

Nuffield Trust Series No. 12



**Dying
to Know**

Public Release
of Information
about Quality
of Health Care

Martin Marshall
Paul Shekelle
Robert Brook
Sheila Leatherman

Foreword by
Robert Brook
and
John Wyn Owen

RAND



The Nuffield Trust
FOR RESEARCH AND POLICY
STUDIES IN HEALTH SERVICES

Co-published by The Nuffield Trust and RAND
© Nuffield Trust 2000

Copies available in the UK from The Nuffield Trust,
59 New Cavendish Street, London W1G 7LP

ISBN: 1 902089 50 2

Telephone: 020 7631 8450

Facsimile: 020 7631 8451

email: mail@nuffieldtrust.org.uk

Website: www.nuffieldtrust.org.uk

Library of Congress Cataloging-in-Publication Data

Dying to know : public release of information about quality of health care.

p. cm.

"MR-1255-NT."

Includes bibliographical references.

ISBN 0-8330-2891-X

1. Medical care—Great Britain—Quality control. 2. Medical care—United States—Quality control. 3. Consumer education—Great Britain. 4. Consumer education—United States. 5. Medical care—Great Britain—Evaluation. 6. Medical care—United States—Evaluation.

RA399.A1 D95 2000

362.r068'5—dc21

00-059110

Copies available in the United States and Canada from RAND

ISBN: 0 8330 2891 X

1700 Main St., Santa Monica, CA 90407

Telephone: 310 393 0411

Facsimile: 310 451 6996

e-mail: order@rand.org

Website: www.rand.org

The Nuffield Trust Publications Committee
Professor John Ledingham, DM,FRCP
Dame Fiona Caldicott, DBE,FRCP,FRCPsych
Philippa Hurd
John Wyn Owen, CB

A REPORT FOR THE NUFFIELD TRUST

Martin N. Marshall M.Sc, M.D., F.R.C.G.P

Clinical Senior Research Fellow, National Primary Care Research and Development Centre, University of Manchester, Williamson Building, Oxford Road, Manchester M13 9PL. Tel: 0161 275 7601, Fax: 0161 275 7600, e-mail: mmarshall@fsl.cpcr.man.ac.uk

Paul G. Shekelle M.D., Ph.D

Senior Natural Scientist at RAND; Associate Professor of Medicine, University of California, Los Angeles; and Senior Research Associate of the Veterans Affairs Health Services Research and Development Service.

Robert H. Brook M.D., Sc.D., F.A.C.P

Vice President, RAND Health Program
Professor of Medicine and Health Services, UCLA Centre for Health Services
RAND Health Program
1700 Main Street, P.O Box 2138, Santa Monica, CA 90407, USA

Sheila Leatherman MSW

President and Founder, Center for Health Care Policy and Evaluation, 9900 Bren Road East, Minnetonka, MN 55343, USA
Senior Associate, University of Cambridge, Judge Institute of Management Studies

CONTENTS

Foreword	9
Acknowledgements	12
Executive Summary	15
Abbreviations	18
1. Introduction	21
1.1 Background	21
1.2 Definitions	22
1.3 Scope of report	22
1.4 Aims	24
2. Methodology	25
3. Public disclosure in the context of health care systems	27
3.1 Structure of the United States (US) health system	27
3.2 Historical development of US reporting systems	29
3.3 Public disclosure in the United Kingdom	35
The purpose of public release of performance data	41
4.1 Overview	41
4.2 Regulation and public accountability	42

CONTENTS

4.3	Consumer choice	43
4.4	Purchasing decisions	44
4.5	Provider behaviour	44
4.6	A coherent rationale for public disclosure	45
5.	Overview of publicly available performance data in the United States	46
5.1	Organisations involved	46
5.2	Types of data available	47
5.3	Publication format	50
5.4	The controversy	51
6.	Evaluation of the impact of public data	54
6.1	General critique	54
6.2	Attitude of physicians	55
6.3	Impact on behaviour of hospitals and other provider organisations	58
6.4	Information that consumers want to access	60
6.5	Impact on decision making of consumers	62
6.6	Explanation of impact on decision making	65
6.7	Information wanted by purchasers	66

6.8	Impact on employers' purchasing decisions	66
6.9	Impact on quality of care outcomes	68
6.10	Mechanisms of action of performance data	70
6.11	Impact on costs	70
6.12	Summary of the evidence	72
7.	Discussion: policy issues for the United Kingdom	73
7.1	Generalisability of the United States research	73
7.2	The purpose of public disclosure:	
	a conceptual model	74
7.3	Public disclosure and quality improvement	76
7.4	Factors influencing the content of the data	78
7.5	Risk adjustment of performance indicators	82
7.6	Factors influencing the release of the data	83
7.7	The role of the medical profession	85
7.8	The adverse consequences of publication	86
7.9	The financial cost of implementing a policy	
	on public disclosure	88
7.10	The research agenda	88
8.	Recommendations	91

CONTENTS

References	99
Appendix 1 - Frequency of conditions/procedures studied in report cards	109
Appendix 2 - Evidence tables	112
Appendix 3 - Conceptual model of public disclosure	121
Appendix 4 - Summary of reporting systems that have been subject to evaluation	122
Appendix 5 - US Diagnosis Related Group frequency data	125
Appendix 6 - UK common causes of hospital bed occupancy	126
Appendix 7 - Illustrative examples of performance indicators in the UK	127

FOREWORD

The modern field of measuring quality began about three decades ago. During this period, we have established that we can measure quality of care and that quality of care varies enormously. We have documented a large gap between the care people should receive and the care they do receive. We have learned that many quality problems are system problems, and that quality problems most often occur at the boundaries of systems. We have established that where you go or whom you see for care affects quality far more than who you are. We have also established that where you live has a strong influence on how much care you use, but that *rate of use* of care is weakly correlated with *appropriateness* of care. Thus, inappropriate overuse of care exists in areas that have a low rate of use of a given procedure and inappropriate underuse exists in areas that have a high rate of use of the procedure.

We have also learned that quality of care problems in the United States and in the United Kingdom are pervasive: They touch every dimension of quality. Some people do not get the care that they need. Some people get more care than they need. Some people get care that is not delivered in a technically excellent manner. There are often problems with the sensitivity and manner in which care is delivered.

We know that quality of care isn't just a U.S. or a UK issue. Every health care system in the world in which quality of care has been measured exhibits the same kinds of problems, along the same dimensions.

During those decades in which we were learning to measure quality of care, people around the world have been trying to improve the system by which medical care is delivered. We have discovered that it is extremely difficult. For instance, despite extensive quality improvement efforts in the United States, a review of quality of care

FOREWORD

studies published in the leading professional journals over the past decade shows that people with acute or chronic conditions receive about two-thirds of the care they need. And people receive only half the preventive care recommended. On the other hand, about one-fifth to one-third of both acute and chronic care is less than appropriate.

Note that these study findings describe care in a country in which

- extensive resources have been spent on accrediting hospitals.
- an accreditation system for health plans has been initiated.
- U.S. physicians have demanding requirements for continuing medical education.
- physicians have increased the amount of time they spend on education.
- the proportion of physicians who are board certified has also increased.
- continuous quality improvement or total quality management has begun to flourish.

In sum, despite dramatic progress in measuring quality of care, despite valiant efforts in many countries to address deficiencies in quality, there is no evidence from any country that systematic approaches to improving care have been successfully implemented.

We would expect a new movement to emerge to confront this failure. The movement involves public disclosure of information about quality at the level of a named health plan, a named doctor, a named hospital, primary care organisation or a named health

authority. In the U.S. this movement is being driven by a variety of stakeholders (e.g. government, business, public). In the UK government with some public support is the primary force behind the movement. At the core of this movement is the concept that producing public information about the quality of care actually provided will complement all of the above mechanisms designed to improve quality. The hope is that information and quality-improvement mechanisms will work synergistically to make improvement in quality more rapid than it otherwise would be.

The purpose of this monograph is to examine the theory behind this assumption of synergy, to identify evidence that supports or refutes the theory, and to suggest the practical and feasible implications of developing a system for public release of information about quality. The recent Institute of Medicine report on medical errors in the United States signals that improving quality of care will be a central political issue for most countries in the developed world in this century. Both the evidence and prior experience suggest that improving quality will be extraordinarily difficult. Thus, it is appropriate and timely that we examine carefully the role that public release of information might play in facilitating more rapid improvement in medical care systems. By examining the contribution of public disclosure of information in the United States, which has experimented most with this technique, we hope to increase our understanding of how quality improvement efforts can be successfully implemented in the UK, and to help quality of care improve more rapidly in the first three decades of the 21st century than it did in the last three decades of the 20th.

Robert Brook, MD
John Wyn Owen, CB
May 2000

ACKNOWLEDGEMENTS

This report has been commissioned as one part of a collaboration between the Nuffield Trust, London, and the RAND Corporation, Santa Monica, California. RAND is a nonprofit institution that helps improve policy and decisionmaking through research and analysis. The authors are indebted to the Trust's Quality Steering Group for their guidance and to the Trustees for their financial support.

Martin Marshall was supported during the writing of this report by a Harkness Fellowship from the Commonwealth Fund, a New York City-based private independent foundation. Paul Shekelle is a Senior Research Associate of the Veteran Affairs Research and Development Service.

The authors would like to thank all of the individuals and organizations who agreed to talk about their experiences of public disclosure and gave generously of their time and expertise.

The authors hope that this report will be of value to policy makers in the United Kingdom and so have used UK-English and spelling.

The views expressed here are those of the authors and not necessarily those of the Nuffield Trust or the Commonwealth Fund.

QUALITY STEERING COMMITTEE

Professor Nick Black	Professor of Health Services Research and Head, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine
Dr Robert Brook	Vice President and Director, RAND Health
Professor Jessica Corner	Professor of Cancer Nursing, Institute of Cancer Research
Jan Filochowski	Chief Executive, The Medway NHS Trust
Professor Leon Fine	Professor of Medicine, Royal Free and University College Medical School
Sheila Leatherman	Founder, Centre for Health Care Policy and Evaluation; Executive Vice President, United Health Group, Minnetonka, US; Senior Associate, Judge Institute of Management Studies, University of Cambridge
Dr Martin Marshall	Clinical Senior Research Fellow, National Primary Care Research and Development Centre, University of Manchester
Ms Ann Monroe	Director, The Quality Initiative, California Health Care Foundation

QUALITY STEERING COMMITTEE

Dr Fiona Moss Associate Postgraduate Dean, NHS Executive London Region; Editor, *Quality in Health Care*

John Wyn Owen Secretary, The Nuffield Trust

Professor Martin Roland Director of Research and Development, National Primary Care Research and Development Centre, University of Manchester

Dr Paul Shekelle Senior Natural Scientist at RAND; Associate Professor of Medicine, University of California, Los Angeles; and Senior Research Associate of the Veterans Affairs Health Services Research and Development Service

Sir Maurice Shock Chairman, The Nuffield Trust

Jane Steele Principal Research Fellow, Public Management Foundation

Observers

Mr Clive Smee Department of Health

Mr Anthony Townsend General Medical Council

(alternate: Mr David Skinner)

EXECUTIVE SUMMARY

Purpose

The aim of this report is to describe and assess the United States' experience of public release of health care performance data in order to help guide future United Kingdom policy.

Background

Public disclosure of the comparative performance of health care providers has been proposed as one mechanism for improving quality and controlling health care costs. This information, often released in the form of report cards, may be used to facilitate regulation and increase public accountability, to inform consumers, purchasers and providers, and to encourage improvements in the quality of care. Performance data have been made public in the US for more than a decade. Because many different public and private organisations have contributed to the process, the content, scientific rigour, measurement methods and publication format are highly variable.

Methods

The content of this paper is based on an extensive review of published and unpublished reports and expert opinion that was conducted between October 1998 and February 1999.

Principal findings

The US experience of public disclosure is presented in the context of its health care system and consumer-orientated culture. The wide-ranging debate centred around the purpose, content and implications of public disclosure is described. Despite a rapidly expanding report card industry, there has been little formal evaluation of its impact on purchasers or providers or quality of care. Current evidence, based largely on descriptive and quasi-experimental studies, suggests:

EXECUTIVE SUMMARY

- Physicians (doctors) and provider organisations are sceptical about report cards and consider them to have minimal utility.
- Hospitals respond with internal changes, to the publication of comparative performance data, especially in a competitive environment.
- Currently available report cards are rarely read by individual consumers or purchasers of care and, even if accessed, have little influence on purchasing decisions.
- Publishing comparative mortality data appears to result in improved outcomes. The mechanism of action is unclear.

Policy implications

On the basis of the US experience of public disclosure of performance data described in this report, the following recommendations can be made to guide UK policy:

- The intended purpose or purposes of public disclosure should be made clear to all stakeholders.
- Public disclosure should be seen as an evolutionary process, becoming progressively more sophisticated and comprehensive over time.
- Public disclosure should be seen as a tool to support all of the quality initiatives in the NHS, including clinical governance, the National Institute for Clinical Excellence, the Commission for Health Improvement, the National Service Frameworks, the National Performance Assessment Framework and the revalidation of health professionals.

- Provider organisations should be a key audience for information about performance.
- The financial cost of implementing a national policy on public disclosure is likely to be significant and should be considered alongside the benefits.
- Specific educational initiatives for target audiences should be implemented alongside public disclosure.
- Health professionals and their representative bodies should be fully involved in the process of public disclosure.
- Both process and outcome measures of quality should be published.
- Outcome indicators must be risk adjusted.
- Public disclosure should be accompanied by a strategy for monitoring the benefits and unintended consequences.
- Public disclosure should be accompanied by possible explanations for the variations reported.
- A research and development programme supporting the generation and evaluation of public performance data should be supported by the NHS R&D Directorate.

ABBREVIATIONS

AMI	Acute myocardial infarction
CABG	Coronary artery bypass graft
CAHPS	Consumer assessment of health plans
CHI	Commission for Health Improvement
CHOP	California Hospitals Outcomes Project
CQI	Continuous quality improvement
CSRS	Cardiac Surgery Reporting System
DRG	Diagnosis related group
GAO	General Accounting Office
GMC	General Medical Council
HCFA	Health Care Financing Administration
HEDIS	Health Plan Employer Data Information Set
HMO	Health Maintenance Organization
ICC	Intra-cluster correlation
NCQA	National Committee for Quality Assurance
NHS	National Health Service
NPAF	National Performance Assessment Framework
NSF	National Service Frameworks

PCG Primary Care Group

PHC4 Pennsylvania Health Care Cost Containment Council

UK United Kingdom

US United States

1. INTRODUCTION

1.1 Background

Efforts to improve the quality of health care are not new but have been strengthened in the last two decades as evidence is produced of poor performance and wide variations in the quality of care. This has resulted in attempts to measure and demonstrate improvements in quality.

Health professionals have traditionally been trusted with primary responsibility for standards of care. Neither individual professionals, nor the organisations within which they work, have had to demonstrate systematically that they were achieving acceptable levels of performance. Many professionals audited their practice but this was used almost entirely for peer review purposes and even significant deficiencies in care were usually unknown by the public. Several factors have caused this situation to change. A general trend towards greater openness in public affairs and a desire for improved value for money have combined with advocacy of public disclosure of health care performance as a mechanism for controlling costs and improving quality. The availability of computerised data and dramatic advances in methods of measuring quality have allowed relevant measures of quality to be developed for public disclosure. Alongside these changes have been some high profile examples of poor quality practice that have dented the public's confidence in professional self regulation.

As a result, there is an increasing expectation that health care providers should collect and report information on quality of care, that purchasers should use this information to make decisions on behalf of their population and that the general public has a right to access this information for individual choices. A variety of terms have been used to describe these data, including 'report cards',

INTRODUCTION

'consumer reports', 'public performance reports' and 'provider/practice profiles'. Examples using mortality data have also been referred to as 'score cards' (Topol and Califf, 1994) and 'death lists' (Vladeck et al, 1988).

1.2 Definitions

Report cards may be defined as standardised, publicly released reports on quality of care (Epstein, 1995). Typically they are used to make comparisons of performance amongst individuals or organisations (Longo et al, 1997), over time or against defined standards of care. This report focuses on comparative performance data that have actively been made public (not simply placed in the public domain).

Certain terms in common usage are unique to, or have different meanings in, the US and UK. The term *physician* refers to any medically qualified practitioner. *Health Plan* refers to the organisational and financial entity which provides for the delivery of health services according to a contract between an insurance carrier and its enrollees. The terms *user*, *consumer*, *patient* and *general public* are used synonymously in this report. *Provider* refers to the organisations and individuals who deliver a health care service and can include organisations like hospitals, groups of health professionals or individual health professionals. *Purchaser* refers to those who buy from providers and in the US can include individuals, employers who buy on behalf of their work-force and organisations who purchase from groups of physicians.

1.3 Scope of report

The report will provide a rationale for public disclosure, describe the US and UK health care context and the potential role of public disclosure alongside other quality improvement initiatives. The

types of data available, format of release and organisations involved in the production of performance reports in the United States will be described. A review of the controversy surrounding the release of report cards in the US will illustrate the range of responses that might be expected from interest groups in the UK. A detailed evaluation of the published research on report cards will include: a critique of current research, a review of physicians' attitudes, the types of information wanted by consumers and purchasers, the impact of public data on purchaser and provider behaviour, effect on quality of care outcomes, evaluation of proposed mechanisms of action, and costs of public disclosure. The policy implications of the US experience of public disclosure for the UK will be considered and policy recommendations will be made.

There are many examples of performance reports in the US, many of which lack rigour and few of which have been formally evaluated. To describe all of these initiatives is not possible, so a small number of high profile examples will be given to illustrate general principles, emphasising those that have been formally evaluated. A brief overview will be provided of the UK experience of putting performance data in the public domain, in order to place the US experience in a British context. However, the aim of this report is not to provide an exhaustive review of past or present British initiatives.

The delivery of health care is an increasingly multi-professional activity and therefore it is desirable for any quality improvement strategy to take a multi-professional approach. However, research conducted on public disclosure in the US has focused almost entirely on the medical profession and this review of the evidence will necessarily reflect that emphasis.

INTRODUCTION

1.4 Aims

The aim of this report is to describe the US experience of public release of health care performance data in order to help guide future UK policy.

Performance data of variable quality have been made public in the US for more than a decade. However, public disclosure of data is still in its infancy: rapid changes are taking place and evaluation of the usefulness of public disclosure lags behind the provision of the information. The unregulated and uncoordinated release of information in the US results in both problems and opportunities for outside observers.

Some of the lessons learnt in the US will not be generalisable to other health systems (see section 6.1) but careful study could enlighten the UK debate in many areas. These include the role of public disclosure as one part of quality improvement programmes, the type and format of data with potential for greatest impact, the response of interest groups (including users, purchasers, provider organisations, health professions and the media) and the likely impact of the release.

2. METHODOLOGY

The evaluation of the impact of public disclosure of performance data and the background information contained in this report is based on an extensive review of US published and unpublished information and expert opinion that was conducted between October 1998 and February 1999.

First, the published international peer-reviewed literature was accessed through Medline and Embase electronic databases. Searches using MeSH headings <report cards> <public performance reports> <provider profiling> <public/consumer/patient information> <consumer reports> were conducted independently by the principal author and by a professional librarian. Original articles and commentaries were reviewed. The Cochrane library was also accessed. The reference lists of all articles were searched. Authors of published studies and other experts in the field were asked to recommend relevant published and unpublished studies.

Second, documents and websites prepared by the Agency for Health Policy and Research, General Accounting Office, Health Care Financing Administration (HCFA), Institute of Medicine, National Committee for Quality Assurance (NCQA) and State organisations were reviewed.

Third, semi-structured interviews were conducted with experts in the field who were asked for their opinions about public disclosure and to recommend other data sources. Key informants included academics, policy advisers and others working in the public and private sector involved in the public release of performance data.

Fourth, media coverage of the public release of performance data was reviewed by studying news and editorial articles written in

METHODOLOGY

response to major disclosures. In particular, report cards published in *Newsweek*, *US News and World Report* and *Consumer Reports* were studied.

Finally a sample of report cards was reviewed. This included report cards produced by not-for-profit coalitions of health care purchasers and by the states of California, Florida, Massachusetts, Minnesota, New Jersey, New York and Pennsylvania.

3. PUBLIC DISCLOSURE IN THE CONTEXT OF HEALTH CARE SYSTEMS

3.1 Structure of the United States (US) health system

The following brief overview describes US health care and the history and context that has promoted public disclosure of measures of performance.

