Performance Measurement for Health System Improvement

Experiences, Challenges and Prospects

Peter C. Smith, Elias Mossialos, Irene Papanicolas
and Sheila Leatherman

HEALTH ECONOMICS,
POLICY AND MANAGEMENT

European
Observatory
on Health Systems and Policies

WORLD HEALTH
ORGANIZATION
EUROPE

CAMBRIDGE

Performance measurement in specific domains

4.1 | *Performance measurement in primary care*

HELEN LESTER, MARTIN ROLAND

Introduction

This chapter explores the value and complexities of measuring perform-ance in primary care. We begin with a definition of primary care and a description of its importance within the wider health-care system. We then explore the importance of measuring performance in this set-ting and provide an overview of some of the quality improvement strategies currently in use. The second part of the chapter describes a conceptual framework for quality measurement and reporting; the qualities of an ideal performance measure; and the relative value of process and outcome measures within primary care. The third part describes three very different primary-care focused systems in which performance measurement has been critical to improving health care: (i) Quality and Outcomes Framework in the United Kingdom; (ii) changes in the Veterans Health Administration in the United States; and (iii) European Practice Assessment. We conclude by highlighting challenges that policy-makers, researchers and clinicians face in future performance measurement in primary care.

Background to performance measurement in primary care

Defining primary care

WHO made the improvement of primary health care a core policy in the Alma-Ata declaration (WHO 1978) and the *Health for All by the Year 2000* strategy. The World Health Assembly renewed the com-mitment to global improvement in health (particularly for the most disadvantaged populations) in 1998 and this led to the *Health for All in the 21st Century* policy and programme.

The term 'primary care' has different meanings in different coun-tries. The providers of primary care may be general practitioners, fam-ily physicians, specialists working in the community, nurses or nurse

practitioners and (perhaps) physicians' assistants. These practitioners may work in solo practices or in large multi-professional groups and may or may not be integrated with social and community services. Some will have a gatekeeper function to secondary care. Methods of funding primary care also vary from payment by the patient to payment by the state, with a variety of combinations in between.

Primary care is better described in terms of its function rather than its location. The American Institute of Medicine (Donaldson et al. 1996) defined primary care as: 'the provision of integrated, accessible health care services by clinicians who are accountable for addressing a large majority of personal health care needs, developing a sustained partnership with patients and practising in the context of family and community'. This builds on Starfield's earlier definition of primary care as 'first-contact, continuous, comprehensive, and coordinated care provided to populations undifferentiated by gender, disease, or organ system' (Starfield 1994).

The critical elements of primary care are:

- first-contact accessible services where demands are clarified and information, reassurance or advice are given and diagnoses made;
- provision of comprehensive services to meet the needs of patients, with focus on generalism rather than specialism;
- provision of patient-centred rather than disease-centred care;
- provision of a longitudinal relationship between an individual patient and his/her health-care provider;
- coordination of care for individual patients;
- integration of biomedical, psychological and social dimensions of a patient's problem;
- focus on health promotion and disease prevention as well as management of established health problems.

In many countries, the primary-care provider also acts as an advocate for patients as they move through often complex health-care systems.

It has been demonstrated both between and within countries that those with a strong system of primary care have more efficient health systems and better health outcomes than those with a strong focus on hospital services (Macinko et al. 2007; Starfield 1998; Starfield et al. 2005). Countries with high primary care physician to population ratios (but not specialist to population ratios) have healthier popula-

tions and fewer social inequalities in the health of their populations. Primary care therefore has an equity-producing effect, at least for those measures of health that are most responsive to primary care (see Box 4.1.1).

Box 4.1.1 Benefits of primary care

Countries with strong primary care:

- have lower overall costs
- generally have healthier populations.

Within countries:

- areas with higher availability of primary-care physicians (but not specialists) have healthier populations;
- higher availability of primary-care physicians reduces the adverse effects of social inequality.

Source: from Starfield 1998

General practice or family medicine is a core discipline within primary care – in Europe, primary care is not easily conceptualized without general practice. However, primary care encompasses considerably more than general practice alone. In countries in which general practice is well-developed, the functions and characteristics of primary care largely overlap with those of general practice and general practice may have a preferred position in primary care. In other countries, specialists in internal medicine, paediatrics and gynaecology also provide primary medical care that is directly accessible.

Importance of measuring performance in primary care

In order to understand the importance of measuring performance in primary care it may be helpful to remember the ecology of medical care. White et al. (1961) published a framework for thinking about the organization of health care. Inspired in part by careful reporting on the part of British general practitioners (Horder & Horder 1954), this conceptualization suggested that in an average month and in a population of 1000 adults – 750 reported an illness, 250 consulted a physician, 9 were hospitalized, 5 were referred to another physician

and just 1 was referred to a university medical centre. Analysis of 1995-1996 data on the use of health care in the United States (Green et al. 2001) had remarkably similar findings although undertaken thirty years later and in a different country. Among 1000 men, women and children they found that (on average each month) – 800 experienced symptoms, 327 considered seeking medical care, 217 visited a physician in the office (113 to a primary-care physician; 104 to other specialists), 65 visited a professional provider of complementary or alternative medical care, 21 visited a hospital-based outpatient clinic, 14 received professional health services at home, 13 received care in an emergency department, 8 were hospitalized and less than 1 (0.7) was admitted to an academic medical centre hospital (Green et al. 2001).

In essence, most people with symptoms manage them within the community; if they do seek help they use the equivalent of primary care, with very few people referred on for specialist care. Primary care is therefore the cornerstone of most health-care systems and measurement of its performance plays a critical part in ensuring that the whole system works effectively, efficiently and for the benefit of patients.

However, professional acceptance of the need to measure performance in primary care is relatively recent. Until the 1980s, there was a widespread notion in most European countries and in the United States that there was little variation in medical practice and that one doctor was much like another. The British government's attempts to introduce measures of performance in 1986 were described as 'political and provocative' (British Medical Association 1986). The international rise of evidence-based medicine (Sackett et al. 1996) and a growing realization of variations in practice meant that measurement of performance became a higher priority for both primary-care practitioners and policy-makers during the 1990s. Studies began to highlight inappropriate overuse, underuse and misuse of procedures in a variety of different fields (McGlynn et al. 1994). Much of the initial research focused on specialist practice but subsequent studies found considerable variation in the quality of primary care (Mangione-Smith et al. 2007; McGlynn et al. 2003; Seddon et al. 2001). This was accompanied by a wider general recognition that medical error can be an important cause of harm to patients (Kohn et al. 2000). In the United Kingdom, a series of well-publicized 'scandals' in primary and secondary care heightened concern that physicians should not be solely responsible for their own clinical governance and professional regulation (Smith 1998).

Over the last decade, many countries have therefore replaced implicit codes governing the health professional/patient relationship with explicit (often government controlled) rules and regulations for performance in primary care. Politicians' and payers' demands for efficiency increases also created pressure on managers to make decisions about which interventions and ways of working provided best value for money. Measuring performance provided one source of evidence for making such judgements.

Conceptual framework for assessing quality of care

It is helpful to have an overall understanding of the meaning of quality before deciding how to measure it. Campbell et al. (2000) describe a framework for assessing quality of care that distinguishes between care for individual patients and care for populations (see Box 4.1.2).

Quality of care for individual patients

For individual patients, the two central domains are access and effectiveness – can patients get to health care and is it any good when they arrive? Effectiveness covers both clinical and interpersonal care. It is not enough to provide good clinical care without good interpersonal care, and good interpersonal skills cannot substitute for poor clinical skills.

Clinical care may be subdivided into preventive care (staying healthy); care for acute illness (getting better); chronic disease management (living with illness or disability); and terminal care (coping with the end of life). The bracketed terms are those used by the Institute of Medicine in the United States. In addition, safety is sometimes included as a specific domain because of its high political profile and importance for patients. Interpersonal aspects of care are most frequently measured using patient questionnaires such as the General Practice Assessment Questionnaire (GPAQ) (www.gpaq.info) and the EUROPEP questionnaire for evaluating patient satisfaction and experience (Grol et al. 2000).

Good care cannot usually be delivered without good organization of care and attention to the environment in which that care is provided. Measuring organizational competence is therefore an important part of overall quality assessment.

Quality of care for populations

There are two additional domains of quality of care for populations of patients – equity and efficiency. Efficiency is an important marker of quality of care for populations as inefficient care (e.g. prescribing expensive but ineffective drugs) may have opportunity costs for the care that can be provided to other patients. Likewise, equity is a key element of quality especially where resources are distributed unevenly across population groups.

Box 4.1.2 Framework for assessing quality of care

Quality of care for individuals is determined by:

Access
Effectiveness of care
- clinical care
- interpersonal care (patient experience)
Organization of care/organizational development

Quality of care for populations is additionally determined by:

Equity
Efficiency

Source: from Campbell et al. 2000

Overview of quality improvement strategies in primary care

Quality improvement methods share three key elements:

1. *Specification of a desired performance* in the form of clinical guidelines, care pathways, review criteria or clinical policies.
2. *Ways of changing clinical practice*. Numerous approaches have been used with varying degrees of success including lectures, small group education, one-to-one educational outreach visits, audit and feedback, reminder systems, computerized decision support, public release of information and financial incentives. Patient mediated interventions include guidelines for patients and training to increase patient assertiveness in consultation.

3. *Measurement*. Performance needs to be measured to determine whether and to what extent improvement has occurred so that further quality improvement strategies can be targeted appropriately (see Fig. 4.1.1).

Research shows that quality improvement strategies in primary care can make a difference but that no single method is always effective. Passive education tends to be least effective and multi-faceted interventions seem to have most effect, especially when sustained over time (Bero et al. 1998).

Baker et al. (2006) describe quality improvement systems that are being introduced into primary care in most European countries although the speed of introduction is dependent on the development of the profession of general practice in individual countries. Broadly, the European Union can be divided into first, second and third wave groups. The first wave includes Denmark, the Netherlands, Sweden and the United Kingdom. These have well-developed primary care systems with respected primary care practitioners and quality improve-
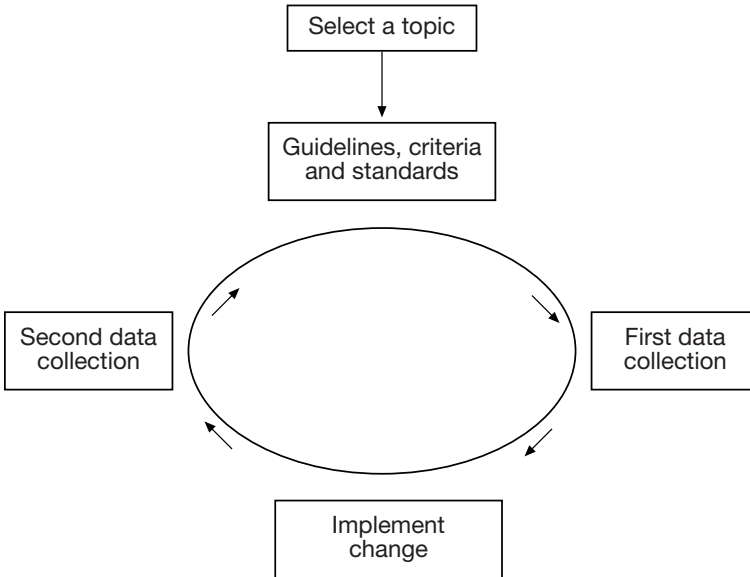

Fig. 4.1.1 Quality improvement cycle

ment systems that are now integral features of the health-care system. The second wave includes Austria, Belgium, France, Germany and Italy. These have made substantial progress since the early 1990s. The third wave is mainly composed of CEE countries. These have limited quality improvement initiatives, often hindered by the low status of general practitioners within the health-care system.

Developing performance measures for primary care

Underlying conceptual framework

The main purposes of a health-care system are to reduce the impact of the burden of illness, injury and disability and to improve the health and functioning of individuals in the population. Measuring the quality of care is one means of assessing how well this aim is being achieved. The Strategic Framework Board was established in the United States in 1999 to design a strategy for national quality measurement and reporting systems and to articulate the guiding principles and priorities for such a system. It produced a dynamic conceptual framework for a national quality measurement and reporting system (see Fig. 4.1.2).

This system aims to evaluate the degree to which the health system is providing safe, effective, timely and patient-centred care. It can also assess whether the delivery of high-quality care is efficient and equitable. It provides accessible information on quality to a variety of audiences including consumers, purchasers and providers to facilitate individual and collective decision-making. It also provides information that regulators, purchasers and providers can use to support continued improvement and achievement of goals (McGlynn 2003).

The Strategic Framework Board outlined a series of criteria and a process by which national goals for quality measurement and improvement could be selected. They suggested that goals should:

• be achievable within the health-care delivery system;
• represent areas in which patients experience a substantial burden of illness, injury or disability or problems with health and functioning;
• be based on evidence that progress on the goal is possible;
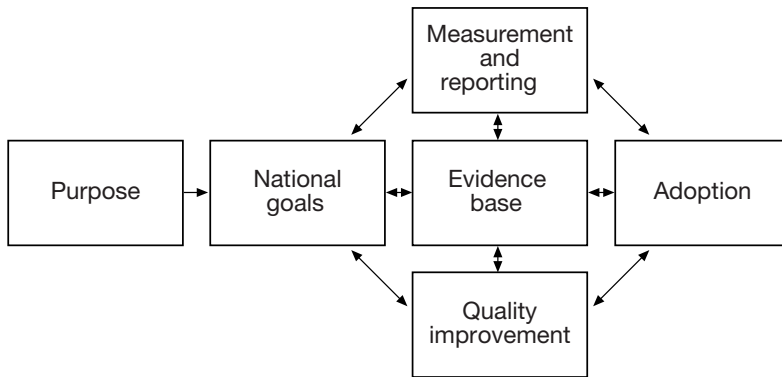• be able to address the quality problems faced by diverse populations;

Fig. 4.1.2 Conceptual map of a quality measurement and reporting system
Source: McGlynn 2003

- be compelling to expert groups and relevant constituents (McGlynn et al. 2003a).

Performance measures can be developed once goals have been set and areas prioritized. Three preliminary issues need to be considered when developing measures.

1. Which aspects of care do you want to assess? Structure (e.g. staff, equipment, appointment systems); process (e.g. prescribing, investigations, interactions between professionals and patients); or outcomes (e.g. mortality, morbidity or patient satisfaction)? (Campbell et al. 2003).
2. Whose perspective is being prioritized? Different stakeholders will have different perspectives on the quality of care (Donabedian 1980). Patients may emphasize good communication skills whereas managers' views are more likely to be influenced by data on efficiency (Campbell et al. 2004).
3. What sort of supporting information or evidence is required? The type of indicator and the method of combining evidence and expert opinion when considering performance measurement are somewhat different in primary care than in other parts of the health system.

Many areas of health care have limited or methodologically weak evidence bases, especially within primary care (Naylor 1995). This

requires performance measures to be developed using evidence alongside expert opinion. However, experts often disagree on the interpretation of evidence so rigorous methods are needed to combine the two. Consensus methods are structured facilitation techniques that explore general agreement amongst a group of experts in order to synthesize evidence with opinion. Group judgements are preferable to individual judgements as they are less prone to personal bias. Several consensus techniques exist including consensus development conferences; Delphi technique; nominal group technique; RAND appropriateness method; and iterated consensus rating procedures (Campbell et al. 2003; Murphy et al. 1998). The ideal qualities of a performance measure are shown in Box 4.1.3.

Box 4.1.3 Ideal qualities of a performance measure

An ideal performance measure has good:

- acceptability: acceptable to both those being assessed and those undertaking the assessment;
- feasibility: valid and reliable consistent data are available and collectable;
- reliability: minimal measurement error, reproducible findings when administered by different raters (inter-rater reliability);
- sensitivity to change: has capacity to detect changes in quality of care;
- predictive value: has capacity to predict quality of care outcomes.

Source: Campbell et al. 2002

Outcome measures are often seen as the gold standard but process measures are often more useful for performance in primary care. Hard outcomes such as mortality may relate to primary care but often occur long after the care has been given. They may be confounded by sociodemographic factors outside the control of primary care staff and also by the availability of secondary care services (Giuffrida et al. 1999). In theory, case-mix adjustment can be used to adjust outcomes for underlying differences in populations (Lilford et al. 2007). However, there is usually insufficient information in the medical record to allow this for primary care populations. Process measures based on scientific evidence which links them to effective outcomes (sometimes referred

to as intermediate outcome measures) are generally recognized as the most useful indicators currently available in primary care. However, the development of methods of case-mix adjustment in primary care, e.g. the use of ambulatory care groups (Weiner et al. 1991), may provide new approaches to this problem.

The relative strengths and weaknesses of process and outcome measures are shown in Tables 4.1.1 and 4.1.2.

In the United Kingdom, coronary heart disease provides a practical example of the appropriate use of different types of performance measures. The Quality and Outcomes Framework has twelve primary care indicators focused on secondary prevention of coronary heart disease. These include producing a register of patients with the condition; a series of process measures aimed at ensuring that patents are given the most appropriate drug treatments; and two intermediate outcome measures that build on process measures of measuring blood pressure and cholesterol levels:

the percentage of patients with coronary heart disease whose last blood pressure reading was 150/90 or less.

and:

the percentage of patients with coronary heart disease whose last measured total cholesterol was 5 mmol/l or less.

In the longer term, there is strong evidence that control of blood pressure and high cholesterol are important in improving survival from coronary heart disease and therefore these intermediate outcomes may be related more closely to health outcomes than pure process measures.

However, for people with coronary heart disease within a secondary care setting, a cardiac surgeon's performance of coronary artery bypass graft **is** measured not only through their activity (process) figures but also through outcome measures. These include their overall mortality rates expressed as a percentage of all operations of that kind undertaken and compared to the national average. Whilst case-mix adjustment is often still necessary, the end result for the patient (death or improved quality of life) is more directly linked to the skill of the surgical team than the blood tests and prescribed medications that form the basis for performance measures in primary care. The use of an outcome measure such as mortality is more justifiable within a secondary care setting.

Table 4.1.1 *Relative advantages and disadvantages of process measures to measure quality*

Advantages	Disadvantages
Readily measured: utilization of health technologies is often measured relatively easily, without major bias or error.	**Salience**: processes of care may have little meaning to patients unless the link to outcomes can be explained.
Easily interpreted: utilization rates of different technologies can often be interpreted by reference to the evidence base without the need for case-mix adjustment or inter-unit comparisons.	**Specificity**: care processes are often quite specific to a single disease or single type of medical care therefore process measures across several clinical areas or aspects of service delivery may be required to represent quality for a particular group of patients.
Smaller sample size: can identify significant quality deficiencies with much smaller sample sizes than outcome indicators.	**Ossification**: focus on process may stifle innovation and the development of new modes of care.
Unobtrusive: care processes can frequently be assessed unobtrusively (e.g. data stored in administrative or medical records).	**Obsolescence**: usefulness may dissipate as technology and modes of care change.
Indicators for action: failures identified in the processes of care provide clear guidance on what must be remedied to improve health-care quality. Also, acted upon more quickly than outcome indicators which often become available only after a long time has elapsed.	**Adverse behaviour**: can be manipulated relatively easily and may give rise to gaming and other adverse behaviours.
Coverage: can capture aspects of care (e.g. speed of access; patient experience), other than health outcomes, that are often valued by patients.	

