufficient time to elapse is needed for proper evaluation to take place and changes to be implemented, monitored and assessed. (House of Commons Select Committee, 2003).

The increasing role that international benchmarking data have in shaping national health policies is evidenced in all recent policy documents in the countries studied (Romanov report, 2002; Wanless, 2002; Wanless, 2003; Projet de loi de santé publique, 2003). The WHO created the *Enhancing Health Systems Performance Initiative* (EHSPI) in parallel with the publication of its 2000 report in order to stimulate dialogues with policy-makers. Through extensive regional consultations with policy-makers, the EHSPI aims to help countries build capacity to address areas identified as priorities in their performance results. This initiative has been very successful in some of the less-developed countries such as Uganda and Iran (WHO, personal communication).

It is worth reflecting on what aspects of benchmarking frameworks may favour the use of data to influence policy:

- *Transparency* about the methods used to derive indicators and possible caveats of data may help prevent inappropriate interpretation of results.
- *A clear classification of target achievement*: For example, the Scottish Executive advocated that targets be reported as ‘achieved’, ‘ongoing’, ‘on track’, ‘delayed’ or ‘may not be achieved’. In many ways, this is the philosophy underlying the French proposed assessment framework in its distinction between targets for information or for evaluative purposes.
- *Limiting the number of indicators and targets* in order to focus measurement on what is considered a priority. Scotland, for example, adopted the ‘*Focus on Four*’ approach to concentrate on areas that require particular attention and set no more than 12 priorities for NHSScotland per year (Scottish Executive, 2003).

- *Specifying the timeframe needed* to allow for meaningful data interpretation may also be helpful.
- *Ensuring timeliness of data* – this was one of the reported failings of the Scottish Clinical outcomes indicators, which were often published 18 months after data collection (Mannion and Goddard, 2001).
- *Presenting clusters of performance indicators* to avoid gaming and facilitate interpretation, as recommended by the Audit Commission in England (Audit Commission, 2003b).

5.6. Does benchmarking stimulate change? lead to better health care system performance? to better care?

Benchmarking is based on the premise that the public reporting of data may influence patients in their choice of provider or treatment, thus allowing them to make better use of available resources and thereby improve the outcomes of care. As attractive as this theory may be intuitively, there is currently little published evidence to validate it, mostly due to the fact that so many other confounding variables enter into play to determine the outcomes of patient care. In fact, evidence from the UK and the United States suggests that the public is fairly apathetic to publicly-disclosed information on performance. Interestingly, this finding was true both to the US, a highly consumer-oriented market, and to the UK, where patient choice is inherently restricted by a state-run system (Marshall *et al.*, 2003).

On the provider side, existing data tend to confirm the perverse effects of benchmarking data on clinical behaviour rather than demonstrate their potential for improving performance. In the UK, an audit of 9 NHS hospital trusts found evidence of manipulation of waiting list data or ‘gaming’ in order to meet government waiting list targets, often resulting in delayed treatment for other patients (Goddard *et al.*, 2000, National Audit Office 2001). Most often however, the publication of benchmarking data is simply met with apathy, just as the publication of clinical guidelines and recommendations alone has been found to bear little impact on changing physician behaviour (Greco and Eisenberg, 1993; Lomas *et al.*, 1989; Cabana *et al.*, 1999). ‘Target fatigue’ is a coined phrase often used to describe the lethargy brought onto hospitals, practitioners and local managers faced with increasing checklists of targets to be met and indicators to be measured and little rationale for how these data will be used.

The Scottish Executive carried out an evaluation of the impact of the Clinical Outcomes Indicators in Scotland in 1997 and found that, despite some positive anecdotal changes, overall the clinical indicators had little impact on clinical behaviour. The reasons stated by surveyed hospitals included lack of credibility of the indicator framework, poor timeliness, low awareness of the data within hospitals, little training and facilitation to interpret the data, no incentives to meet targets and no external accountability for the data (Mannion and Goddard, 2001). Similarly, Walshe evaluated the impact that external review mechanisms, including performance assessment, had on hospital trusts in a region of England and concluded: “We know little about the beneficial or adverse impacts that external review actually has on the performance of reviewed organizations, how much external reviews cost,

which review methods, approaches or measurement techniques are most valid, reliable and effective in particular contexts.” (Walshe *et al*, 2001). This comment may be applied as much to national as to international benchmarking efforts.