The US health system has been described as bewildering not only to international observers but to Americans as well (Reinhardt, 1998). The system is driven by market forces and is based more on temporary compromises between powerful vested interests than on any agreed national policy. The end result is a system characterised by marked contradictions. America leads the world in technological innovations yet over 40 million citizens have no health insurance coverage at any one time. The US spends almost twice as much of its gross domestic product on health care as does the UK and yet for some sectors of the population, health status measures are worse than those in some developing countries (Anderson, 1998). In addition, the satisfaction of the American people with its health care system is lower than for most English-speaking countries (Blendon, 1998).

A mixture of public and private health insurance supports the 84 percent of the population with coverage. The largest public systems are Medicare and Medicaid. Medicare is a social insurance programme for the elderly, some of the disabled under 65, and those with end-stage renal failure. It is administered by the federal government and financed through a combination of payroll taxes, general federal revenues and premiums. It covers 13 percent of the population and accounts for 20 percent of total health care expenditure. Medicaid is an entitlement programme for the poor, administered by the states within broad federal guidelines. It covers 12 percent of the population and accounts for 14 percent of total health care expenditure. Private health insurance, provided by

PUBLIC DISCLOSURE IN THE CONTEXT OF HEALTH CARE SYSTEMS

more than 1200 for-profit and not-for-profit insurance companies (regulated by state insurance commissioners), is purchased by individuals or employers. In the latter case it is funded by voluntary premium contributions shared by employers and employees on a company-specific basis. Private insurance covers 58 percent of the population and accounts for 33 percent of total health expenditure. Individuals may be covered by a combination of public and private insurance policies and cost sharing is common. Some policies may cover basic care but out-of-pocket expenses can be significant - estimated at about 17 percent of national health expenditure.

Concern about costs has resulted in the rapid expansion of Managed Care. This is an imprecise term. Managed care was introduced with the aim of improving quality, accountability and controlling costs by creating health plans to assume responsibility for individual and population health needs on a pre-paid or capitation basis. In practice it usually involves methods to influence clinical decisions made by the providers and users of health services in order to achieve greater adherence to standards and congruence with cost-effective decision making. Practice guidelines and disease management programs may be a key part of the way that services are provided. Plans that contract with independent providers may impose constraints by refusal to pay for health services that they judge to be inappropriate. There is controversy as to whether the introduction of managed care has started to focus the debate not only about costs but also about quality (Brook, 1997) or whether intrusive constraints on clinical decision making are unwarranted.

Concerns about the quality of health care in the United States have been expressed for decades and numerous studies have demonstrated

significant deficiencies (Winslow et al, 1988; Bernstein et al, 1993). Recent attempts to address the problem have focused on the use of market forces. The argument goes that if individuals or group purchasers are provided with evidence that quality varies among health plans, they will take this information into account, alongside cost and other factors, when they purchase their coverage and therefore drive improvement in the health care market. For the market to work, information about quality will need to be made public. This fits in with the belief in the impact of public opinion and the consumer-oriented model that predominates in the US. Antagonists argue that buying health coverage is not the same as buying a car or loaf of bread and that the power of health care purchasers is small in comparison with that of provider interest groups who might resist or attempt to modify the purchaser demands. Furthermore, the ability of consumers to choose their health plan, hospital or physician is sometimes limited.

3.2 Historical development of US reporting systems

Publication of data about performance in the US is not a new phenomenon. In 1754 a Pennsylvania hospital released mortality data tabulated by diagnostic groups (Lansky, 1998) and in 1917 the senior surgeon of Massachusetts General Hospital stated:

"Our charitable hospitals do not consider it their duty to see that good results are obtained in their treatment of patients ...It is against the individual interests of the medical and surgical staff of hospitals to follow up, compare, analyse, and standardise all their results"
(Codman, 1917)

He claimed that this was because of concern that the public would not be impressed with poor results, that the process was difficult

and time-consuming and because no one was willing to pay for the data to be produced.

Information about quality has been collected by US health care organisations for many years but it has usually been for internal use and has rarely been made available to the general public. According to Longo and colleagues (1997), the first modern call for greater openness was made in 1982 (Anderson and Shields, 1982). The authors reviewed methods of changing physician behaviour and concluded that neither clinical audit nor utilisation review had much impact. Their call to make the process and outcomes of care more explicit was answered in 1987 when the HCFA started publishing annual mortality rates for hospitalised Medicare patients. The report studied all causes of hospitalisation, used administrative data and made minimal case mix adjustments. The lack of a sophisticated risk-adjustment system ultimately led to the demise of the report in 1992.

Subsequent attempts to publish valid, reliable and useful information about quality have become increasingly more sophisticated. The following account briefly summarises the historical development of four high profile examples: the New York State Cardiac Surgery Reporting System (CSRS), the Pennsylvania Health Care Cost Containment Council (PHC4), the Health Plan Employer Data Information Set (HEDIS) and the California Hospitals Outcomes Project (CHOP).

New York State Cardiac Surgery Reporting System (CSRS)

In 1989 the New York Department of Health highlighted mortality after coronary artery bypass graft operations (CABG) as a focus for quality improvement. In conjunction with a group of cardiac surgeons, cardiologists, internists and consumers, the health

department developed a register to collect clinical data on patients undergoing coronary artery bypass surgery in New York State hospitals. Data on age, gender, type of coronary artery disease, presence of myocardial ischaemia, level of ventricular function, presence of cardiac or non-cardiac diagnoses, severity of the atherosclerotic process, previous heart operations and whether the procedure was elective or an emergency, were collected prospectively. A multivariate risk-adjustment model was constructed to compare mortality rates amongst hospitals and surgeons. The actual number of deaths for each hospital was divided by the expected number, given the hospital's patients' risk factors, as compared with the risk factors present in the state as a whole.

In 1990 the anonymised 1989 data were made available to the public but a newspaper, *Newsday*, sued them under the state's Freedom of Information Act to gain access to named surgeon-specific data. The state resisted the action on the grounds that low numbers would invalidate the data but it lost the case and published surgeon-specific results in December 1991. Initial press accounts were alarmist and misleading and clinicians were furious. They agreed only to submit data to the Department that could not identify individual surgeons. After detailed discussions, an agreement was reached whereby operative mortality data from the previous three years would be released only for surgeons performing at least 200 operations during that period. The Department made considerable efforts to educate journalists and recent press reports have been more balanced. In subsequent years the risk-adjustment procedure and systems to ensure data reliability have been improved. The CSRS is now regarded as one of the foremost examples of public data disclosure and is the most evaluated system in the US. The risk-adjusted mortality rate is provided to hospitals and surgeons on a regular basis to allow them

PUBLIC DISCLOSURE IN THE CONTEXT OF HEALTH CARE SYSTEMS

to compare levels of performance. The data have been cited as a significant factor in a dramatic reduction in post operative mortality following CABG in the state of New York (Chassin et al, 1996).

Pennsylvania Health Care Cost Containment Council (PHC4)

PHC4 was created by the legislature to constrain costs and improve quality by producing public information about the performance of health care providers (Sirio and McGee, 1996). It was based on the premise that current and accurate data about the costs and quality of care would encourage group and individual purchasers to drive down costs and improve quality through the use of market forces. It was also thought that the data would help shape health-related policies and programmes. The original motivation behind publication of data by the PHC4 was therefore broader than that of the New York CSRS.

PHC4 produced a report card *A consumer guide to coronary artery bypass graft surgery* in 1992. It listed, by surgeon and hospital, the number of CABG procedures performed per year and the risk-adjusted inpatient hospital mortality, compared with expected rates drawn from the risk-adjustment model. The risk-adjustment system, called Medisgroups, includes seven risk adjusters, including presence of a myocardial infarction, age, type of bypass (artery versus vein), presence of cardiogenic shock, presence of congestive heart failure, gender, and severity of illness at the time of admission. It has been patented and is marketed by a for-profit company. A grade is assigned to hospitals according to whether the actual mortality is higher than expected (i.e. is greater than two standard deviations away from the expected value), lower or within the normal range. In addition, the report publishes data on the costs charged by each hospital and compares charges with outcomes. The guide is distributed free to hospitals, surgeons,

public libraries, business groups, legislature, the media and any individual who requests it (Schneider and Epstein, 1996; Schneider and Epstein, 1998). Regular updates have been published since 1992 and the risk-adjustment mechanism has been modified. The PHC4 initiative has also been evaluated but not in as much detail as the New York State CSRS.

Health Plan Employer Data Information Set (HEDIS)

HEDIS is the most commonly used database for assessing performance at the level of the health plan. It is managed by the National Committee for Quality Assurance (NCQA), a not-for-profit accreditation organisation. HEDIS represents an attempt to standardise how plans measure and report performance data and is based on both administrative and clinical data. Since its introduction in 1991, the first set of indicators has become larger. The 1995 version, HEDIS 2.5, contained nine measures directly related to quality whilst the current version, HEDIS 3.0 contains 14 measures and a pilot set of new indicators contains a further 25 measures of quality.

Comparative HEDIS data from volunteering health plans are published as the *Quality Compass*. The second edition published in 1997 contains information from over 330 plans, representing three quarters of all Health Maintenance Organisation enrollees.

The HEDIS indicators have not escaped criticism. They represent a considerable cost and administrative burden to health plans and because participation is voluntary, concern has been expressed that only plans with above average performance would be willing to provide information. Concern has been expressed that the NCQA may be subject to conflicts of interest because of the nature of the organisations contributing to indicator development (Epstein,

1995). The data are not risk adjusted, emphasise process over outcome measures of quality, preventative over curative indicators, and the data collection methods are not standardised (Epstein, 1998). The NCQA has taken a pragmatic approach to indicator development and the data set is going through a process of refinement, attempting to address many of the criticisms that have been levelled against it (Corrigan, 1995).

California Hospitals Outcomes Project (CHOP)

CHOP was established by a state law that was passed in 1991 in response to purchaser demands for lower cost and higher quality health care. The project analysed and disseminated data on risk-adjusted hospital outcomes. It differs from the New York CSRS and PHC4 in that it is based on routinely collected data extracted from hospital discharge summaries. Three years of debate prior to 1991 led to impasse between the state and hospital association and a threat by the state to impose the Pennsylvania proprietary risk-adjustment system unless an alternative was found. The cost of collecting and reporting data for the Pennsylvania system was considerable, so agreement was reached to use existing data and make prospective incremental improvements in the quality of the data (Romano et al, 1995). For each hospital discharge, hospitals are obliged to code procedures, diagnostic categories and basic demographic data and send this information to the state.

The first report was released in 1993 and included inpatient mortality rates for acute myocardial infarction and complication rates for cervical and lumbar discectomy. Results were published in two categories, better or not better than expected. A second report, released in 1996, classified acute myocardial infarction mortality rates as better, worse or not significantly different from expected. Proposals to include discectomy complication rates and

postpartum readmission rates were dropped because of concerns about their validity. A third report on acute myocardial infarction mortality was published in December 1997. All reports are sent to the providers prior to publication and their comments are appended to the final report.

Each CHOP contains different sections. The *Users Guide* contains details about the methods used to produce the data and numerical and graphical comparative results by hospital. The *Technical Guide* provides greater methodological detail and *Detailed Statistical Tables* provide results in depth. A *Hospital Guide* explains to providers how to use the spreadsheet and interpret the results. The timeliness of the data release has been criticised - the 1996 report contained data derived from 1990-92 (Rainwater et al, 1998). The impact of CHOP has not been evaluated to the same degree as had the New York State CSRS or the PHC4.

3.3 **Public disclosure in the United Kingdom.**

Perhaps the first systematic reporting of comparative performance data in the world took place in England in the 1860s, when Florence Nightingale highlighted differences in mortality rates of patients in London hospitals (Nightingale, 1863). Since that time, a variety of methods and tools has been used to improve quality in the UK but there has been little emphasis on the public disclosure of performance data.

In the last decade clinical audit has been a central component of the drive to improve quality. When introduced in 1990, it was accepted and in many areas implemented by the medical profession, in part because it was unthreatening - the process was confidential and the data was used for internal purposes only. In addition to conventional medically-led audit, isolated examples

exist of using specific performance indicators to promote quality improvement. The Maryland Hospital Quality Indicator Project, started in the United States but now operational in several countries including the UK, is one such example of using explicit indicators to make comparisons among voluntarily participating hospitals (Thomson et al, 1997). Despite the enthusiasm for audit, recent evidence suggests that the expected improvements in clinical care resulting from audit activity have not materialised, at least in terms of value for money (Davis et al, 1995; Bero et al, 1998).

Alternative mechanisms are now being considered and the use of public disclosure of performance data is one option. One of the stated objectives for the NHS in 1996/7 was "to improve the quality and quantity of information given to enable patient choice about treatment options" (NHS Executive,1995). Outcomes data have been available in some limited fields for many years in the form of the Confidential Inquiries, but the detail of these inquiries has not been made public and there has been no rigorous evaluation of their impact. Throughout the 1990s the NHS has been encouraged to become more accountable to the public and more open with information. Initially the information provided had minimal direct relevance to quality. It included the provision of largely structural data on services available, the description of some processes and outcomes in annual reports, which in theory are public documents but in practice are not widely disseminated and did not encourage comparisons among different providers. There have been only a small number of examples of public disclosure of performance information in the UK in recent years. Most of these have attracted only minimal public interest, in part because they addressed very specialised areas of expertise, such as *in-vitro* fertilisation or renal transplant success rates. Perhaps the best example of a more generic system for public disclosure is the Clinical Outcomes

Working Group project that compares hospital outcomes data across different hospitals and health areas in Scotland (Dillner, 1994). The project was established by the Clinical Resource and Audit Group which is responsible to the Scottish NHS Management Executive. Thirty indicators have been published in four reports since 1994. Outcomes, including rates of teenage conception, suicide, cancer survival and postoperative emergency readmission are published for hospitals treating a minimum number of patients in each category. There is minimal risk adjustment of the data and the emphasis of the reports is very much on raising awareness of variation, rather than making judgements about performance. Waiting list data are also published widely and used to pursue government policy to reduce waiting times. Neither the impact of the Scottish outcomes data nor the waiting list data has yet been rigorously evaluated, though the Centre for Health Economics, University of York, is currently conducting such a study which is funded by the Department of Health.

The UK government intends to use the publication of quantitative information on performance as a key tool to improve quality (NHS Executive, 1998). Public disclosure will therefore become an integral part of a coordinated and systematic approach to quality improvement in the NHS, including the following initiatives:

The National Institute for Clinical Excellence (NICE).

NICE is responsible for identifying new and existing health technologies that would benefit from appraisal, collecting evidence to assess the clinical and cost-effectiveness of the interventions, producing and disseminating guidelines and coordinating a national strategy to ensure equitable and effective health interventions across the NHS. Explicit indicators of performance will be an integral part of the guidelines produced.

National Service Frameworks (NSF)

The NSFs set national standards and define service models for specific diseases, services or care groups. In addition, they are responsible for ensuring that the models are implemented in a coordinated fashion across the different sections of the NHS and for establishing performance measures against which progress can be measured. The Calman-Hine NSF for cancer services is already established and a NSF for mental health services was published recently. Service frameworks for coronary heart disease and diabetes will be published in the near future. Again, explicit performance indicators will be a key component of each of the frameworks.

Clinical Governance

Clinical governance is a framework through which health care organisations are accountable for continuously improving the quality of their services and safeguarding high standards of care. The aim is to create an environment in which excellence can flourish. The process is led in the main by health professionals but includes all relevant stakeholders. A variety of mechanisms can be used to implement and monitor clinical governance and the explicit use of performance indicators is likely to be one important tool.

The Commission for Health Improvement (CHI)

CHI is an independent 'watch-dog' which will be used to monitor the performance of health care provider organisations. It is proposed that the Commission will combine the roles of inspection and regulation with consultation and guidance. CHI will ensure that clinical governance processes are in place, carry out a rolling programme of inspections of NHS organisations and intervene if local quality assurance mechanisms have not been effective. Public performance data will be used, alongside other types of evidence, to make judgements about performance.

The National Performance Assessment Framework (NPAF)

The NPAF sets targets for six different areas suitable for performance measurement:

- Health improvement.
- Fair access.
- Effective delivery of appropriate health care.
- Efficiency.
- Patient/carer experience.
- Health outcomes of NHS care.

Any designated clinical topic chosen for the framework should include all six areas for performance measurement. Standards for each of the specific indicators within these areas are agreed between the NHS Executive Regional Offices and Health Authorities, between Health Authorities and Primary Care Groups and between Primary Care Groups and Trusts. In addition, Local Health Improvement Programmes will have to take them into consideration. Following a period of consultation and in response to some specific criticisms (Thomson, 1998; McColl et al, 1998), the framework was revised and the first effectiveness indicators for hospital outcomes were published for Wales and England in the spring of 1999. The results of the first patient experience survey were published later in 1999. The hospital outcome data showed wide variations among geographical regions and specific hospitals but the public response was relatively balanced and somewhat muted. At this stage it is too early to make judgements about the

impact of the data release on the various stakeholders, or on quality of care.

Revalidation and appraisal of doctors

The final part of the UK government strategy for quality improvement in the NHS is the revalidation and appraisal of doctors. Recent high profile examples of failure of self-policing by the medical profession have resulted in demands for an explicit link between fitness to practice and the maintenance of a doctor's name on the medical register. Formal mechanisms for three yearly revalidation of all doctors are being developed by the specialist societies, adapting a structure devised by the General Medical Council (GMC, 1998). Use of explicit performance indicators might well be part of the revalidation process. It is likely that a formal process for revalidation will be in place by 2002. In addition, an annual appraisal for doctors has been recommended by the Secretary of State for Health.

In conclusion, there is no significant history or culture of public disclosure in health care in the UK. However, the government's quality agenda for the NHS incorporates a number of different initiatives which are all likely to use public reporting of performance data as a tool to inform, promote regulation and accountability and encourage quality improvement.

4. THE PURPOSE OF PUBLIC RELEASE OF PERFORMANCE DATA

4.1 Overview

The theoretical foundation in support of the use of performance data to make judgements about quality is based on sparse and generally weak empirical data. The arguments for and against public release of data from identifiable providers are often ideological or largely anecdotal and often highly polarised. Advocates are inclined to promote release in the name of openness or to promote consumer choice without considering the untoward effects. They sometimes portray those who question the merits of public disclosure (usually perceived as health professionals closing ranks) as defensive and secretive. In addition, there has been minimal agreement amongst the various stakeholders about the expected gains from the release of comparative performance data. It has therefore proved difficult to judge whether the benefits of public disclosure of performance data outweigh the disadvantages.

The absence of a commonly embraced rationale for the public release of performance data is illustrated by the wide variety of purposes and audiences for which commentators believe the data might be used. The most common expectation is that it will promote an efficient market economy in health care (Edgman-Levitan and Geary, 1996; Bentley and Nash, 1998), usually in the belief that information about performance will encourage consumers to choose to access high quality providers (Hibbard and Weeks, 1989; Hannan et al, 1994; Lansky, 1998; Schneider and Epstein, 1998). Some authorities suggest that the information could be used by providers as a marketing tool (Longo et al, 1997). Others suggest that it will help to control costs (Berwick and Wald, 1990; Sirio and McGee, 1996), or at least counter the influence of cost as the principal determinant of purchaser decision making (Brook, 1997; Hibbard et al, 1997b; Mukamel and Mushlin, 1998). In addition, public information about performance has been

THE PURPOSE OF PUBLIC RELEASE OF PERFORMANCE DATA

proposed as a tool to regulate the health system, a method of ensuring accountability of provider organisations (Hannan et al, 1994; Rosenthal et al, 1998) or of making judgements about the performance of individual professionals (Kassirer, 1994). Finally, some perceive public disclosure to be a mechanism to promote quality improvement, by informing purchasers (Hibbard et al, 1997) or by encouraging providers to focus on quality problems (Hannan et al, 1994; Scheider and Epstein, 1996; Longo et al, 1997; Rainwater et al, 1998).

The lack of a clear conceptual framework can be illustrated by the following examples of the diverse motivations behind some of the principal reporting systems in the US. The original motivation of the *Pennsylvania Consumer Guide to Coronary Artery Bypass Graft Surgery* was to use the information to drive down costs - as was clear from the name of the organisation that released the report, the Pennsylvania Health Care Cost Containment Council (Bentley et al, 1998). The rationale for releasing the Health Care Financing Administration (HCFA) mortality data was to inform the public about the quality of hospitals in their region, in the belief that the public would then avoid low quality institutions (Mennemeyer et al., 1997). The explicit aim of the California Hospitals Outcomes Project (CHOP) was to stimulate quality improvement by motivating the providers to improve care (Rainwater et al, 1998).