Source: Davies 2005

Table 4.1.2 *Relative advantages and disadvantages of outcome measures to measure quality*

Advantages	Disadvantages
Focus: directs attention towards the patient (rather than the service) and helps nurture a 'whole system' perspective.	**Measurement definition**: relatively easy to measure some outcome aspects validly and reliably (e.g. death) but others are notoriously difficult (e.g. wound infection).
Goals: represent the goals of care and the NHS more clearly.	**Attribution**: may be influenced by many factors outside the control of a health-care organization.
Meaningful: tend to be more meaningful to some of the potential users of clinical indicators (patients, purchasers).	**Sample size**: requires large sample sizes to detect a statistically significant effect even when there are manifest problems with the processes of care.
Innovation: focus on outcomes encourages providers to experiment with new modes of delivery to improve patient care and experience.	**Timing**: may take a long time to observe.
Far sighted: focus on outcomes encourages providers to adopt long-term strategies (e.g. health promotion) that may realize longer-term benefits.	**Interpretation**: observed outcomes may be difficult to interpret if the processes that produced them are complex or occurred distant to the observed outcome.
Manipulation: less open to manipulation than process indicators although providers can influence risk-adjusted outcome by exaggerating the severity of patients' conditions (upstaging).	**Ambiguity**: good outcomes can often be achieved despite poor processes of care (and vice versa).

Case studies of performance measurement in primary care

Three case studies are presented below, each chosen to illustrate a different way of developing and implementing quality improvement schemes that include measuring performance in primary care. Each

describes the political and clinical context in which the measures were introduced; the measures themselves; the known intended and unintended consequences; and the critical factors that influenced implementation and changes to health and health care.

Case study 1: Quality and Outcomes Framework

The Quality and Outcomes Framework is a pay-for-performance scheme introduced in the United Kingdom in April 2004 as part of a new General Medical Services contract for general practitioners. Its introduction was facilitated by the alignment of a series of factors during the previous decade, including public disquiet over the quality of health-care services; the rise of evidence-based medicine; a change in the culture of the profession that enabled recognition of variations in the quality of primary care; and recognition of serious underfunding of health care in the United Kingdom in comparison to other countries (Roland 2004).

In these circumstances, professional representatives (General Practitioners Committee of the British Medical Association) were able and willing to negotiate with the government to provide elements of primary care through a system of performance related pay. The government was willing to invest up to 20% of the primary care budget, 90% of which was new money, in order to develop a series of incentivized evidence-based indicators across a range of clinical and organizational areas in primary care.

The Quality and Outcomes Framework consists of approximately 140 measures based on evidence or professional consensus. The majority (65%) of indicators are focused on clinical areas although the use of a balanced scorecard approach is reflected in a range of clinical, organizational and patient focused elements in the framework (see Box 4.1.4).

Points for individual indicators are awarded in relation to the level of achievement (e.g. percentage of people with diabetes with blood pressure below a defined target). A graduated scale of payments starts above a minimum threshold (25% initially but 40% since 2006) and ends at a maximum threshold (usually 90%). The framework is revised on a biennial basis – new clinical areas are added and issues that have become a standard part of primary care (usually within the organizational domain) are removed.

Box 4.1.4 Quality and Outcomes Framework: performance measure domains (2008)

Clinical: coronary heart disease, heart failure, stroke and transient ischaemic attacks, atrial fibrillation, hypertension, diabetes mellitus, chronic obstructive pulmonary disease, asthma, epilepsy, hypothyroidism, cancer, mental health, depression, dementia, learning disability, palliative care, chronic kidney disease, obesity, ethnicity coding.

Organizational: records and information, information for patients, education and training, practice management, medicines management.

Patient experience: length of consultations, patient surveys, patient experience of access to primary care.

Additional services: cervical screening, child health surveillance, maternity services, contraceptive services.

General practitioners can exclude patients from the quality calculation for a number of broadly defined reasons (exception reporting). This excludes them from both the numerator and the denominator of the quality calculation. Reasons for exception reporting include:

- patient is on maximum tolerated therapy
- patient refuses to participate
- patient is newly diagnosed or recently registered
- not clinically appropriate to include the patient.

Almost all practices in the United Kingdom now use an electronic medical record, a critical factor for successful implementation of the performance measurement system. Data on performance on each of the measures is collected at practice level through a national IT system. The Quality Management and Analysis System (QMAS) is used to calculate payments and as a public source of information on quality of care in individual practices. Practices can benchmark themselves against their performance in previous years and against other practices locally and nationally. Data are easily accessible on the Internet and patients can look up their own practice scores for each individual indicator (http://www.qof.ic.nhs.uk/index.asp).

The Quality and Outcomes Framework is a voluntary system but it has been taken up by over 99% of practices in the United Kingdom. During the first year, the levels of achievement exceeded those anticipated by the government – an average of 83.4% of the available incentive payments were claimed (Doran et al. 2006). Achievements were similarly high in the second and third years.

The indicators, particularly those in clinical areas, represent a mixture of process measures and intermediate outcome measures. Intermediate outcome indicators generally have more points attached to them as they are more difficult to achieve and represent a greater workload. The Quality and Outcomes Framework contains no pure outcome indicators since one of its central tenets is that the measure has to be within the control of primary care. This inevitably means that a majority of the clinical measures are process in nature (registers, improving systems). However, many of the clinical areas include a series of intermediate measures for which there is evidence that improvements in these parameters lead to better long-term outcomes, e.g. lowering blood pressure, lipid and glucose levels in conditions such as heart disease, stroke, hypertension, diabetes and kidney disease.

The process of developing new indicators involves multiple stakeholders. Every other year the general public, patients, national organizations, the Department of Health and health-care professionals submit ideas for inclusion. These are prioritized by representatives from the Department of Health and the medical profession. Evidence in each area is then reviewed by a panel of academic experts and summarized in a series of reports that are available for viewing by the general public once negotiations have been completed. Indicators in the reports are developed through a two-stage modified RAND process with primary care practitioners (Brook et al. 1986) and commented on by a national patient organization and by IT experts. The final set of evidence-based performance measures represents a negotiated compromise between the government (needing to ensure the best possible use of Treasury resources for public health benefit) and the British Medical Association (representing the medical profession). The negotiation is important for establishing a level of professional ownership.

Data on the impact of financial incentives in the Quality and Outcomes Framework are available from a study of forty-two representative practices in England – detailed data on quality of care were collected at a series of time points (1998, 2003, 2005, 2007),

including some that predated the financial incentives. The results of the study show that the quality of care for the three major diseases studied (coronary heart disease, asthma, type 2 diabetes) was improving rapidly between 1998 and 2003 prior to the introduction of the incentives. Improvements continued after the introduction of financial incentives and the rate of improvement increased for asthma and diabetes. Care for coronary heart disease was increasing most rapidly before the financial incentives and continued to improve at the same rate. Overall, the results suggest that the introduction of pay for performance was associated with a modest acceleration in improvement in the quality of care (Campbell et al. 2007).

The findings of the study are consistent with previous work. This suggests that financial incentives can have a modest effect in changing professional behaviour (Epstein et al. 2004) and that patients receive higher-quality care in geographical areas where performance measures and monitoring have been established (Asch et al. 2004).

However, such schemes also have potential unintended consequences (McGlynn 2007). These include possible myopia (pursuit of short-term targets at the expense of legitimate long-term objectives) or misrepresentation (deliberate manipulation of data so that reported behaviour differs from actual behaviour) (Smith 1995). There is concern, as yet largely unfounded in the United Kingdom (Doran et al. 2006), that pay for performance may also increase racial and ethnic disparities (Casalino et al. 2007).

In the United Kingdom, family practitioners have expressed concerns that the financial incentives will produce adverse effects including reductions in continuity of care; fragmentation of care as a result of specialization within practices; and neglect of conditions for which financial incentives are not provided (Roland et al. 2006). More broadly, the introduction of the pay-for-performance programme has been associated with a general trend away from placing implicit trust in NHS health-care professionals and toward more active monitoring of their performance (Checkland et al. 2004). Despite these concerns, overall job satisfaction among family physicians was higher in 2004 than in 2001 (Whalley et al. 2006) and a recent report from the United States suggests that targeted quality improvement programmes have not resulted in any deterioration in quality of care in untargeted disease areas (Ganz et al. 2007). The results generally support the Institute of Medicine's view that pay-for-performance programmes

can make a useful contribution to improving quality (Fisher & Davis 2006), particularly when part of a comprehensive quality improvement programme.

The size of the gains in quality in relation to the costs of pay for performance remains a political issue in the United Kingdom. The government now accepts that it paid more than expected for the improvements in performance (BBC 2007; National Audit Office 2008) – investing over £ 3 billion in primary care in the first three years of operation of the Quality and Outcomes Framework. General practitioners appear to have increased the proportion of practice income taken as profit since the new contract was introduced, suggesting that gains in quality could have been achieved at a lower cost. Payment is made at practice rather than individual physician level in order to reflect the significant degree of teamwork required to achieve a high level of performance and achievement. However, few non-physicians have received substantial pay rises as a result of the Quality and Outcomes Framework.

Case study 2: Veterans Health Administration

There has been health and social support for aged or injured soldiers in the United States since colonial times. However, a national programme for American veterans was consolidated with the establishment of the Veterans Administration (VA) in 1930. As resources were expanded following the Second World War the VA was elevated to Cabinet status and became the Department of Veterans Affairs in 1989. Its healthcare system has grown from 54 hospitals in 1930 and now includes 155 medical centres with at least one in each state, Puerto Rico and the District of Columbia. VA operates more than 1400 sites of care, including 872 ambulatory care and community-based outpatient clinics, 135 nursing homes, 45 residential rehabilitation treatment programmes, 209 Veterans Centers and 108 comprehensive home-care programmes. Almost 5.5 million people were treated in VA healthcare facilities in 2006 http://www1.va.gov/vetdata/docs/4X6_fall07_sharepoint.pdf).

Until the mid 1990s, the VA operated largely as a hospital system providing general medical and surgical services and long-term care. Medical centres and facilities were relatively independent of each other and even competitively duplicated services. In the late 1980s and early 1990s, the VA became increasingly criticized as an expen-

sive and poor quality system with its failings publicized widely in the media, including popular movies. Members of Congress argued that the organization needed new management or even that funding should be discontinued. In 1996, the Veterans' Health Care Eligibility Reform Act enabled the system to be restructured from a hospital to a health-care system. Two documents – *Vision for Change* (Kizer 1995) and *Prescription for Change* (Kizer 1996) – outlined the challenges facing the VA and served as a strategic outline for organizational restructuring and a new strategy for systemizing quality and value.

There were three key reforms (Perlin et al. 2004).

1. Eligibility – broadly expanded the eligibility of veterans who could use the VA.

2. Operational – major structural change that established the Veterans Integrated Service Networks (VISNs) to move away from a hospital-centric service. Twenty-two regional networks assumed responsibility for the performance of all medical centres and clinics within their area. Resources were allocated according to the capitation formula and networks became responsible for coordinating care in order to reduce duplication and incentivize care coordination. At the same time, the VA began to expand the provision of primary care which was legalized and mandated by legislation in 1994. The VA also expanded and updated its IT system to allow better coordinated care, with the eventual introduction of a single electronic medical record across the whole system.

   Between 1995 and 1996 the VA closed 52% of its acute care hospital beds; ambulatory care visits increased by 43%; over 200 new outpatient clinics were funded by the redirected savings; and a pharmacy benefits programme and a national formulary were instituted. The VA introduced a new electronic medical record with tools for assessment and improvement such as reminders to carry out certain services and documentation of patient care that could be accessed first within the VISN and then nationwide. The VA implemented computerized order entry for medication, tests and consultation. The electronic medical record also enabled better integration of care and communication across providers, since all providers had access to it.

   Quality transformation – performance measurement of key indicators of chronic and preventative care and, more recently, acute

hospital and palliative care are the cornerstones of this reform. To further motivate improvement, the VA has forged partnerships with health services researchers to measure quality and evaluate quality improvement interventions. The VA established nine quality enhancement research initiatives to help assess and improve quality in prevalent conditions like diabetes and heart failure and expanded the funding available for all VA health services researchers to focus on quality improvement. The VA also instituted annual patient surveys to assess access, satisfaction and health status.

3. Quality Transformation – quality measures are selected through an external peer review programme. Most of the measures come from major American quality monitoring organizations such as the NCQA but they also include measures of particular relevance to veterans. Data are collected quarterly by an external contractor who audits medical records from a sample in each facility. This is relatively expensive as the external contractor is paid several million dollars per year. Currently, there are approximately fifty quality measures within the system, collected with a level of clinical detail that makes them meaningful to clinicians (see Box 4.1.5).

To motivate improvement on the measures, VISN directors are accountable through a performance contract that either offers an incentive or withholds roughly 10% of salary. The VISN directors hold facilities and providers accountable through clear expectations of performance rather than direct individual monetary incentives. However, the VA administration is currently looking at ways to stimulate quality improvement through more direct use of pay for performance. The results at VISN and facility levels are publicized and recognized throughout the VA and stronger performances are recognized with awards. Much of the motivation therefore rests upon professional pride, on being recognized as a high-performing facility.

Within ten years, the VA moved from a reputation for providing poor quality care to being lauded for the provision of the best care within the United States (Longman 2005). Influenza vaccination rates rose from 28% in 1994 to 78% in 2000. Annual measurement of glycated haemoglobin in patients with diabetes rose from 51% to 94% and beta-blocker treatment following myocardial infarction rose from 70% to 95% in the same period (Jha et al. 2003). The absolute level of

---

Box 4.1.5 Veterans Administration performance measurement areas

Chronic and acute care	Preventive care
Diabetes e.g. low density lipoprotein cholesterol (LDLC) controlled (<130 mg/dl or 3.4 mmol/l)	Influenza vaccination
Acute myocardial infarction e.g. LDLC less than 130 mg/dl after heart attack and beta blocker on discharge after heart attack	Pneumococcal vaccination
Obstructive lung disease	Tobacco screening
Obesity	Mammography
Hypertension	Cervical cancer screening
Pain assessment	Colorectal cancer screening
Major depression	Hyperlipidaemia screening
Tobacco treatment	Alcohol screening
Community acquired pneumonia	Prostate screening
Heart failure	
Substance use disorders	

---

quality of care for veterans was also higher than for patients covered by Medicare. Kerr et al. (2004) showed that the quality of diabetes care in 2000-2001 was higher in the VA than in geographically matched commercial managed care plans for almost every aspect studied including timely eye screening, testing glucose and lipids concentrations and glucose and lipid control. Although overall care was higher for veterans than the community, the advantage was greatest for the measures that the administration was using to monitor quality (e.g. retinal screening for people with diabetes) and spilled over beyond the targeted measures to the conditions covered by the performance monitoring (e.g. diabetes). However, veterans had no advantage for conditions outside the performance monitoring system (Asch et al. 2004).

In summary, the change in quality of care in the VA over a relatively short time demonstrates the value of organizational change. This includes reorganization into networks; the shift to ambulatory settings; and the value of a high-quality information system. The VA's experience has shown that well-constructed and clinically detailed

measures of performance play a valuable role in improving quality of care in the community even without large monetary incentives for individual doctors (Conrad et al. 2006). Extrinsic motivation of competition between regions and small financial incentives to regional directors helped to drive the change (as did an enabling environment) but the cornerstone for improving quality was the systematic use of data-driven measures to monitor performance.

Case study 3: European Practice Assessment

The European Practice Assessment Practice Management (EPA-PM) framework was developed between 2002 and 2004 as part of the TOPAS-EUROPE Association, in collaboration with the Bertelsmann Foundation (Engels et al. 2005). The framework was designed for use across a wide group of European countries. It aims to measure the quality of the management and organization of general practices in order to contribute to the assessment of, and improvements in, the quality of primary care and to enable comparisons to be made between primary-care practices, both within and between countries.

EPA-PM is based around a conceptual framework for practice management with five domains (see Box 4.1.6). The indicators relating to practice management were collated from published sets of indicators and literature; the conceptual framework was then used to organize the indicators into relevant dimensions. The indicators were rated in a systematic selection process by six national expert panels, taking account of both evidence and professional opinion – 62 out of 171 indicators met the criteria for validity across all countries (Engels et al. 2006). All the questionnaires and checklists in the EPA instrument are derived from these indicators. The instrument was piloted in 273 practices across 9 European countries in 2004 and resulted in the present version: EPA 2005.

EPA-PM has been used widely in Germany and Switzerland and integrated within existing accreditation systems in the Netherlands. It has also been used in Australia, Belgium, Canada, Denmark, Romania, Saudi Arabia and Slovenia. EPA-PM combines measurement and feedback tools to enable individual practices to monitor progress continuously against benchmarks. A trained facilitator visits each practice and conducts the EPA-PM process. This emphasizes an educational approach to encourage practice staff to conduct self-

---

Box 4.1.6 EPA-PM: performance domains

Domain	No. of indicators	Example of indicator
Infrastructure	27	Sufficient seating in the waiting room.
People/staff	7	Responsibilities within the team are clearly defined.
Information	16	Practice has computerized medical record system.
Finance	4	Practice produces annual financial report that includes all income and expenditure.
Quality and safety	8	All staff involved in quality improvements. Practice has sterilizer or autoclave.

---

assessments; to reflect on their own strengths and weaknesses; and to identify areas for quality improvement. There are also questionnaires for the practice manager, general practitioners and all other staff and a separate questionnaire (EUROPEP) for patients. Individual practice feedback is given on the same day. Each assessment is benchmarked so that practices can compare their performance with others and observe changes in their own practice over time. Benchmarks in Europe are available online (http://www.ru.nl/topas-europe/index.php?idcatside=13).

Unlike the Quality and Outcomes Framework and VA systems, EPA-PM is focused solely on organizational issues in primary care and is formative in nature, iteratively linking assessment with improvement. It is intended to promote an educative and reflective approach with team-based learning and practice-specific feedback. Like the Quality and Outcomes Framework it is voluntary but the levers for change are professional development rather than financial incentives. The enablers for change are largely systems based and motivations are intrinsic (professional) rather than extrinsic (financial rewards).

EPA-PM is still at a relatively early stage of development with few data on implementation and longer-term effect at practice level. Its ethos and the collaborative consensual nature of its piloting suggest that it may represent a future model of developing performance measures for and in primary care. Its ability to cross international borders

echoes the wider political agenda of European unification and makes sense in a world of increasing economic migration of both patients and health professionals (Grol & Wensing 2007). However, it is also important to remember that comparisons across health systems can be misleading and that successful approaches will not necessarily work in the same way when transplanted to another system (Sheldon 2004).