5.7. Is the policy process underpinning benchmarking coherent?

“Developing sound and consistent measures of the quality of Canada’s health care system is important work but to have an impact, it must be linked to effective mechanisms and policy changes aimed at improving quality and outcomes.”

Romanow report, 2002

The concept of the policy process was evoked in several of the policy models described in **Part IV**. The effectiveness of performance management systems may be measured in terms of i) coherence, ii) capacity and iii) clinical engagement (Smith 2002). Clinical engagement has been discussed previously. Capacity suggests that the health care system has the structures and skills in place to accommodate a performing benchmarking system and ensure that it meets its objectives. Coherence implies that benchmarking initiatives fit in with other external review (accreditation, audit, evaluation, regulation) systems in place within the health care system and that their roles are complementary and not overlapping.

All of the health care systems studied have recognized the need to bring more coherence to their regulatory and evaluative frameworks. In Canada, the Romanow report calls for the creation of a Canadian Health Council, which would align policy prioritization, indicator development and performance assessment within a single organization. In France, significant changes to the public health bill (*Code de santé publique*) are being proposed in parallel with the Public Health framework in order to align organizations responsible for different facets of policy-setting, regulation and evaluation and create a more efficient policy process. In the UK, the intense focus on external review and the plethora of organizations taking responsibility for different aspects of it have contributed to a climate of “inspectorial overload” at the frontlines within the NHS (Walshe *et al*, 2001). The consolidation of roles within the *Commission for Healthcare Audit and Inspection* (CHAI) as of 2004 promises to improve coherence and coordination between different evaluation mechanisms and realign policy objectives. Similar consolidation is being evidenced in Scotland, with the joining of the Clinical Standards Board for Scotland, Health Technology Board Scotland and other organisations under the single umbrella of the NHS Quality Improvement Scotland as of 2003 (www.nhshealthquality.org).

Recent policy initiatives have also focused on ensuring that suitable capacity exists to accommodate the collection of reliable and valid benchmarking data, and all of the health care systems studied are boosting the capacity of their health information infrastructure. Wales outlined an ambitious strategy to build a new health information infrastructure (Health Information in Wales, 2003). In France, the proposed indicator framework outlines explicitly the needed changes in capacity alongside each indicator and target, clearly suggesting that these efforts must be driven simultaneously to ensure reliable data

collection. Canada's early investment in a sophisticated national health information infrastructure has lent significant credibility to benchmarking efforts and allowed for the comparative assessment of data and changes over time (Canadian Health Services Research Foundation, personal communication).

5.8. Does benchmarking follow a top-down or a bottom-up approach?

In most of the benchmarking frameworks explored in this report, objectives and priorities are set nationally and trickled down to regional or local levels for implementation. This centrally-driven performance management culture tests the trust and relationships between different players in the decentralized system and raises issues of governance, accountability and trust. As a remedy to this tension, policy-makers have been advocating 'a new localism' in benchmarking (House of Commons Select Committee, 2003). The Audit Commission in England speaks of a 'differentiated approach' that starts with clear national aspirations and then uses clusters of national performance indicators combined with locally-set targets to foster improvement (Audit Commission, 2003b). The Labour Government has promised to allow 'constrained discretion' to filter into the NHS performance frameworks. In Scotland, whilst standards are nationally-set, accountability for achievement falls on each of the local NHS Boards (Scottish Executive, 2003). In France, the regions are explicitly defined as the appropriate level of implementation and accountability for public health policies, and regional public health groupings of stakeholders (*Groupement d'intérêt public régional*) have responsibility for the coordination of regional programmes and the implementation of national policy objectives.