The principal theoretical reasons for disclosing performance data and benefits and untoward effects of each may be summarised as follows:

4.2 Regulation and public accountability

Securing central control has been cited as one of the principal reasons for publishing performance data (Smith, 1995). The focus

of control may be classified as political (as required by elected representatives or taxpayers) or managerial (as sought by managers within an organisation) (Hofstede, 1981).

In the UK, the need for accountability has been increased by the replacement of the unitary NHS with semi-autonomous business units in the form of Trusts and Primary Care Groups. The NHS is highly regulated, chiefly because the government is deemed to be responsible for the efficient use of public money. The US is beginning to experience increasing constraint and regulation of the free market, in part because of fears that it is not delivering in the interests of consumers and because the growth of managed care is perceived to be threatening patient choice (Blendon et al, 1998). Regulation is best based on good data and if valid information on performance is made public it is more likely to be used by all interest groups. Public release is compatible with the trend in western countries towards open government and release of data extends beyond health to areas such as education and law and order. It is thought that visible public accountability of government will strengthen the democratic process.

There is a trend towards increased regulation of health care quality in both the US and the UK. Some of the benefits of regulation are clear but antagonists argue that if it is based on poor quality or misleading information then it may be misused by government or misinterpreted by the general public and may result in more problems than it solves.

4.3 Consumer choice

The most commonly cited reason for public disclosure in the US is to promote consumer choice, based on the theory that a market model is best driven by an informed consumer. Most of the

available information in the past has been about cost and therefore market driven changes have been financially focused. In theory, making relevant quality information available will encourage consumers to take quality into account, alongside other factors, and encourage the markets to compete on quality and perhaps drive out low quality providers. Even in the UK, consumer choice is encouraged and could in theory be supported by information on quality (Entwistle et al, 1996). Some argue that only a small proportion of informed consumers are necessary to drive the market. Antagonists argue that consumers will not or cannot use information (for example, as a result of lack of real choice or difficulty with comprehension), that health care markets are not sensitive to the same forces as are non-health markets, or that the market will be driven inappropriately by consumers.

4.4 Purchasing decisions

Whilst individual consumers might have little power to drive large health care markets, large purchasers, such as businesses, government or managed care organisations in the US and health authorities, fundholders or Primary Care Groups in UK, could in theory use performance data to make their purchasing decisions and influence providers. Antagonists argue that purchasers' fiscal responsibilities in both the US and the UK encourage decisions based on cost over quality and that they have little interest or motivation to change their behaviour.

4.5 Provider behaviour

The potential of public disclosure to influence provider behaviour is based on the theory that organisations and professionals have an intrinsic desire to improve practice but that barriers (time, competing priorities, lack of knowledge) prevent the expected or desired improvement. Since using data for internal purposes has

not produced expected benefits, it is argued that publishing the information will 'turn up the heat' by reminding, refocusing or shaming them into action. For this to happen, the published data have to relate to areas within the control of the providers and be appropriate to their aims. Antagonists argue that the degree of disclosure required to influence behaviour positively is unknown and that demoralised or defensive providers will react negatively rather than positively to the public release of performance data. They argue that there is good evidence of negative effects such as gaming, inappropriate prioritisation and fraud (Smith, 1995).

4.6 A coherent rationale for public disclosure

The four reasons described above for implementing a policy on public disclosure of performance data are neither discrete nor mutually exclusive. It is reasonable to assume that they might all operate, to a variable extent, in any country and for any reporting system. However, the potential for conflict needs to be recognised, for example between promoting consumer choice in a free market and central regulation, or between externally-made judgements and internally-driven quality improvement. It follows that a clear and explicit purpose for introducing public disclosure is fundamental to its design, implementation and evaluation.

5. OVERVIEW OF PUBLICLY AVAILABLE PERFORMANCE DATA IN THE UNITED STATES

5.1 Organisations involved

Various organisations produce report cards, with diverse motivations and little communication or collaboration among them. Some regulatory bodies produce data to stimulate quality improvement and to demonstrate public accountability. Health plans and hospitals use performance data to distinguish themselves in the competitive market. Even physician groups are beginning to use comparative data to gain a market advantage. Some of the organisations which produce the data start out with independent funding sources but later form coalitions with health care providers, purchasers or other commercial interests.

Government report card initiatives have included both Federal and State supported programmes. The principal federal example was the HCFA hospital mortality data published from 1987 to 1992. State programmes are responsible for some of the most rigorous data and include the New York State Department of Health's *Cardiac Surgery Reporting System*, the Californian Office of Statewide Health Planning and Development which supports the *California Hospitals Outcome Project*, the Florida Agency for Health Care Administration, which produces *Checkup*, the Massachusetts Health Quality Partnership which produces the *Massachusetts Acute Care Hospital Statewide Patient Survey Project* and the Pennsylvania Health Care Cost Containment Council's *Consumer Guide to Coronary Artery Bypass Surgery*.

Most report cards are produced by non-governmental coalitions. Two of the most active initiatives are those involving managed care organisations and those organised by purchasers. An example of the former is the National Committee for Quality Assurance (NCQA) which produces *Quality Compass* based on HEDIS data. In addition, some managed care organisations have

produced and publicly released their own performance data. Many coalitions are emerging to provide data for group and individual purchasers such as the Pacific Business Group on Health, which produces *Health Scope* and the Foundation for Accountability, which has produced report cards on depression, diabetes and asthma.

Finally, the media have played an important part in the production and dissemination of comparative data about health care provider performance. *Newsweek's* annual survey publishes data on overall plan ranking, satisfaction and accreditation, as well as measures of 'staying healthy', 'getting well' and 'living with illness', separately for adults and children. The satisfaction scores are produced by subtracting the number of members who stated they were highly dissatisfied with their plan from the number who said they were highly satisfied. *US News and World Report* also publishes regular performance reports. Methodological details for all these newspaper reports are sketchy or absent and they may not be based on rigorous methods.

Professional societies have made little practical contribution to the production of report cards. The American Medical Association intends to develop and publish a national report on physician accreditation which will require specific measures of performance. The aim is to reduce the fragmentation and duplication of current sources of information on physician quality.

5.2 Types of data available

Data on a wide variety of dimensions of quality and performance of health plans, provider organisations and individual providers are made public in the United States. Most of the information is based on routinely available administrative or clinical data, or

OVERVIEW OF PUBLICLY AVAILABLE PERFORMANCE DATA IN THE UNITED STATES

specially collected survey data. The validity and reliability of much of the data are highly variable.

Information is available about the structural characteristics of provider organisations and the processes and outcomes of care for many acute and chronic conditions and preventative interventions. The publication of information about consumer satisfaction and experiences of care is becoming particularly commonplace. One of the driving forces for this trend is a reorientation away from medically dominated measures of quality (Lansky, 1993; Lansky, 1998) and a perception that asking users is the most reliable way of assessing quality of interpersonal skills and access to care (Epstein, 1998). The Agency for Health Care Policy and Research has funded the development of a new instrument, the *Consumer Assessment of Health Plans* (CAHPS) to measure generic and specific aspects of satisfaction with care and this instrument is now becoming widely adopted.

The scope of information provided in report cards is broad and increasing at a rapid rate but one notable feature of the reports is that the difference among plans or providers is often small. For example, the overall satisfaction ratings of physician groups reported in *Health Scope* range from 78 percent to 83 percent. Some report cards contain statistically significant differences from the mean score but whether these represent valid differences that are useful to purchasers and fair to plans and providers is often unclear.

It is difficult to conduct an accurate survey of all report cards because of the many and disparate organisations involved in their production. The most recent survey that could be found was conducted in 1994 by the California Office of Statewide Health

Planning and Development (Richards et al, 1994). The authors identified two national published report cards, 30 statewide or regional examples, three metropolitan examples and seven corporate examples. Most report cards described performance for more than one indicator. One hundred and eighteen indicators reported on medical or surgical in-patient care and 26 reported on out-patient or ambulatory care (Appendix 1). In the five years since this survey was conducted, the number of both publicly and privately produced report cards has increased dramatically.

Examples of information that has been made available in report cards include:

- In-hospital mortality data for all causes.
- Mortality data for specific operations (e.g. coronary artery bypass surgery, carotid endarterectomy, hip replacements).
- Mortality following myocardial infarction, pneumonia and stroke.
- Cardiac surgery intervention rates.
- Cervical and breast screening rates.
- Immunisation rates.
- Diabetic eye examination rates.
- Rates of advice to quit smoking.
- Percentage of consumers reporting blood pressure and cholesterol well controlled.

OVERVIEW OF PUBLICLY AVAILABLE PERFORMANCE DATA IN THE UNITED STATES

- Provision of care after hospitalisation for a mental illness.
- Check ups for new mothers.
- Overall patient satisfaction rates.
- Ease of getting referrals.
- Doctor communication skills.
- Ease of finding a personal doctor.
- Referral rates to specialists.
- Rate of complaints against providers or legal action against individual physicians.
- Recommendation of plans to family or friends.

5.3 Publication format

Data on performance are available in hard copy and on the internet. The presentation of most reports is highly professional. Many have gone through a process of modification as a result of user feedback.

The length of the reports is variable. Some are very short summaries - a single side of paper: others are very long, with detailed justifications of criteria selection and methods used to collect and analyse the data. The reports use a range of different methods to present the data. Some reports simply state whether a provider is statistically different from the mean score for a specific indicator, others use pictorial presentations of

performance, including bar charts, stars and even happy (or unhappy) faces.

Electronic publication is increasingly popular in the US, where nearly half of all households have access to the internet. Websites allow access at different levels, from brief summary information to detailed methodological justification and statistical data. Examples of websites include those of the New York Department of Health (www.health.state.ny.us), the Pennsylvania Cost Containment Council (www.phc4.org) and the Pacific Business Group on Health (www.healthscope.org) A site developed by a private business venture (www.HealthCareReportCards.com) is producing increasingly sophisticated data. The company purchases Medicare mortality data from the HCFA, which it then risk-adjusts and rates hospitals that treat 30 or more Medicare patients in each disease or intervention category.

5.4 The controversy

The public disclosure of performance data has been described as essential (Lansky, 1998), desirable (Epstein, 1998) inevitable (Kassirer, 1994) and potentially dangerous (Topol and Califf, 1994; Ziegenfuss, 1996). No commentators have totally rejected publication, despite sometimes vehement criticisms of current initiatives (Schneidman, 1993). The debate in the literature is often highly polarised and both advocates and antagonists often use the same data to support their opposing arguments. Not surprisingly, many people are confused by the contrary evidence.

The controversy is best illustrated by describing the conflicting studies and resulting correspondence to the New York Cardiac Surgery Reporting System. The conclusions of the quasi-experimental study, describing dramatic improvements in post-CABG

mortality following publication of performance data in New York state (Hannan et al, 1994), have been questioned. Similar improvements in outcome were reported in North New England where performance data were kept confidential and disclosed only to the participating surgeons (O'Connor et al, 1996). Others have suggested that the improvement in CABG mortality in New York state was the result of out-migration of high-risk individuals to other states (Omoigui et al, 1996) or due to the refusal of New York surgeons to operate on those with highest risk of death (Schneider and Epstein, 1996). Some have questioned the reliability of the data collection (Ziegenfuss, 1996; Jollis and Romano, 1998). Other studies have raised doubts about the validity of the risk-adjustment mechanism (Schneider and Epstein, 1996; Jollis and Romano, 1998) or suggested that there has been overreporting of risks by providers in order to reduce their published mortality rates. The potential of an inadequate risk-adjustment mechanism to deter surgeons from accepting high-risk patients has also been highlighted.

These criticisms have been addressed by the team who produced the original study (Chassin et al, 1996). Their data did not suggest that there had been an out-migration of patients and they have described the mechanisms that were put in place to ensure data reliability. Overall reliability was found to be high after an independent audit, though two hospitals were asked to re-code some of their data. The team explains the increase in the number of reported risk factors in terms of previous underreporting and the inevitable re-definition of risk factors as the adjustment mechanism was refined. They have also demonstrated that their risk-adjustment model more than compensates for surgeons operating on high-risk patients (Hannan et al, 1997). Out-of-state transfers and reduced access for high-risk patients have also been discounted by an independent team of researchers (Peterson et al, 1998).

Other general criticisms of public disclosure have been voiced. Some regard report cards as backward looking, unable to predict future performance, judgmental and incompatible with the principles of continuous quality improvement (Green and Wintfield, 1995; Goddard et al, 1998). Others have questioned the timeliness of the reports in relation to the date of data collection, and have questioned the medical model underlying many reports that focuses on mortality as the outcome (Rainwater et al, 1998). The unintended consequences of publication, including the concentration on areas included in the report at the expense of other objectives, as well as gaming the system have also been highlighted (Romano et al, 1995; Goddard et al, 1998). Finally, the risk of misrepresentation of the data, particularly in the media, has been described in detail (Romano et al, 1995; Chassin et al, 1996).

The US General Accounting Office commissioned a review of report cards in 1994 and at that time some experts were advising caution because of the shortage of valid quality measures and the deficiencies in data quality (GAO, 1994). They suggested that it might be 10-15 years before reliable report cards could be produced whilst others suggested that making a start would encourage debate and faster progress. Not surprisingly, the debate about these issues continues today.

6. EVALUATION OF THE IMPACT OF PUBLIC DATA

6.1 General critique

The lack of rigorous evaluation of the impact of public disclosure of performance data is perhaps surprising after a decade of experience. The first comprehensive review of report cards in 1994 commented on the paucity of evidence (GAO, 1994) and several commentators since then have called for evaluative research (Hibbard et al, 1997a; Epstein, 1998). Despite this there is still little evidence to answer even the most fundamental questions relating to the most effective type of data to make publicly available and the impact of currently available report cards on quality of care.

There are several possible reasons for this apart from the most obvious explanation that there has to be a lag period between the introduction and evaluation of a new initiative. It is possible that some people regard public disclosure like motherhood and apple pie, that it is so obviously the right thing to do that formal evaluation of its impact is not necessary. In addition, the 'political incorrectness' of challenging a tool of informed consumerism may discourage candid appraisal. Others might be fearful of risking a negative evaluation of the currently available (and largely inadequate) data and are waiting to evaluate better quality indicators. Vested business interests may also feel threatened by a formal evaluation of an area in which there is considerable business potential. Also, funding bodies may not see it as a priority, or may consider that the methodological problems of researching the effectiveness of public disclosure, in comparison with other quality improvement initiatives, are insurmountable.

Evaluation of report cards requires a clear theoretical framework to identify the purpose of publication and an understanding of the strengths and weaknesses of the data that are being made public.

Most published studies have ignored or skimmed over both of these issues. One of the most fundamental questions that remains unanswered is whether the public release of performance data is more effective at improving quality of care than using the same data solely for internal purposes?

With the above reservations in mind, only tentative conclusions should be drawn from the following literature review, summarised as evidence tables in Appendix 2. The evidence is categorised into common themes which are presented as a conceptual model in Appendix 3. A brief summary of the reporting systems which have been the subject of a formal evaluation is provided in Appendix 4.

6.2 Attitude of physicians

Four studies have investigated the attitude of the medical profession to the publication of performance data (Hannan et al, 1997b; Schneider and Epstein, 1996; Vladeck et al, 1988; Borowsky et al, 1997). The two most detailed studies used the New York and Pennsylvania cardiac data and found that physicians are interested in but sceptical about the data and that they consider it to have minimal utility.

Hannan et al (1997b) surveyed all cardiologists in New York state belonging to the American College of Cardiologists to examine whether the published performance data influenced their referral practices. Only one third of those sent a questionnaire replied. Ninety four percent of these found the reports easy to read and 67 percent considered the data to be very or somewhat accurate. Twenty two percent routinely discussed the data with their patients and 38 percent considered it to have influenced their referral pattern. The most common reason for disliking the reports was a perception that it discouraged cardiac surgeons from operating on

EVALUATION OF THE IMPACT OF PUBLIC DATA

high-risk patients. The authors concluded that the data in their current format have minimal impact on established referral patterns. The low response rate suggests a lack of interest in the subject but it is unclear how the attitudes of respondents and non-respondents would have differed and therefore how valid are the conclusions.

Schneider and Epstein (1996) asked a different set of questions in their structured survey of cardiologists and cardiac surgeons in Pennsylvania. This makes direct comparisons with the New York survey difficult. They surveyed half of all registered cardiovascular specialists in the state to determine the awareness, utility, limitations of, and influence on practice of the PHC4 performance reports. All surgeons and 84 percent of cardiologists were aware of the report but only 10 percent perceived the data to be important when assessing the performance of a particular surgeon. Less than 10 percent discussed the data with more than 10 percent of eligible patients. Eighty seven percent of cardiologists stated that the reports had minimal or no influence on referral patterns and two percent stated that it had a significant influence. Concerns that were expressed about the data included a desire for published outcomes other than mortality (78 percent), concerns about the adequacy of the risk adjustment (79 percent) and fears that the data are easily manipulated by hospitals or physicians (53 percent). Almost two thirds of cardiologists reported increasing problems finding surgeons to operate on high-risk patients and the same proportion of cardiac surgeons reported that they were less willing to operate on such patients. The authors concluded that cardiovascular specialists had not been integrated into the report card movement. The relationship between self reported behaviour and actual behaviour was not examined in this study.

Vladeck et al (1988) used an indirect method to determine the potential influence of HCFA hospital mortality data on referral practices. The authors performed a before-and-after quasi-experimental study of all New York acute care hospitals, divided into higher than average, average and lower than average mortality rates. The authors found that there was no statistical difference in bed occupancy rates among the three groups of hospitals before and after publication of the data. They concluded that publication of data had minimal effect on admission patterns. Although the study was conducted in a single state and lacked controls, it confirmed the findings of the physician self reports that were described above.

Borowsky et al (1997) highlighted an indirect and unpredicted effect of publishing report cards on physician behaviour. Apparently, despite the negative reaction to public disclosure described above, physicians were galvanised into producing their own report cards on the health plans. The authors surveyed 100 physicians in each of three plans in Minneapolis-St Paul, Minnesota, to determine their views on competing plans and compare ratings among plans. A good response rate in this study suggested a high degree of physician interest, in contrast to Hannan's study. The physicians focused on the deficiencies of their plans and appeared to differentiate among plans - 24, 64 and 92 percent of physicians in each of three plans would recommend their plan to their own families. The trend towards physician ratings of plans may be regarded as defensive on the part of the profession but it has been supported by a key commentator on report cards (Epstein, 1998).

In conclusion, the physicians said that they were interested in report cards but were sceptical about the validity of current examples and were unwilling to use them in practice, either in

terms of sharing the information with patients or using the data to influence their own referral patterns. There was no evidence to determine whether this response is changing over time, nor whether high quality adjusted data are more acceptable or useful to physicians than are other forms of data.

6.3 Impact on behaviour of hospitals and other provider organisations

Evidence from three studies suggested that report cards can have a positive impact on provider behaviour (Bentley and Nash, 1998; Rosenthal et al, 1998; Longo et al, 1997), whilst two studies suggested a more mixed impact on providers (Berwick and Wald, 1990; Rainwater et al, 1998).

The impact on hospitals of the Pennsylvania consumer guide to CABG surgery was studied using a survey that was based on qualitative interviews (Bentley and Nash, 1998). The authors surveyed a random sample of key informants from both the hospitals and the purchasers. As a result of the publication of the consumer guide, the organisations stated that they put more effort into marketing their products, that they were more likely to monitor clinician performance and benchmark this activity against other hospitals in the area. They claimed that costs increased as a result but that the report promoted greater collaboration among clinicians. The evidence was based on self reports.

The impact of a profile report produced by Cleveland Health Quality Choice on hospital behaviour was evaluated by Rosenthal et al (1998). Using a poorly described methodology, the authors conducted four case studies in which they claimed that the report led to the development of successful hospital programmes to reduce length of stay, caesarean section rates and hospital mortality

rates. The descriptive study did not consider other possible explanations for the observed improvements but did describe common characteristics for successful use of the data, including strong leadership, interdisciplinary team work, data sharing and the development of consensus guidelines.