The group that developed organizational indicators for use in general practice across Europe has used a similar approach to develop a set of indicators focusing on the prevention of cardiovascular disease – EPA-Cardio. An initial review of the literature was followed by selection of candidate indicators that were rated for validity by panels of informed general practitioners. Again, separate panels were convened in each of the nine participating countries. Overall, 44 out of 202 indicators (22%) were rated valid for inclusion on a 'European set'. These focused predominantly on secondary prevention and management of established cardiovascular disease and diabetes. There was less agreement on indicators of preventive care or for patients without established disease. Although 85% of 202 potential indicators assessed were rated valid by at least one panel, lack of consensus among panels meant that a smaller set was agreed by all panels. This was probably caused by a mixture of differences in health systems, cultures and attitudes to prevention and shows some of the problems in achieving agreement about the measurement of quality across different health-care systems (Campbell et al. 2008)

Conclusions

This chapter has outlined the central importance of performance measurement as a prerequisite for improving primary health care. Common themes have arisen across the different implementations described – the complexity of developing meaningful evidence-based measures that work in primary care; and the expense of setting up and maintaining a performance measurement system. We conclude with reflections on the issues raised by current schemes and some of the challenges that lie ahead for policy-makers, researchers and clinicians.

Where should performance measures be used?

The focus for performance measurement in primary care will vary by

health economy but generic underpinning priorities would include health conditions with:

- high prevalence;
- significant morbidity or mortality;
- recognized gap between actual and potential performance;
- good evidence that introduction of a measure will lead to an improvement in care;
- political importance.

Is there an optimal way of improving performance?

There is no agreed optimal combination of methods to improve performance in health care and it is important to recognize that measurement is one of a series of levers that policy-makers and funders can use. Numerous approaches have been used, with varying degrees of success. These include educational programmes directed at the community and/or health workers; audit and feedback; reminder systems; computerized decision support; public release of information; and financial incentives.

Public reporting of performance data has been championed during the last decade as a mechanism for increasing accountability to payers and patients, though with limited evidence of its effectiveness (Fung et al. 2008). As yet, there is no evidence to suggest that patients change their medical provider if differences in quality are demonstrated (Galvin & McGlynn 2003). Rather, it seems that provider behaviour is stimulated by public release of information on quality of care (Marshall et al. 2000). One reason why the Quality and Outcomes Framework has stimulated general practitioner activity in the United Kingdom is that detailed results for every practice (down to individual indicators) are available on the Internet.

Financial incentives (pay for performance) are used increasingly commonly as a method of quality improvement. We reviewed some of the evidence behind this approach in the section on the Quality and Outcomes Framework. Pay for performance is far from a panacea and the results from most well-designed evaluations show only modest benefits. There is also a series of fundamental questions about which elements of primary care could or should be financially incentivized. Financial incentives are most likely to be effective in influencing professional behaviour when performance measures and rewards are aligned

to the values of the staff being rewarded (Marshall & Smith 2003). Indeed, external incentives may crowd out motivation – the desire to do a task well for its own sake – if they clash with the professional's perceptions of his/her role or identity and of quality care (Gagné & Deci 2005). If measures and underlying data are not viewed as valid then physicians may see them as unfair or inappropriate (Bokhour et al. 2006).

Overall, Oxman et al's (2005) conclusion that there is 'no magic bullet' for quality improvement still stands. Single interventions are often disappointing and the best evidence for quality improvement comes from systems that have used multiple and sustained interventions designed to improve quality. However, it should not be inferred that nothing works. Examples such as the VA show that major system-wide change can be achieved with effective leadership which focuses on quality improvement as a key part of the delivery of health care.

Unintended consequences of performance measurement

It is important to monitor potential adverse effects of any quality improvement scheme that might selectively bring benefits to populations which are already advantaged. A clear example of a perverse and unintended consequence is the incentive designed to reduce waiting times to see general practitioners in the United Kingdom. Unexpectedly, this made it more difficult for patients to book appointments in advance (Salisbury et al. 2007).

There is also concern that financial incentives may lead to neglect of non-incentivized conditions (McGlynn 2007). This concern does not appear to have been realized in two recently published studies, from the United States (Ganz et al. 2007) and the United Kingdom (Steel et al. 2007), respectively. However, this type of study inevitably compares quality of care for those aspects that can be measured readily. Much of the criticism of the Quality and Outcomes Framework in the United Kingdom relates to the potential loss of the caring aspects of a general practitioner's work (Mangin & Toop 2007). There is a danger that measurement of isolated aspects of performance may fundamentally alter the concept of quality in primary care and begin to redefine what is important within it. There is a sense of urgency here since it may not be too long before the senior practitioners within primary care become those who have grown up in a climate that values what

can be measured easily above less definable aspects of care (Lester & Roland 2007). We need to guard against this and remember that the science of performance measurement is just one element of the art of primary care.

One potential problem with quality improvement initiatives is that groups which are compliant or easy to treat may selectively benefit – because they present for treatment; their doctors selectively give them more attention; doctors or health plans selectively disenrol patients from disadvantaged groups for whom it may be more difficult to reach quality targets. This is an example of the inverse equity hypothesis (Victora et al. 2000) which suggests that public health interventions may produce an initial widening of inequalities. However, this effect was not seen when incentives were introduced for cervical cytology in the United Kingdom in 1990 as there was progressive narrowing of inequality in the delivery of health care (Baker & Middleton 2003; Middleton & Baker 2003). The introduction of the Quality and Outcomes Framework in the United Kingdom also appears to be associated with a reduction in inequality (Doran et al. 2008). Nevertheless, the issue remains important, especially in health-care systems in which doctors have a disincentive to enrol patients who may not reach quality targets.

Removing and refreshing measurement sets

Those thinking of adopting performance measures might do well to think through the rules for removing these measures beforehand. In the United States, 'the percentage of patients with acute myocardial infarction who receive a prescription for beta-blockers within seven days of hospital discharge' has been used to evaluate managed care plans since 1996. A decade ago, only two thirds of the patients who survived acute myocardial infarction received beta blockers; today, nearly all do. As the curve representing the tenth percentile crept above 90%, the NCQA found little variation among health plans and therefore retired the measure (Lee 2007). This methodology could be adopted and adapted to suit different measures and health expectations.

Future challenges

As population demography changes, patients are increasingly likely to

present with more than one condition. Currently, 65% of Medicare beneficiaries have more than one condition and almost 20% have four or more (Berenson & Horvath 2003). Primary care will provide the majority of ongoing care for this growing population within most health systems. There is therefore a need to develop and validate sets of measures that make sense to primary care by taking account of the number and severity of conditions at an individual level. This may require piloting of new measures that are focused at patient level and can take into account the complexity of differing evidence bases for different conditions within the same patient. Indeed piloting of new performance measures is fast becoming the norm in both the United Kingdom and the United States and may provide an opportunity to experiment with new types of indicators, thresholds and the effects of differing financial incentives.

The consequences of co-morbidity will almost certainly include the potential for increasingly fragmented care, with the possibility of poor informational and interpersonal continuity. Coordination of care at the level of the individual patient pathway will present a series of challenges to clinicians and policy-makers and may well become a central focus of future performance measurement.

However, perhaps the greatest challenge facing primary performance measurement is to find the point of equipoise between trust and control (O'Neill 2003). In a system based on trust, it is a professional responsibility to measure performance and improve quality of care. Currently, many health-care systems appear to have a greater focus on control, accountability and public reporting – performance measurement is seen more as a societal or government responsibility. Is it possible that, in the longer term, this emphasis will erode an important part of the very medical professionalism that enabled quality improvement initiatives to flourish in the first place? Performance measurement, and the process of continuous quality improvement that it encourages, has enormous potential to improve the quality of primary care. The challenge is to develop more trust-promoting approaches that make sense to all actors and produce the greatest benefit for patients.

References

Asch, SM. McGlynn, EA. Hogan, MM. Hayward, RA. Shekelle, P. Rubenstein, L. Keesey, J. Adams, J. Kerr, EA (2004). 'Comparison of quality of care for patients in the Veterans Health Administration and patients in a national sample.' *Annals of Internal Medicine*, 141(12): 938–945.

Baker, D. Middleton, E (2003). 'Cervical screening and health inequality in England in the 1990s.' *Epidemiology of Community Health*, 57(6): 417–423.

Baker, R. Wensing, M. Gibis, V (2006). Improving the quality of performance. In: Saltman, R. Rico, A. Boerma, W (eds.). *Primary care in the driver's seat?* Maidenhead: Open University Press.

BBC (2007). *GP pay raise 'was mistake'*. BBC News Health. London: British Broadcasting Corporation (http://news.bbc.co.uk/player/nol/newsid_6280000/newsid_6280000/6280077.stm?bw=bb&mp=rm).

Berenson, RA. Horvath, J (2003). 'Confronting the barriers to chronic care management in Medicare.' *Health Affairs (Millwood)*, Suppl. web exclusives: W3-37-53.

Bero, L. Grilli, R. Grimshaw, JM. Harvey E. Oxman, AD. Thomson, MA (1998). 'Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings.' *British Medical Journal*, 317(7156): 465–468.

Bokhour, BG. Burgess, JF. Hook, J. White, B. Berlowitz, D. Guildin, MR. Meterko, M. Young, GJ (2006). 'Incentive implementation in physician practices: a qualitative study of practice executive perspectives on pay for performance.' *Medical Care Research and Review*, 63(Suppl. 1): 73–95.

British Medical Association (1986). *Report to Special Conference of Representatives of Local Medical Committees*. 13 November 1986. London: British Medical Association.

Brook, RH. Chassin, MR. Fink, A. Solomon, DH. Kosecoff, J. Park, RE (1986). 'A method for the detailed assessment of the appropriateness of medical technologies.' *International Journal of Technology Assessment in Health Care*, 2(1): 53–63.

Campbell, SM. Roland, MO. Buetow, SA (2000). 'Defining quality of care.' *Social Science and Medicine*, 51(11): 1611–1625.

Campbell, SM. Braspenning, J. Hutchinson, A. Marshall, MN (2002). 'Research methods used in developing and applying quality indicators in primary care. *Quality and Safety in Health Care*, 11(4): 358–364.

Campbell, SM. Braspenning, J. Hutchinson, A. Marshall, MN (2003). 'Research methods used in developing and applying quality indicators in primary care.' *British Medical Journal*, 326(7293): 816–819.

Campbell, S. Shield, T. Rogers, A. Gask, L (2004). 'How do stakeholder groups vary in a Delphi technique about primary mental health care and what factors influence their ratings?' *Quality and Safety in Healthcare*, 13(6): 428–434.

Campbell, S. Reeves, D. Kontopantelis, E. Middleton, E. Sibbald, B. Roland, M (2007). 'Quality of primary care in England with the introduction of pay for performance.' *New England Journal of Medicine*, 357(2): 181–190.

Campbell, S. Ludt, S. Van Lieshout, J. Boffin, N. Wensing, M. Petek, D. Grol, R. Roland, M (2008). 'Quality indicators for the prevention and management of cardiovascular disease in primary care in nine European countries.' *European Journal of Cardiovascular Disease Prevention and Rehabilitation*, 15(5): 509–515.

Casalino, LP. Elster, A. Eisenberg, A. Lewis, E. Montgomery, J. Ramos, D (2007). 'Will pay for performance and quality reporting affect health care disparities?' *Health Affairs* (*Millwood*), 26(3): 405–414.

Checkland, K. Marshall, M. Harrison, S (2004). 'Re-thinking accountability: trust versus confidence in medical practice.' *Quality and Safety in Health Care*, 13(2): 130–135.

Conrad, DA. Saver, BG. Court, B. Heath, S (2006). 'Paying physicians for quality: evidence and themes from the field.' *Journal on Quality and Patient Safety*, 32(8): 443–451.

Davies, H (2005). *Measuring and reporting the quality of health care: issues and evidence from the international research literature*. Edinburgh: NHS Quality Improvement Scotland.

Donabedian, A (1980). *Explorations in quality assessment and monitoring. Vol. 1: The definition of quality and approaches to its assessment*. Ann Arbor, MI: Health Administration Press.

Donaldson, MS. Yordy, KD. Lohr, KN. Vanselow, NA (eds.) (1996). *Primary care: America's health in a new era*. Washington DC: National Academy Press.

Doran, T. Fullwood, C. Gravelle, H. Reeves, D. Kontopantelis, E. Hiroeh, U. Roland, M (2006). 'Pay-for-performance programs in family practices in the United Kingdom.' *New England Journal of Medicine*, 355(4): 375–384.

Doran, T. Fullwood, K. Kontopantelis, E. Reeves, D (2008). 'Effect of financial incentives on inequalities in the delivery of primary care in England: analysis of clinical activity indicators for the quality and outcomes framework.' *Lancet,* 372(9640): 728–736.

Engels, Y. Campbell, S. Dautzenberg, M. van den Hombergh, P. Brinkmann, H. Szécsényi, J. Falcoff, H. Seuntjens, L. Kuenzi, B. Grol, R. EPA Working Party (2005). 'Developing a framework of, and quality indicators for, general practice management in Europe.' *Family Practice,* 22(2): 215–222.

Engels, Y. Dautzenberg, M. Campbell, S. Broge, B. Boffin, N. Marshall, M. Elwyn, G. Vodopivec-Jamsek, V. Gerlach, FM. Samuelson, M. Grol, R (2006). 'Testing a European set of indicators for the evaluation of the management of primary care practices.' *Family Practice,* 23(1): 137–147.

Epstein, AM. Lee, TH. Hamel, MB (2004). 'Paying physicians for high-quality care.' *New England Journal of Medicine*, 350(4): 406–410.

Fisher, ES. Davis, K (2006). 'Audio interview: pay for performance – recommendations of the Institute of Medicine.' *New England Journal of Medicine,* 355(13): e14 (http://content.nejm.org/cgi/content/full/NEJMp068216/DC1).

Fung, C. Lim, Y-W. Mattke, S. Damberg, C. Shekelle, PG (2008). 'Systematic review: the evidence that publishing patient care performance data improves quality of care.' *Annals of Internal Medicine*, 148(2): 111–123.

Gagné, M. Deci, EL (2005). 'Self-determination theory and work motivation.' *Journal of Organizational Behavior,* 26(4): 331–362.

Galvin, RS. McGlynn, EA (2003). 'Using performance measurement to drive improvement. A road map for change.' *Medical Care*, 41(3): 48–60.

Ganz, DA. Wenger, NS. Roth, CP. Kamberg, CJ. Chang, JT. Maclean, Ch. Young, RT. Solomon, DH. Higashi, T. Min, L. Reuben, DB. Shekelle, PG (2007). 'The effect of a quality improvement initiative on the quality of other aspects of health care: the law of unintended consequences?' *Medical Care*, 45(1): 8–18.

Giuffrida, A. Gravelle, H. Roland, M (1999). 'Measuring quality of care with routine data: avoiding confusion between performance indicators and health outcomes.' *British Medical Journal*, 319(7202): 94–98.

Green, LA. Fryer, GE Jr. Yawn, BP. Lanier, D. Dovey, SM (2001). 'The ecology of medical care revisited.' *New England Journal of Medicine*, 344(26): 2021–2025.

Grol, R. Wensing, M (2007). 'Measuring performance quality in general practice: is international harmonization desirable?' *British Journal of General Practice,* 57(542): 691–692.

Grol, R. Wensing, M. Mainz, J. Jung, HP. Ferreira, P. Hearnshaw, H. Hjortdahl, P. Olesen, F. Reis, S. Ribacke, M. Szecsenyi, J (2000). 'Patients in Europe evaluate general practice care: an international comparison.' *British Journal of General Practice*, 50(460): 882–887.

Horder, J. Horder, E (1954). 'Illness in general practice.' *Practitioner*, 173(1034): 177–185.

Jha, AK. Perlin, JB. Kizer, KW. Dudley, RA (2003). 'Effect of the transformation of the Veterans Affairs Health Care System on the quality of care.' *New England Journal of Medicine*, 348(22): 2218–2227.

Kerr, EA. Gerzoff, RB. Krein, SL. Selby, JV. Piette, JD. Curb, JD. Herman, WH. Marrero, DG. Narayan, KM. Safford, MM. Thompson, T. Mangione, CM (2004). 'Diabetes care quality in the Veterans Affairs Health Care System and commercial managed care: the TRIAD study.' *Annals of Internal Medicine*, 141(4): 272–281.

Kizer, KW (1995). *Vision for change: a plan to restructure the Veterans Health Administration*. Washington, DC: Department of Veterans Affairs.

Kizer, KW (1996). *Prescription for change: the guiding principles and strategic objectives underlying the transformation of the Veterans Health Administration*. Washington, DC: Department of Veterans Affairs.

Kohn, LH. Corrigan, J. Donaldson, MS (eds.) (2000). *To err is human: building a safer health system*. Washington, DC: National Academies Press.

Lee, TH (2007). 'Eulogy for a quality measure.' *New England Journal of Medicine*, 357(12): 1175–1177.

Lester, HE. Roland, MO (2007). 'Measuring quality through performance. Future of quality measurement.' *British Medical Journal*, 335(7630): 1130–1131.

Lilford, RJ. Brown, CA. Nicholl, J (2007). 'Use of process measures to monitor the quality of clinical practice.' *British Medical Journal*, 335(7621): 648–650.

Longman, P (2005). 'The best care anywhere.' *Washington Monthly*, 37(1/2): 38–48.

Macinko, J. Starfield, B. Shi, L (2007). 'Quantifying the health benefits of primary care physician supply in the United States.' *International Journal of Health Services Research*, 37(1): 111–126.

Mangin, D. Toop, L (2007). 'The Quality and Outcomes Framework: what have you done to yourselves?' *British Journal of General Practice*, 57(539): 435–437.

Mangione-Smith, R. DeCristofaro, AH. Setodji, CM. Keesey, J. Klein, DJ. Adams, JL. Schuster, MA. McGlynn, EA (2007). 'The quality of ambu-

latory care delivered to children in the United States.' *New England Journal of Medicine*, 357(15): 1515–1523.

Marshall, M. Smith, P (2003). 'Rewarding results: using financial incentives to improve quality.' *Quality and Safety in Health Care*, 12(6): 397–398.

Marshall, MN. Shekelle, PG. Leatherman, S. Brook, RH (2000). 'The public release of performance data: what do we expect to gain? A review of the evidence.' *Journal of the American Medical Association*, 283(14): 1866–1874.

McGlynn, EA (2003). 'Introduction and overview of the conceptual framework for a national quality measurement and reporting system.' *Medical Care*, 41(Suppl. 1): 1–7.

McGlynn, EA (2007). 'Intended and unintended consequences. What should we really worry about?' *Medical Care*, 45(1): 3–5.

McGlynn, EA. Asch, M. Adams, A. Keesey, J. Kicks, J. DeCristofaro, A. Kerr, EA (2003). 'The quality of health care delivered to adults in the United States.' *New England Journal of Medicine*, 348(26): 2635–2645.

McGlynn, EA. Cassel, CK. Leatherman, ST. DeCristofaro, A. Smits, HL (2003a). 'Establishing national goals for quality improvement.' *Medical Care*, 41(Suppl. 1): 16–29.

McGlynn, EA. Naylor, CD. Anderson, GM. Leape, LL. Park, RE. Hilborne, LH. Bernstein, SJ. Goldman, BS. Armstrong, PW. Keesey, JW. Pinfold, SP. Damberg, C. Sherwood, MJ. Brook, RH (1994). 'Comparison of the appropriateness of coronary angiography and coronary artery bypass surgery between Canada and New York State.' *Journal of the American Medical Association*, 272(12): 934–940.