In Canada, tension between national governance and implementation at the provincial or territorial levels is a major challenge facing the health care system (Romanow report, 2002). Interestingly, health care system goals are defined at the provincial and territorial, as opposed to the national, level in Canada. A recent study found that whilst provincial/territorial objectives and goals provide a philosophical backdrop to local health planning and policies, they bore less influence than did locally-defined goals and objectives. The authors suggest that their findings support the view that rational models of policy-making are naïve, and that the bottom-up models may more realistically reflect the complex nature of policy development and implementation (Williamson *et al*, 2003). Nonetheless, they also state that centrally-defined system objectives and goals play an essential symbolic role in setting a cohesive framework for priority-setting and policy-making at all levels of the health care system. This observation supports Sabatier's theory that the importance of overarching system goals should not be underestimated, as they filter to local levels of policy-making even in seemingly 'bottom-up' decentralised systems (Sabatier, 1997).

Probably the most important and striking common feature of all the policy documents evaluated for the purposes of this research was their focus on public engagement as a priority for the future of health policy. Some policy documents state that there should be public involvement "at every stage in the design, delivery and review of health services" (Scottish Executive, 2003). Broad public consultations played a key role in informing the proposed strategy and priorities advocated in the Romanow report on the future of the

Canadian health care system, the Wanless report on the future of the NHS in England and in Wales (Wanless 2002; 2003), and the *Projet de loi de santé publique* in France. Many policy-makers advocate public involvement to determine the goals and priorities for the health care system and set expectations for health care system performance. It seems to be accepted as fact that public engagement will lead to more relevant policy-making and that public pressure will play a significant role to stimulate improved system performance (Marshall *et al*, 2003).

Whilst the idea of public engagement in health policy and specifically in benchmarking is both appealing and laudable, it is difficult to conceive of how to achieve it. The mechanisms or changes needed to move the public away from the current culture of apathy towards benchmarking data have yet to be defined. Moreover, one must acknowledge that true devolution coupled with public engagement in performance assessment would significantly transform the existing system of governance within most health care systems. It will be of interest to see what the impact of current policy developments such as the creation of the Office for Information on Health Care Performance in the UK may have on increasing public awareness of and interest in benchmarking data. Realistically, several years may be needed within health care systems and society in general before this new paradigm of health policy reaches full maturity.

PART VI: DISCUSSION

Benchmarking in health stands at a conceptual and practical crossroads. The focus so far has been on achieving measurement, yet the contribution of this measurement to encouraging evidence-based policy and improving health system performance has yet to be ascertained (Walshe, 2001; Goddard, 2000).

The challenge with evaluating the impact of benchmarking is that indicator frameworks are changed continually, thus precluding the possibility of observing incremental changes in results over time. Moreover, confounding factors make it difficult to isolate the effect that benchmarking may have on health system performance. The rapid pace of health care reforms does not allow for sufficient time to observe the impact of given policies through indicator data, and similarly, to allow the knowledge gained from indicator data to filter back into the policy-making process.

This report has attempted to draw from the theoretical and policy literature as well as from the experience of international and national benchmarking initiatives to propose a framework that may be used to evaluate benchmarking efforts in the future. The proposed framework will be applied to the analysis of benchmarking efforts in specific areas of policy, namely public engagement and cancer policy, in subsequent phases of this research. It is the hope that these analyses will allow to further test and develop the proposed framework and thereby provide further insights into how to best assess the relationship between health policies, their stated goals and objectives and the measurement systems used to evaluate their impact.

The concepts and practical experience of benchmarking explored in this report allow to some tentative conclusions about the policy relevance of benchmarking:

- The pursuit of meaningful indicators to evaluate and compare health system performance within and across health care systems is essential in a health policy environment that aims to foster accountability and stewardship for health.
- Problems of data availability remain a significant challenge to benchmarking within and across countries. More focus needs to be given to developing rigorous health information systems capable of ensuring systematic, high-quality data collection.