A more rigorous before-and-after quasi-experimental design was used to examine the impact of an obstetrics consumer report on hospital behaviour in Missouri (Longo et al, 1997). Half of the hospitals that did not have a car seat programme, formal transfer arrangements or breast feeding nurse educators prior to publication of the report, instituted or planned these services for the institution after publication. Hospitals in competitive markets were twice as likely to implement changes as were those who had a monopoly. All clinical outcome indicators improved after publication, including satisfaction, caesarean section rates and newborn mortality rates. Because of the study design, causality cannot be definitely asserted and in particular linking publication to reduced mortality over the relatively short time period of the study seems somewhat implausible.

Both of the studies reporting a more mixed impact of report cards on provider behaviour used self-reported survey data. Berwick and Wald (1990) conducted a postal survey of hospital leaders to determine their attitude to HCFA hospital mortality data; and they compared the responses by whether the leader was from a hospital with a higher or lower than expected mortality rate. The majority expressed negative views of the usefulness of the data to themselves and to consumers and doubted the accuracy of the data. Thirty one percent used it for quality improvement purposes, although leaders in hospitals with high mortality were more likely to use the data than were those with low mortality rates. Twenty percent reported

EVALUATION OF THE IMPACT OF PUBLIC DATA

that the data had caused problems for them, mostly in the form of poor publicity. The authors concluded that the attitudes of hospital leaders resulted in significant barriers to the use of published outcome data to encourage quality improvement.

Rainwater et al (1998) conducted a telephone survey of 39 key informants in acute care hospitals in California to assess the impact of CHOP report cards on hospital behaviour. Most respondents did not read the reports in detail. Three-quarters found some aspect to be useful, principally for benchmarking performance and one third instituted change as a result of the publication. The reports were criticised for focusing on mortality as the outcome and for the long time lag between data collection and publication. The authors concluded that the CHOP report card has had little impact on quality improvement in hospitals but the small sample and self-reporting nature of the study mean that the validity of this conclusion is questionable.

In conclusion, and taken in conjunction with the improvements initiated by the hospitals following the publication of the New York CSRS data, it seems that provider organisations are responsive to the release of comparative performance data.

6.4 Information that consumers want to access

Three studies were reviewed that investigated the specific data that consumers wanted to see in report cards (Edgman-Levitan and Geary, 1996; Hibbard and Jewitt, 1997; Robinson and Brodie, 1997). Some of the requests were contradictory but patterns emerge.

A detailed investigation using patient surveys, focus groups and interviews with advocacy and dissemination groups was conducted by Edgman-Levitan and Cleary (1996). Patient satisfaction,

experience with current data and expectations of new data were validated by interviews with health care managers dealing with patient requests for information. Patients seemed to want information on costs, benefits covered, quality of care, overall satisfaction, technical competence, the evaluation provided by physicians, and information on coordination and access to care. Even low income patients with minimal education questioned the accuracy of the reported data. Greater weight was given to information provided by family, friends and to what "people like them" think. The authors reported that the type and format of the presentation wanted varied according to the age and health of the individual. One format will not suit all: for example some patients wanted a summary whereas others wanted detailed information. The authors concluded that informed patients might be willing to make trade-offs between cost, access and quality.

Many of these findings are confirmed by a telephone survey of over 2000 patients conducted by the Princetown Survey Research Associates on behalf of the Kaiser Foundation and Association for Health Care Policy and Research (Robinson and Brodie, 1997). The survey also described the reliance on family and friends for information but in addition it described the importance of intermediaries, such as the referring physician. Thirty nine percent of respondents had seen data on performance and one third of these had used the data to influence a health care decision. Most considered that the data was aimed at, and most useful to, group purchasers and providers of care. The authors concluded that publication of data can be a useful and an acceptable adjunct to decision making but will not replace other sources of information.

One study has attempted to explain why patients rate some types of data over others. Hibbard and Jewitt (1997) used a survey and

EVALUATION OF THE IMPACT OF PUBLIC DATA

focus groups to develop an understanding of the meaning of quality indicators to users. They found that indicators that are poorly understood by users are also rated as not useful, irrespective of the scientific rigour or epidemiological utility of the indicator in question. Consumers often do not understand indicators because they have no understanding of the health care context within which the indicator operates. The authors concluded that importance alone is not a sufficient reason to include indicators in report cards that are designed for consumer use.

In conclusion, consumers expressed a desire for a wide range of information on quality but did not necessarily know how to use it and wanted intermediaries to make sense of it on their behalf.

6.5 Impact on decision making of consumers

The fact that consumers are willing and able to state their requirements for publicly released information does not necessarily mean that the available data have an impact on their decision making. Six studies addressed this issue: five concluded that it had minimal influence on consumer decision making (Hibbard and Weeks, 1989; Robinson and Brodie, 1997; Schneider and Epstein, 1998; Vladeck et al, 1988; Mennemeyer et al, 1997) and one suggested that it may have some impact (Mukamel and Mushlin, 1998).

Hibbard and Weeks (1989) looked at the impact of comparative data on physician fees on consumer use of services. Whilst not strictly quality data, this study is the only example of a randomised controlled trial looking at the impact of public disclosure on consumer choice. The authors randomly allocated 717 Medicare enrollees and 658 state employees in Salem, Oregon to receiving cost data. They measured the doctor visit rate, expenditures on ambulatory care and costs per visit and no statistically significant

difference in utilisation was observed between the intervention and control groups. It appeared that lack of information is not the only determinant of consumer choice.

The other four studies were all descriptive, using self reports as the basis to assess impact on consumer decision making but they all confirmed the above findings. Robinson and Brodie (1997), using methods described above, found that one in ten consumers reported using performance data to make decisions. Vladeck et al (1988), also described above, found that consumers continued to access hospitals with high published mortality rates.

The use by consumers of the Pennsylvania report card on CABG mortality was assessed using a telephone survey of 474 patients who had undergone CABG surgery in the previous year in four Pennsylvania hospitals, two of which had lower and one higher than expected mortality rates (Schneider and Epstein, 1998). Sixty percent responded and 12 percent of these were aware of the guide at the time of their surgery and less than a quarter of these stated that it had any significant impact on their choice of surgeon. Most of these patients were unable to correctly specify the data on which they made their decision of which surgeon to access. An awareness of the report was associated with younger age, college education, high preoperative health status and the existence of heart disease for longer than one year. When the guide was described to them, 56 percent reported an interest in seeing it and a similar percentage thought that they would probably have changed their surgeon if he or she had a higher than expected mortality rate. There was, however, a low level of willingness to pay for the report.

These results were confirmed by a poorly described study looking at the effect of the HCFA mortality data on hospital utilisation, as

EVALUATION OF THE IMPACT OF PUBLIC DATA

measured by hospital discharge rates (Menemeyer et al, 1997). Published adjusted mortality data had minimal impact on hospital utilisation, whereas there were large and significant effects produced by anecdotal press reports of untoward deaths in hospitals. Reports of the unfortunate death of a patient who fell off a trolley in one hospital resulted in reductions in discharges of approximately nine percent. The authors concluded that the HCFA data had no significant impact on consumer decision making.

The only study suggesting that public disclosure had some impact on consumer decision making evaluated the impact of the New York CSRS report on hospital and physician market share and on price changes before and after publication of the data (Mukamel and Mushlin, 1998). The quasi-experimental study indicated that hospitals and physicians with better outcomes experienced higher rates of growth in market share and that physicians with better outcomes had higher rates of growth of charges for the procedure. The magnitude of the association varied geographically, possibly reflecting socio-demographic differences, and declined over time, suggesting that the market responds primarily to new information. No attempt was made to control for the impact of cost on the market and the authors maximised the chances of a positive result by studying only fee-for-service systems and excluding managed care data from the analysis.

In conclusion, it appears that the currently available performance data had minimal impact on consumer choice. Whether this finding is a feature of the consumers' access to the data, their ability or willingness to use it or the nature of the data currently available is unclear from these studies. The following section will review the published studies attempting to explain this phenomenon.

6.6 Explanation of impact on decision making

Several studies have helped to explain the lack of impact of report cards on consumer and purchaser decision making. Some of these studies have already been described in detail and will be reviewed briefly.

Hibbard et al, (1998) hypothesised that it was consumer understanding of the information that lay at the root of the problem. The authors surveyed 1673 Medicare enrollees from five geographic areas across the US with a high penetration of managed care to determine their understanding of the difference between managed care and fee for service. One third of respondents knew almost nothing about HMOs and only one in ten had adequate knowledge to make informed decisions using the information provided. The authors concluded that consumers lacked a useful framework to put performance data into the context of the health system.

Employers have some of the same problems making sense of performance data. In-depth interviews with representatives from 33 large employers in California, Cleveland, Pennsylvania and New York revealed that HEDIS data and patient satisfaction data were regarded as biased, inaccessible, too detailed and ambiguous (Hibbard et al, 1997b). They felt that quality was the responsibility of the Managed Care Organisations, rather than the purchasers.

Some consumers did not understand whether high or low performance rates are good (Jewitt and Hibbard, 1996) and rated anecdotal evidence more highly than empirical evidence (Menne-meyer et al, 1997; Robinson and Brodie, 1997). In part, this was due to a lack of trust in information provided by both providers and purchasers who may have vested interests (Robinson and Brodie, 1997). There was also evidence that the time from

accessing to having to make a decision based on the data was too short for some people to make use of the information (Schneider and Epstein, 1998).

6.7 Information wanted by purchasers

The data requirements of employers as purchasers of care on behalf of their employees are likely to be different from those of individual consumers. Three studies were found in this area (Gabel et al, 1998; Gold et al, 1995; Hibbard et al, 1997b).

Factors affecting employer choice of plan were examined in a national survey of 1502 employers (Gabel et al, 1998). More than two thirds of employers offering managed care plans stated that employee satisfaction, cost and administrative efficiency were the most important factors in plan selection. Information about physicians was also found to be important in this study and this finding was supported by the attention given by managed care plans to the characteristics of new physicians prior to their recruitment. A telephone survey of 138 managed care plans from twenty nationwide metropolitan areas in the US demonstrated that plans used board certification and even utilisation data to choose their physicians (Gold et al, 1995). Hibbard et al (1997b) showed that employers found process data more useful than outcome data and that they preferred brief summary information to detailed reports.

In conclusion, from the limited evidence it seems that employers primarily want non-clinical information to guide their purchasing decisions.

6.8 Impact on employers' purchasing decisions

Since employers are major purchasers of care in the US, they have the potential to use performance data to improve quality. Only two

studies were found which sought to determine whether currently available data influenced employers purchasing decisions (Hibbard et al, 1997b; Gabel et al, 1998).

The impact of NCQA accreditation and HEDIS data on employer choice of plan was examined by surveying 1502 employers across the US with more than 200 workers (Gabel et al, 1998). The results were compared with a similar survey conducted by the same research team the previous year. The percentage of employers familiar with accreditation increased from 29 percent to 35 percent. The improvement was far more marked for large employers. Eleven percent considered NCQA accreditation to be very important and five percent considered HEDIS data to be very important. Less than ten percent required NCQA accreditation for plan selection and one percent provided the HEDIS data to their employees to help them to choose a plan. It therefore seems that performance data has a small but increasing impact on purchaser decisions.

A similar study examined the use of a variety of different performance measures by 33 large employers in California, Cleveland, New York and Pennsylvania (Hibbard et al, 1997b). In-depth interviews were conducted with key informants. Seventy eight percent reported that HEDIS data was available to them and 75 percent had access to patient satisfaction data. Fifty four percent of employers reported using HEDIS data to choose plans and 59 percent reported using consumer satisfaction data. Although these figures are better than those reported by Gabel et al, (1998), the authors concluded that use of data was still limited, largely because they were inadequately packaged and targeted for employers.

Both studies therefore demonstrated limited use of performance data by health care purchasing employers.

6.9 Impact on quality of care outcomes

The most fundamental question is whether public disclosure of performance data influences outcomes of care. Only three studies have attempted to address this issue, all of which have used observational designs (Hannan et al, 1994; Peterson et al, 1998; Longo et al, 1997).

The New York CSRS is the most rigorously studied system for public disclosure of performance data. Hannan et al (1994) studied all (57,187) patients undergoing isolated CABG surgery who were discharged from the 30 New York hospitals performing the procedure between 1989 and 1992. A clinical database was used to identify significant independent risk factors and to assess risk-adjusted in-hospital mortality rates. Outcome measures were the actual, expected (from a logistic regression model) and risk-adjusted mortality rates. The actual mortality decreased from 3.52 percent to 2.78 percent over the period of the study. The illness severity of the patients being operated upon increased during this time, so the risk-adjusted mortality decreased even further - from 4.17 percent in 1989 to 2.45 percent in 1992; a reduction of 41 percent which was considerably greater than the national average. The risk-adjustment model proved to be sensitive at all ten levels of patient severity. This study resulted in considerable debate among academics and several further studies attempted to support or refute the authors' conclusions that publication made a significant contribution to the observed improvement (Hannan et al, 1995; Omoigui et al, 1996; Hannan et al, 1997; Peterson et al, 1998).

The main criticism of the release of data in New York was that it could have reduced access to CABG surgery by forcing sicker patients to seek surgery outside the State or by surgeons

refusing to operate on high-risk patients. Peterson et al, (1998) used national Medicare data to examine trends in the percentage of New York residents aged 65 years or more who received out-of-state surgery before and after the initiation of the provider profiling programme. They also examined procedure use by elderly patients with myocardial infarctions (MI) within the state to determine whether high-risk patients were being refused treatment. Contrary to a previous single centre study (Omoigui et al, 1996), they found that the percentage of New York residents receiving out-of-state bypass operations decreased between 1987 and 1992 from 12.5 percent to 11.3 percent ($P<0.01$ for trend). They also found that the likelihood of bypass surgery following an MI actually increased. The authors confirmed a reduction in 30 day mortality in New York well above the national average during the period of study (33 percent reduction in New York versus 19 percent nationally, $p<0.001$). They confirmed that New York had one of the lowest mortality rates and largest improvements of all states studied. The only other area with similar figures was Northern New England, which also produced provider profile reports, though for internal use rather than for publication (O'Connor et al, 1996).

The third study demonstrating improved outcomes examined obstetric care in Missouri hospitals and has been described earlier in the report (Longo et al, 1997). The authors reported several improvements in structural, process and outcome measures of obstetric care as a result of provider profiling.

In conclusion, there is some evidence from quasi-experimental studies that publication of comparative performance data may contribute to improved outcomes.

6.10 Mechanisms of action of performance data

The evidence of improvements in mortality as a result of provider profiling is reasonably convincing, though the added impact of releasing the data to the public rather than using them solely for internal purposes is less clear. The mechanisms of action are also unclear and only one empirical study has attempted to test an explanatory hypothesis.

Hannan et al, (1995) hypothesised that the improvement was related to management of surgeon volume, based on the knowledge that low volume surgeons had higher mortality rates. Using the same data described in a previous study (Hannan et al, 1994), they found that low volume surgeons (≤ 50 operations per year) experienced a 60 percent reduction in risk-adjusted mortality, whilst high volume surgeons (>150 operations per year) experienced a 34 percent reduction. The percentage of operations performed by low volume surgeons decreased by 25 percent, from 7.6 percent in 1989 to 5.7 percent in 1992. The authors concluded that the overall decline was in part the result of an exodus of the low volume and high-risk surgeons, probably as a result of hospitals restricting operating privileges. They also considered the markedly better performance of surgeons new to the system and improved performance of non-low-volume surgeons as other explanations. However, none of the above explanations accounted for the bulk of the 41 percent decline in mortality and the true mechanisms of impact on mortality following CABG are unclear.

6.11 Impact on costs

There has been little evaluation of the costs of publishing performance data, in part because much of the administrative data is routinely collected and some of the clinical data might be collected for internal review purposes. Initial costs of the

development of measures, analytical methods and data management systems and ongoing costs of data collection, analysis, auditing, dissemination and management of the responses are likely to be significant but have not been reported.

In a study of how Pennsylvania hospitals responded to the PHC4 *Consumer Guide to Coronary Artery Bypass Graft Surgery* (Bentley et al, 1998), the authors found that the majority of hospitals devoted a larger share of their financial resources to their bypass programme but exact figures are not given. Gabel et al, (1998) looked at the cost of NCQA accreditation and found that accredited plans actually cost four percent less than non-accredited ones. This does not mean that the costs of data collection to obtain accreditation were insignificant and is probably explained by accredited plans tending to be larger and more able to spread their costs. A US General Accounting Office report (GAO, 1994) describes two Pennsylvania hospitals that have estimated the cost of reporting data to the PHC4. One estimated that it spent \$26.5 million, or \$14.20 per patient discharge to collect and report the data in 1991. Another estimated the cost at \$17.43 per patient. Larger hospitals spent about half as much as did smaller hospitals because they were able to spread their costs over more patient discharges.

In conclusion, the costs of public disclosure per patient discharge are not clear but the cost of extraction of data from clinical records for two hospital-based systems has been estimated to be in the region of \$16 per record at 1991 prices (GAO, 1994). Given increases in staff costs and the time required to review records for more complicated systems, this figure may be an underestimate of the real costs today. How much of an underestimate is not known.

6 .12 Summary of the evidence

The following conclusions can be drawn from the above:

- There is insufficient research on which to make evidence-based policy decisions and cost benefit analyses.
- Currently available report cards are rarely read by individual consumers or group purchasers and even when read, appear to have little influence on purchasing decisions.
- Physicians and provider organisations are critical of performance reports although provider organisations seem to be the most responsive to publicly disclosed data.
- Publishing comparative data is associated with improved outcomes, at least in the limited case of post CABG mortality.

7. DISCUSSION: POLICY ISSUES FOR THE UNITED KINGDOM

The following section describes the important conceptual and technical issues arising from the United States' experience of public disclosure of performance data.

7.1 Generalisability of the United States research

The different historical development, health systems and cultures of the US and UK mean that the experience of public disclosure of performance data may not be directly transferable. The purpose of this section is to consider the applicability of the above findings to the UK health care system.

In broad terms, the potential of public disclosure as a mechanism for regulation, incorporating judgements of quality and as one facet of quality improvement, is similar in both countries. So too is the requirement for high quality data on performance. The nature of the information that consumers want, with an emphasis on simple, non-technical measures, is probably similar for British and American citizens. The types of data needed for contracting and purchasing decisions are also likely to be similar.

However, there are some important differences. UK society is not as consumer orientated as that of the US and the demand for health care information from the general public is unlikely to be as great. There is little evidence that the internal health care market in the UK operated in any effective way and as competition is replaced by cooperation, it is increasingly unlikely that providers will be motivated by market pressure to respond to comparative data in the same way as do US organisations. The relatively high level of central regulation in the NHS could work either way on the effectiveness of public disclosure. On one hand, purchasers may lack the authority to act on performance data because this might threaten the existence of under-performing individuals or organisations; but

on the other hand, the opportunity to implement public disclosure in a controlled and sensitive way is greater in the UK than it is in the US.

The forces influencing the behaviour of the medical profession may also be different. There is evidence in both countries that financial incentives have an impact on behaviour but the different systems of reimbursement put US physicians at higher risk and so anything that might influence their income may have a greater impact on their behaviour. Finally, the same conditions do not exist in the UK for private enterprise to participate in public disclosure and so one of the key players in the US will be absent from the UK scene.

7.2 The purpose of public disclosure: a conceptual model

Three models are proposed to provide a clear conceptualisation of the purpose of public release of performance data.

First, the *Public Accountability* model sees public disclosure as a public responsibility, independent of the consequences. Proponents argue that the public good will be served by openness and that the release of data, in conjunction with appropriate education and the ensuing debate, will help clarify important societal issues. The broad benefits of involving the public in this way may be as important as any measurable impact on quality of care. For this model, the scientific rigour of the data may be of less importance than the fact that it is being released and the reporting level can be high. With no clear strategic purpose, this model may have little impact on quality of care. However, it will also be least likely to be perceived as threatening to professionals.

The second model, a *Market Orientated* one, assumes that the provision of comparative data on quality will allow informed and

willing consumers to drive quality improvement through selective purchasing or utilisation behaviours. To make valid and fair comparisons, the data would have to be standardised. The limited evidence described in this report suggests that the data might be most effective if published at the level of a provider organisation (for example, hospitals, primary care groups or group general practices), though increasingly in the US consumers voice an interest in individual provider level data.