Murphy, MK. Black, NA. Lamping, DL. McKee, CM. Sanderson, CF. Askham, J. Martineau, T (1998). 'Consensus development methods and their use in clinical guideline development.' *Health Technology Assessment*, 2(3): 1–88.

National Audit Office (2008). *NHS pay modernisation: new contracts for general practice services in England*. London: National Audit Office.

Naylor, CD (1995). 'Grey zones of clinical practice: some limits to evidence-based medicine.' *Lancet*, 345(8953): 840–842.

O Neill, O (2003). 'Trust with accountability?' *Journal of Health Services Research and Policy,* 8(1): 3–4.

Oxman, AD. Thomson, MA. Davis, DA. Haynes, RB (1995). 'No magic bullets: a systematic review of 102 trials of interventions to improve professional practice.' *Canadian Medical Association Journal*, 153(10): 1423–1431.

Perlin, JP. Kolodner, RM. Roswell, RH (2004). 'The Veterans Health Administration: quality, value, accountability, and information as transforming strategies for patient-centered care.' *American Journal of Managed Care*, 10(11 Pt 2): 828–836.

Roland, M (2004). 'Linking physicians' pay to the quality of care – a major experiment in the United Kingdom.' *New England Journal of Medicine*, 351(14): 1448–1454.

Roland, MO. Campbell, SM. Bailey, N. Whalley, D. Sibbald, B (2006). 'Financial incentives to improve the quality of primary care in the UK: predicting the consequences of change.' *Primary Health Care Research and Development*, 7(1): 18–26.

Sackett, DL. Rosenberg, WMC. Gray, JAM (1996). 'Evidence-based medicine: what it is and what it isn't.' *British Medical Journal*, 312(7023): 71–72.

Salisbury, C. Goodall, S. Montgomery, AA. Pickin, DM. Edwards, S. Sampson, F. Simons, L. Lattimer, V (2007). 'Does advanced access improve access to primary health care? Questionnaire survey of patients. *British Journal of General Practice,* 57(541): 615–621.

Seddon, ME. Marshall, MN. Campbell, SM. Roland, MO (2001). 'A systematic review of the quality of clinical care provided in general practice in the UK, Australia and New Zealand.' *Quality in Health Care*, 10(3): 152–158.

Sheldon, T (2004). 'Learning from abroad or policy tourism?' *British Journal of General Practice*, 54(503): 410–411.

Smith, P (1995). 'On the unintended consequences of publishing performance data in the public sector.' *International Journal of Public Administration*, 18(2/3): 277–310.

Smith, R (1998). 'All changed, changed utterly.' *British Medical Journal*, 316(7149): 1917–1918.

Starfield, B (1994). 'Is primary care essential?' *Lancet,* 344(8930): 1129–1133.

Starfield, B (1998). '*Primary care: balancing health needs, services and technology.* Oxford: Oxford University Press.

Starfield, B. Shi, L. Macinko, J (2005). 'Contribution of primary care to health systems and health.' *Milbank Memorial Fund Quarterly*, 83(3): 457–502.

Steel, N. Maisey, S. Clark, A. Fleetcroft, R. Howe, A (2007). 'Quality of clinical primary care and targeted incentive payments: an observational study.' *British Journal of General Practice,* 57(539): 449–454.

Victora, CG. Vaughan, JP. Barros, F. Silva, AC. Tomasi, E (2000). 'Explaining trends in inequities: evidence from Brazilian child health studies.' *Lancet,* 356(9235): 1093–1098.

Weiner, JP. Starfield, BH. Steinwachs, DM. Mumford, LM (1991). 'Development and application of a population-oriented measure of ambulatory care case-mix.' *Medical Care,* 29(5): 452–472.

Whalley, D. Bojke, C. Gravelle, H. Sibbald, B (2006). 'GP job satisfaction in view of contract reform: a national survey.' *British Journal of General Practice*, 56(523): 87–92.

White, KL. Williams, TF. Greenberg, BG (1961). 'The ecology of medical care.' *New England Journal of Medicine,* 265: 885–892.

WHO (1978). *Primary health care.* Report of the International Conference on Primary Health Care, Alma-Ata, USSR, 6–12 September 1978. Geneva: World Health Organization.

4.2 | *Chronic care*

MARTIN MCKEE, ELLEN NOLTE

Introduction

This chapter examines the challenges inherent in assessing how health systems perform in response to chronic diseases. These are diseases that persist over an extended time and require a complex response involving coordinated inputs from a wide range of health professionals, access to essential medicines and (where appropriate) monitoring equipment. Ideally this is embedded within a system that promotes patient empowerment. There are many chronic diseases but in this chapter we draw extensively on experience with diabetes. The reasons for this are three-fold. First, diabetes was the first example of an acute disease that was transformed into a chronic disorder by the introduction of effective treatment. Second, it exemplifies the complex nature of chronic disease as its complications affect many different bodily systems and call upon the expertise of a wide range of specialists. Third, it provides a lens through which to view the performance of the overall health system.

Health system performance is the focus of the chapter and this volume. However, before looking specifically at performance it is necessary to understand the specificities of chronic diseases, many of which pose substantial challenges for performance measurement. It may also be helpful to reflect on the rapidly increasing contribution of chronic diseases to the overall burden of disease, a development that has important consequences for the assessment of health system performance more generally.

Growing importance of chronic disease

The discovery and subsequent purification of insulin in 1921 marked a fundamental transformation in the nature of health care. Until then,

there was extremely limited scope for therapeutic intervention in the event of illness. Essentially, the physician could offer sympathy and symptomatic relief – perhaps using aspirin, first manufactured some twenty-five years previously – while the patient either recovered or died. The treatments available were largely useless and in some cases harmful. For the first time it was possible to treat patients who would otherwise die with effective, life-sustaining treatment.

For some years it seemed that the problem of diabetes had been solved. Certainly, people with insulin-dependent diabetes had to make significant changes to their lifestyles and adopt what are now seen as overly rigid diets. However, the complexity of diabetes was not yet apparent. By the 1950s the first generation of children whose lives had been saved by insulin were reaching middle age and manifesting a range of unexpected complications that affected vision, renal function and cardiovascular systems. Some complications (e.g. diabetic retinopathy) were quite new conditions; others (e.g. ischaemic heart disease) were also seen in the non-diabetic population but appeared earlier and more frequently in people with diabetes.

These developments posed major challenges. People with diabetes had typically developed long-term relationships with an individual physician or a small team of physicians specializing in diabetes. However, they now needed additional specialist care from ophthalmologists, renal physicians and vascular surgeons, among others. They also needed help from a range of paramedical staff such as dietitians and podiatrists. This was a new and very different model of care. Essentially, patients embarked on a journey to obtain appropriate specialized care at multiple destinations but often without either a map or a navigator. Inevitably, many perished along the way.

Diabetes is a simple biological problem (the inability to produce a particular hormone) that gives rise to a multi-system disease process. Yet, it is far from unique. A revolution in chemical engineering in the 1960s made available an increasing number of new classes of pharmaceuticals, many of which had the ability to transform the management of disease processes if they were taken indefinitely. Thiazide diuretics were joined by beta blockers and calcium antagonists in the management of hypertension. Inhaled beta sympathomimetics and steroids similarly transformed obstructive airways disease. Other classes of pharmaceuticals had a major impact on conditions such as arthritis, Parkinson's disease and epilepsy.

These new opportunities had profound consequences for the delivery of health care as a prescription was only the beginning of the process. These medicines required monitoring, first to ensure that parameters such as blood pressure or (for obstructive airways disease) respiratory function was being controlled adequately; second, to detect any side-effects at the earliest opportunity. The greatest changes were seen in the field of mental health, where the development of antidepressants and antipsychotics made it possible to close large psychiatric hospitals and replace them with community-based services.

Other changes have been less obvious but still profound. By the 1980s the advent of new chemotherapeutic agents had transformed many cancers from brief, fatal illnesses (like diabetes prior to 1922) into long-term chronic disorders which people died with, rather than from. More recently, the availability of life-sustaining treatment has similarly transformed the management of AIDS. In an unexpected parallel with diabetes it is only now that the long-term consequences are becoming clear. People on long-term treatment for AIDS are developing a range of complications, some of which relate to the underlying disease process (e.g. some malignancies) and others that are a consequence of the treatment (e.g. ischaemic heart disease linked to the atherogenic effects of antiretrovirals).

However, medical care is not the only factor driving increases in the numbers of people surviving with chronic diseases. The other is the ageing of populations. As the proportion of older people in the population grows so does the likelihood of developing a potentially disabling chronic condition because of accumulated exposure to chronic disease risk factors over a lifetime (Ben-Shlomo & Kuh 2002; Janssen & Kunst 2005). Data from Germany, the Netherlands and the United States suggest that about two thirds of those who have reached pensionable age have at least two chronic conditions (Deutsches Zentrum für Altersfragen 2005; van den Akker et al. 1998; Wolff et al. 2002).

To understand this phenomenon fully it is necessary to consider the ageing process. Populations are ageing rapidly in all industrialized countries but few commentators expect the maximum lifespan observed (currently 122 years) to increase significantly. They do expect that life expectancy at birth will continue to increase as it has in a linear fashion for over 150 years – those who would once have died young now survive for longer. At least in industrialized countries, much of this earlier gain was due to a marked decline in deaths in infancy and

childhood. This now offers limited scope for further progress and future gains are expected to arise from the delay in deaths among adults.

Fries (1983) examined the process of ageing in depth and distinguished two processes, both involving the progressive loss of physiological function. The first set is essentially unmodifiable (although subsequent research has suggested that this may not be entirely true in all cases) and includes formation of cataracts and the loss of glomeruli in the kidneys that leads to a decline in renal function. Less importantly, this set also includes the greying of hair. The second set includes glucose intolerance, physical strength, cardiac reserve and cognitive function. These processes can be delayed by appropriate lifestyle changes and can also be compensated for by appropriate treatments. Fries proposed the compression of morbidity theory – while the maximum lifespan was unlikely to increase substantially, as populations adopted healthier lifestyles and as therapeutic advances continued, the period of illness (morbidity) that individuals would experience prior to their deaths would be compressed.

There is now considerable evidence that this has happened. Studies in several countries reveal that healthy life expectancy has increased at a faster rate than overall life expectancy. For example, a recent systematic review demonstrated how disability and limitations among older adults in the United States declined consistently during the 1990s (Freedman et al. 2002). However, accumulating evidence suggests that at least part of this improvement is a consequence of therapeutic advances, as complex combinations of treatment increasingly enable older people to function with multiple disorders. For example, Freedman et al. (2007) report that between 1997 and 2004 a rising prevalence of chronic conditions among older Americans (aged sixty-five and over) was accompanied by declines in the proportion reporting disability as a result of those conditions. This was supported by an analysis of the Swedish population which also reported an ageing population with a decline in disability over time but an increase in health problems among survivors (Parker & Thorslund 2007).

A typical 75-year-old may have disorders affecting multiple body systems (e.g. hypertension, arthritis, chronic airways disease, heart failure, Parkinson's disease). He/she may be undertaking treatment with perhaps ten different medications, all potentially interacting with each other and with a metabolism influenced by coexisting impairments in liver and kidney function. Such combinations of illnesses,

treatments and physiological function are of such complexity that they
are unlikely to become the subject of the randomized controlled tri-
als that give rise to the evidence on which treatment decisions should
be made. This situation poses severe problems for those seeking to
assess the ability to respond to chronic disease and limits the scope of
evaluations.

The ageing of populations is thus an important driver of increases
in chronic disease but it is important to remember that these diseases
are not limited to the older population. Especially in countries experi-
encing rising levels of obesity, increasing numbers of young and mid-
dle-aged people are developing some form of chronic health problem.
It has been estimated that in 2002, 60% of all DALYs attributable
to non-communicable diseases in Europe were lost before the age of
sixty (WHO 2004). Recent evidence from the United States points to
a rapid increase in the number of children and youths with chronic
health conditions over the past four decades (Perrin et al. 2007), in
particular as a response to growing levels of obesity. Rising rates of
childhood chronic conditions imply subsequent higher rates of related
conditions among adults (van der Lee et al. 2007).

This section, and the one preceding it, demonstrates clearly how the
burden of disease is changing, with a transition from acute to chronic
disease. The next section examines some of the implications for health
systems.

Implications of the growth in chronic disease

The effects of the transition from acute to chronic disease are not triv-
ial. In 2006, approximately 30% of the population in the European
Union aged fifteen years and over reported a long-standing health prob-
lem, and one in four currently receives long-term medical treatment
(TNS Opinion & Social 2007). Surveys undertaken in England and
the United States suggest that one third and 45%, respectively, of the
adult population has some form of chronic health problem (Hoffman
et al. 1996; Wilson et al. 2005). People with chronic diseases are more
likely to utilize health care, particularly when they have multiple prob-
lems. For example, in England people with chronic illness account for
80% of general practice consultations and about 15% of people who
have three or more problems account for nearly 30% of inpatient days
(Wilson et al. 2005). Estimates for the United States place the costs of

chronic illness at around three quarters of the total national health expenditure (Hoffman et al. 1996). Some individual chronic diseases (e.g. diabetes) account for between 2% and 15% of national health expenditure in some European countries (Suhrcke et al. 2005).

This changing context has profound implications for policy-makers in the health sector. Health care is still largely built around an acute, episodic model of care that is ill-equipped to meet the requirements of those needing chronic care (Table 4.2.1). Experience in many countries shows that the responses required and their multiple interlinkages are very complex and it cannot be assumed that a model appropriate to these needs will simply emerge.

Table 4.2.1 *Features differentiating acute and chronic disease*

	Acute illness	Chronic illness
Onset	Abrupt	Generally gradual and often subtle
Duration	Limited	Lengthy and indefinite
Cause	Usually single	Usually uncertain
Diagnosis and prognosis	Usually accurate	Usually uncertain
Technological intervention	Usually effective	Often indecisive, adverse effects are common
Outcome	Cure possible	No cure
Uncertainty	Minimal	Pervasive
Knowledge	Professionals knowledgeable, patients inexperienced	Professionals and patients have complementary knowledge and experience

Source: Adapted from English Department of Health 2004

Health systems based on networks of semi-autonomous professionals and organizations struggle to ensure that the right combination of services is in the right place at the right time. In the past the standard response to complex illness was to restrict patients' movements by confining them to hospital beds to wait patiently for the appropriate services. This approach is still used in some countries but it is incompatible with a world in which patients with chronic diseases live and

work in the community and go to the services they need rather than waiting for those services to come to them.

In these circumstances it is perhaps inevitable that those with chronic health problems often receive less than optimal quality of care. Chronic conditions frequently go untreated or are poorly controlled until more serious and acute complications arise. Where those conditions are recognized, there is often a large gap between evidence-based treatment guidelines and current practice. McGlynn et al. (2003) demonstrated that only about 45% of individuals with diabetes in the United States at the end of the 1990s had received the recommended package of care. The proportion was somewhat higher for patients with congestive heart failure (64%) but was still suboptimal. Similarly, a systematic review of the quality of clinical care in general practice in Australia, New Zealand and the United Kingdom found that only 49% of patients with diabetes had undergone routine foot examinations and only 47% of eligible patients had been prescribed beta blockers after a heart attack, even in the highest-achieving practices (Seddon et al. 2001).

Change will require the institution of new managerial and organizational skills, backed up by effective information systems, but this can happen only if the role of health systems is re-conceptualized. This is of particular importance for monitoring performance. Too often, the discourse surrounding health systems is based on a model of acute care that is relatively much less important than it was. This is apparent in the ways that many politicians judge the performance of health systems. Their focus on waiting lists and the numbers of procedures undertaken recalls the statement attributed to Einstein: "not everything that can be counted counts, and not everything that counts can be counted."

The challenges of assessing how well health systems respond to chronic illness are examined in the next section.

Assessing performance: different dimensions

Before looking at the specific issues that arise with chronic diseases, it is helpful to recall that performance assessment of health systems has multiple dimensions – the nature of the assessment undertaken will depend on the dimension in which the proposed question lies.

The first dimension is the level at which assessment takes place. For example, the different levels of decision-making within a health-care system can range from the primary process of patient care (micro level) to the organizational context (meso level) to the financing and policy or health system context (macro level) (Plochg & Klazinga 2002). This can be illustrated with reference to diabetes.

Beginning with the primary process of patient care, an assessment of performance may focus on doctor-patient interaction to communicate inevitably complex messages about the natural history of the disorder and to set out the options to manage the disease in ways that are appropriate for the patient's lifestyle and aspirations. Such an assessment might draw on, for example, techniques based on conversational analysis (Maynard & Heritage 2005).

At the meso level, assessment might focus on the extent to which different aspects of the disease process are managed by the appropriate member of the clinical team or organization. Ideally the clinical management of diabetes will be located on a related measure of the quality of primary care – the extent to which admissions (for complications and diabetic emergencies) to hospital are avoided. This measure of avoidable hospitalization has been shown to vary with access to effective care (Billings et al. 1996).

At the macro level, the rate of diabetes-related blindness or amputation among a population of people with diabetes may serve as an indicator of the performance of the whole health-care system. For example, the OECD Health Care Quality Indicators Project has identified amputation rates in people with diabetes as a potential key indicator for international comparisons of health-care quality across OECD countries (Armesto et al. 2007). These end results capture the performance of many different health professionals, including those who manage the underlying disease, those who identify complications at an early stage and those who treat them once they arise.

Finally, much of the growing epidemic of type II diabetes is fuelled by rising levels of obesity. This can be ameliorated by healthy public policies directed at the relative price, availability and marketing of energy-dense foods (reflecting, for example, restrictions on advertising or the use of 'fat taxes') and opportunities for energy expenditure through physical activity (reflecting, for example, construction of recreational facilities and cycle lanes). Thus, the mortality from type II

diabetes might be considered a measure of the performance of govern-
ment as a whole, with high rates signifying a failure to enact appropri-
ate intersectoral health-promoting policies.

A second dimension differentiates the process and outcome of care.
A typical process measure – used in many structured diabetes disease
management programmes – is control of the metabolic disorder that
characterizes diabetes. This is undertaken by monitoring HbA1c levels
among patients to capture blood glucose levels over the preceding few
weeks (Knight et al. 2005). For example, the United Kingdom's Quality
and Outcomes Framework (Department of Health 2003) includes the
frequency of undertaking regular HbA1c tests on patients with diabe-
tes as a measure of quality of care. Many structured programmes use
a related outcome measure – the proportion of patients with diabetes
whose last HbA1c result was below a certain level.

Plochg and Klazinga (2002) argue for the necessity of considering
each of the three levels of decision-making in the health-care system
as each is characterized by distinct rationales addressing different
dynamics. Thus, decision-making at the micro level (where patient
care is delivered) is facing growing complexity due to the growth in
available knowledge and technologies; an increase in the managerial
complexity involved in the delivery of multidisciplinary health care;
and, especially, patients' increasing engagement in decision-making.