- Further research is still needed to identify the most reliable and valid indicators to measure different dimensions of health system performance. The public, healthcare managers and clinicians, policy-makers and the media need to be made aware of the limitations of existing indicator data to avoid misinterpretation. Moreover, benchmarking frameworks should make explicit the time and capacity conditions that must be met to facilitate data interpretation.
- The rational policy model suggesting that clear policies and objectives should drive indicator measurement and that evidence collected should guide policy-making is not to be entirely dismissed. On the contrary, clearly-stated objectives and a solid conceptual framework underlying benchmarking lend credibility to the indicator and target data being collected and may contribute to the permeation of a true performance culture throughout the health care system.
- Frameworks for evaluating the performance of health care systems should be policy-driven, as opposed to data-driven. The starting point for any benchmarking framework should be the policies and objectives that one is wishing to evaluate. The collection of data that may appropriately and reliably reflect progress against these goals should be prioritised.
- The interpretation of indicator data should not lose sight of the policy context within which they are measured, the players involved in formulating and implementing policy, the time lag needed to assess the impact of different policies, and aspects of health care that remain unmeasured by available data.
- The policy relevance of benchmarking hinges on the overall cohesiveness of the policy process. Evaluating indicator data across countries without any understanding for the regulatory and evaluative policies underpinning them may hamper data interpretation.
- The move to further decentralisation within health care systems is likely to have a significant impact on the future of benchmarking. For true devolution to be achieved, local players responsible for data collection will need to accept greater ownership for performance data, whilst policy-makers will need to shift policy priorities to reflect local needs. The significant shifts in accountability and governance that true devolution would bring about within health care systems should not be underestimated.
- Policy-makers have fully embraced the notion that public engagement in all policy-making will act as a key driver to improving health system performance. Whether policies may truly be effective in engaging the public and how they may translate into improved performance and health outcomes has yet to be determined.

In summary, reliable comparisons of indicator data within and across countries may provide a powerful tool to guide policy and stimulate learning both within and between health care systems. The recent experience of international and national benchmarking initiatives has brought to light the challenges inherent in collecting, evaluating and interpreting indicator data to guide policy-making. It illustrates the need to define policy objectives within health care systems, ensure coherence between different evaluative and regulatory processes, and move away from a culture of measurement towards a culture that embraces performance and is truly evidence-driven. There are encouraging signs that benchmarking may yet achieve its policy objectives. However, sufficient time must be allowed to assess its policy relevance. In the meantime, further research is encouraged to help policy-makers use comparative data appropriately to help understand discrepancies in health care and ultimately, improve the quality of care.

REFERENCES

Almeida C, Braveman P, Gold MR, *et al.* Methodological concerns and recommendations on policy consequences of the World Health report 2000. *Lancet* 2001; 357: 1692-7.

Appleby J, Boyle S. "Tony's Billions", *Health Care UK* (Spring 2000): 64-70.

Arah OA, Klazinga NS, Delnoij DM, ten Asbroek AH, Custers T. Conceptual frameworks for health systems performance: a quest for effectiveness, quality, and improvement. *Int J Qual Health Care*. 2003; 15(5):377-98.

Audit Commission. *Patterns for improvement. Learning from comprehensive performance assessment to achieve better public services*. London: Audit Commission (2003). Weblink: www.audit-commission.gov.uk.

Audit Commission. *Targets in the public sector*. October 2003 (2003b). Available on: www.audit-commission.gov.uk

Berwick DM. Continuous improvement as an ideal in health care. *N Engl J Med* 1989; 320: 53-56.

Berwick DM, Godfrey AB, Roessner J. *Curing health care*. San Francisco: Jossey Bass, 1990.

Berrino *et al* (eds). *Survival of cancer patients in Europe. The Eurocare study*. International Agency for Research on Cancer no. 132. Lyon, 1995.

Black N. Evidence based policy: proceed with care. *BMJ* 2001 Aug 4; 323(7307):275-9.

Blalock AB. Evaluation research and the performance management movement. From estrangement to useful integration? *Evaluation* 1999; 5(2): 117-149.

Brennan TA. The role of regulation in quality improvement. *Milbank Quarterly* 1998; 76: 709-31.