The third option, a ***Professional Orientated model***, assumes an intrinsic desire on the part of health professionals to improve their practice, given the appropriate environment. This may be motivated in part by a desire to retain autonomy in the face of greater governmental regulation. Providing data on variations in practice aids this process and making it public increases responsiveness. The data act as a catalyst to identify and solve problems and publication turns up the heat to enable the catalyst to work. Standardisation is not the highest priority, since provision of rigorous data is only one small part of the quality improvement process which will also include audit, educational programmes and benchmarking. The benefits of detailed risk adjustment may not be worth the costs. In essence, this model is a publicised clinical audit.

The purpose of publication therefore dictates the reporting level, the targeted audience and the content of the data. Identifying extremely poor performers and shifting the mean level of performance may require different levels of data rigour. The US experience suggests that managing all processes of public disclosure will optimise the chances of positive outcomes. To publish performance data and leave the consequences to chance is not enough. For example, the publication of the New York CABG mortality data might not have influenced outcomes to the same

extent if the State Department of Health had not taken action, such as suspending one hospital from operating until its programme had been restructured (Chassin et al, 1996).

73 **Public disclosure and quality improvement**

One purpose for public disclosure advocated in this report is the promotion of quality improvement and this is central to the concept of clinical governance, introduced as part of the 1997 NHS reforms. Quality improvement requires both changes in the health care system and changes in stakeholder behaviour. The following section briefly summarises the concept of continuous quality improvement (CQI), describes a theoretical framework for behaviour change and describes the contribution that public disclosure, alongside other strategies, could make to quality improvement.

CQI represents a philosophical approach characterised by continual improvement of the processes of care, focused on the outcomes that are important to the users. It is based on the understanding and use of explicit information about performance and requires effective leadership and teamwork throughout the organisation (Shortell et al, 1998). CQI has been widely used in business and was first applied to health care managerial systems a decade ago (Berwick, 1989). Only in the last few years has its relevance to clinical practice been considered and there is currently little empirical evidence that its application improves quality of care across whole institutions (Blumenthal and Epstein, 1996). This may in part be a reflection of its recent introduction to health care or the methodological difficulties of studying cultural change using traditional research methods. It may also reflect the cultural and structural barriers to implementing an approach that requires a complete re-engineering of health systems (O'Brien et al, 1995).

Whilst data about performance is a fundamental component of CQI, the impact on organisations of making this information widely available has received little attention to date.

To take a uni-dimensional view of quality improvement is clearly inappropriate. No single intervention, whether it be public disclosure, audit or incentives, will influence all people all the time or to the same extent. A flexible, multi-faceted and targeted approach is likely to be most effective (Bero et al, 1995) and this requires a conceptual model of behaviour change.

The focus for behaviour change may be on internal processes or external influences (Grol, 1997). Public disclosure of performance data could in theory contribute to both of these but is often perceived to damage the former and is seen as a punitive tool of the latter.

Internal processes include:

- Educational interventions aimed at intrinsic professional motivation (for example small group or problem-based learning).
- Epidemiological approaches based on rational information seeking and decision making (for example evidence-based guidelines).
- Marketing, focusing on an attractive and targeted product (for example using the mass media for health promotion messages).

External influences include:

- Behavioural approaches using external stimuli (for example reminder systems or economic incentives).

- Social interaction using role models (for example peer review).
- Organisational models (for example continuous quality improvement), and
- Coercive influences which focus on control and pressure (for example legislation or complaints).

There is a risk that external interventions, such as performance management using public disclosure of specific indicators, may tend to displace internal motivation for quality improvement (Sheldon, 1998). Use of performance indicators is more likely to be effective if it builds on established formal and informal professionally-based quality improvement strategies.

In summary, we know little about the impact of public disclosure of performance data on the processes of continuous quality improvement or behaviour change. It is likely to have both positive and negative effects and is best seen as one of many quality improvement strategies available within a health care system.

7.4 Factors influencing the content of the data

A review of public disclosure in 1995 highlighted several methodological problems with the data contained in report cards, including incomplete measures, lack of standardisation and inadequate risk adjustment (Epstein, 1995). Many of these problems have been addressed in recent years but there is still disagreement about the content of disclosed data. To illustrate this, the debate may be classified into three groups.

First, the purists claim that only well recognised, scientifically tested indicators should be made available for external use, i.e. acceptable levels of validity and reliability for the population

under study, rigorously risk adjusted, and focused on outcome measures because improving outcomes is the ultimate aim of health care. Since mortality is the only consistently recorded outcome, public reports should concentrate on mortality data. They also argue that clinical data collected specifically for the purpose should be used in preference to routinely collected administrative data and that publishing the results of evaluation of the indicators in peer-reviewed scientific journals adds to their credibility. The underlying premise of those who advocate such evidence based indicators (McColl et al, 1998) is that only top quality data will be credible, and therefore acceptable, to clinicians and useful to purchasers.

At the other end of the spectrum, some argue that the quality of data is not as important as the principle of openness, that publication is just one small part of the quality improvement process and that information should be made public as quickly as possible, even if that means compromising on the scientific properties of the data. They argue that process quality measures will be more useful than outcomes, because they are more comprehensible, within the control of practitioners and more immediate in their impact. They argue that if outcomes are to be used then measures other than mortality should be considered (Topol and Califf, 1994). For example, providing mortality data for patients undergoing a coronary artery bypass operation may not be meaningful because patients may not consider death to be the most relevant outcome. They may assume that they will survive and be more interested in whether they will be able to walk to the shops after the operation. If process measures are used, risk adjustment has been shown to be less relevant (McGlynn et al, 1998) and would anyway be difficult given the quality of data available, particularly in the UK (Shaw et al, 1998). They consider the face validity of the indicators to be

DISCUSSION: POLICY ISSUES FOR THE UNITED KINGDOM

more important than their statistical validity and point out the inadequacies of current risk-adjustment mechanisms (Jollis and Romano, 1998).

The third group are those who take a middle line. They state that public indicators should be as good as possible but do not have to be perfect. Waiting for the best will retard the process and delay the potential benefits of publication (Epstein, 1995; Hannan, 1998). They argue that processes are reasonable measures for public release, as long as they can clearly be linked to outcomes, preferably by research evidence (Davies and Crombie, 1995). They consider the costs as well as the benefits of risk-adjusting outcome measures and point out that the gain of using data extracted from medical records over routinely available data is small and probably not worth the considerable cost. Some advocate releasing the performance data to the providers well before going public in order to give them the right of reply, with responsible challenges published with the reports as an addendum. This approach has been used in the California Hospitals Outcomes Project.

Choosing the most appropriate diseases or procedures for publicly released performance measures requires careful consideration (McGlynn and Asch, 1998). There is a danger that subjects are chosen by default, because the information is available or of high quality rather than for any rational reason. Subjects should satisfy certain criteria:

- That the area for performance measurement should be important to the interest group for which it is produced, should be common and known to be problematic in terms of quality.
- That data should be available or easily collected.

- That the health impact and financial cost of poor quality is significant, and finally
- That there are significant variations among providers because demonstration of minimal variability will not encourage improvement.

An example of Diagnosis Related Group frequency data for Californian hospitals is provided in Appendix 5 to illustrate how the number of hospital discharges and mortality rate for each group might influence the choice of indicators for public disclosure. Similar frequency data for English NHS hospital services is provided in Appendix 6. This illustrates that cerebrovascular disease, pneumonia and myocardial infarction are common causes of hospital bed occupancy in the UK and might be appropriate conditions for publication of comparative mortality data. Risk-adjustment systems have already been developed for these conditions in the US and could be applied in the UK but they would require collection of a small number of clinical data elements in addition to the routine data that are currently available.

The quality of the data available in the UK for public disclosure is of particular concern. Both administrative and clinical data have been collected in the US for many years as a consequence of the system of payment. Whilst the quality of these data has been criticised, it is a significant improvement on the amount and quality of the data available in the National Health Service. This is particularly true of data in general practice (Marshall, 1999) and also true of secondary diagnosis coding of hospital clinical data, which is required for risk adjustment. The publication of valid comparative performance data will be severely hampered by these problems.

7.5 Risk adjustment of performance indicators

The issue of risk adjustment warrants further brief mention. If published performance data are to be used primarily to compare outcomes then risk adjustment is necessary. The rationale is to remove sources of variation that are not directly related to quality of care. Characteristics known to affect risk of poor outcomes in hospitalised patients include age, sex, acute physiological status, reason for hospitalisation, severity of condition, presence of co-morbidity, functional status, psychosocial and cultural factors, socioeconomic factors and patient preferences (Iezzoni, 1997a).

Despite the desirability of risk adjustment, the process is not without problems (Iezzoni, 1997b). The collection and collation of risk factors is time consuming, expensive and presents practical difficulties - for example identifying whether the risk was pre-existing or occurred during the hospital stay. The ability to adjust risk for less objective characteristics, such as patient preferences, is limited. There are many risk-adjustment systems available that use diverse sources of data and possess different levels of complexity. When applied to the same data, different systems do not always rank hospitals in the same way. The implication is that providers in the United States could shop around for the system that shows them in the best light. Therefore, the same risk-adjustment system should be applied at a regional or national level to data derived from different hospitals in the UK.

The relative merits of using administrative data (i.e. data collected routinely for other purposes) and clinical data have also been debated (Black, 1999). Some authorities claim that the statistical performance of risk-adjustment systems using the two forms of data is not significantly different (Iezzoni, 1997b). However, the performance of systems using discharge abstract data may be

artificially improved by erroneously including codes for events that did not exist at the time of hospital admission (Iezzoni et al, 1995). The purposeful increased coding of catastrophic events for dying patients has been referred to as 'death code creep' (Iezzoni, 1997b). Other authorities suggest that risk-adjustment mechanisms using clinical data are better than those using administrative data. However, in a study comparing clinical and administrative databases for CABG mortality rates the addition of only three clinical risk factors to the administrative database seemed to account for much of the difference in performance of the two systems (Hannan et al, 1992). The use of clinical data is generally agreed to be more credible, and therefore more acceptable to physicians. The addition of clinical data elements to an administrative risk-adjustment database would probably alter the apparent relative performance of different hospitals in the UK.

7.6 Factors influencing the release of the data

Having addressed the purpose and content of the information for public disclosure, the next task to consider is how it should be published. The following section will consider some of the issues, including the level of public disclosure, the format, the process and the timing of release.

Performance data should be published at the level most likely to stimulate improvement. Indicators published as part of the English and Welsh National Performance Assessment Framework have been published at the health authority and hospital level. High level reporting, such as for health authorities, may not be the level most likely to promote significant quality improvement. Health authorities have less control over the processes and outcomes of care than providers, such as hospitals or general practices. Data release at the level of Primary Care Groups (PCGs) is also

DISCUSSION: POLICY ISSUES FOR THE UNITED KINGDOM

proposed but how quickly these new groups will function as organisational units is unclear. If PCGs exercise little control over their constituent practices, publication of results at this level may have little impact on the quality of care provided by the practices. Publication at the level of individual practitioners has its risks, as already mentioned, but is theoretically possible if condition prevalence rates for each individual practitioner are high enough. However, it will be expensive.

Problems do exist with targeting individual practitioners. First, many components of quality are not solely within the control of the practitioner - even mortality following cardiac surgery is dependent on the anaesthetist, surgical team, general hospital care, and the patient. Second, because health care delivery is increasingly a team activity, rather than an individual one, and the focus of continuous quality improvement is on the system rather than on the individual, it is not rational to monitor the performance of individuals independently of their operational team. There is therefore a strong argument for releasing data at the level of the provider organisation or system.

The timing of public release of data needs to be considered. The acceptance of public disclosure requires a culture change which is best managed sensitively. Some organisations in the US have provided the data for internal use for a year or two before making them publicly available.

Aggregation of report card indicators across diseases or conditions is becoming a more common practice and has been recommended for the UK National Performance Assessment Framework. This is considered to address the ability of the human brain to take in only a limited number of facts and reduces the risk of tunnel vision as

described in section 6.5.(para 2) (McGlynn et al, 1997). However, composite indicators may be difficult for consumers to understand (Edgman Levitan and Cleary, 1996) and whilst aggregation may produce useful summary statistics, its utility as a way of reducing the total number of indicators requires further study.

Finally, the format of the final publication is likely to have an impact on its effectiveness. There is little empirical evidence to advocate one format over another but readability and brevity are relevant factors. Readability does not necessarily mean that the reports have to be highly professional in their appearance - the Scottish Clinical Outcome Indicators were purposefully produced informally, so as to avoid excessive credibility being attached to the results (Kendrick et al, 1998). Most report cards rank providers, but because the variation among the results of some published indicators is not large, the result can be misleading and therefore some advocate an alphabetical or geographical order (Institute of Medicine, 1994). Finally, some reports come with multiple caveats about abuse of the information but the extent to which they are heeded is unclear.

7.7 The role of the medical profession

To impose the public disclosure of performance data on a reluctant medical profession would be possible but the consequences of this action, in terms of resistance, demoralisation and detrimental effects on trust have already been highlighted. Professions are innately resistant to change and will respond defensively when their autonomy is threatened (Freidson, 1970).

Self-policing is one of the defining features of professions and one that the medical profession fights hard to preserve. Public disclosure provides the tools for those outside the medical profession to judge those within. It may be perceived as a threat to

professional autonomy, a lack of trust in the standards that doctors achieve, or an attempt by others to make definitive judgements in areas of inevitable uncertainty. Some might argue that public disclosure threatens one of the defining characteristics, and therefore even the identity, of the medical profession. The profession has proved itself to be able to adapt in the past and may need to redefine its core features. In the future, the profession is likely to be characterised by greater accountability to government, health care managers, fellow professionals and the general public with less individual clinical autonomy.

The extent to which the medical profession accepts the principles and practicalities of public disclosure is central to its success. Using publication as a stick to beat the profession is likely to have significant adverse consequences, but encouraging the profession to take the initiative will increase the opportunity to use public disclosure as part of a quality improvement strategy. This might mean that the pace and the content of public disclosure is less than might be desired by government. Pushing the agenda too far or too fast may, however, result in loss of morale amongst an important but vulnerable part of the NHS workforce.

7.8 The adverse consequences of publication

Whilst the potential benefits of public disclosure are receiving some attention from the research community, reports of negative consequences are largely anecdotal. It is, however, important to consider them if the benefits of public disclosure are to be maximised.

The unintended consequences have been classified into five areas (Goddard et al, 1998; Smith, 1995). First, *tunnel vision* causes organisations to concentrate on the areas that are being measured, to the exclusion of other important issues. This has been observed

in the US, where the publication of performance data resulted in behaviour changes in the measured processes of some organisations (Longo et al, 1997; Bentley and Nash, 1998). This effect could be managed by utilising a broad range of quality indicators and sophisticated patient sampling techniques.

Second, publication can result in the pursuit of narrow local objectives at the expense of broad organisational goals. This has been referred to as *sub-optimisation* and is a particular issue when the implementation of performance measurement is the responsibility of junior managers who may not be aware of the larger agenda of their organisation. Third, publication can result in *myopia*, or focusing on short-term issues at the expense of long-term strategies.

Fourth, public disclosure can result in *misrepresentation* of performance results. This can be a deliberate and malicious manipulation by the provider organisation or an implicit 'massaging' of the data. This problem was observed in New York where an external audit revealed the data quality of a small number of hospitals to be suspect (Chassin et al, 1996). Misrepresentation can also be manifest in media reporting of report cards, partly as a result of sensationalism and partly as a result of ignorance of statistical issues such as probability, confidence intervals and normal ranges. Finally *gaming* describes altered behaviour as a result of public disclosure so as to obtain a strategic advantage. This is particularly common when individuals may be made vulnerable as a result of poor performance data (Schneider and Epstein, 1996).

A further potential consequence of public disclosure is the effect that it might have on public confidence and trust in health professionals and the health care system. (Davies and Lampel, 1998). This has been the subject of much debate but little empirical

research. Explicit information about under-performance may have adverse consequences, particularly in a public system with relatively little opportunity to seek care elsewhere, and there is a risk that public disclosure could cause more problems than it attempts to rectify. The Institute of Medicine has highlighted the vulnerability of health professionals and institutions to harm as a result of public disclosure, resulting in loss of reputation and even livelihood. Individuals are probably more vulnerable than institutions and this should be taken into account when considering the level of reporting (Institute of Medicine, 1994).

Many of these problems could be minimised by ensuring that public disclosure is conducted sensitively and fairly, occurs in a supportive environment, and is managed as a dynamic part of the quality improvement process.

7.9 The financial cost of implementing a policy on public disclosure

The resources required to implement a policy on public reporting are likely to be significant and extend beyond the cost of simply placing performance data in the public domain. Resources are required to develop the indicators, use the indicators to measure performance, report the results, support and educate the key stakeholders to ensure that they make the best use of the information and then act upon any deficiencies that have been highlighted. The opportunity cost of allocating resources to the public reporting of performance data in place of other quality improvement strategies or direct patient care should be considered.

7.10 The research agenda

The public disclosure of performance data is in its infancy and much further research is required to determine whether its benefits

outweigh its risks and whether it is worth the cost. The answer to this question may differ by the condition studied, the setting and the population of patients affected.

The most fundamental research question is whether publishing performance data leads to greater benefits and fewer risks than keeping the information only for the internal use of health professionals and organisations. Notwithstanding the merits of quasi-experimental designs, this question would best be addressed using a large controlled trial randomised by geographical district, with the intervention groups (provider organisations or health districts) exposed to public release of performance measures and the control groups collecting the same data but using it for internal quality improvement purposes only. An assessment of all significant measures of quality, including those relating to process (e.g. what physicians do to patients) and outcomes (e.g. mortality, morbidity, patient and physician satisfaction), would be required. The results of the trial would have to be seen in the context of the quality of data being released. The policy direction of the New NHS and the publication of the National Performance Assessment Framework in the UK presents a unique opportunity, as well as a compelling need, to provide experimental evidence of the benefits and disadvantages of public disclosure.

In addition to this fundamental question, the negative impact of publication of performance data on public trust and professional morale needs to be investigated further. This would best be addressed using a qualitative case study design.

The procedure of risk adjustment requires further investigation. The relative effectiveness of simple and cheap risk adjustment using routinely collected data against sophisticated and expensive

DISCUSSION: POLICY ISSUES FOR THE UNITED KINGDOM

mechanisms that require additional data collection needs to be assessed, as does the validity of comparing performance data derived from different risk-adjustment mechanisms.

Finally, the content and presentational format of published data for the different audiences should be considered. There is increasing evidence of what information consumers and purchasers say they want to receive in principle but this might be different from what is acceptable and effective in practice.

A policy on public disclosure is most likely to be effective if guided by empirical evidence. The paucity of evidence in the United States has been highlighted in this report and there has been even less research in the United Kingdom. Government policy would benefit from a focused and adequately funded research and development programme, for which the NHS R&D Directorate is the most obvious funding body.

Illustrative examples of performance indicators for the United Kingdom

The development and reporting of performance indicators can be a complex and problematic process. Examples of indicators that could be used in the United Kingdom, with particular reference to the associated statistical issues, are described in Appendix 7.

8. RECOMMENDATIONS

The following policy recommendations are based on the review of the US experience of public disclosure. They were presented by the Nuffield Trust to Mr Frank Dobson, the then Secretary of State for Health, in July 1999 and circulated to key policy makers and advisers in the UK.

8.1 The intended purpose or purposes of public disclosure should be made clear to all stakeholders

Those who work in the NHS may feel threatened by public disclosure and may question the resources required to collect and report performance information. Making a clear statement about the expected benefits will help these people to understand the rationale for greater openness. In addition, the intended purpose will dictate the content and process of public disclosure. There is some evidence to suggest that disclosure can facilitate public accountability, improve the decision making of consumers and purchasers, inform decisions about resource allocation and regulation and promote quality improvement. An explicit definition of the goals and objectives will also help identify the evaluation criteria that are used to assess whether and how public disclosure is improving health care processes and outcomes.

8.2 Public disclosure should be seen as an evolutionary process, becoming progressively more sophisticated and comprehensive over time

Public disclosure represents a major culture challenge for health professionals and organisations who to date have had little obligation to demonstrate accountability for quality of care. There is a danger that public disclosure may be perceived as threatening professional autonomy and therefore work against the creation of an environment where systematic evaluation and improvement

RECOMMENDATIONS

can flourish. Change will take time and the quality of the performance data will be an important determinant of the acceptability of the public information and its ability to promote change. The state of the art of performance indicators and the information technology to support them need to be continuously refined. Quality indicators do not have to be perfect but do need to be good enough to achieve 'buy-in' from stakeholders. The National Performance Assessment Framework represents a good starting point both in the articulation of a long-term policy direction and as a classification of measures for development. Its credibility, however, will depend upon the demonstration of year-on-year improvements in the selection of indicators and their measurement and reporting. The framework will need to be responsive to constructive criticism. The Department of Health should work closely with academics, clinicians and other stakeholders to refine the framework.