The importance of involving patients fully in their own care was
highlighted in the 1989 St. Vincent Declaration which set out a widely
accepted set of goals and principles for the prevention, diagnosis and
management of diabetes and its complications. This considers people
with diabetes to be members of a therapeutic partnership in which
they are linked with the various health professionals to whom they
look for advice as they negotiate an appropriate therapeutic regime.
Thus, performance measures must take account of the need to balance
the evidence that imposing a strict and inflexible regime of diet and
exercise will minimize the risk of complications against the knowledge
that this comes at the cost of precluding the patient from leading a
'normal' life.

Different rationales prevail at the macro level, largely related to
the question of how to allocate scarce resources in health care. For
example, policy-makers faced with competing demands may have to
decide whether the finite sums available are to be invested in the care
of people with one or other chronic disease, or whether they will be

used for the management of chronic disease or the reduction of waiting times for acute care. If the choices made at each level are not coordinated they can result in ambiguous goals, conflicting interests and excessive bureaucracy and ultimately limit the effectiveness of efforts to improve performance.

It is equally important to consider both process and outcome measures as they provide different, yet complementary, insights into the care process. Ultimately, the outcome of care is most important (e.g. in the amount of blindness, amputations and premature deaths avoided) but it is also important that those patients for whom these may be long-term outcomes receive care that is humane and reflects their expectations and lifestyles.

Finally, assessment of performance must take a broad perspective not least because the implementation of performance measures will change the behaviour of health-care providers, especially when supported by sanctions or incentives. This is an area that is fraught with the risk of unintended consequences as those whose performance is being assessed concentrate on what is being measured rather than what may be important.

The health system perspective

The preceding section showed how a comprehensive assessment of the ability to respond to chronic disorders necessarily requires evaluations of both process and outcomes, with inquiry at different levels. In this section the focus is on the level of the overall health-care system involved. Chronic disorders are complex and involve inputs from a wide range of health professionals equipped with appropriate knowledge and access to effective technology and pharmaceuticals. Hence, chronic disease is an ideal lens through which to assess the overall performance of the health-care system.

We propose a diagnostic hierarchy that involves a step-wise evaluation of health system performance. This approach begins by using existing data to identify potential problems. Normally this will not provide information on the precise reasons for any problem identified – this will require further steps using additional data. Once again we use the example of diabetes as an illustration.

Effective treatment reduces the risk of the disabling and potentially fatal complications of diabetes (Diabetes Control and Complications

Trial Research Group 1993; United Kingdom Prospective Diabetes Study Group (UKPDS) 1998; Writing team for the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group 2002) and the risk of premature cardiovascular disease (Diabetes Control and Complications Trial Research Group 1995; Gaede et al. 2003). For this reason, several commentators have argued that any death from diabetes in a young person is a sentinel health event that should raise questions about the quality of health-care delivery at the level of the organization concerned (Connell & Louden 1983; McColl & Gulliford 1993; Nolte et al. 2002). However, such deaths occur in all health-care systems although the rates vary substantially between countries.

The Diabetes Epidemiology Research International (DERI) study monitored cohorts of young people with type I diabetes in the United States, Japan, Israel and Finland. It found large differences in ten-year survival with the worst outcomes in the United States and Japan, and the best in Israel (DERI Mortality Study Group 1995). A separate study conducted by the British Diabetic Association (Laing et al. 1999) found that survival in the United Kingdom was comparable to that in Israel; the death rate in Japan was between four and five times higher than those in the United Kingdom or Israel.

A subsequent study drew on data collected in a standardized form during the WHO DiaMond and EURODIAB studies (Nolte et al. 2006). This data on the incidence of type I diabetes among children aged 0–14 was combined with data on mortality at ages 0–39 (selected to capture ages where certification of deaths attributable to diabetes was likely to be relatively reliable) to generate a mortality-incidence ratio. This study covered twenty-nine countries and confirmed the existence of very great differences in outcomes. Again, the worst results were obtained for a number of eastern European countries and Japan. The best outcomes were seen in some European countries with national health services, including the United Kingdom, Sweden, Spain, Italy and Greece. Clearly, such studies are dependent on the quality of recording of mortality and thus can only be undertaken in high- and some middle-income countries (see Chapter 2.1 on population health). It is also necessary to use only data on deaths at young ages as, although diabetes is often a contributory factor in deaths at older ages, there is considerable variation in recording practices.

These studies demonstrate that there is a remarkable variation in diabetes outcomes across countries. This suggests that there are gross differences in health systems' ability to provide adequate care for people with chronic diseases but gives little indication of why such differences exist. The next step therefore involves the study of data that can shed light on the immediate causes of death that drive the differences in order to highlight possible underlying organizational and system failures.

In the DERI study, much of the observed excess mortality in the Japanese cohort was attributable to diabetic renal disease (Diabetes Epidemiology Research International Mortality Study Group 1991). This reflected the higher incidence of end-stage renal disease and less access to dialysis than in the United States (Matsushima et al. 1995). Another study demonstrated how lower survival among individuals with type I diabetes in Estonia and Latvia (in comparison to Finland) was driven by much higher rates of the acute complications of diabetes (Podar et al. 2000).

These findings suggest the need to examine the specificities of the health systems in question, so the next step is more detailed assessment of the actual processes of care. For example, Tabak et al. (2000) compared the management of diabetes in Hungary and the United States and found that American patients were less likely to receive education about their condition; to see an ophthalmologist or diabetologist; or to perform self-monitoring of blood glucose. Hungarian patients had a lower prevalence of retinopathy, registered blindness and albuminuria (an indicator of kidney damage) but were more likely to experience severe hypoglycaemia (suggesting over-restrictive treatment). Again, this highlights the need to look holistically at processes and outcomes at all levels of care.

A holistic approach was demonstrated in a series of studies in the former Soviet Union, following an observation that death rates from diabetes among young people had risen markedly since 1991 – as much as eight-fold in some countries such as Ukraine (Telishevska et al. 2001). An analytical framework was developed in which four sets of inputs were identified as being essential for the delivery of effective care at the whole-system level: (i) human resources, in the form of an appropriate combination of skilled professionals and informed patients; (ii) physical resources, in the form of pharmaceuticals (e.g.

insulin and oral hypoglycaemics) and equipment (e.g. glucometers and reagent strips); (iii) knowledge resources, in the form of evidence-based clinical guidelines; and (iv) social resources, in the form of social support for patients. For patients to survive, the right combination of resources must be brought together in the right place and at the right time.

This framework was operationalized to create an instrument that could be used to undertake a rapid appraisal of a health system and was applied in Kyrgyzstan (Hopkinson et al. 2004) and Georgia (Balabanova et al. 2009). The studies identified an array of individual weaknesses but the overriding problem concerned integration. For example, individual health professionals would be trained abroad in methods of foot care but would be unable to obtain the inexpensive equipment required to provide it on their return. Patients would have glucometers but not the reagent strips required to use them. Newly diagnosed patients would be discharged from hospital without a supply of insulin and would become ill while they waited for the distribution system to make it available in their local pharmacy. The studies clearly highlight the multiple challenges that these two systems face in providing comprehensive diabetes care. They demonstrate how a single intervention (e.g. training health professionals in foot care, providing adequate supplies of insulin) to address a key problem in low-income settings (Beran et al. 2005) may be necessary but by no means sufficient to improve diabetes care in these settings.

This chapter has focused on diabetes for several reasons, chiefly because it is the easiest to study among the common chronic disorders. Diabetes is a very common condition: worldwide prevalence is estimated to be 2.8% (2000) and expected to increase to 4.4% by 2030 (Wild et al. 2004). The onset of type I diabetes is relatively acute and the diagnosis is unambiguous. This contrasts with conditions such as hypertension or chronic airways disease in which the onset of disease is more insidious and where many of those affected will not be identifiable. The required treatment of diabetes is largely uncontroversial and the natural history of the condition is both well-understood and modifiable by effective care, as outlined earlier. However, the system response to diabetes involves the delivery of integrated individualized care and thus is essentially the same as that required for patients with any (or multiple) chronic disorders. As such, it is uniquely placed to act as a marker of health system performance in the field of chronic

care. Essentially, a health-care system that is unable to deliver effective and timely care for patients with diabetes is unlikely to be able to do so for other chronic disorders (McKee & Nolte 2004).

Towards high-performing health systems

The preceding sections highlight the many challenges that exist in assessing the performance of health systems with regard to chronic disease. International comparisons of outcomes indicate clearly that health systems do matter and studies of the process of care identify the critical importance of coordinating the elements of care. Proposed models that seek to ensure the coordination of care have proven extremely difficult to evaluate – in part because they are often implemented in different ways in different settings (Wagner et al. 1999). The problems that need to be addressed may also differ between settings and make comparison problematic. Finally, those evaluations that have been undertaken have often been conducted in settings that cannot easily be generalized. Notwithstanding these problems, it is possible to propose some broad principles that are likely to underpin the delivery of optimal care for patients with chronic diseases (Singh 2005; Zwar et al. 2006).

The presence of appropriately skilled and motivated health professionals who have access to appropriate pharmaceuticals and technology and continuing professional development is a prerequisite for the delivery of optimal care. However, the challenge is how to organize them once they are in place.

Primary care plays a critical role. The complexity inherent in chronic disease means that patients will require assistance to navigate their path through the system in all but the simplest cases. This is best achieved by a partnership between the patient and his/her primary-care provider, with the latter able to take a holistic view of the patient's problems and propose solutions that are consistent with his/her lifestyle and expectations.

Multi-professional teams are important. The precise combination of skills required will vary with a patient's individual needs but will almost always include physicians, nurses and a range of other health professionals (e.g. dietitians, podiatrists, physiotherapists). There is now compelling evidence that physicians are not always the most appropriate providers of much of the routine care for chronic diseases

(Sibbald et al. 2004) and nurse-led clinics are becoming increasingly common in many countries (McKee et al. 2005). However, integrated care requires mechanisms that ensure strong linkages between all those involved in the delivery of care (Ouwens et al. 2005).

Patient self-management has been described as a 'cornerstone of treatment' (American Diabetes Association 2003) although the extent to which it is possible varies among different disease processes and in relation to the patient's functional ability, especially in terms of cognitive skills. Effective self-management gives patients greater motivation, skills and information. One study of diabetes identified this as the single most important factor in determining outcomes such as good metabolic control, reduced complication rates and hospitalization (Stam & Graham 1997). The means of supporting self-management are complex and can be resource intensive, requiring regular access to appropriate levels of care. Determination of the patient's needs, goals and treatment requires negotiation and not instruction (Fisher et al. 2005). It is much more than just patient education. Patient empowerment also requires strong health system governance structures that can secure patients' rights and protect vulnerable individuals.

Care should be responsive to the needs of patients and their carers, rather than trying to fit within rigid structures and models. It should be patient centred – 'respectful of and responsive to individual patient preferences, needs, and values and ensuring that patient values guide all clinical decisions' (National Diabetes Education Program 2005). The care of chronic disorders involves a partnership and therefore it should be delivered in ways that are acceptable to both patients and practitioners, ensuring that patients can participate fully in decision-making.

Care should also be evidence-based. The individual elements of the care process should be demonstrably effective on the basis of careful evaluations within representative samples of patients. This evidence should be available to practitioners in the form of guidelines and standards that should be sufficiently flexible to accommodate new technologies. However, this alone will not be sufficient to ensure high-quality care.

References

American Diabetes Association (2003). 'Standards of medical care for patients with diabetes mellitus.' *Diabetes Care,* 26(Suppl. 1): 33–50.

Armesto, S. Lapetra, M. Wei, L. Kelley, E (2007). *Health care quality indicators project*. 2006 data collection update report. Paris: OECD.

Balabanova, D. McKee, M. Koroleva, N. Chikovani, I. Goguadze, K. Kobaladze, T. Adeyi, O. Robles, S. (2009). 'Navigating the health system: diabetes care in Georgia.' *Health Policy Planning*, 24: 46–54.

Ben-Shlomo, Y. Kuh, D (2002). 'A life course approach to chronic disease epidemiology: conceptual models, empirical challenges, and interdisciplinary perspectives.' *International Journal of Epidemiology,* 31(2): 285–293.

Beran, D. Yudkin J. de Courten, M (2005). 'Access to care for patients with insulin-requiring diabetes in developing countries: case studies of Mozambique and Zambia.' *Diabetes Care*, 28(9): 2136–2140.

Billings, J. Anderson, GM. Newman, LS (1996). 'Recent findings on preventable hospitalizations.' *Health Affairs (Millwood)*, 15(3): 239–249.

Connell, FA. Louden, JM (1983). 'Diabetes mortality in persons under 45 years of age.' *American Journal of Public Health,* 73(10): 1174–1177.

Department of Health (2003). *Investing in general practice. The new general medical services contract*. (February 2003). London: Department of Health (http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4071966).

Department of Health (2004). *Chronic disease management: a compendium for information*. London: Department of Health.

DERI Mortality Study Group (1995). 'International analysis of insulin-dependent diabetes mellitus mortality: a preventable mortality perspective.' *American Journal of Epidemiology,* 142(6): 612–618.

Deutsches Zentrum für Altersfragen (2005). *Gesundheit und Gesundheitsversorgung. Der Alterssurvey: Aktuelles auf einen Blick, ausgewählte Ergebnisse* [Health and health provision. The Alterssurvey: News and views, selected results]. Bonn: Bundesministeriums für Familie, Senioren, Frauen und Jugend.

Diabetes Control and Complications Trial Research Group (1993). 'The effect of intensive treatment of diabetes on the development and pro-

gression of long-term complications in insulin-dependent diabetes mellitus.' *New England Journal of Medicine,* 329(14): 977–986.

Diabetes Control and Complications Trial Research Group (1995). 'Effect of intensive diabetes management on macrovascular events and risk factors in the Diabetes Control and Complications Trial.' *American Journal of Cardiology,* 75(14): 894–903.

Diabetes Epidemiology Research International Mortality Study Group (1991). 'International evaluation of cause-specific mortality and IDDM.' *Diabetes Care,* 14(1): 55–60.

Fisher, EB. Brownson, CA. O'Toole, ML. Shetty, G. Anwuri, VV. Glasgow, RE (2005). 'Ecological approaches to self-management: the case of diabetes.' *American Journal of Public Health,* 95(9): 1523–1535.

Freedman, V. Martin, L. Schoeni, R (2002). 'Recent trends in disability and functioning among older adults in the United States.' *Journal of the American Medical Association,* 288(24): 3137–3146.

Freedman, V. Schoeni, R. Martin, L. Cornman, J (2007). 'Chronic conditions and the decline in late-life disability.' *Demography*, 44(3): 459–477.

Fries, JF (1983). 'The compression of morbidity.' *Milbank Memorial Fund Quarterly Health & Society,* 61(3): 397–419.

Gaede, P. Vedel, P. Larsen, N. Jansen, G. Parving, H. Pedersen, O (2003). 'Multifactorial intervention and cardiovascular disease in patients with type 2 diabetes.' *New England Journal of Medicine,* 348(5): 383–393.

Hoffman, C. Rice, D. Sung, H (1996). 'Persons with chronic conditions. Their prevalence and costs.' *Journal of the American Medical Association,* 276(18): 1473–1479.

Hopkinson, BD. Balabanova, D. McKee, M. Kutzin, J (2004). 'The human perspective on health care reform: coping with diabetes in Kyrgyzstan.' *International Journal of Health Planning & Management,* 19(1): 43–61.

Janssen, F. Kunst, A (2005). 'Cohort patterns in mortality trends among the elderly in seven European countries, 1950–99.' *International Journal of Epidemiology,* 34(5): 1149–1159.

Knight, K. Badamgarav, E. Henning, J. Hasselblad, V. Gano, A. Ofman, J. Weingarten, S (2005). 'A systematic review of diabetes disease management programs.' *American Journal of Managed Care,* 11(4): 242–250.

Laing, SP. Swerdlow, AJ. Slater, SD. Botha, JL. Burden, AC. Waugh, NR. Smith, AW. Hill, RD. Bingley, PJ. Patterson, CC. Qiao, Z. Keen, H (1999). 'The British Diabetic Association Cohort Study, I: all-cause mortality in patients with insulin-treated diabetes mellitus.' *Diabetic Medicine,* 16(6): 459–465.

Matsushima, M. Tajima, N. LaPorte, RE. Orchard, TJ. Tull, ES. Gower, IF. Kitagawa, T (1995). 'Markedly increased renal disease mortality and incidence of renal replacement therapy among IDDM patients in Japan in contrast to Allegheny County, Pennsylvania, USA.' *Diabetologia*, 38(2): 236–243.

Maynard, DW. Heritage, J (2005). 'Conversation analysis, doctor–patient interaction and medical communication.' *Medical Education*, 39(4): 428–435.

McColl, AJ. Gulliford, MC (1993). *Population health outcome indicators for the NHS. A feasibility study*. London: Faculty of Public Health Medicine and the Department of Public Health Medicine, United Medical and Dental Schools of Guy's and St. Thomas' Hospitals.

McGlynn, E. Asch, S. Adams, J. Keesey, J. Hicks, J. DeCristofaro, A. Kerr, E. (2003). 'The quality of health care delivered to adults in the United States.' *New England Journal of Medicine*, 348(26): 2635–2645.

McKee, M. Nolte, E (2004). 'Responding to the challenge of chronic diseases: ideas from Europe.' *Clinical Medicine*, 4(4): 336–342.

McKee, M. Dubois, C-A. Sibbald, B (2005). Changing professional boundaries. In: McKee, M. Dubois, CA. Nolte, E. *Human resources for health in Europe*. Buckingham: Open University Press: pp.63–78.

National Diabetes Education Program (2005). *Making systems changes for better diabetes care*. Bethesda, MD: National Diabetes Education Program (http://www.betterdiabetescare.nih.gov/WHATpatientcenteredcare.htm).

Nolte, E. Bain, C. McKee, M (2006). 'Diabetes as a tracer condition in international benchmarking of health systems.' *Diabetes Care*, 29(5): 1007–1011.

Nolte, E. Scholz, R. Shkolnikov, V. McKee, M (2002). 'The contribution of medical care to changing life expectancy in Germany and Poland.' *Social Science & Medicine*, 55(11): 1907–1923.

Ouwens, M. Wollersheim, H. Hermens, R. Hulscher, M. Grol, R (2005). 'Integrated care programmes for chronically ill patients: a review of systematic reviews.' *International Journal for Quality in Health Care*, 17(2): 141–146.

Parker, MG. Thorslund, M (2007). 'Health trends in the elderly population: getting better and getting worse.' *Gerontologist*, 47(2): 150–158.

Perrin, J. Bloom, S. Gortmaker, S (2007). 'The increase of childhood chronic conditions in the United States.' *Journal of the American Medical Association*, 297(24): 2755–2759.

Plochg, T. Klazinga, NS (2002). 'Community-based integrated care: myth or must?' *International Journal for Quality in Health Care*, 14(2): 91–101.

Podar, T. Solntsev, A. Reunanen, A. Urbonaite, B. Zalinkevicius, R. Karvonen, M. LaPorte, RE. Tuomilehto, J (2000). 'Mortality in patients with childhood-onset type 1 diabetes in Finland, Estonia, and Lithuania.' *Diabetes Care,* 23(3): 290–294.

Seddon, M. Marshall, M. Campbell, S. Roland, M (2001). 'Systematic review of studies of quality of clinical care in general practice in the UK, Australia, and New Zealand.' *Quality in Health Care,* 10(3): 152–158.