The Bristol Royal Infirmary Inquiry (2001), *Learning from Bristol: the report of the public inquiry into children's heart surgery at the Bristol Royal Infirmary 1984-1995*. Command paper: CM 5207. The Bristol Royal Infirmary Inquiry, Bristol.

- Busse R, Schlette S (eds). *Health Policy Developments. International Trends and Analyses. Issue 1/2003*. Bertelsmann Foundation Publishers. Available on: www.bertelsmann-foundation.de/publications.
- Cabana MD, Rand CS, Power NR, Wu AW, Wilson MH, Abboud PA, Rubin HR. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA* 1999; 282(15): 1458-65.
- Canadian Council on Health Services Accreditation. *A guide to the Development and Use of Performance Indicators*. Ottawa: Canadian Council on Health Services Accreditation, 1996.
- Canadian Institute for Health Information (2001). A portrait of Canada's health care system. In *Health Care in Canada 2001*. Ottawa: Canadian Institute for Health Information.
- Canadian Institute for Health Information (2000). Statistics Canada. *Roadmap Initiative...Launching the Process*. Ottawa: Canadian Institute for Health Information.
- Commission for Health Improvement (2003). NHS performance ratings. Weblink: <http://www.chi.gov.uk/eng/ratings/index.shtml> (accessed 20 August 2003).
- Commission on the Future of Health Care in Canada (2002), *Building on Values: The Future of Health Care in Canada – Final Report*. (Romanow report), Ottawa. Weblink: http://www.healthcarecommission.ca/Suite247/Common/GetMedia_WO.asp?MediaID=1162&Filename=HCC_Final_Report.pdf.
- Dargie, C., in association with Dawson, S. and Garside, P. (1999) Policy futures for UK health: Pathfinder: a consultation document. London: Nuffield Trust.
- Dargie, C. in association with Dawson, S. and Garside, P. (2000) Policy futures for UK health: 2000 report. London: Stationery Office.
- Department of Health (1997). *The New NHS: Modern, Dependable*. London: Stationery Office.
- Department of Health (2000). *The NHS Plan: A Plan for investment, a plan for reform*. London, Stationery Office.
- Department of Health (2001a), *Priorities and Planning Framework 2002/2003*. London: Stationery Office.
- Department of Health (2001b). *NHS Performance Indicators: A Consultation*. London: Stationery Office.
- Department of Health (2001c). *NHS Performance Ratings, Acute Trusts 2000/01*. London: Stationery Office.
- Department of Health, Social Services and Public Safety for Northern Ireland (2002). Draft programme for government: 2002-2003. Available on: <http://www.pfgni.gov.uk/draftsept02/master.pdf>
- Department of Health, Social Services and Public Safety for Northern Ireland (2002b). *Investing for Health*. Available on: <http://www.dhsspsni.gov.uk/ihc>.