8.3 Public disclosure should be seen as one component of clinical governance

The principles of clinical governance have been largely accepted by health professionals and managers. Accountability for continuous quality improvement is a defining feature of clinical governance and in future should be based at least partially on publicly available information about performance. Providing evidence of deficiencies in quality, or evidence of best practice, for internal use alone does not appear to have produced the expected or desired level of improvement. Public disclosure may be seen as a way of focusing the attention of both clinicians and managers on specific areas. It will be most effective if integrated into other quality improvement strategies, for example educational initiatives, the use of professional and financial incentives, organisational change and regulation.

8.4 Provider organisations should be a key audience for information about performance

Assuming that the encouragement of quality improvement is one of the intended purposes of public disclosure, the reporting level of performance data needs to be carefully considered. Current evidence suggests that individual consumers are the least responsive to performance data, even in the consumer-orientated US. Therefore, whilst users should contribute to the process of public disclosure, they are not necessarily the prime audience for the data. Provider organisations appear to be the most responsive of the stakeholders. This is because they are sensitive to their public image and because they have the authority to act on sub-optimal levels of performance and promote better standards of practice. Most of the evidence is based on hospitals in the US and it is likely that UK hospitals and Primary Care Groups will respond in a similar way. Reporting at the provider organisation level is consistent with the policy direction of clinical governance. Reporting at a higher level, for example at the level of Health Authorities, is also required and is an appropriate place to start for practical and statistical reasons. However, if information is only reported at this high level it is less likely to have a direct impact on the quality of patient care. Reporting at a lower level, such as Trust Directorates or Primary Care Groups, is statistically more difficult but will be effective. Reporting at the lowest level of individual doctors may be possible in selected cases, such as high volume surgeons, but does not take into account or promote team work and is methodologically difficult.

8.5 The financial cost of implementing a national policy on public disclosure is likely to be significant and should be considered alongside the benefits

An accurate assessment of the financial cost of public disclosure has not been conducted in the US but the resources required to

RECOMMENDATIONS

develop, measure, report and most importantly improve performance are likely to be significant. The opportunity cost of allocating resources to public disclosure in place of direct patient care needs to be defended. If public disclosure is regarded as a necessity in a public service, irrespective of its potential to promote quality improvement, the costs may not be considered to be a significant factor. If, however, the principal aim is to improve quality, public disclosure will have to be judged alongside other quality improvement strategies and a full and formal cost-benefit analysis should be conducted.

8.6 Specific educational initiatives for target audiences should be implemented alongside public disclosure

There is evidence from the US of a defensive response or a lack of response to performance data from all the stakeholder groups. The chances of a constructive response could be increased by informing and educating the target audiences through initiatives such as:

- Education of the public through the use of mass media.
- Education of all health professionals from the start of their basic training and as a component of a continuing professional education programme.
- Release of data as part of an educational package aimed at providers to promote quality improvement. This could include an explicit statement that the level of performance revealed by public disclosure should not only be seen as a function of the effectiveness of individual practitioners but also the team within which they function, the organisation within which they work and the resources available to them.

- Development of a strategy to promote greater collaboration and sharing of information amongst organisations.
- Proactive education of the media, which has proved to be an important component of successful public reporting programmes in the US.

8.7 Health professionals and their representative bodies should be fully involved in the process of public disclosure

Experience in both the UK and US highlights the importance of involving health professionals and their professional bodies in the selection, implementation, monitoring and evaluation of the indicators that are to be made public. Specific mechanisms or processes for active participation should be defined. For example, allowing professionals and their provider organisations a period of time, both to respond to performance data and to put mechanisms in place to improve performance prior to publication, is one way of integrating them into the process of public disclosure. Some reporting systems in the US gave providers a period of one year prior to the public release of the first data about a specific condition. Providers might be encouraged to send in written responses to the data, which would then be published alongside the performance reports.

8.8 Both process and outcome measures of quality should be published

Health outcomes are intuitively appealing but have inherent problems when used to measure and compare quality of care. Outcomes are often the result of factors outside the control of the health system and focusing on outcomes gives no insight into how providers can improve the processes of care. Some outcomes, for example mortality, occur infrequently in comparison with the

RECOMMENDATIONS

processes that prevent them and in health care there is often a long period of time between the action of a provider and the consequences of that action. The use of outcome measures is more applicable to some areas of practice than others but in general the use of process measures overcomes many of these problems. In particular the use of process measures can be justified when there is solid evidence that they are strongly linked to health outcomes. There is an increasing body of evidence that process measures are a more sensitive and more feasible measure of quality of care than outcome measures.

8.9 Outcome indicators must be risk adjusted

Comparisons of outcomes that are valid and credible can only be made if the sources of variation amongst providers that are not directly related to quality of care are removed by risk adjustment. The level of sophistication of the risk-adjustment mechanisms currently being used is highly variable. A balance must be achieved between complex systems (with associated implications for cost and feasibility) and little or no risk adjustment which may penalise those providers who accept high-risk patients, result in gaming of the system or reduce the credibility of the whole process of public disclosure. The level of risk adjustment should evolve alongside other aspects of public disclosure. Current experience suggests that the proposed adjustment of indicators in the National Performance Assessment Framework could be significantly improved upon by incorporating additional important risk factors, for example adjusting for social deprivation as one factor that influences emergency hospital admissions for asthma. Process quality measures may not need to be adjusted if they are constructed so that the patients to whom they are applied are described with precise clinical detail.

8.10 Public disclosure should be accompanied by a strategy for monitoring the benefits and unintended consequences

Public disclosure has both risks and unintended consequences. Published evidence of deficiencies in the care provided by professionals who already feel over-burdened can be demoralising and may adversely effect public trust in the health service. Misinterpretation of information, manipulation of data and an inappropriate focus on what is being measured, to the detriment of other areas of activity, have all been described. Some of these effects are inevitable but virtually all can be prevented, predicted or managed to optimise the benefits of public disclosure. The Commission for Health Improvement should play an important role in the monitoring, evaluation and policy assessment of public disclosure.

8.11 Public disclosure should be accompanied by possible explanations for the variations reported

It is inevitable that performance data will be of great public interest and may be misinterpreted or over-interpreted by the public, the media, health professionals and managers. This will have adverse consequences for the credibility and potential impact of future data. The risks could be reduced by accompanying performance reports with expert analysis and interpretation of the data. This commentary could then be used by providers as a catalyst for internal discussion and further action or could be used by government officials when addressing NHS resource allocation.

8.12 A research and development programme focusing on the generation and evaluation of public performance data should be supported by the NHS R&D Directorate

A policy on public disclosure is likely to be most effective if guided by empirical evidence of the associated merits and risks. The evidence is currently lacking, particularly in the UK, and would

RECOMMENDATIONS

benefit from a focused and adequately funded research and development programme. Information is needed about the content and presentation format of information most useful to consumers, providers and regulators, the impact of disclosure on professional morale and public trust in the NHS, the unintended consequences and the most appropriate risk-adjustment mechanisms. The introduction of the National Performance Assessment Framework provides a unique opportunity to provide experimental or quasi-experimental evidence of the relative merits of public disclosure versus the use of the same data for internal quality improvement purposes.

REFERENCES

Anderson G. *Multinational comparisons of health care: Expenditure, coverage and outcomes*. The Commonwealth Fund, New York; 1998

Anderson OW, Shields MC. Quality measurement and control in physician decision making: State of the art. *Health Services Research* 1982;17:125-155

Bentley JM, Nash DB. How Pennsylvania hospitals have responded to publicly released reports on coronary artery bypass graft surgery. *The Joint Commission Journal on Quality Improvement* 1998;24:40-49

Bero LA, Grilli R, Grimshaw JM, Harvey E, Oxman AD, Thomson MA on behalf of the Cochrane Effective Practice and Organisation of Care Review Group. Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. *BMJ* 1998;317:465-468

Bernstein SJ, McGlynn EA, Siu AL, Roth CP, Sherwood MJ, Keesey JW, Koseoff J, Hicks N, and Brook RH. The appropriateness of hysterectomy: A comparison of care in seven health plans. *JAMA* 1993;269:2398-2402

Bero L, Freemantle N, Grilli R, Grimshaw J, Oxman A. *Implementing the findings of medical research: a review of published research in the health sector*. Aberdeen, University of Aberdeen, 1995

Berwick DM. Continuous Quality Improvement as an ideal in health care. *NEJM* 1989;320:53-56

Berwick DM., Wald DL. Hospital leaders' opinions of the HCFA mortality data. *JAMA* 1990;263:247-249

Black N. High-quality clinical databases: breaking down barriers. *The Lancet* 1999;353:1205-1206

Blendon RJ. *Commonwealth Fund's 1998 International Health Policy Survey*, The Commonwealth Fund, New York;1998

REFERENCES

Blendon RJ, Brodie M, Benson JM, Altman DE, Levitt L, Hoff T, Hugick L. Understanding the managed care backlash. *Health Affairs* 1998;17: 80-110

Blumenthal D, Epstein AM. Quality of Health Care Part 6: The role of physicians in the future of quality management. *NEJM* 1996;335:1328-31

Borowsky SJ, Davis MK, Goertz C, Lurie N. Are all health plans created equal? The physician's view. *JAMA* 1997;278:917-921

Brook RH. Managed care is not the problem, quality is. *JAMA* 1997;278:1612-1614

Campbell MK, Grimshaw JM. Cluster randomised trials: time for improvement. The implications of adopting a cluster design are still largely being ignored [editorial]. *BMJ* 1998; 317:1171-2

Chassin MR, Hannan EL, DeBuono BA. Benefits and hazards of reporting medical outcomes publicly. *NEJM* 1996;334:394-398

Codman EA. *A study in hospital efficiency* (Boston, 1917), reprinted by the Joint Commission on Accreditation of Healthcare Organisations, Oakbrook Terrace, Illinois, 1996;53

Corrigan JM. How do purchasers develop and use performance measures? *Medical Care* 1995;33:JS18-JS24

Davies HTO, Crombie IK. (Ed) Assessing the quality of care. *BMJ* 1995;311: 766

Davies HTO, Lampel J. Trust in performance indicators. *Quality in Health Care* 1998;7:159-162

Davis DA, Thomson MA, Oxman AD, Haynes B. Changing physician performance: A systematic review of the effect of continuing medical education strategies. *JAMA* 1995; 274:700-705

Dillner L. Scottish death rates published with health warning. *BMJ* 1994;309:1599-1600

Edgman-Levitan S, Cleary PD. What information do consumers want and need? *Health Affairs* 1996;15:42-56

Entwistle VA, Sheldon TA, Sowden AJ, Watt IS. Supporting consumer involvement in decision making: what constitutes quality in consumer health information? *International Journal for Quality in Health Care* 1996;8:425-437

Epstein A. Performance reports on quality - prototypes, problems and prospects. *NEJM* 1995;333:57-61

Epstein A. Rolling down the runway; the challenges ahead for quality report cards. *JAMA* 1998;279:1691-1696

Freidson E. *Profession of medicine: A study of the sociology of applied knowledge*. Harper and Row, New York; 1970

Gabel JR, Hunt KA, Hurst K, KPMG Peat Marwick. *When employers choose health plans: Do NCQA accreditation and HEDIS data count?* The Commonwealth Fund, New York;1998

GAO. *Health care reform: "Report cards" are useful but significant issues need to be addressed*. United States General Accounting Office. GAO/HEHS-94-219;1994

GMC. *Good Medical Practice*. General Medical Council, 1998

Goddard M, Mannion R, Smith PC. *The NHS Performance Framework: Taking account of economic behaviour*. The York series on the NHS White Paper - a research agenda. Centre for Health Economics, York; 158;1998

Gold MR, Hurley R, Lake T, Ensor T, Berenson R. A national survey of the arrangements managed care plans make with physicians. *NEJM* 1995;333:1678-1683

REFERENCES

- Green J, Wintfeld N. Report cards on cardiac surgeons; assessing New York State's approach. *NEJM* 1995;332:1229-1232
- Grol R. Beliefs and evidence in changing practice. *British Medical Journal* 1997;315:418-421
- Hannan EL, Kilburn H, Lindsey ML, Lewis R. Clinical versus Administrative Data Bases for CABG Surgery. Does it Matter? *Medical Care* 1992;30:892-907
- Hannan EL, Kilburn H, Racz M, Shields E, Chassin MR. Improving the outcomes of coronary artery bypass surgery in New York State. *JAMA* 1994;271:761-766
- Hannan EL, Siu AL, Kumar D, Kilburn H, Chassin MR. The decline in coronary artery bypass graft mortality in New York State. *JAMA* 1995;273:209-213
- Hannan EL, Siu AL, Kumar D, Racz M, Pryor DB, Chassin MR. Assessment of coronary artery bypass graft surgery performance in New York; is there a bias against taking high-risk patients? *Medical Care* 1997a;35:49-56
- Hannan EL, Stone CC, Biddle TL, DeBuono BA. Public release of cardiac surgery outcomes data in New York: What do New York State cardiologists think of it? *American Heart Journal* 1997b;134:55-61
- Hannan EL. Measuring hospital outcomes: don't make perfect the enemy of good! /. *Health Serv. Res. Policy* 1998;3:67-69
- Hibbard JH, Jewett JJ. Will quality report cards help consumers? *Health Affairs* 1997;16:218-228
- Hibbard JH, Weeks EC. Does the dissemination of comparative data on physician fees affect consumer use of services? *Medical Care* 1989;27:1167-1174

Hibbard JH, Slovic P, Jewett JJ. Informing consumer decisions in healthcare: Implications from decision-making research. *The Millbank Quarterly* 1997a;75:395-414

Hibbard JH, Jewett JJ, Legnini MW, Tusler M. Choosing a health plan: Do large employers use the data? *Health Affairs* 1997b;16:172-180, 1997b.

Hibbard JH, Jewett JJ, Engelmann S, Tusler M. Can Medicare beneficiaries make informed choices? *Health Affairs* 1998;17:181-193

Hofstede G. Management control of public and not-for-profit activities. *Accounting, Organisations and Society* 1981;6:193-211

Institute of Medicine. *Health Data in the Information Age; Use, Disclosure and Privacy*. National Academy Press, Washington D.C.;1994

Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernan YD. Predicting who dies depends on how severity is measured. *Ann Intern Med.* 1995;123:763-770

Iezzoni LI. The Risks of Risk Adjustment. *JAMA* 1997a;278:1600-1607

Iezzoni LI. *Risk Adjustment for Measuring Health Care Outcomes*. 2nd Edition. Health Administration Press, Chicago, IL; 1997b

Jewett JJ, Hibbard JH. Comprehension of Quality of Care Indicators: Differences among the Privately Insured, the Publicly Insured and the Uninsured. *Health Care Financing Review* 1996;18.1:75-94

Jollis JG, Romano PS. Pennsylvania's Focus on Heart Attack - Grading the scorecard. *NEJM* 1998;338:983-987

Kassirer JP. The use and abuse of practice profiles. *NEJM* 1994;330:634-635

Kendrick S, Cline D, Finlayson A. *Clinical Outcome Indicators in Scotland: Lessons and prospects*. Presented at the Third International Conference on Strategic Issues in Health Care Management, University of St Andrews;1998

REFERENCES

- Kerry SM, Bland JM. The intracluster correlation coefficient in cluster randomisation. *BMJ* 1998a;316:1455
- Kerry SM, Bland JM. Sample size in cluster randomisation. *BMJ* 1998b; 316:549
- Lansky D. The new responsibility: Measuring and reporting on quality. *Journal on Quality Improvement* 1993;19:545-551
- Lansky D. Measuring what matters to the public. *Health Affairs* 1998;17:40-41
- Longo DR, Land G, Schramm W, Fraas J, Hoskins B, Howell V. Consumer reports in health care; Do they make a difference in patient care? *JAMA* 1997;278:1579-1584
- McColl A, Roderick P, Gabbay J, Smith H, Moore M. Performance indicators for primary care groups: An evidence based approach. *BMJ* 1998;317:1354-1360
- McGlynn EA, Damberg C, Kerr E, Schuster M. *Quality of Care for Children and Adolescents: A Review of Selected Clinical Conditions and Quality Indicators*. Report No. DRU-1721-HCFA. RAND, Santa Monica;1997
- McGlynn EA, Asch SM. Developing a Clinical Performance Measure. *American Journal of Preventative Medicine* 1998;14(3S):14—21
- McGlynn EA, Damberg CL, Kerr EA, Brook RH. Health Information Systems: design issues and analytical applications. Santa Monica: RAND, 1998
- Mant D, Hicks N. Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction. *BMJ* 1995, 311:793-796
- Marshall MN. Improving quality in general practice: qualitative case study of barriers faced by health authorities. *BMJ* 1999;319:164-167

Marshall MN, Shekelle PG, Leatherman S, Brook RH and Nuffield Trust Quality Steering Group. Policy recommendations on public disclosure. Presented to Secretary of State for Health, August 1999, Nuffield Trust, London.

Mennemeyer ST, Morrisey MA, Howard LZ. Death and Reputation: How consumers acted upon HCFA mortality information. *Inquiry* 1997;34:117-128

Mukamel DB, Mushlin AI. Quality of Care Information Makes a Difference: An Analysis of Market Share and Price Changes After Publication of the New York State Cardiac Surgery Mortality Reports. *Medical Care* 1998;36:945-954

NHS Executive, *Priorities and Planning guidance 1996/7*, NHS Executive, London;1995

NHS Executive, *The new NHS. Modern Dependable: A National Framework for Assessing Performance. Consultation Document*. NHS Executive, London; 1998

Nightingale F. *Notes on hospitals. 3rd ed*. Longman, Green, Longman, Roberts and Green, London; 18 63

O'Brien JL, Shortell SM, Hughes EF et al. An Integrative Model for Organisation-wide Quality Improvement: Lessons from the Field. *Quality Management in Health Care* 1995;3:19-30

O'Connor GT, Plume SK, Olmstead EM, Morton JR, Maloney CT. A regional intervention to improve the hospital mortality associated with coronary artery bypass graft surgery. *JAMA* 1996;275:841-846

Omoigui NA, Miller DP, Brown KJ, Annan K, Cosgrove D, Lytle B, Loop F, Topol EJ. Outmigration for coronary bypass surgery in an era of public dissemination of clinical outcomes. *Circulation* 1996;93:27-33

Peterson ED, DeLong ER, Jollis JG, Muhlbaier LH, Mark DB. The effects of New York's bypass surgery provider profiling on access to care and patient outcomes in the elderly. / *Am Coll Cardiol* 1998;32:993-999

REFERENCES

- Rainwater JA, Romano PS., Antonius DM. The California Hospital Outcomes Project: How useful is California's report card for quality improvement? *The Joint Commission Journal on Quality Improvement* 1998;24:31-39
- Reinhardt UE. *Recent developments in American health policy: from "unmanaged care" to "managed costs"* The Commonwealth Fund International Symposium on Health Care Policy; Costs Quality and Equity: Current Issues and Future Prospects. Briefing documents. Commonwealth Fund, New York; 1998
- Richards T, Blacketer B, Rittenhouse D. *Statewide, Metropolitan, Corporate and National Efforts in Monitoring and Reporting Quality of Care*. California Office of Statewide Health Planning and Development, California; 1994
- Robinson S, Brodie M. Understanding the quality challenge for health consumers: The Kaiser/AHCPR survey. *Journal on Quality Improvement* 1997;23:239-244
- Romano PS, Zach A, Luft HS, Rainwater J, Remy LL, Campa D. The California Hospitals Outcomes Project: Using administrative data to compare hospital performance. *The Joint Commission Journal on Quality Improvement* 1995;21:668-682
- Rosenthal GE, Hammar PJ, Way LE, Shipley SA, Doner D, Wojtala B, Miller J, Harper DL. Using hospital performance data in quality improvement: The Cleveland Health Quality Choice experience. *The Joint Commission Journal on Quality Improvement* 1998;24:347-360
- Schneider EC, Epstein AM. Influence of cardiac-surgery performance reports on referral practices and access to care; a survey of cardiovascular specialists. *NEJM* 1996;335:251-256
- Schneider EC, Epstein AM. Use of public performance reports: a survey of patients undergoing cardiac surgery. *JAMA* 1998;279:1638-1642

Schneidman D. The public dissemination of physician-specific mortality data: Who has a right to know? *American College of Surgeons Bulletin* 1993;78:32-37

Shaw A, Somerset M, Dunn L, Peters T, Faulkner A, Sharp D. Can we trust the quality of routine hospital outpatient information in the UK? *Journal of Health Services Research and Policy* 1998;3:203-206

Sheldon T. Promoting health care quality: what role performance indicators? *Quality in Health Care* 1998;7(Suppl):S45-S50

Shortell SM, Bennett CL, Byck GR. Assessing the impact of Continuous Quality Improvement on clinical practice: What it will take to accelerate progress. *The Millbank Quarterly* 1998;76:593-624

Sirio CA, McGee JL. Public reporting of clinical outcomes - The data needs of health care stakeholders. *American Journal of Medical Quality* 1996;11:S78-S81

Smith P. On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration* 1995;18:277-310

Thomson RG, McElroy H, Kazandjian VA. Maryland Hospital Quality Indicator Project in the United Kingdom: an approach for promoting continuous quality improvement. *Quality in Health Care* 1997;6:49-55

Thomson R. Quality to the fore in health policy - at last (Editorial). *BMJ* 1998; 317:95-96

Topol EJ, Califf RM. Scorecard cardiovascular medicine; Its impact and future directions. *American College of Physicians* 1994;120:65-70

Vladeck BC, Goodwin EJ, Myers LP, Sinisi M. Consumers and hospital use: The HCFA "death list". *Health Affairs* 1988;7:122-125

REFERENCES

Winslow CM, Kosecoff J, Chassin MR, Kanouse DE, Brook RH. The appropriateness of performing coronary artery bypass surgery. *JAMA* 1988;260:505-9

Ziegenfuss JT. Healthcare quality report cards receive grade-incomplete (Editorial). *American Journal of Medical Quality* 1996;11:55-56

This book summarises a larger body of work conducted by the authors relating to the public disclosure of comparative performance data. Readers might wish to access the following additional information published in peer reviewed scientific journals.