Sibbald, B. Shen, J. McBride, A (2004). 'Changing the skill-mix of the health care workforce.' *Journal of Health Services Research & Policy,* 9(Suppl. 1): 28–38.

Singh, D (2005). *Evidence about improving care for people with long-term conditions.* Birmingham: University of Birmingham, Surrey and Sussex PCT Alliance.

Stam, DM. Graham, JP (1997). 'Important aspects of self-management education in patients with diabetes.' *Pharmacy Practice Management Quarterly,* 17(2): 12–25.

Suhrcke, M. McKee, M. Sauto Arce, R. Tsolova, S. Mortensen, J (2005). *The contribution of health to the economy in the European Union.* Brussels: European Commission.

Tabak, AG. Tamas, G. Zgibor, J. Wilson, R. Becker, D. Kerenyi, Z. Orchard, TJ (2000). 'Targets and reality: a comparison of health care indicators in the US (Pittsburgh Epidemiology of Diabetes Complications Study) and Hungary (DiabCare Hungary).' *Diabetes Care,* 23(9): 1284–1289.

Telishevska, M. Chenet, L. McKee, M (2001). 'Towards an understanding of the high death rate among young people with diabetes in Ukraine.' *Diabetic Medicine,* 18(1): 3–9.

TNS Opinion & Social (2007). *Health in the European Union.* Special Eurobarometer 272e. Brussels: European Commission.

United Kingdom Prospective Diabetes Study Group (UKPDS) (1998). 'Intensive blood-glucose control with sulfonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS33).' *Lancet,* 352(9131): 837–853.

Van den Akker, M. Buntinx, F. Metsemakers, J. Roos, S. Knottnerus, J (1998). 'Multimorbidity in general practice: prevalence, incidence, and determinants of co-occurring chronic and recurrent diseases.' *Journal of Clinical Epidemiology,* 51: 367–375.

Van der Lee, J. Mokkink, L. Grootenhuis, M. Heymans, H. Offringa, M (2007). 'Definitions and measurement of chronic health conditions in childhood.' *Journal of the American Medical Association,* 297(24): 2741–2751.

Wagner, EH. Davis, C. Schaefer, J. von Korff, M. Austin, B (1999). 'A survey of leading chronic disease management programs: are they consistent with the literature?' *Managed Care Quarterly,* 7(3): 56–66.

WHO (2004). *Health statistics and health information systems. Revised Global Burden of Disease (GBD) 2002 estimates*. Geneva: World Health Organization (http://www.who.int/healthinfo/global_burden_disease/ estimates_regional_2002_revised/en/).

Wild, S. Roglic, G. Green, A. Sicree, R. King, H (2004). 'Global prevalence of diabetes.' *Diabetes Care,* 27(5): 1047–1053.

Wilson, T. Buck, D. Ham, C (2005). 'Rising to the challenge: will the NHS support people with long-term conditions?' *British Medical Journal,* 330(7492): 657–661.

Wolff, J. Starfield, B. Anderson, GF (2002). 'Prevalence, expenditures, and complications of multiple chronic conditions in the elderly.' *Archives of Internal Medicine,* 162(20): 2269–2276.

Writing team for the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group (2002). 'Effect of intensive therapy on the microvascular complications of type 1 diabetes mellitus.' *Journal of the American Medical Association,* 287(19): 2563–2569.

Zwar, N. Harris, M. Griffiths, R. Dennis, S. Roland, M. Powell Davies, G. Hasan, I (2006). *A systematic review of chronic disease management*. Canberra: Australian Primary Health Care Research Institute.

4.3 | *Performance measurement in mental health services*

ROWENA JACOBS, DAVID MCDAID

Introduction

Mental health warrants a dedicated chapter within this book as it accounts for 14% of the global burden of disease. An estimated 450 million people worldwide are affected by mental health problems at any given time and one in five people will experience a psychiatric disorder (excluding dementia) within any given year (Horton 2007; WHO Regional Office for Europe 2003). Moreover, as we will indicate, assessment of the performance of mental health services presents challenges that may be unique within health care.

Within Europe, mental health problems account for approximately 20% of the total disability burden of ill health but often appear to be a lower policy priority than many other areas of health. This is despite the fact that nearly all countries readily admit that poor mental health has major impacts, not only on health but also on many other sectors of the economy (Taipale 2001).

The costs of poor mental health are conservatively estimated to account for 3%-4% of GDP in the European Union (EU) alone, yet none of these countries actually spends much more than 1% of GDP on mental health (Knapp et al. 2007). Differences in the boundaries between health and social care make cross-country comparisons difficult but health system funding for mental health in the EU ranges from almost 14% in England to much less than 4% in other countries including Bulgaria, the Czech Republic, Poland and Portugal.

One challenge for performance measurement is that many of the impacts of mental health go well beyond economic consequences – poor mental health has seriously marginalizing social consequences for individuals. These problems are compounded by deeply rooted stigma, fear, prejudice and discrimination; in some parts of Europe it remains effectively taboo to discuss the challenges that mental health raises for governments (Sayce & Curran 2007). Fundamental human

rights can also be affected as mental health is almost unique in its potential for compulsory detainment and treatment of individuals.

Another challenge arises because the organization and management of mental health services varies greatly within health-care systems across countries. A growing evidence base supports a community care centred approach, with substantial developments in pharmaceutical and psychosocial therapies and in services to help individuals reintegrate into the community. Many of these interventions appear to be cost effective in a variety of settings (Chisholm et al. 2004, Gutierrez-Recacha et al. 2006). This changing evidence base means that different countries are now at very different stages in rebalancing their mental health systems to make community based care the mainstay of the system. This principle was reaffirmed in the Mental Health Declaration for Europe and the Mental Health Action Plan for Europe endorsed by all fifty-two Members of the WHO European Region in 2005 (WHO Regional Office for Europe 2005).

Nearly all of western Europe has seen a shift in the balance of care with the closure of many psychiatric hospitals and the transfer of other beds to general hospitals. In much of northern Europe this has been accompanied by investment in social and community care based services. Mediterranean countries such as Italy, Portugal and Spain have made little investment in community based alternatives and much of the responsibility for support now rests with families. However, those services that are available often have very fragmented funding and delivery structures (McDaid et al. 2007), potentially leading to substantial variations in the type and quality of care provided (Hermann et al. 2006).

In contrast, very large and often isolated long-stay psychiatric hospitals and social care homes (*internats*) still dominate in much of central and eastern Europe. There are few incentives to change the balance of care, particularly where local communities rely on them for employment. The abuse of human rights within these institutions remains a key concern despite pressure from civil society organizations, the Council of Europe and judgements from the European Court of Human Rights (Parker 2007; Taipale 2001).

These challenges have caused the formal development of performance assessment procedures for mental health to lag behind that observed in many other sectors of the health system. Where aspects of performance have been assessed, measurement can be problematic.

Different countries have differences in social and cultural tolerance of what constitutes acceptable behaviour which in turn leads to differences in the size of the population deemed to have mental health problems. Assessment of the utilization of mental health services also needs to take account of the use of compulsory detention and treatment orders.

Some quality development initiatives sought better measurement of quality of life assessments by focusing initially on the cost effectiveness of some interventions, (Faria 1997), as well as monitoring the protection of human rights. However, the measurement of effectiveness can be complicated by difficulties with the reliability of psychiatric diagnoses and lack of consensus on the aetiology and treatment of many psychiatric illnesses (Evers et al. 1997). Moreover, in some limited circumstances, service users whose cognition is affected may find it difficult to express opinions and/or place a value on services received. Also, as with chronic conditions, the success of treatment may vary over time. In some circumstances it may be difficult to estimate the costs of treatment because of a lack of appropriate criteria for defining poor mental health. Crucially, as poor mental health can be stigmatizing, there is also a need to liaise with other sectors to measure key non-health outcomes such as changes in contact rates with the criminal justice system; levels of homelessness; and return to employment (Evers et al. 2007).

In this chapter we discuss some of the key developments in mental health performance measurement and provide international examples of how this has progressed. We reflect on the principal developments in the use of routine outcome and clinical process measurement. We also consider concerns about monitoring inequalities in mental health, looking at particular challenges for risk adjustment, attribution and causality. We end with a discussion of the key issues for mental health; the development of information technology and information management systems; and the policy implications of developments.

Performance measurement in mental health

As with other areas of health care, there are a number of potential dimensions for performance measures for mental health. Data on key outcomes and processes of care can facilitate improvement within pro-

vider organizations and provide insights into the quality and levels of performance that are feasible (Hermann et al. 2006). Many performance measures assess a range of aspects around the success of treatment, continuity, access, coordination and prevention; others may measure the treatment of specific disorders (Hermann et al. 2004b). In addition, there may be a set of useful performance measures specifically focused on carers.

Outcome measures in mental health can include health status (decrease in symptoms), social functioning, size of social network, quality of life, mortality, suicide, relapse and readmission. Non-health outcomes such as employment and housing status can be important. Process measures might include user satisfaction; rate of engagement and missed contacts; unplanned admissions or admissions under mental health legislation; length of stay; staff recruitment, retention and morale; as well as use of services and caseloads (Jenkins et al. 2000). More recently, some countries (e.g. Scotland) have begun to develop performance measures relating to mental health promotion and mental disorder prevention that incorporate measures of mental well-being or happiness (Health Scotland 2006; Tennant et al. 2007).

In principle, hundreds of performance indicators could be proposed for mental health system assessment but there may be huge variations in their evidence base, operational development, collection burden, availability, acceptability, reliability and validity. Stakeholders in the mental health-care system (e.g. payers, providers, regulators, clinicians, people with mental health problems and their families) often lack consensus on which aspects of performance should be used but several dimensions are considered to be of increasing importance. These include service access and integration and more user-focused standards of care such as responsiveness of service delivery, cultural appropriateness, consistency of services across a country and public protection (Clarkson & Challis 2002).

Recent work in Scotland investigated what would be the minimum requirements to help inform performance assessment. It was observed that, in the interim, systems do not have to be perfect. The "challenge is to develop good enough recording and reporting systems in the first instance that may only partially meet the needs of all the stakeholders, whilst developing a clear vision of the final shape of what is needed to support benchmarking and continuous improvement" (Donnelly 2008).

There have been a number of developments internationally – both to collect data on relevant performance indicators and to make use of these data within the context of performance measurement systems. Different dimensions of performance can be presented individually; form elements of a balanced scorecard comprising a range of measures across different domains; or be synthesized into a composite score or index of quality. For example, the reporting card systems being developed in Scotland use quality, efficiency, finance and future capability as the key dimensions (Donnelly 2008).

Reporting cards have long been used routinely in the United States (e.g. within VA-funded services) and have had a substantial impact on the types of care available and length of treatment (Rosenheck & Fontana 1999). Also, since 1986 the Colorado Division of Mental Health has implemented a performance contracting model to monitor a wide range of activity at both divisional and community mental health centre level. Indicators in the Colorado scheme are grouped around five dimensions considered important at the local level – financial viability, productiveness, responsiveness, comprehensiveness of services, outcomes. A standardized outcome measure is used to check compliance with standards.

In Australia, progress on the implementation of the National Mental Health Plan is assessed though examination of the delivery of services. For example, in the state of Victoria a number of different performance dimensions are monitored and a mental health dataset is collected. This covers information to support clinical standards at local level and planning and service standards at higher levels. Higher-level indicators include needs assessment, population indices, socio-economic status, homelessness and service utilization data, all of which are used for resource allocation purposes (Clarkson & Challis 2002). Supply-side indicators (e.g. number of beds and staff numbers per population) are used to monitor the shift towards more community-based care. Outcome indicators are also routinely collected.

Thus far, systems have tended to focus on administrative measures of quality because the data are more readily available and have lower collection costs (Druss et al. 1999). They also tend to be more developed for working age adult populations than for services for children and adolescents or older people. It can also be difficult to identify measurement approaches that specifically assess whether mental health systems meet the needs of minority populations. We now describe

some of the principal developments in outcome and process measurement for mental health.

Outcome measures

Challenges in measuring health outcomes

Outcome measures can be used as a performance measure if they are summarized across the service users of a particular provider or across providers (Manderscheid 2006). A conventional definition of an outcome in mental health care is, 'the effect on a patient's health status attributable to an intervention by a health professional or health service' (Andrews & Peters 1994, p.4).

This definition raises a number of concerns as the link between health service interventions and outcomes is far from straightforward. Firstly, outcomes can also improve as a result of self-help, environmental changes or support from professionals outside the health sector. Moreover, maintaining (rather than improving) an individual's health status may be viewed as a positive outcome in some circumstances. Outcomes may also vary with different perspectives (e.g. of the clinician, person with mental health problems, their family or professional carer). Mental health interventions may also be delivered at different levels, for example using specific treatments, combinations of treatments or population-wide interventions. Outcomes may vary at these different levels and make outcome measurement in mental health extremely complex (Gilbody & Whitty 2002).

Routine outcome assessment requires either the clinician or the service user to monitor and rate changes in health status. Such outcome assessment reflects service-user reports of internal psychic phenomena which cannot be observed or verified externally. Classification systems such as the ICD diagnose illness according to the presence or absence of mental symptoms that are 'subjective' in their nature. This is not to say that there has not been significant work in producing standardized instruments to diagnose psychiatric disorders in a reliable manner and quantify the degree of severity of a disorder. The range of measures available tend to measure the frequency and intensity of specific psychiatric symptoms (psychopathological rating scales) or are instruments that judge a disorder's impact on the individual (measures of social functioning and global measures of outcome, or quality of life

assessment). A wide number of these rating scales are used in psychiatric research or clinical trials but few are used routinely in clinical practice – too few to allow performance monitoring.

Clinicians complete most rating scales in psychiatry as the user voice has largely been ignored in the development of various instruments to rate health outcomes. Recently there has been more attention on the importance of the user voice and patient choice in decision-making (Ford 2006). Ideas of 'partnership' and 'shared decision-making' are becoming key in service delivery in some settings (Bower & Sibbald 1999). A multidimensional approach to rating which could incorporate user, clinician and family reports has been suggested (Dickey & Sederer 2001). However, clinicians and users have shown little agreement in ratings between different scales or even when using the same instrument (Garcia et al. 2002; Kramer et al. 2003). Nonetheless, clinician-, family- and user-rated instruments are now used routinely and successfully alongside each other in a number of settings. These are discussed in the next section.

International efforts towards routine health outcome assessment

Routine outcome measurement has been undertaken using a range of instruments and assessment scales internationally. Much of this work had been led by initiatives in Australia and the United States.

Australia

Australia has the most coherently developed approach to treatment-level routine outcome assessment. The first national mental health strategy included a systematic review of patient outcomes (Andrews & Peters 1994) which led to proposals for specific instruments for routine use. These instruments were independently field-tested for their utility; the resulting recommendations informed Australian practice in routine outcome assessment (Meehan et al. 2002).

The use of standard outcome measures for all mental health service users was mandated (Brooks 2000). All Australian states have signed agreements to submit routinely collected outcomes and casemix data to the Australian government on a regular basis (Callaly et al. 2006). This has involved a substantial commitment of resources by mental health providers and has produced a large national dataset.

Table 4.3.1 *Mandated outcome measures in Australia*

	Adult services	Child and adolescent services (CAMHS)	Older people's services
Clinician-rated	HoNOS Abbreviated Life Skills Profile (LSP)	Health of the Nation Outcome Scales for Children and Adolescents (HoNOSCA) Children's Global Assessment Scale (CGAS) Factors Influencing Health Status (FIHS)	Health of the Nation Outcome Scales for Elderly People (HoNOS 65+) Abbreviated Life Skills Profile (LSP) Resource Utilization Groups – Activities of Daily Living (RUG-ADL)
User-rated	BASIS-32 K-10+ MHI-38	Strengths and Difficulties Questionnaire (SDQ)	BASIS-32 K-10+ MHI-38

The measures mandated for use in Australia are listed in Table 4.3.1. Different combinations of indicators are used for those in receipt of adult, older people's or child and adolescent mental health services (CAMHS) (Callaly et al. 2006). All groups make use of the Health of the Nation Outcome Scales (HoNOS)[1] that include specialist variants for children and older people. Originally developed by the Royal College of Psychiatrists in England, the basic form of this instrument contains twelve items measuring behaviour, impairment, symptoms and social functioning on a five-point severity scale (Wing et al. 1996).

In addition to HoNOS, all adult and older people's mental health services are required to offer consumers one of three user-rated (self-report) instruments. Victoria, Tasmania and the Australian Capital Territory use the Behavior and Symptom Identification Scale (BASIS-32); New South Wales, South Australia, the Northern

[1] HoNOS is mandatory in Australia, England and New Zealand. There are also substantial programmes of use in Nova Scotia, Canada, the Netherlands, Norway and Italy.

Territory and Western Australia use the Kessler 10 plus (K-10+); and Queensland uses the Mental Health Inventory (MHI-38). All CAMHS are required to use the same self-report measure – the Strengths and Difficulties Questionnaire (Callaly et al. 2006).

There are mixed perceptions of the value of the outcome measurement system in Australia (Meehan et al. 2006). User-rated outcome measures are well-valued when they are seen to help service users to identify their own needs while allowing for better dialogue with clinicians and helping them to see the service-user point of view (Callaly et al. 2006). In practice, the greater the severity of illness the lower the likelihood that a service user will be offered the chance to complete the self-report measure. Those with more severe symptoms may also be more likely to decline to use the measure.

In contrast to user-rated outcome measures, the collection of clinical outcome data has received a much more lukewarm response. Initially, the majority of clinicians have perceived the Australian government's primary objective for introducing the measures to be financial management rather than to ensure the quality of services. Another limitation is that HoNOS cannot also be used to measure mental health outcomes in general practice. However, some acknowledge that national data collection could support the ability to compare services and treatment types and thus lead to more efficient and effective services (Callaly & Hallebone 2001, Callaly et al. 2003). This resistance to the use of outcome measures is not unique to Australia; the dominant driving force for the use of outcome measurement has been the need for aggregate data for management and accountability purposes rather than a desire to improve direct clinical utility.

England
In the early 1990s, the government's health strategy set the improvement of health and social functioning of people with severe mental health problems as its first mental health target and proposed that success against this target should be quantified (Department of Health 1992). The *Health of the Nation* led to the creation of the HoNOS instrument. A National Service Framework was also introduced in 1999. This put an emphasis on clinical governance and practice guidelines, service-user experience and the need to collect outcome data (Department of Health 1999). This framework and the increased focus

on performance management were both intended to make managers more accountable through routine inspections; audit and publication of comparative data; and by encouraging engagement in activities that previously may not have been taken seriously (Rea & Rea 2002).

In 2002, 49% of all English mental health service providers were using HoNOS in at least one service delivery site; only 11% were routinely using the instrument in all service settings; and 34% were using the instrument routinely in more than half of their service settings. Collection of the Mental Health Minimum Dataset (MHMDS) for England, including HoNOS, became mandatory for all mental health provider organizations in the NHS in April 2003 (Appleby 2004).