- Department of Health, Social Services and Public Safety for Northern Ireland (2003). *Building on progress: priorities for 2003-2004*. Available on: <http://www.pfgni.gov.uk/plans.pdf>
- Dixon J, Inglis S, Klein R. Is the English NHS underfunded? *BMJ* 1999; 318: 522-6.
- Dror Y. Muddling through – science or inertia? *Public Administration Review* 1964; 24: 153-7.
- Freeman T. Using performance indicators to improve health care quality in the public sector: a review of the literature. *Health Services Management Research* 2002; 15: 126-137.
- Girard J-F and Minville E. (2002), “Measuring up: lessons and potential” in: Smith, P. (ed.), *Measuring up. Improving health system performance in OECD countries*, OECD, Paris, pp. 337-47.
- Goddard M, Mannion R, Smith P. Enhancing performance in health care: a theoretical perspective on agency and the role of information. *Health Economics* 2000; 9: 95-107.
- Goldstein H, Spiegelhalter DJ. League tables and their limitations: statistical issues in comparisons of institutional performance. *J Royal Statistical Society Series A*. 1996; 159(3): 385-409.
- Gordon I, Lewis J, Young K. Perspectives on policy analysis. *Public Administration Bulletin* 1977; 25: 26-30.
- Greco PJ, Eisenberg JM. Changing physicians’ practices. *New England Journal of Medicine* 1993; 329(17): 1271-3.
- Health Information in Wales (2003). *Informing Healthcare. Transforming healthcare using information and IT*. Available on: <http://www.wales.nhs.uk/sites/documents/365/ihc-strategy-e.pdf>
- Herten van L. and Gunning-Schepers LJ. *Historical perspectives on European health and policy*. In: *Health targets in Europe: polity, progress and promise*. Marshall Marniker, Ed. BMJ books: London, 2002.
- Herten van LM, van de Water HPA. New global Health for All targets. *BMJ* 1999; 319: 700-3.
- Her Majesty’s Treasury (2000). *Spending Review 2000: New Public Spending Plans, 2001-2004*. Cm 4808. London: Stationery Office.
- Hogwood BW, Gunn LA. *Policy Analysis for the Real World* (1984): Oxford: Oxford University Press.
- House of Commons Public Administration Select Committee (2003). *On Target? Government by measurement*. Fifth Report of session 2002-2003. Available on: <http://www.publications.parliament.uk/pa/cm200203/cmselect/cmpublicadmi/62/6203.htm>.
- Hurst J and Jee-Hughes M. *Performance measurement and performance management in OECD health systems*. Labour market and social policy – Occasional papers 47 (Paris: OECD, 2000).
- Institute of Medicine (2001). *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington: National Academy Press.

- Institute of Medicine (2001b). *Envisioning the National Health Care Quality Report*. Hurtado MP, Swift EK, Corrigan JM, eds. Washington: National Academy Press.
- Jenkins B. *Policy analysis: models and approaches*. In: M. Hill (ed). *Policy process: a reader*. 2nd edition. Harlow: Prentice Hall (1997): 30-40.
- Kavarnos, P., and E. Mossialos. 1996. *The Methodology of International Comparisons of Health Care Expenditures: Any Lessons for Health Policy*. London.
- La liste noire des hôpitaux*. Paris, France: Sciences et Avenir, October 1997.
- Lane J-E. *Implementation, accountability and trust*. In: *The Policy Process: A Reader*. Michael Hill (Ed): 2nd edition, Prentice Hall, 1997: pp. 296-313. Also: *Eur J Political Research* 1987; 15(5): 527-46.
- Leatherman S, McCarthy D. *Quality of health care in the United States: a chartbook*. The Commonwealth Fund, NY, 2002.
- Leatherman S, Sutherland K. Evolving quality in the new NHS: policy, process, and pragmatic considerations. *Quality In Health Care* 1998; 7(Suppl): S54-S61.
- Lindblom CE. The science of “muddling through”. *Public Administration* 1959; 19: 79-99.
- Lomas J, Anderson GM, Dominick-Pierre K, Vayda E, Enkin MW, Hannah WJ. Do practice guidelines guide practice? The effect of a consensus statement on the practice of physicians. *New England Journal of Medicine* 1989; 321(19): 1306-1311.
- Lomas J, Woods J, Veenstra G. Devolving authority for health care in Canada's provinces: an introduction to the issues. *CMAJ* 1997; 156: 371-377.
- Mannion R, Goddard M. Impact of published clinical outcomes data: case study in NHS hospital trusts. *BMJ* 2001; 260-263.
- Marinker M. *Evidence and imagination*. In: *Health targets in Europe: polity, progress and promise*. Marshall Marniker, Ed. BMJ books: London, 2002.
- Marshall MN, Shekelle PG, Davies HTO, Smith PC. Public reporting on quality in the United States and the United Kingdom. *Health Affairs* 2003; 22(3): 134-148.
- Matillon Y, Loirat P, Guiraud-Chaumeil B. Les rôles de l'ANAES dans la régulation du système de santé français. *Actualité et Dossier en Santé Publique*. Dec 2001 ; 37 : 46-50.
- McGinnis JM. Targeting progress in health, *Public Health Reports* 1982; 97(4): 295-307.
- McIntyre S, Chalmers I, Horton R, Smith R. Using evidence to inform health policy: case study. *BMJ* 2001; 322: 222-225.
- McKee M. *Values, beliefs and implications*. In: *Health targets in Europe: polity, progress and promise*. Marshall Marniker, Ed. BMJ books: London, 2002.
- McKee M, James P. Using routine data to evaluate quality of care in British hospitals. *Medical Care* 1997; 35(10) Suppl: OS102-OS111.