Davies HTO, Marshall MN. Public disclosure of performance data: does the public get what the public wants? *The Lancet* 1999; 353: 1639-1640

Marshall MN. Time to go public on performance? (editorial). *British Journal of General Practice* 1999; 49: 691-692

Leatherman S, McCarthy D. Public Disclosure of health care performance reports: experience, evidence and issues for policy. *International Journal for Quality in Health Care* 1999; 11: 93-105

Marshall MN, Shekelle PG, Leatherman S, Brook RH. Public reporting of performance: lessons from the USA. *Journal for Health Services Research and Policy* 2000; 5: 1-2

Marshall MN, Shekelle PG, Leatherman S, Brook RG. Public disclosure of performance data: learning from the US experience. *Quality in Health Care* 2000; 9: 53-57

Marshall MN, Shekelle PG, Leatherman S, Brook RH. What do we expect to gain from the public release of performance data? A review of the evidence. *JAMA* 2000; 283: 1866-1874

Marshall MN. Turning up the heat on performance. *British Journal of General Practice* 2000; 50: 419

APPENDIX 1

Table 1.1. Frequency of each type of condition or procedure studied in report cards (Richards et al., 1994)

Medical condition or surgical procedure	Number of studies
In-patient Health Care	
caesarean delivery	9
coronary artery bypass graft surgery	8
acute myocardial infarction	7
vaginal delivery	6
cholecystectomy	5
vaginal birth after caesarean section	5
asthma	4
chronic obstructive pulmonary disease	4
congestive heart failure	4
hysterectomy	4
pneumonia	4
stroke	4
total hip replacement	4
trauma	4
heart transplant	4
prostatectomy	3
transient ischaemic attack	3
angioplasty	3
carpal tunnel syndrome	2
cardiac catheterization	2
gastrointestinal resections	2

APPENDIX 1

Table 1.1. Frequency of each type of condition or procedure studied in report cards (Richards et al., 1994) (Cont.)

laminectomy	2
unstable angina	2
adverse drug reactions	1
appendectomy	1
carotid endarterectomy	1
cataract removal	1
cervical and lumbar disc excision	1
computerised axial tomography contrast media	.
diabetes	1
fracture reduction	1
gastrointestinal haemorrhage	1
hip fracture	1
influenza	1
initial pacemaker insertion	1
intensive care patients	1
knee replacement	1
lumbar spine surgery	1
open reduction of hip fracture	1
post-operative infections	1
sepsis	1
percutaneous transluminal coronary angioplasty	

Table 1.1. Frequency of each type of condition or procedure studied in report cards (Richards et al., 1994) (Cont.)

thoracic aneurysm	1
tonsillectomy	1
valve surgery	1
vascular repairs	1

Outpatient Health Care

asthma	4
prenatal care	4
immunisations and vaccines	3
lower back pain	3
mental health	3
hypertension	2
breast cancer	2
cervical cancer	2
colorectal cancer	1
diabetes	1
otitis media	1

APPENDIX 2 - EVIDENCE TABLES

Table 2.1 Attitude of physicians to data release

Studies	Design	Public data being studied	Aim	Setting and subjects	Results	Comments
Hannan et al, 1977b	descriptive - postal survey	New York CSRS - risk adjusted CABG mortality	to assess use of data by referring cardiologists	NY state - all cardiologists belonging to American College of Cardiology	94% found reports easy to read, 67% regarded content as very/somewhat useful, 33% not at all accurate, 22% routinely discussed data with patients, 38% said data had very/somewhat affected referral patterns. Concern expressed that data deterred operation on high risk patients	low response rate, self reports
Schneider and Epstein, 1996	descriptive - postal survey	Pennsylvania PHCA - risk adjusted CABG mortality	to assess awareness and views of cardiologists and cardiac surgeons of data	Pennsylvania - random 50% of cardiologists and cardiac surgeons in state	82% of cardiologists and all surgeons aware of report, 10% perceived it to be very important, less than 10% discussed it with more than 10% of patients, 87% of cardiologists said it had little influence on referral practice, 2% had significant influence. Absence of outcomes other than mortality, inadequate risk adjustment, unreliable data were main criticisms. 59% of cardiologists and same % of surgeons reported patient access problems.	self reports
Vladeck et al, 1988	pre-post	HCFA hospital-specific mortality rates for all Medicare patients - min. risk adjustment	effect of data on bed occupancy	NY state - all NY general acute hospitals	no statistical difference after publication	
Borowsky et al, 1997	descriptive - telephone survey	N/A	to determine physicians views on plans and compare ratings of plans	Minneapolis St Paul - 100 physicians in each of 3 plans	<20% gave plans highest rating for delivery of high quality care, main problems were insufficient time with patients, benefits and co-payments, utilisation management techniques. Significant difference in way physicians rated different plans	small sample, forced answers to complex questions

Table 2.2 Impact on behaviour of hospitals and other providers

Studies	Design	Public data being studied	Aim	Setting and subjects	Results	Comments
Bentley and Nash, 1998	descriptive - key informant interviews to develop questionnaire, then postal survey	Pennsylvania PHC4 - risk adjusted CABG mortality	effect of consumer guide on hospital behaviour	Pennsylvania and New Jersey - random sample of hospitals, payers and purchasers	organisations put extra efforts into marketing products as result of publication, more likely to monitor clinicians performance, use other hospitals as benchmark and promotes better collaboration within organisation	pilot, small sample, no statistical analysis
Longo et al, 1997	pre-post	obstetrics consumer report	examine impact of published report on hospital behaviour	Missouri - all hospitals providing obstetric care	50% of hospitals implemented process improvements as result of publication, esp. if in competitive environment. All outcomes improved - satisfaction, appropriateness of charges, Caesarian rate, high risk infant transfer rate, vaginal delivery after caesarian rate, very low birth rate and new born death rate.	study design means causal relationship between publication and outcomes is not certain
Rainwater et al, 1998	descriptive - telephone survey	CHOP risk adjusted AMI mortality.	to assess acceptability of report cards to hospitals	sample of all acute care hospitals in California	75% found some aspect useful, esp for benchmarking, improving coding and educating physicians. Timeliness and emphasis on mortality criticised	small sample, viewpoint of hospital leaders only
Rosenthal et al, 1998	descriptive - case studies	Cleveland Health Quality Choice - multiple adjusted and unadjusted hospital outcomes	to determine whether public data changes hospital practice	Cleveland hospitals - no detail provided	reported changed processes of care resulting in some improved outcomes	poorly described case studies
Berwick et al, 1990	descriptive - postal survey	HCFA hospital-specific mortality rates for all Medicare patients - min risk adjustment	to determine reaction of hospital leaders to data release	2 hospitals from each state with higher and lower than expected mortality	negative views about utility, accuracy. Reported problems with data. High mortality hospitals more likely to use data.	appropriateness of postal survey questionable

Table 2.3 Information individual consumers want to access

Studies	Design	Public data being studied	Aim	Setting and subjects	Results	Comments
Edgman-Levitan and Cleary, 1996	descriptive - patient surveys and focus groups	N/A	to investigate consumer information needs	setting not described. Medicare enrollees	request broad range of information, question accuracy of currently available data, greater weight given to informal sources of information	minimal details of methods
Robinson and Brodie 1997	descriptive - telephone survey	N/A	to investigate role of data in consumer health care decision making	setting not described 2006 adults	reliance on family, friends and referring physician for information. 39% saw information on health plans and 33% of these used it. Regard information as not primarily aimed at consumer	minimal details of methods

Table 2.4 Impact on decision making of individual consumers

Studies	Design	Public data being studied	Aim	Setting and subjects	Results	Comments
Hibbard and Weeks, 1989	Experimental - RCT	cost data	to determine whether cost information on medical encounter influences consumer choice	Salem, OR, 717 medicare enrollees, 658 state employees randomly allocated	cost has no statistically significant impact on utilization	potential contamination of intervention and control groups
Mulkam and Mushlin, 1998	pre-post	New York CSRS risk adjusted CABG mortality	to determine impact of data on hospital and physician market share and physician charges	New York State, all 30 hospitals performing CABG, 80% of surgeons, FFS patients only	hospitals and physicians with better outcomes have higher rates of growth or market share and physicians have higher rate of growth of charges	no control for confounders
Scheider, Epstein, 1998	descriptive - tele-phone survey	Pennsylvania PHCA - risk adjusted CABG mortality	to examine awareness and use of consumer guide amongst post-CABG patients	4 Pennsylvania hospitals, 2 with lower than expected mortality, 1 middle rank, 1 higher. 474 of 673 patients who had undergone bypass in previous year	20% (56) aware of guide, 18 knew hospital rating and 7 surgeon ratings, 11 said information had moderate/major impact. 6 discussed ratings with physician, younger age, college education, higher pre-operative health status and heart disease > 1 year associated with increased awareness. On describing report 56% very/somewhat interested in seeing it, 58% would probably change surgeon as result, low level of willingness to pay for report	fails to address complexity of decision making process
Vladeck et al, 1988	pre-post	HCFA hospital-specific mortality rates for all Medicare patients - min risk adjustment	effect of data on bed occupancy	NY state - all NY general acute hospitals	no statistical difference after publication	no controls
Mennemeyer et al, 1997	pre-post	HCFA hospital-specific mortality rates for all Medicare patients - min risk adjustment	compare HCFA data with press reports of disasters on hospital utilization	all hospitals treating medicare patients	small but significant impact of HCFA data, large impact of press reports	

Table 2.5 Explanation of impact on decision making

Studies	Design	Public data being studied	Aim	Setting and subjects	Results	Comments
Hibbard et al 1998	descriptive - telephone survey	N/A	to determine whether medicare beneficiaries understand the difference between managed care and fee for service	5 geographical areas across the US with high penetration of managed care. 1673 medicare enrollees	30% knew almost nothing about HMOs, only 11% had adequate knowledge to make informed decisions. 1673 decisions.	
Hibbard and Jewitt, 1997	descriptive - patient surveys and focus groups	N/A	to assess relationship between importance of information and how well it is understood	not described	poorly understood indicators not perceived to be important, poor understanding of health care system	setting and details of methods
Robinson and Brodie, 1997	descriptive - telephone survey	N/A	to investigate role of data in consumer decision making	setting not described 2006 adults	reliance on family, friends and referring physician for information. 39% saw information on health plans and 33% of these used it. Regard not information as not primarily aimed at consumer	described
Jewitt and Hibbard, 1996	descriptive	N/A	to determine understanding of quality of care indicators amongst privately insured, publicly insured and uninsured	not described	indicators poorly understood, unclear whether high or low ratings are good, different health beliefs, aggregated indicators cause particular problems	No comment
Scheider and Epsrein, 1998	descriptive - telephone survey	Pennsylvania PHCA - risk adjusted CABG mortality	to examine awareness and use of consumer guide amongst post-CABG patients	4 Pennsylvania hospitals, 2 with lower than expected mortality, 1 middle rank, 1 higher. 474 of 673 patients who had undergone bypass in previous year	20% (56) aware of guide, 18 knew hospital rating and 7 surgeon rating; 11 said information had moderate/major impact on decision. 6 discussed ratings with physician, younger age, college education, higher pre-operative health status and heart disease > 1 year associated with increased awareness. On describing report very/somewhat interested in seeing it, 58% would probably change surgeon as result, low level of willingness to pay for reports	fails to address complexity of decision making process

Table 2.6 Information wanted by employers as purchasers

Studies	Design	Public data being studied	Aim	Setting and subjects	Results	Comments
Gabel et al, 1998	descriptive - surveys	HEDIS data, NCCQA accreditation	to determine impact of NCCQA accreditation and HEDIS data on employer choice of plans	National, random 1502 employers with more than 200 workers	% of employers familiar with data increased from 29-35% over 2 years, used more by larger employers, 11% rated NCCQA data and 5% HEDIS data as very important. 9% required NCCQA accreditation for selection and 1% provided HEDIS data to their employees. Fully accredited plans cost 4% less than non-accredited plans.	Minimal details of methods
Gold et al, 1995	descriptive - telephone survey	physician specific utilization and certification data	to describe recruitments, compensation, risk-sharing and oversight of physicians by Managed Care Organizations	National, 20 metropolitan areas. 138 Managed Care plans	used data to choose physicians, board certification important, 13% stated utilization data had no impact on choice, 26% moderate influence, 61% little influence	
Hilbard et al, 1997b	descriptive - indepth inter-views	HEDIS data	to determine whether large employers use HEDIS data to guide purchasing decisions	Pennsylvania, Cleveland, CA and NY, interviews with 8-9 key informants from large employers in each area, responsible for 1.8 million lives	78% reported HEDIS data was available to them, 75% patient satisfaction data, 54% used HEDIS data to choose plans, 59% used satisfaction data. Barriers to use were concern about bias, quality of care not their responsibility, information inaccessible, unclear about significance, data overload. Prefer process to outcomes data.	Poor response rate and sampling strategy

Table 2.7 Impact on employers' purchasing decisions

Studies	Design	Public data being studied	Aim	Setting and subjects	Results	Comments
Hibbard et al, 1997b	descriptive - in-depth interviews	HEDIS data	to determine whether large employers use HEDIS data to guide purchasing decisions	Pennsylvania, Cleveland, CA and NY, interviews with 8-9 key informants from large employers in each area, responsible for 1.8 million lives	78% reported HEDIS data was available to them, 75% patient satisfaction data and 25-71% outcome data, 54% used HEDIS data to choose plans, 59% used satisfaction data. Barriers to use were concern about bias, quality of care not their responsibility, information inaccessible, unclear about significance, data overload. Prefer process to outcomes data.	Poor response rate and sampling strategy
Gabel et al, 1998	descriptive - surveys	HEDIS data, NCQA accreditation	to determine impact of NCQA accreditation and HEDIS data on employer choice of plans	National, random 1502 employers with more than 200 workers	% of employers familiar with data increased from 29-35% over 2 years, used more by larger employers, 11% rated NCQA data and 5% HEDIS data as very important. 9% required NCQA accreditation for selection and 1% provided HEDIS data to their employees. Fully accredited plans cost 4% less than non-accredited plans.	Minimal details of methods

Table 2.8 Impact on quality of care outcomes

Studies	Design	Public data being studied	Aim	Setting and subjects	Results	Comments
Hannan et al, 1994	pre-post	New York CSRS - risk adjusted CABG mortality	to assess changes in mortality following publication of mortality data	all 30 NY State hospitals performing bypass operations, 57,187 patients undergoing CABG from 1989-92	actual mortality decreased from 3.52% - 2.78%, risk adjusted mortality from 4.17% - 2.45%	
Peterson et al, 1998	pre-post	New York CSRS - risk adjusted CABG mortality	to examine effects of provider profiling on recruitment and outcomes of bypass surgery	NY state, medicare beneficiaries 65 years and over	proportion of NY state residents receiving out of state surgery declined, likelihood of post-MI bypass increased. Unadjusted 30 day mortality declined by 33% in NY and 19% nationwide	no reference to change of provider within state, data only to 1992
Longo et al, 1997	pre-post	obstetrics consumer report	examine impact of report on hospital behaviour	Missouri - all hospitals providing obstetric care	50% of hospitals implemented process improvements as result of publication, especially if in competitive environment. All outcomes improved - satisfaction, appropriateness of changes, Caesarian rate, high risk infant transfer rate, vaginal delivery after caesarian rate, very low birth rate and new born death rate.	Study design means causal relationship between publication and outcomes can not be definitely established

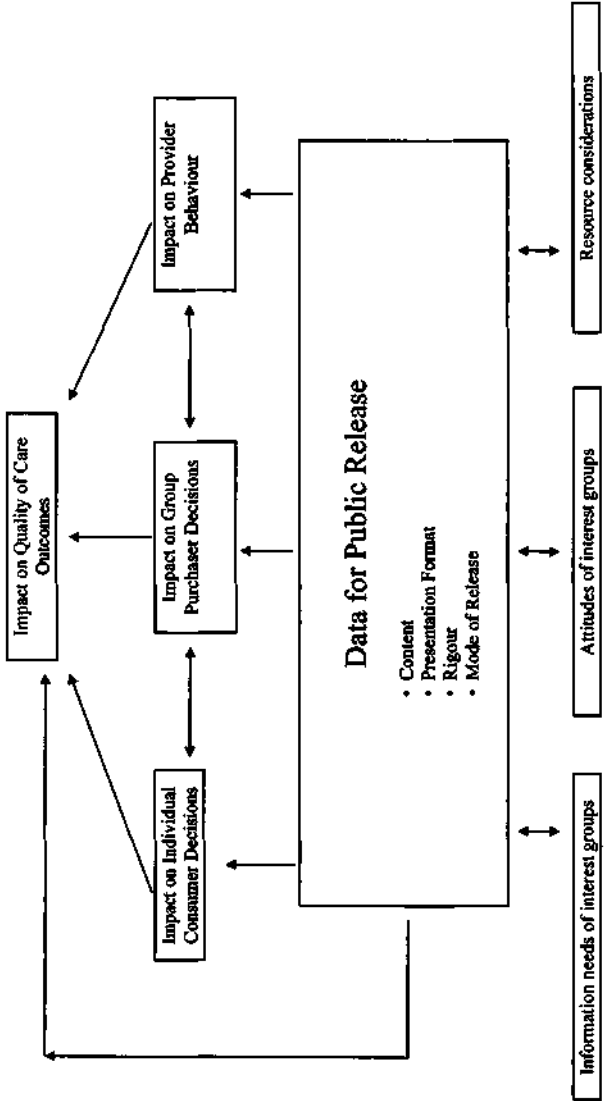
Table 2.9 Mechanisms of action of performance data

Studies	Design	Public data being studied	Aim	Setting and subjects	Results	Comments
Hannan et al, 1995	observational	New York CSRS - risk adjusted CABG mortality	to explain mechanisms by which surgeon volume of CABG operations influences risk adjusted mortality rates	all 30 NY State hospitals performing bypass operations, 57,187 patients undergoing CABG from 1989-92	low volume surgeons had greater reductions in risk adjusted mortality than high volume surgeons (60 v 34%). % of operations performed by low volume surgeons reduced by 2.5%	explanation accounts for only small part of total observed improvement in mortality