The Mental Health Minimum Dataset is not specifically an indicator format but it can support the use of patient-centred indicators and is used by the Healthcare Commission (the regulator, now called the Care Quality Commission) at a more aggregate level for performance monitoring. A review by an outcomes advisory expert group concluded that local providers would need to develop expertise and systems to make effective use of the newly available outcomes data in order for the new system to inform local service delivery in England (Fonagy et al. 2004). However, work undertaken in Canada suggests that access to improved support materials and the use of initiatives to increase completion rates (including timely feedback to clinicians) can be useful at individual, team and service levels to significantly improve the uptake and ease of use of HoNOS (Kisely et al. 2008).

Netherlands

Overall assessment of health system performance in the Netherlands in 2006 includes a section devoted specifically to the mental health (including substance abuse) system, based on core indicators on mental health related outcomes. These include the uptake of prevention measures by target groups; changes in mental and social functioning (using the Global Assessment of Functioning – GAF); suicides and suicide attempts; discharge rates from the mental health system; and the percentage of the target population reached by professionals (Westert & Verkleij 2006). A mental health-care thermometer, a twenty-question instrument recording service-user satisfaction with involvement in treatment and care decisions has been introduced. In future this will allow service-user views of the system to be incorporated into the analysis.

United States of America

In the United States, the focus on outcome measurement as a measure of success has been driven largely by cost containment efforts. As in several European countries, difficulties in accurately quantifying the resources needed for DRGs for mental health and the increasing proportion of health expenditure devoted to mental health have led to a growing emphasis on outcome measures (Slade 2002). Purchaser-driven pressures have driven activity in routine outcome assessment here more than anywhere else. Outcomes measurement is increasingly being implemented in both public (e.g. VA) and private programmes.

Payers have variable mandates for outcomes measures and they are used more widely in specialist rather than generic managed-care organizations. Clinician ratings are used in some state hospitals (Ford 2006) and also within the VA mental health system where clinicians use the GAF tool to assess all mental health inpatients at discharge and all outpatients at least every ninety days of active treatment (Greenberg & Rosenheck 2005). The VA chose to use this tool because it had been used routinely for inpatient discharges since 1991 and therefore training needs were limited. Further implementation was incentivized by introducing a national performance measure on GAF recording compliance, with monitoring published monthly. Implementation was supported by national training initiatives.

User-rated instruments are used in the commercial public sector (for instance, in Medicaid carve-outs by some private psychiatric hospitals) and within some public mental health systems. Mental health service users have also been involved in the development of some outcome measurement systems, as illustrated in Ohio (Ohio Department of Mental Health 2007) (see Box 4.3.1).

Other outcome measures

Readmission rates

Measures other than specific outcome scales can be used to assess outcomes. These include rates of readmission to inpatient care services. The reductions in average length of stay observed in many high-income countries are more likely to be effective if appropriate levels of community based care and support are in place. Any increase in readmission rates might thus be seen as a potential indicator of poor

> **Box 4.3.1 Ohio Mental Health Consumer Outcomes System**
>
> A development task force commissioned by the Ohio Department of Mental Health focused on identifying what mattered to service users and their families. Pilot projects found that consumers liked being asked about their lives and seeing their outcomes instruments used in discussions with staff about their treatment plans.
>
> The final approach, the Consumer Outcomes System, uses three instruments for adults and three for children and their families. The adult instruments include two service-user orientated outcome measurement instruments; those for children have one instrument targeted at young service users and a second targeted at parents/guardians.
>
> In 2003, the state introduced a rule requiring service providers to implement the Consumer Outcomes System. Implementation has been supported by training, technical support and subsidies. As of March 2005, reports were being generated by 277 provider agencies with records for 211 000 service users (Ford 2006).

quality initial treatment (including premature discharge) or it might reflect failure in the provision and quality of community based services (Lyons et al. 1997). However, several reviews have concluded that readmission rates are not a suitable indicator of quality of care in psychiatric hospitals, although appropriate discharge planning and follow-up visits may be associated with lower rates of readmission (Durbin et al. 2007; Lien 2002).

Readmission data require careful interpretation. Some studies suggest that a co-morbid substance-related disorder is the best predictor of readmission in a public hospital setting (Haywood et al. 1995, Lyons & McGovern 1989). Across countries there are often significant barriers in the cross-referral of patients with dual diagnoses to mental disorder and substance abuse treatment programmes. Readmission rates may also offer useful information for service providers on general admission policies and thresholds for admission. Subsequent analysis of the medical necessity of admissions might also be undertaken.

The availability of crude data on readmission rates in many countries can be misleading. There are a number of reasons why it may be

problematic to determine accurately the rate of readmission. One key challenge in identifying whether treatment has been ineffective is that service users may be free to move between different public (and private) hospitals. This requires data to have unique patient identifiers that can be tracked not only over time but also to link each discharge with subsequent readmission in any facility for the same condition. Many national datasets are unable to meet these requirements. Moreover, individuals may also be re-institutionalized in facilities outside the health-care system, for example in social care facilities or within the prison system (Priebe et al. 2005). Such facilities are often not included in data collection systems. Another practical problem is that individuals who are readmitted may be treated primarily for a physical rather than a mental health problem. This reflects not only the high rate of physical co-morbidity in people with mental health problems but also the fact that tariffs set for health conditions may not cover the full costs of care (Halsteinli et al. 2006).

Suicide

Rates of suicide and deaths from unidentified causes are another commonly used measure for looking at the performance of both mental health treatment services and population-wide mental health strategies. For instance, suicide rates are used as a key indicator in assessing mental health performance against the National Service Framework for Mental Health in England (Department of Health 1999).

The majority of suicides are linked to mental health problems (Wilkinson 1982). Many people who ultimately complete suicide have come into contact with health (and other) care services. Appropriate suicide awareness training for front-line staff can be effective in reducing suicides by helping to identify individuals who may be particularly at risk (Mann et al. 2005). This suggests that some cases are potentially avoidable through appropriate early intervention from health and other services. Suicide rates may therefore be a good indicator of how well health and other local services in general are meeting the needs of people with mental health problems. High rates of suicide or undetermined death might suggest further investigation into areas such as access to treatment and the level of training for professionals at primary care level; integration of primary, secondary and social care services; clinical, organizational, staffing and resource management

in psychiatric services; and follow-up procedures for service users (Renvoize & Clayden 1990).

Data on suicides are available in virtually all high-income countries but there are major challenges in using suicide rates as an indicator of a health system's effectiveness in dealing with mental health problems. Many factors well beyond the health system may influence rates of suicide, including changes in the economic climate, social isolation and rapid societal change as seen (for instance) in central and eastern Europe (Berk et al. 2006). This suggests the need for adequate risk adjustment for some of these factors.

At a statistical level some groups in the population have high suicide rates (e.g. older people, young men) but, even when including deaths from undetermined causes, the absolute number of deaths from suicide is often too low to assess change over time. This problem can be addressed to some extent by using data over a longer time period, for example over three years instead of one.

Another potential confounder in using suicides as a possible performance indicator for mental health is differences in the procedures for recording the cause of death in different countries. For instance, some require a coroner's investigation but may still have different legal definitions of suicide (Renvoize & Clayden 1990); others require police reports (e.g. at the site of a motor vehicle crash) before determining whether a suicide is recorded. Cultural and religious taboos may also discourage physicians and others from recording a death as suicide (Kelleher et al. 1998).

Physical health problems

One major gap in assessing changes in outcome for people receiving treatment for mental health problems are impacts on their physical health status. The evidence base consistently indicates that the mortality rates from many physical illnesses, most notably cardiovascular disease and diabetes, are significantly higher for people living with enduring mental illness than for those in the general population (Harris & Barraclough 1998; Fleischhacker et al. 2008). This is observed regardless of the type of mental health problem. People living with psychoses such as schizophrenia and those with more common problems (e.g. anxiety and depressive disorders) can be at greater risk of physical health problems (Osborn et al. 2007).

Moreover, the adverse effects of most antipsychotic medications for people with severe mental health problems include excessive weight gain (Allison et al. 1999; Newcomer 2005). People with depression and anxiety-related disorders are also at increased risk of weight gain – there is good evidence that long-term use of many older antidepressants (tricyclics) and of newer generation heavily prescribed selective serotonin reuptake inhibitors (SSRIs) can result in weight gain (Demyttenaere & Jaspers 2008; Gartlehner et al. 2008; Ness-Abramof & Apovian 2005).

There are strong links between poor mental and poor physical condition. To date, performance indicators have typically looked neither at changes in physical health status nor at whether individuals with mental health problems are treated for co-morbid physical health problems or receive advice and support to help minimize potential adverse health impacts of some treatments.

Is there any evidence that outcome measurement leads to service improvement?

There is consensus that outcomes should be routinely measured but is there any evidence that this is effective in improving services in any way? Overall evidence from various reviews seems scant (Gilbody et al. 2003) or mixed at best (Gilbody et al. 2001). The latter systematic review found only nine studies that looked at the addition of outcome measurement to routine clinical practice in both psychiatric and non-psychiatric settings. The results show that routine feedback of instruments had little impact on the recognition of mental disorders or longer term psychosocial functioning. Clinicians welcomed the information gained from the instruments but rarely incorporated these results into routine clinical decision-making. Given that routine outcome measurement can be costly the authors concluded that there was no robust evidence to suggest that it is of benefit in improving psychosocial outcomes in non-psychiatric settings (Gilbody & Whitty 2002).

Similarly, studies suggest that one-off outcome measurements do very little to shift clinical practice or change clinician behaviour (Ashaye et al. 2003). A more recent randomized controlled trial (Slade et al. 2006) on the effectiveness of standardized outcome measurement indicated that monthly outcome monitoring markedly reduced

psychiatric admissions. However, it was not shown to be effective in improving primary outcomes of patient-rated unmet need and quality of life, nor did it improve other subjective secondary outcome measures. The study was longitudinal in nature and had more regular outcome measurement for patients (month on month assessment) and showed that this can prompt earlier intervention by clinicians to avert relapse which would otherwise lead to hospitalization, thus reducing admissions. The intervention therefore reduced psychiatric inpatient days and resulting service use costs and proved cost effective.

More evidence can be found in a six-country European study (Priebe et al. 2002) that examined how service-users' views could be fed into treatment decisions. The MECCA (Towards More Effective European Community Care for Patients with Severe Psychosis) trial tested the hypothesis that intervention would lead to better outcomes in terms of quality of life over a one-year period. A better outcome was assumed to be mediated through more appropriate joint decisions or a more positive therapeutic relationship. Results showed that while the intervention added time to clinical appointments it did lead to a significant improvement in quality of life.

The key message from these studies appears to be that one-off (or infrequent) outcome measurement seems to have equivocal results in terms of actually improving subjective outcomes. However, outcome measurement that is performed longitudinally and more regularly using a broad range of measures (ideally collected routinely in databases and backed up by regular monitoring) can significantly improve quality of life and/or reduce psychiatric admissions.

Process measures

A number of process measures related to mental health services can help to track performance variations within and between different providers. Typically process measures are used because they are more readily available in administrative datasets. Indicators of input (i.e. the level of resources invested in mental health) are a key component of many process measures.

Typical process measures include indicators such as length of stay and various measures of bed use or occupancy rates (Glover et al. 1990). These can include trends in very long stay service users (i.e. those living in institutions for more than one year). Other hospital-

centric input measures can include the size of the hospital (number of inpatients) and staffing throughput measures, for example the number of service users per consultant, per nurse or per therapist (Geddis 1988). These crude ratios may provide useful information on staffing mixes, dependency levels and workload.

In Norway, for example, several process indicators for mental health are collected within the national system for measurement of quality within the health system – proportion of treatment undertaken compulsorily; waiting times for first outpatient consultation; duration of untreated psychosis; and the number of children and adolescents who have been diagnosed as having a mental health problem. In addition, in 2009 the government has commissioned the independent research organization SINTEF to publish information on service utilization, the number of therapists per service use and the skill mix/balance between psychologists and psychiatrists (Halsteinli 2008).

Community and ancillary service inputs that may be measured include quantification of the activities of community mental health teams supporting people to live in their homes; the provision of emergency out-of-hours services; and access to occupational rehabilitation services, sheltered housing and day care services (Jenkins & Glover 1997). Inputs from primary care services (e.g. general practitioners, nurses, health visitors, counsellors) also need to be counted on some notional basis, for example –the average number of patients presenting in primary care with a mental health problem. Other measurable indicators recently identified as important to quality assessment in Scotland include reducing and changing the pattern of antidepressant prescribing and then assessing whether or not any savings from these actions are reinvested in effective psychological therapies (Donnelly 2008). Box 4.3.2 provides an example of how traditional inpatient focused process indicators are being supplemented by additional community service indicators in Ireland (Health Research Board 2008).

Service-user experiences

In addition to data on inputs into the mental health system, data recording levels of service-user satisfaction are being used increasingly to help assess quality of care. The interest in assessing service-user satisfaction has been driven by a number of concerns. Service-user satisfaction with care has been found to be associated with better concordance

---

Box 4.3.2 Collection of mental health system process indicators in Ireland

In Ireland, the Health Research Board's Mental Health Research Unit collects a range of information. This includes the National Psychiatric Inpatient Reporting System that has recorded all admissions and discharges to inpatient psychiatric hospitals and units throughout the country – as well as related socio-demographic, diagnostic and service related information – over forty years.

WISDOM is a new system being developed to gather information on the use of both community based and inpatient mental health services. Also, part of the Health Research Board's 2007–2011 research programme will work towards the further development of mental health specific performance indicators; an objective of the national mental health strategy – A Vision for Change.

A proof of concept phase of WISDOM will be tested in the Donegal Local Health Area and comprehensively evaluated before the system is implemented more widely throughout the country. Evaluation of the proof of concept phase began in January 2008 with a review of evaluation literature, with a specific focus on the evaluation of information systems, user-focused evaluation and evaluation of training.

---

with treatment and outcomes. Also, there has been a shift towards greater consumer rights and a growth in mental health user movements (Callan & Littlewood 1998; Rose & Lucas 2007). Satisfaction measures may be useful to clinicians and managers because they can provide information on processes (e.g. satisfaction with treatment) as well as outcomes of care (e.g. a perspective on the success of treatment – see section on outcome measures).

Early studies seemed to report consistently high levels of user satisfaction with mental health services, often surpassing professionals' expectations (Kalman 1983). They also suggested that service users might have been reluctant to voice critical comments for fear of damaging the therapeutic relationship (Warner et al. 1994). Certainly there is a vocal community of individuals who regard themselves as 'survivors' of the psychiatric system (Rose & Lucas 2007).

The evidence on whether patient demographics are associated with satisfaction appears mixed (Lebow 1982) although some studies show some correlation with age, gender, legal status and ethnicity. Women, younger people, those involuntarily detained and ethnic minority service users historically may have had lower levels of satisfaction with the care that they received (Greenwood et al. 1999; Hansson 1989; Leavey et al. 1997; Perreault et al. 1996). Service users who were dissatisfied also tended to report more adverse experiences (Greenwood et al. 1999). Again, the reasons for different levels of patient satisfaction are complex – certain diagnostic categories (such as drug abuse or diagnoses of schizophrenia) tend to be associated with lower levels of satisfaction but other studies have found social problems to be more important than diagnosis in influencing satisfaction (Babiker & Thorne 1993).

As with other areas of the health system, there are a number of concerns when collecting what can be costly and time-consuming data on service-user satisfaction (Druss et al. 1999). For instance, there are risks that surveys suffer from both response and recall bias and it is not clear to what extent expressions are associated with prior expectations (Babiker & Thorne 1993; Callan & Littlewood 1998). Some questionnaires have also been too reductionist – it is not sufficient to know that service users are dissatisfied without knowing why. Many instruments have also been criticized for asking patients to rate only those aspects of care that the provider deems important rather than those which are important to service users (Rose et al. 2006). In addition, performance measures are usually conducted at provider level while data are collected at individual patient level and therefore require satisfaction scores to be aggregated to the provider level.

The detailed survey used in England and Wales is one example of an instrument that has been tailored to look at a range of issues. As Box 4.3.3 indicates, this gathers data on a number of different dimensions of service use that are of importance not just to service providers but also to service users.

Use of guidelines

It has been suggested that guidelines can help to improve quality of care by advocating evidence-based practice models with a view to improving patient outcomes and reducing variations in treatment

---

Box 4.3.3 Service-user satisfaction surveys in England and Wales

The Healthcare Commission has conducted a detailed survey of community mental health service users in England and Wales since 2004. This looks at the quality of care; communication with health professionals, crisis care and psychotherapy; and access to other support including help for family carers and social inclusion. The results of the survey are fed back to NHS providers with the aim of helping them to improve performance. In 2007, 75% of 15 900 service users in the survey reported care received to be good, very good or excellent; 81% indicated that their psychiatrist was 'definitely listening to them' (Healthcare Commission 2007). Reports are also prepared for the sixty-nine individual primary care providers, comparing service-user satisfaction against national benchmarks.

(Weinmann et al. 2007). The development and use of guidelines and national service plans for the promotion of mental health and for the treatment and rehabilitation of people with mental health problems are now considered of great importance in many countries. Well-developed guidelines and strategies are available (e.g. National Service Framework for Mental Health in England and Wales) but many guidelines and national service plans remain of low quality, leading some commentators to argue for the creation of institutions to support pan-national development of guidelines (Stiegler et al. 2005).

As with other areas of the health system, evidence also suggests that guideline implementation tends at best to have a modest impact on patient outcomes for a limited duration. Ongoing support or feedback has been identified as important in changing physician behaviour and improving patient outcomes on the back of guideline implementations (Bero et al. 1998; Grol 2001). Even if the performance of mental health professionals can be influenced, improving guideline adherence may not necessarily lead to better outcomes. Guidelines may be too artificial if the external validity of the trials on which they are based is limited by select patient groups (Weinmann et al. 2007). A corollary is that guideline adherence may be a poor performance measure for providers and a poor proxy measure for patient outcomes.

Table 4.3.2 *Twelve-month service use by severity of anxiety, mood and substance disorders in WMH Surveys (%)*

	Severe	Moderate	Mild	None
Belgium	60.9	36.5	13.9	6.8
France	48.0	29.4	21.1	7.0
Germany	40.0	23.9	20.3	5.9
Italy	51.0	25.9	17.3	2.2
Netherlands	50.4	31.3	16.1	7.7
Spain	58.7	37.4	17.3	3.9
USA	59.7	39.9	26.2	9.7

Source: Adapted from Wang et al. 2007

Inequalities in access and utilization

Inequalities in mental health care raise particular challenges, not only for the organization and management of services but also for how systems are able to monitor such inequalities in order to improve performance. The majority of those with mental disorders do not come into contact with mental health services (Thornicroft 2008). The challenge can be illustrated by looking at World Mental Health (WMH) Survey data on the use of services for anxiety, mood and substance abuse disorders. Conducted across seventeen countries, this survey reported that overall only around one third of those who could benefit from treatment actually made use of services (Wang et al. 2007).

Table 4.3.2 provides data on seven of the countries included in the WMH Surveys. Among individuals with the most severe of these mental disorders at least 39% (Belgium) and at most 60% (Germany) did not receive any treatment. Table 4.3.3 also indicates that no more than 42% of those who actually received services obtained what was deemed to be a minimally adequate level of treatment for their disorder. There were also substantial variations in the proportion of those with more severe disorders who received adequate treatment.