Milward L, Kelly M, Nutbeam D. Public health intervention research: the evidence. London: Health Development Agency, 2001.

Murray CJ, Frenk J. A framework for assessing the performance of health systems. *Bull World Health Organ.* 2000;78(6):717-31.

Musgrove P. Judging health systems: reflections on WHO's methods. *Lancet* 2003; 361: 1817-20.

National Service Frameworks. Weblink: <http://www.doh.gov.uk/nsf/>

Navarro V. Assessment of the World Health Report 2000. *Lancet* 2000; 356: 1598-1601.

NHS Modernisation Agency. *The NHS Plan – a progress report.* The NHS Modernisation Board's Annual Report 2003.

Nutley S, Smith PC. League tables for performance improvement in health care. *J Health Serv Res Policy* 1998; 3(1): 50-57.

OECD (1985), *Measuring health care, 1960-1983 expenditure, costs and performance*, OECD Social Policy Studies No. 2. OECD, Paris.

OECD (1990), *Health care systems in transition.* OECD Social Policy Studies No. 7, OECD, Paris.

OECD (2003), *OECD Health Data 2003: a Comparative Analysis of 30 OECD Countries*, OECD, Paris.

O'Leary DS. Measurement and accountability: taking careful aim. *Jt Comm J Qual Improv* 1995; 21: 254-357.

O'Neill B. Doctor as murderer. Death certification needs tightening up, but it still might not have stopped Shipman. *BMJ* 2000; 320:329-330 (Editorial 5 February)

Pringle M, Wilson T, Grol R. Measuring "goodness" in individuals and healthcare systems. *BMJ* 2002; 325: 704-7.

Projet de loi sur la politique de santé publique. Rapport d'objectifs de santé publique. Weblink accessed on 21/5/03: http://www.sante/gouv.fr/htm/actu/sante_publicue/3sante_publicue.htm.

Richardson J, Wildman J, Robertson IK. A critique of the World Health Organisation's evaluation of health system performance. *Health Economics* 2003; 12: 355-366.

Ritsatakis A. "Experience in setting targets for health in Europe", *European Journal of Public Health* 2000, vol. 10 Suppl. 4, pp. 7-10.

Sabatier PA. *Top-down and bottom-up approaches to implementation research.* In: *The Policy Process: A Reader.* Michael Hill (Ed). Prentice Hall, 2nd Edition 1997.

Sabatier P and Mazmanian D. The conditions of effective implementation. *Policy Analysis* 1979; 5(Fall): 481-504.