Table 2.10 Impact on costs

Studies	Design	Public data being studied	Aim	Setting and subjects	Results	Comments
Bentley and Nash, 1998	descriptive - key informant inter-views to develop questionnaire, then postal survey	Pennsylvania PHCA - risk adjusted CABG mortality	effect of consumer guide on hospital behaviour to determine impact of NCOA accreditation and HEDIS data on employer choice of plans	Pennsylvania and New Jersey - random sample of hospitals, payers and purchasers	organisations put extra efforts into marketing products as result of publication, more likely to monitor clinicians performance, use other hospitals as benchmark and promotes better collaboration within organisation	pilot, small sample, no statistical analysis
Gabel et al, 1998	descriptive - surveys	HEDIS data, NCOA accreditation	to determine impact of NCOA accreditation and HEDIS data on employer choice of plans	National, random 1502 employers with more than 200 workers	% of employers familiar with data increased from 29-35% over 2 years, used more by larger employers, 11% rated NCOA data and 5% HEDIS data as very important. 9% required NCOA accreditation for selection and 1% provided HEDIS data to their employees. Fully accredited plans cost 4% less than non-accredited plans.	Unclear methodology
GAO, 1994	review of evidence	N/A	overview of role of report cards, including possible costs of data collection	N/A	cost of data collection from clinical record approximately \$16 per hospital record at 1991 prices	no formal cost analysis

APPENDIX 3 - Conceptual model of public disclosure



APPENDIX 4 - Summary of reporting systems that have been subject to evaluation

Name	Subject	Data source	Reporting Risk Level	Evaluation
California Hospitals Outcome Project (CHOP)	1. Acute myocardial infarction	30 day in- routinely collected administrative discharge data	hospital	Rainwater et al., 1998 Aim: to assess acceptability to hospitals
	2. patient mortality		yes - developed own system	
	3. readmission following delivery mortality following hip fracture		2. yes - APACHE III	
			3. yes - developed own system	Rosenthal et al., 1998. Aim: to determine whether publication of data changed hospital practice
Cleveland Health Quality Choice	1. Patient satisfaction	1. postal survey of random sample of patients	hospital	
	2. mortality and length of stay in Intensive Care Units	2. review of medical records		
	3. mortality and length of stay for selected medical and surgical conditions	3. as for 2		
	4. cesarean section rates in obstetric units	4. as for 2		
Health Care Financing Administration (HCFA) hospital mortality	30 day mortality rates of medicare patients after hospital admission	routinely collected administrative discharge data	hospital	Vladeck et al., 1988. Aim: to determine effect on bed occupancy. Berwick and Wald, 1990. Aim: to determine reaction of hospital leaders to data release Menneneyer et al., 1997. Aim: to compare impact of data with impact of press reports of one-off disasters on hospital discharges
			yes - Medicare Mortality Predictor System	

Name	Subject	Data source	Reporting Risk Level	Evaluation
	- cervical and breast cancer screening			
	- prenatal care in 1 st trimester			
	- low birth weight infants			
	- follow up after hospital admission for mental illness			
Health Plan Employer Data and Information Set (HEDIS version 3.0)	- childhood and adolescent immunization status	mostly administrative data, some from clinical records	no - mostly process measures of quality	Gabel et al, 1998 Aim: to determine impact on employer choice of plan Hubbard et al, 1997 Aim: to determine use by large employers
	- eye exams for diabetics			
	- advising smokers to quit			
	- E blocker treatment post.MI			
	- check up after delivery			
	- treating children's ear infections			
	- flu shots for older adults			
	- health of seniors			
	- availability of obstetrics services			
	- patient satisfaction			
	- appropriate charges			
Missouri Obstetrics report	- cesarean section rates	administrative data and telephone survey	ultrasound rates only, others not adjusted	Longo et al, 1997 Aim: to examine impact on hospital behaviour
	- high risk infant transfer			
	- ultrasound rates			
	- vaginal birth after cesarean section			
	- very low birth rate			
	- new born death			

APPENDIX 4 - Summary of reporting systems that have been subject to evaluation

Name	Subject	Data source	Reporting Risk Level	Risk adjustment	Evaluation
					Hannan et al, 1997 Aim: to assess use of data by referring physicians
					Hannan et al, 1994 Aim: to assess impact on mortality rates
New York State Cardiac Surgery Reporting System (CSRS)	in hospital mortality following coronary artery bypass surgery	specialty collected clinical and administrative data	hospital and individual surgeon	yes - highly detailed	Peterson et al, 1998 Aim: to assess impact on recruitment and mortality rates Hannan et al, 1995 Aim: to explain possible mechanisms of action Mukamel and Mushlin, 1998 Aim: to determine impact on hospital and physician market share and physician charges
Pennsylvania Health Care Cost Containment Council (PHCC)	in hospital mortality following coronary artery bypass surgery	specialty collected clinical and administrative data	hospital and individual surgeon	yes - highly detailed	Schneider and Epstein, 1996 Aim: to assess attitude of cardiac surgeons and cardiologists to data Schneider and Epstein, 1998 Aim: to examine awareness and views of consumer guide amongst post-CABG patients Bentley and Nash, 1998 Aim: to assess impact on hospital behaviour

APPENDIX 5 - 1996 Californian Hospitals discharge rate. DRG ranking by number of discharges (aged >18 years)

Rank	DRG	No. of discharges	Mortality rate
1	vaginal delivery	278050	0.0%
2	pneumonia	75750	5.6%
3	congestive heart failure	73743	4.8%
4	cesarean section	62432	0.0%
5	stroke	57259	7.5%
6	psychoses	51890	0.0%
7	miscellaneous digestive disorders	43389	0.4%
8	chronic obstructive, airway disease	36457	1.5%
9	myocardial infarction	35322	10.5%
10	gastrointestinal bleed	34624	2.4%
11	chest pain	32564	0.0%
12	cardiac arrhythmia	29599	1.9%
13	nutritional/metabolic disorder	28395	2.4%
14	septicaemia	24848	13.3%
15	trauma	23475	6.1%
16	atherosclerosis	22552	0.4%
17	percutaneous cardiac procedure	20607	0.9%
18	renal tract infection	18765	1.4%
19	asthma	17851	0.1%
20	angina pectoris	17730	0.2%
21	diabetes	16946	0.9%
22	circulatory disorder	15997	0.7%
23	respiratory disease with ventilation	15184	31.7%
24	prenatal hospitalisation	14747	0.0%
25	pancreatic disorder except malignancy	13951	1.3%
26	HIV	13148	9.9%
27	cellulitis	13024	0.5%
28	major bowel procedures	12926	6.6%
29	peripheral vascular disorder	12789	2.9%
30	hip/femur procedure	12019	2.0%

APPENDIX 6 - 1994/5 English hospital services bed days and average length of stay (days) for leading causes per 10,000 population

Rank	Condition	Bed days	Length of stay
1	Mental retardation	1983	223
2	Schizophrenic psychoses	922	130
3	Organic psychoses	620	66
4	Cerebro-vascular disease	429	27
5	Affective psychoses	344	48
6	Other psychoses	293	33
7	Fracture neck of femur	242	20
8	Arthropathies excluding rheumatoid arthritis	228	12
9	Pneumonia	212	14
10	Osteoarthritis and allied disorders	203	13
11	Neurotic and personality disorders	203	30
12	Acute myocardial infarction	168	8
13	Dorsopathies excluding ankylosing, spondylitis	105	8
14	Abdominal pain	105	3
15	Bronchitis, emphysema, asthma	96	4
16	Malignancies of trachea, bronchus and lung	86	9
17	Cardiac dysrhythmias	82	6
18	Fracture tibia and ankle	82	10
19	Rheumatism, excluding back	79	5
20	Other joint disorders	71	5

APPENDIX 7 - ILLUSTRATIVE EXAMPLES OF PERFORMANCE INDICATORS FOR THE UNITED KINGDOM

The purpose of this appendix is to illustrate some of the practical implications of developing valid comparative performance measures suitable for public disclosure. Examples will demonstrate the importance of the sample size required to determine statistically significant differences in the quality of care provided at the level of the health authority, Primary Care Group, hospital and individual practitioner, for different quality indicators. The first series of examples illustrates the implications of different levels of reporting, of using process versus outcome indicators and of reporting indicators for which there are likely to be large differences in quality among providers, in comparison with indicators for which there may be small differences in quality. The second example illustrates the importance of accounting for cluster effects when calculating sample size.

The calculations for tables 7.1 to 7.5 assume that the hypothetical providers, A and B, at each level of reporting are identical in every way apart from the quality of care that they provide for the specific indicator under study. The sample size calculations assume that simple random samples have been drawn from the whole population, i.e., the samples are not clustered in any way, and do not require statistical adjustment for cluster effects. In addition, no adjustments have been made where the sample size is large in comparison with the population size. The population size at each level represents reasonable approximations, in the case of the general practice, representing a practice with five or six full time partners. For each indicator, the number of eligible patients for an intervention, or the mortality rate for AMIs, is based on previous published studies (Mant and Hicks, 1995; McColl et al., 1998). The estimated difference in performance among providers A and B represents the levels (in some cases, extreme levels) that might be found in practice.

APPENDIX 7 - ILLUSTRATIVE EXAMPLES OF PERFORMANCE INDICATORS FOR THE UNITED KINGDOM

The results illustrate three key issues. First, if simple random sampling is used, it is easier to compare providers at a higher reporting level because a smaller proportion of eligible patients needs to be sampled. Second, comparing process indicators requires smaller sample sizes than comparing outcome indicators and therefore process indicators are more sensitive than outcome indicators as measures of differences in quality between providers. Third, indicators for which there are likely to be large differences in quality between providers require smaller sample sizes than those for which the differences between providers is likely to be small.

Primary Care Indicators

Table 7.1: Aspirin for patients at high risk of coronary or ischaemic cerebrovascular events (process indicator)

reporting level	population size ^a	number of eligible patients for intervention ^b	Number of patients given intervention ^c		sample size needed to detect significant difference at 80% power and 5% significance (sample size as % of eligible patients)
			Provider A (50% uptake)	Provider B (70% uptake)	
Health Authority	500 000	15 000	7 500	10 500	103 ^b (0.68 %)
Primary Care Group	100 000	3 000	1 500	2 100	103 ^b (3.3 %)
Practice	10 000	300	150	210	103 ^b (34.3 %)
General Practitioner	2 000	60	30	42	103 ^b (171.6 %)

Table 7.2: Influenza vaccination for those aged >65 years (process indicator)

reporting level	population size*	number of eligible patients for intervention*	number of patients given intervention*		sample size needed to detect significant difference at 80% power and 5% significance (sample size as % of eligible patients)
			Provider A (50% uptake)	Provider B (70% uptake)	
Health Authority	500 000	78 500	39 250	54 950	103 % (0.13 %)
Primary Care Group	100 000	15 700	7 850	10 990	103 % (0.65 %)
Practice	10 000	1 570	785	1 099	103 % (6.5 %)
General Practitioner	2 000	314	157	220	103 % (32.8 %)

Table 7.3: Patient satisfaction (outcome indicator)

reporting level	population size*	number of patients giving rating "very satisfied"*		sample size needed to detect significant difference at 80% power and 5% significance (sample size as % of population)
		Provider A (75% "very satisfied")	Provider B (78% "very satisfied")	
Health Authority	500 000	375 000	390 000	3201 % (0.64 %)
Primary Care Group	100 000	75 000	78 000	3201 % (3.2 %)
Practice	10 000	7 500	7 800	3201 % (32 %)
General Practitioner	2 000	1 500	1560	3201% (160%)

Secondary Care Indicators

Table 7.4: Mortality from acute myocardial infarction (AMI) (outcome indicator)

reporting level	number of AMI admissions per year	number of deaths per year*		sample size needed to detect significant difference at 80% power and 5% significance (sample size as % of nos. of admissions per year)
		Provider A (mortality rate 20%)	Provider B (mortality rate 15%)	
Health Authority (catchment population 500 000)	750	150	112	945* (126.0%)
District General Hospital (catchment population 300 000)	450	90	67	945* (210.0 %)
Physician (assuming 10 physicians responsible for AMIs per hospital)	45	9	7	945* (2100.0 %)

**Table 7.5: Administration of fibrinolytic medication post AMI
(process indicator)**

reporting level	number of AMI admissions per year ⁻	number of eligible patients for intervention per year (90%) ⁻	number of patients given intervention [*]		sample size needed to detect significant difference at 80% power and 5% significance (sample size as % of no. of eligible patients per year)
			Provider A (90% uptake)	Provider B (95% uptake)	
Health Authority (catchment population 500 000)	750	675	608	641	474 [*] (70.2 %)
District General Hospital (catchment population 300 000)	450	405	365	385	474 [*] (117.0 %)
Physician (assuming 10 physicians responsible for AMIs per hospital)	45	40	36	38	474 [*] (1185.0 %)

- Average or estimates given as examples
- * Estimates based on McColl et al., 1998
- * figures given as examples
- Estimates based on Mant and Hicks, 1995
- ⊗ no adjustment made for finite population

The above examples illustrate sample size calculations based only on the difference in performance between providers and the probability of finding a real difference for a given level of statistical power. They also assume that a 'simple random sample' - that is all members of the population have an equal chance of being sampled - has been taken across the entire population. In practice, a simple random sample is often not feasible or cost-effective.

APPENDIX 7 - ILLUSTRATIVE EXAMPLES OF PERFORMANCE INDICATORS FOR THE UNITED KINGDOM

Instead, it is often necessary to sample from groups, or 'clusters', within the population. For example, suppose one wants to compare the performance between two PCGs based on medical record data. To draw a simple random sample from each PCG would be expensive and time-consuming. Each PCG sample would probably contain patients from a large number of the practices in that PCG, necessitating large travel and other sample accrual costs such as those needed to meet the individual approval requirements at each sampled practice. Instead, a sample of perhaps half the practices in each PCG might be drawn, and then the medical records abstracted for all patients of each sampled practice. Such a sample is clustered, where the clusters are the practices, and the cluster size for each cluster is the number of patients treated in each practice.

A clustered sample would cause no statistical problems, that is it would be as efficient and informative as a simple random sample, if the quality of care provided by a practice to a particular patient was independent of the quality of care provided by the same practice to another patient in that practice. In reality, however, there is likely to be greater similarity between patients who attend the same practice than between patients who attend different practices because a practice tends to treat its patients similarly. For example, all hypertensive patients within the same practice will be treated more similarly than the manner in which hypertensive patients are treated by all the different practices because of practice-specific shared guidelines and interaction between caregivers within practices. The level of dependence is measured by a quantity called 'the intra-cluster correlation (ICC)': an ICC of 0 represents complete independence and an ICC of 1 represents complete dependence, i.e., all patients treated in each practice are treated exactly the same.

Not only the ICC is important in sample size calculations, but also the average cluster size impacts the calculations. The larger the cluster size, the greater the effect of the clustering as more sampled patients are seen by the same doctor and are dependent. The ICC and the cluster size are used to calculate a quantity known as the 'design effect'. The design effect is used to adjust the actual sample size drawn in the clustered sample to the effective sample size.' For example, if the design effect is two, and the actual sample size drawn in the clustered sample is 200, the effective sample size is 100 ($=200/2$). The effective sample size is the number of patients who would have to be drawn via a simple random sample (unclustered) to achieve the same statistical precision, that is to provide as much information, as the clustered sample. In the example, a simple random sample of size 100 will produce estimates of the same precision, and will provide the same statistical power, as a clustered sample of size 200. If the costs saved are worth the loss in precision, then a clustered design may be the logical choice. For example, if the cost of a clustered sample of size 200 is equal to the cost of a simple random sample of size 100 due to the fact that clustering saves money which can subsequently be used to sample more respondents, and the design effect is two, the decision-maker is indifferent between the two designs: they cost the same, and result in the same effective sample size.

Intra-cluster correlation can occur at more than one level in a clustered sample. Suppose one was interested in the quality of care delivered within a Health Authority. In a Health Authority, patients are clustered within practices, who are in turn clustered within PCGs. To determine the quality of care delivered in a Health Authority, one might randomly sample practices across the Health Authority. In this sample design, the clustering at the PCG level is ignored as the patients are sampled randomly across practices.

APPENDIX 7 - ILLUSTRATIVE EXAMPLES OF PERFORMANCE INDICATORS FOR THE UNITED KINGDOM

Only the effect of clustering of sampled patients within sampled practices needs to be taken into account. However, a less costly design might be to sample PCGs, then practices within sampled PCGs, and then patients within sampled practices. This design might reduce the number of PCGs that are sampled, and thereby reduce travel and other costs. In terms of this practice level of clustering, practices in the same PCG, in theory at least, provide care that is more similar than that delivered in practices in other PCGs. Generally, the dependence that results from the higher level of clustering (the PCG) is less severe than the dependence that results from the lower level of clustering (the practice). Thus in doing sample size calculations, one typically adjusts for the lowest level of clustering only. In addition, we note that we have ignored an even lower level of clustering that might be hypothesised: that within doctor, as we have assumed that patients see a variety of doctors belonging to a practice depending on who is available. That is, patients do not consistently receive treatment from the same physician, if not we would need to consider that dependence as well.

The ICC, cluster size and resulting design effect can have a significant effect on the power of a study to identify real differences. If the design effect is large, the effective sample size is reduced and the actual sample size required to detect real differences is increased.

The importance of adjusting for the effects of clustering during both the design and analysis phases of a research project has been recognised for many years in the survey design literature, and recently in the health services literature (Campbell and Grimshaw, 1998; Kerry and Bland, 1998a; Kerry and Bland, 1998b). The ICC due to clustering within a doctor's list has been estimated to be in the range 0.05-0.3 (J. Grimshaw, personal communication, 1999).

The ICC is higher for process measured than for outcome measures, probably because of the greater biological variability associated with the measurement of outcomes compared with the measurement of behaviour. It is also higher for secondary care than for primary care, possibly because hospital practice is more consistent than general practice.

In order to illustrate the impact of clustering on power calculations, we will consider a particular example. In these calculations, we assume that the ICC ranges are as follows (based on estimates from actual data from J. Grimshaw, personal communication, 1999):

0 to 0.1 for primary care process measures such as aspirin for high risk patients or influenza vaccination;

0 to 0.05 for primary care outcome measures such as patient satisfaction;

0 to 0.3 for secondary care process measures such as fibrinolysis post-AMI;

and 0 to 0.15 for secondary care outcome measures such as AMI mortality.

We wish to compare two PCGs that consist of ten practices each. We sample five practices per PCG. We then determine the difference that we will be able to detect with 80% power, assuming a two-sided test of level 0.05 if we sample 25 patients per practice, 50 patients per practice, and 75 patients per practice respectively. We produce calculations for the most extreme ICC value hypothesised for each outcome

Table 7.6: Effect distinguishable between two PCGs with 80% power assuming a 0.05 level of significance under different sampling designs.

		Indicators				
		Aspirin for high risk patients (assume 50% in one PCG)	Influenza Vaccination for > 65 years (assume 50% in one PCG)	Patient satisfaction (assume 75% in one PCG)	AMI Mortality (assume 15% in one PCG)	Fibrinolytic post AMI (assume 99% in one PCG)
ICC		0.10	0.10	0.5	0.15	0.3
Number of patients sampled in each of 5 practices	Total numbers of patients in sample	Difference distinguishable				
25	125	50% v. 83%	50% v. 83%	75% v. 96%	14% v. 54%	50% v. 90%
50	250	50% v. 81%	50% v. 81%	75% v. 94%	15% v. 52%	52% v. 59%
75	375	50% v. 80%	50% v. 80%	75% v. 93%	15% v. 51%	52% v. 99%

Table 7.6 demonstrates that as we increase the number of patients sampled per cluster, the effect that is distinguishable does not decrease much. This results because additional patients per cluster (practice) are not contributing much new information due to the clustering effect.

In conclusion, the sample size required to identify statistically significant differences among providers is influenced by the level of reporting, the nature of the data, the degree of difference in performance, the required power to identify a difference, the probability of identifying a real difference, the intra-cluster correlation coefficient and cluster size. All these factors should be taken into account before valid comparisons can be made about the relative performance of providers. This implies that it will be easier and more efficacious to disclose data about process rather than

outcome measures of quality and at high reporting levels, such as health authorities, rather than low levels, such as groups of physicians. Valid comparisons of performance using outcomes data, or at lower reporting levels would require aggregation of indicators across diseases or conditions. The most effective way of aggregating indicators, and the implications of doing so, are currently being studied in the US. (McGlynn et al, 1997).