Saltman and Figueras 1998, European Health Care Reform. WHO series no. 72

- Schieber GJ. *Financing and delivering health care. A comparative analysis of OECD countries*, OECD Social Policy Reports No. 4, OECD, Paris (1997).
- Scottish Executive (2003). NHS Scotland. *Partnership for care. Scotland's Health White Paper*. Available on: <http://www.scotland.gov.uk/library5/health/pfcs-00.asp>.
- Scottish Office Department of Health (1998). *Acute Services Review (Carter Report)*. Edinburgh.
- Scottish Office Department of Health (1999). *Working Together for a Healthier Scotland*. A consultation document. Edinburgh. Available on: <http://www.archive.official-documents.co.uk/document/scotoff/hscot/wthscot.htm>
- Shaw C, Kalo I. A background for national quality policies in health systems. WHO Regional Office for Europe. Geneva: 2002.
- Sheldon T. Promoting health care quality: what role performance indicators? *Quality in Health Care* 1998; 7(Suppl): S45-50.
- Smee C. The Performance Assessment Framework: where did it come from and where is it going? *Health Care UK. The King's Fund review of health policy*. Spring 2000; 3-8.
- Smith G and May D. The artificial debate between rationalist and incrementalist models of decision making. In: M. Hill (ed). *Policy process: a reader*. 2nd edition. Harlow: Prentice Hall (1997): 163-174.
- Smith PC. Performance management in British health care: will it deliver? *Health Affairs* 2002; 21(3): 103-115.
- Smith PC. On the unintended consequences of publishing performance data in the published sector. *Int J Public Administration* 1995; 18: 277-310.
- Solberg LI, Mosser G, McDonald S. The three faces of performance measurement: improvement, accountability and research. *Journal on quality improvement* Mar 1997; 23(3): 135-147.
- Walshe K. International comparisons of the quality of health care; what do they tell us? *Quality Safety Health Care* 2003; 12: 4-5.
- Walshe K, Freeman T. Effectiveness of quality improvement: learning from evaluation. *Qual Saf Health Care*. 2002 Mar;11(1):85-7.
- Walshe K, Wallace L, Freeman T, Latham L, Spurgeon P. The external review of quality improvement in health care organizations: a qualitative study. *Int J Quality Health Care* 2001; 13(5): 367-74.
- Walt G, Gilson L. Reforming the health sector in developing countries: the central role of health policy. *Health Policy and Planning* 1994; 9(4): 353-370.
- Wanless D. *Securing our Future Health: Taking a Long-Term View*. April 2002. Available on: <http://www.hm-treasury.gov.uk/wanless>.

Wanless D. *The review of health and social care in Wales. The report of the project team advised by Derek Wanless.* June 2003. Available on:
<http://www.wales.nhs.uk/documents/wanless-review-e.pdf>.

Welsh Office (1998). *Better Health, Better Wales.* Available on: <http://www.archive.official-documents.co.uk/document/cm39/3922/3922.htm>

Wildavsky A. *Speaking truth to power: the art and craft of policy analysis.* Boston: Little. Brown (1979).

Williams, A. Science or Marketing at WHO? A commentary on "World Health Report 2000", *Health Economics* 2001; 10(2): 93-100.

Williamson DL, Milligan CD, Kwan B, Franklin CJ, Ratner, PA. Implementation of provincial/territorial health goals in Canada. *Health Policy* 2003; 64: 173-191.

World Health Organisation. *The WHO Report 2000, Health Systems: Improving Performance* (Geneva: WHO, 2000).

World Health Organisation's Evidence for Health Policy: Health system performance, Geneva. Weblink: www.who.int/health-systems-performance.

World Health Organisation Regional Office for Europe. *Health 21: the Health for All policy for the WHO European Region. 21 targets for the 21st century.* Copenhagen: WHO Regional Office for Europe, 1998.

APPENDIX 1:

LIST OF PERSONS INTERVIEWED

(in alphabetical order)

Professor Hubert Allemand, Caisse Nationale d'Assurance Maladie, France

Professor Gwyn Bevan, London School of Economics

Professor Nick Black, London School of Hygiene and Tropical Medicine

Dr Ruth Bonitar, WHO

John Bullivant, Audit Commission for Wales

Sir David Carter, Former Chief Medical Officer for Scotland

Dr Elizabeth Docteur, OECD

Professor Christopher Ham, Strategy Unit, Department of Health.

Jeremy Hurst, OECD

Peter Hussey, OECD/ Commonwealth Fund

Dr Kei Kawabata, WHO

Sir Alan Langlands, University of Dundee, Scotland

Dr Iciar Larizgoitia-Jauregui, WHO

Dr Colin Mathers, WHO

Dr Soeren Mattke, OECD

Dr Jonathan Perlin, Veterans Administration, USA

Professor Victor Rodwin, Wagner School of Public Policy, New York University

Dr Tessa Tan-Torres, WHO

Professor Peter Smith, University of York

Dr Phyllida Travis, WHO

Nicole Valentine, WHO

Dr Kieran Walshe, Manchester University

Professor Gill Walt, London School of Hygiene and Tropical Medicine

Dr Ke Xu, WHO