Again, there are complex reasons for low utilization of mental health services. The stigma surrounding poor mental health appears to be a major contributor to a lack of contact with services (Schomerus & Angermeyer 2008) and anticipated discrimination appears to deter

Table 4.3.3 *Minimally adequate treatment use for respondents using services in the WMH Surveys in previous twelve months (% of people by degree of severity)*

	Any severity	Severe	Moderate	Mild	None
Belgium	33.6	42.5	35.5	-	29.4
France	42.3	57.9	36.5	41.5	40.2
Germany	42.0	67.3	53.9	-	35.4
Italy	33.0	-	33.4	-	31.0
Netherlands	34.4	67.2	34.1	-	20.8
Spain	37.3	47.5	43.6	48.5	29.2
USA	18.1	41.8	24.8	4.9	-

Source: Adapted from Wang et al. 2007

people from coming into contact with services (Corrigan & Wassel 2008). Individuals may be fearful of being discriminated against if they are labelled as having a mental health problem. This under-utilization of services is reported even in those countries that require no out-of-pocket payments to access services. As members of the general population, these individuals are also exposed to common misconceptions surrounding mental disorders – for instance that they cannot be cured or that drug treatments do not work.

Contact rates also differ by mental health problem – highest for severe psychotic conditions (e.g. schizophrenia) but much lower for conditions perceived to be less serious (e.g. depression) (Wittchen & Jacobi 2005). Again this may be due to a lack of knowledge about mental health problems. People with psychosis may be more likely to come to the attention of services during the acute phases of their condition but there is some evidence to suggest that the general public do not believe that conditions such as depression always require intervention from mental health services. It is believed that these are caused by socio-environmental events or may reflect individual weakness – individuals just need to 'get a grip'(Thornicroft 2007). Troubling patterns of interaction with mental health services tend to include under-representation in outpatient care and over-representation in inpatient and emergency care. Failure to receive outpatient care may be associated with higher rates of hospitalization and longer lengths of stay.

Rates of contact with mental health services may also be lower in specific population groups than in the general population. The stigma of mental illness may be particularly acute in young people with mental health problems – one study reported that only 4% of these young people contacted their primary care practitioner about their problems (Potts et al. 2001). Performance measures need to be able to identify differences by population subgroups. One approach used in assessing Oregon's Medicaid State Plan compared population-based average health utilization data against normative benchmarks or performance guidelines for particular mental disorders and then examined outliers or unusual behaviour among provider organizations (McFarland et al. 1998). Guidelines were then risk adjusted to take account of co-morbidity in the target population and the outcome measured was the level of functioning. It was found to be a major challenge to incorporate outcomes data into administrative and claims databases which measured treatment processes.

Racial and ethnic disparities have also been demonstrated to lead to differences in the rates and patterns of treatment in mental health services. Many studies show that the probability of being diagnosed with schizophrenia is much higher among minority populations (Chow et al. 2003; Tapsell & Mellsop 2007). Afro-Caribbean people have been at higher risk of involuntary commitment and are likely to be referred by legal means, for example under the United Kingdom's Mental Health Act (Callan & Littlewood 1998; Fearon et al. 2006; Mohan et al. 2006), making the use of services more coercive.

There may also be a lack of cultural sensitivity in the provision of care, or taboos within the community. In some sections of the population there may be a tendency to attribute mental health problems to religious and other culturally sanctioned belief systems and lack of access to receptive culturally sensitive providers (Chow et al. 2003). People with mental health problems tend to be over-represented in poor neighbourhoods with high rates of unemployment, homelessness, crime and substance abuse and members of racial and ethnic minorities tend to be disproportionately represented in poor areas. The relationship between ethnicity, poverty and mental health service use is therefore complex.

These findings suggest the need to tailor services more carefully to meet the needs of minority groups; ensure fewer disparities in service access and use; and carefully monitor appropriate pathways in

care. All of these concerns raise challenges for performance measurement within mental health systems. The issue is of particular interest in many western European countries experiencing recent new inward economic migration from countries in central and eastern Europe as these new migrants can be highly vulnerable to mental health problems. In addition, refugees present very different challenges to mental health systems as individuals may experience severe post-traumatic stress disorders. Yet, not one of eighteen OECD countries recently surveyed had the most basic of data on service follow-up for ethnic minority groups (Garcia-Armesto et al. 2008).

Reviews of services in Europe suggest that few mental health services are yet equipped to meet these needs (Watters 2007; Watters & Ingleby 2004). In New Zealand, culturally specific measures of mental health status (Hua Oranga) are being used to help develop appropriate outcome measures and performance indicators integral to the National Mental Health Information Strategy. This experience may be of use to those seeking to develop equally culturally appropriate indicators in other countries (Ministry of Health 2006).

Productivity measurement

The literature on price indices for mental health care in the United States is particularly relevant for measurement of the productivity of mental health services. Rising expenditure on mental health has generated considerable interest in constructing price indices, in particular for major depression, schizophrenia and bipolar disorder. The literature indicates that it is important to focus on the direct medical costs of treating an episode of illness rather than changes in the prices of the inputs used in treatment. For all three disorders, studies suggest that the price of treating an episode or individual have declined in recent years. This is contrary to many of the officially reported figures, for example those from the Bureau of Labour Statistics.

This literature improves on previous methods by attempting to define the units of output of medical care that reflect the changing bundles of inputs required to treat these problems. Output is also defined in a way that incorporates measures of the quality of treatment. Outputs had been considered solely in terms of services used in the treatment of disease, for example physician visits, hospital stays, prescriptions. The newer approach views these as inputs into the treat-

ment of mental health problems. Output is viewed as a course of treatment over a specified period, combining a number of treatment inputs which produce health benefits. The studies focus on the episodes of poor mental health. This involves pooling a number of treatments into bundles that are ex ante expected to lead to similar outcomes. This conception of output allows for a change in the composition of inputs or substitution among inputs as a result of technological change.

Many of the studies show that changes in the composition of treatment enable treatment episode costs to fall, even when input costs are rising. Berndt (2004) argues that this can be explained by the fact that official (Bureau of Labour Statistics) statistics do not make allowances for changes in the mix of treatment over time. Studies have also reported a considerable shift over time in the composition of treatment for depression (Berndt et al. 1998; Berndt et al. 2001; Berndt et al. 2002; Frank et al. 1998; Frank et al. 1998a), schizophrenia (Frank et al. 2004; Frank et al. 2006) and bipolar disorder (Ling et al. 2004).

For example, the studies found a shift in the mix of treatment for depression over recent years. The combination of psychotherapy and tricyclic antidepressants (TCAs) is being replaced by the use of newer selective SSRIs, sometimes in combination with psychotherapy. The move away from more costly psychotherapy-intensive treatment to less costly psychopharmacological treatments has had a significant impact on the average cost of treating an episode of acute phase major depression. Since expenditures on depression were thought to have increased over the study period, the source of this increase was likely to be an increase in volume rather than price as the cost of treating an episode of depression fell. Quality also improved because episodes that met guideline standards increased over the period (Berndt 2004).

Similarly, for schizophrenia, one study constructed treatment bundles which consisted of both single treatments (e.g. any antipsychotic medication) and more than one form of treatment such as medication and psychotherapy (Frank et al. 2004). Output was defined as the course of treatment over an entire year, given that schizophrenia is a severe and persistent mental disorder. The study found significant compositional changes in treatment with various forms of psychosocial therapy and older pharmaceutical treatments being replaced by newer atypical antipsychotics, in line with guidance. It was concluded that, as the cost of treating an individual per annum had declined, the

observed increase in overall expenditure indicated that there had been an increase in the number of individuals being treated. Compositional changes in the types of treatment for bipolar disorder have been more gradual than those for either depression or schizophrenia. Four treatment bundles were defined: no treatment; psychotherapy only; mood stabilizers only; and psychotherapy and mood stabilizers combined (Berndt 2004; Ling et al. 2004).

Taking the evidence from the above studies, one recent study examined the level and composition of all mental health spending in the United States (Berndt et al. 2006). Quality-adjusted price indices for several major mental disorders (anxiety; schizophrenia; bipolar disorder; major depressive disorders; and all others) were applied to national mental health expenditure account estimates to examine changes in real output for the whole mental health sector. The study used estimates on depression, schizophrenia and bipolar disorder from previous research and aggregated results across all categories of mental health problem to arrive at overall price indices. These price indices reveal large gains in real output (70%–75%) relative to those used by the Bureau of Labour Statistics (16%–17%).

An alternative to calculating price and output indices for productivity calculations is to use a non-parametric approach such as data envelopment analysis (DEA) to calculate a productivity index. DEA was used to calculate a Malmquist productivity index for Norwegian psychiatric outpatient clinics to examine whether any change is related to personnel mix, budget growth or financial incentives (Evers et al. 2007). Bootstrapping methods were used to construct confidence intervals for the technical productivity index and its decomposition. A second stage regression was run on the productivity index to examine variables that may potentially be statistically associated with productivity growth. Overall the study reported substantial technical productivity growth. Personnel growth had a negative impact on productivity growth but a growth in personnel with university education increased productivity. Other than taking staff education as a proxy for staff quality on the input side, this study did not take account of any other changes in the quality of the output or interventions over time. The researchers call for more research to explore this. Further data on productivity in the Norwegian mental health system will be published in 2009 (Halsteinli 2008).

Risk adjustment

Comparisons of performance across different providers and over time rely on the assumption that organizations have similar basic characteristics and structures. This is seldom the case in mental health as services can be highly diverse. Moreover, there is a strong association between poor mental health and socio-economic deprivation. This greatly increases the need to make more equitable comparisons between mental health providers serving different populations. Risk adjustment in performance measures can be used to take account of differences in factors that are beyond facilities' control (Schacht & Hines 2003). One objection to statistical methods of risk adjustment is that the confounding cannot be completely removed as groups may differ on a number of characteristics other than the risk-adjustment variable used (Dow et al. 2001). Risk adjustment is only ever a partial fix but it allows more equitable and valid comparisons.

Statistical adjustment is not expected to make groups more comparable on all confounding variables but rather to make them more equal than they would have been with no adjustment (Hendryx & Teague 2001). The goal is to reduce the risk of drawing incorrect conclusions about the performance of some providers. Variables used to take account of group differences in the mental health context include age, gender, legal status and admission-referral source. It is often particularly challenging to control for casemix in mental health – DRGs (and their equivalents in other countries) are typically used for casemix and are based on diagnosis but they have been shown to be problematic and poor predictors of service use (Halsteinli et al. 2006; McCrone 1995).

There has been a lot of work on the risk adjustment of outcomes for specific interventions in mental health and some on risk adjustment for the development of payment systems (Ettner et al. 1998). However, there has been very little work on the risk adjustment of indicators for the purpose of comparing the performance of multiple providers (Dow et al. 2001).

Hendryx et al. (1999) developed models for risk adjusting outcome data to compare provider performance. Demographic and diagnostic data were used to risk adjust client functional status, quality of life and satisfaction ratings. Risk adjustment resulted in somewhat differ-

ent rankings of provider performance although there was no statistical comparison of rankings with and without adjustment. Dow et al. (2001) risk adjusted two outcome measures (global rating of functioning; consumer satisfaction measure) using data on 7000 individuals over a three-year period from 24 state-funded providers in Florida. There was significant variation between providers on the two outcome measures but the risk adjustment had a fairly small impact on their overall rank ordering. However, it had a major effect for a few specific providers, particularly those with small caseloads.

Data comparability across providers and data quality largely determines whether these types of risk-adjustment models can be implemented in practice. The Behavioral Healthcare Performance Measurement System (BHPMS) for state psychiatric inpatient facilities in the United States is one example of the use of risk adjustment to facilitate benchmarking (see Box 4.3.4).

There is very little use of such risk adjustment mechanisms outside the United States and a number of challenging questions must be answered in order to facilitate their development and greater use. For example, does the collection of service-user self-report and clinician-rated variables make a difference to models built exclusively on the demographic and clinical indicators available in administrative databases? Investment in resources to collect additional data may not be merited if models from administrative databases perform as well (Hendryx & Teague 2001). Inappropriate or ineffective risk adjustment raises the possibility that providers will treat performance comparisons with scepticism, mistrust or even active opposition, thereby jeopardizing any performance measurement system. On the other hand, valid risk-adjustment models may encourage providers to use comparative findings as an opportunity for improvement.

Expanding the dimensions of performance assessment

Potentially important indicators of performance may lie outside the health system yet are influenced by inputs from it. A major report on social inclusion and mental health in England highlighted the importance of reintegration into employment. It reported that health service professionals were reluctant to encourage individuals to seek employment for fear that they might be unsuccessful and would have diffi-

---

Box 4.3.4 Making use of risk adjustment in performance measurement

The United States National Association of State Mental Health Program Directors Research Institute developed the BHPMS for state psychiatric inpatient facilities. The programme covers around 240 psychiatric facilities in 50 states and is approved by the Joint Commission.[2]

A standardized set of data definitions and reporting requirements allows the development of benchmarks. A risk-adjustment method using logistic regression is applied using individual and organizational characteristics that show significant relationships. Monthly performance data allow the models to be updated if necessary (Schacht & Hines 2003). A time-series graphical display with confidence intervals is developed for each indicator for each organization and sent to providers in a confidential report.

Risk-adjustment models have now been developed for readmission, seclusion and restraint. The characteristics used in the models include age, gender, race, marital status, diagnoses, living arrangements, legal status and referral source on admission. Institutional characteristics include unit mission (expected length of hospitalization) and specialty, bed capacity, security level and locked status.

Each organization's rate of performance is now compared to a predicted risk-adjusted rate for the specific population that it serves. This represents an improvement on the previous system in which each service was simply compared against the average.

---

culty in regaining social welfare benefits (Social Exclusion Unit 2004). Yet employment has been shown to be a protective factor for mental health. One randomized controlled trial in six European countries has shown that supported employment schemes (in which health professionals work alongside specialist employment staff) are highly effective in helping people with severe mental health problems to return to work (Burns et al. 2007).

The promotion of reintegration into the workplace is a specific goal of mental health policy in England. However, it is challenging to mea-

---

[2]   Previously known as Joint Commission on Accreditation of Healthcare Organizations and Affiliates.

sure the performance of the mental health system by taking account of inputs from outside the health service – namely the workplace. There are inputs within the workplace where employers may contribute to the promotion, prevention or treatment of mental health problems (Jenkins & Glover 1997). It is extremely challenging for the public mental health system to gauge accurately the contributions made by managers, human resources teams and occupational health teams in private companies. Indeed, few countries are able to measure these inputs accurately amongst public sector employers.

Mental health services can have inputs in partnership with other sectors including housing and education. For example, potential mental health inputs in a school setting might include a notional share of the contribution made by teachers or educational psychologists (Jenkins & Glover 1997). This could be calculated by looking at the epidemiology of mental health problems in schools or the extent of specific help given to pupils in schools.

Performance data and IT

Information systems and the development of databases and informatics in mental health remain one of the biggest challenges for performance measurement. Information systems and databases provide vital information for performance assessment and performance management for: assessing needs; resource management and planning; joint working between health and social care professionals; ensuring the effective delivery of appropriate care; measuring the effectiveness of different treatments and different settings; clinical audit and research; more refined contracting; and assessing costs (Jenkins & Glover 1997).

The measurement of performance in mental health is often opportunistic and piecemeal, reflecting the availability of data rather than performance dimensions that should be measured and monitored. The shift from hospital-based to community care; hospital closures; and the reconfiguration of services have largely not been accompanied by investment in computing systems. This makes it difficult to evaluate policies and develop services on a sound basis for decision-making (Glover 1995).

The geographical dispersal of many services to smaller community sites requires the development of wider computer networks. Furthermore, the nature of care is changing significantly – moving

towards an integrated care pathway that is multidisciplinary in nature. Typically, integrated care cannot be identified readily as datasets still tend to be episodic and based on hospital care alone. Many information systems were not appropriately networked and datasets that have been available have tended to produce data that are inappropriate or unhelpful (Glover 1995).

Data analysis for performance management purposes still tends to be focused at the macro level; it is less common for individual teams or clinicians to use electronic data collection systems to guide decision-making at the micro level (Clarkson & Challis 2002). Moreover, policy-makers, providers and purchasers require different types of information to make decisions about the numbers of service users to treat; range of clinical problems; outcomes of care; and value of the services provided. Rea and Rea (2002) suggest that there should be a distinction between performance management and the management of performance and their very different informational requirements. Performance management requires information after the event and is used to make comparisons and devise league tables between different organizations for central government purposes. The management of performance requires users and practitioners to be involved in the development of systems and routines.

Routine collection of data requires careful and explicit definition of which data items are to be collected and the points in the care pathway at which data returns are to be made. Historically, hospital admission and discharge have been the main triggers for data returns but these systems of data collection are no longer suitable. Clinical staff tend to be more accurate at data recording than administrative staff but will have significant involvement in the data gathering process only if it has some clinical value. Computerized information systems should be designed to ensure that they meet the information requirements of clinical professionals and can safely replace a paper-based system (Jenkins & Glover 1997). An audit of information systems and their local use can help to identify gaps in systems that may be addressed as part of a performance measurement system (Donnelly 2008).

Collection of the Mental Health Minimum Dataset for England has been mandatory within the NHS since April 2003. Information on mental health service use stored within an electronic record has been recognized to be critical to the usefulness of this. When electronic records are fully implemented it will be possible to monitor outpatient

attendances which may extend over many years as well as hospital, community and day care attendances which may commonly overlap. For each institution it should be possible to track the characteristics of the patient; health organizations involved; nature of the problems, including their range and severity; amounts of different interventions delivered to the patient; the way these interventions are combined as packages and scheduled over time; and changes in the patient's condition over time. Cost data are not included.

Outside the United Kingdom, there is still very limited use of unique identifiers for individual service users to enable system performance to be tracked. A recent survey reported that individual service-user records could be linked to different output measures in only six out of seventeen countries (Garcia-Armesto et al. 2008). Denmark is one such country, collecting highly detailed administrative data on health service use by people with mental health problems. Such information is absent in Australia where it has proved difficult to develop computer systems that reliably collect useful data and provide feedback and reports that are of sufficient quality to help clinicians and managers to guide service development (Callaly et al. 2005). Nonetheless, there has been a tremendous effort to develop an electronic medical record and to reduce duplication of data collection by different health agencies involved with the same patient (i.e. to integrate electronic health records between service providers) (Callaly et al. 2005).

In contrast, the routine datasets that provide activity data for Medicaid billing in the United States are extremely well-kept, up to date and almost entirely accurate (Huxley & Evans 2002). When a capitation scheme was introduced in Colorado State it was feared that data quality would decline because of the lack of direct financial incentive, however the State countered this by offering mental health providers a cash incentive for the best outcomes (Huxley & Evans 2002).

Conclusions

Poor mental health is one of the principal causes of disability and morbidity worldwide. It has a major impact on economies and public health but typically has not received the requisite level of policy priority in comparison to other areas for health action. Of course, additional resources cannot be invested in mental health (or any other

aspect of health) without ensuring that the proposed interventions are of high quality; meet the needs of service users; are distributed fairly; lead to improvements in health and other outcomes; and are likely to be cost effective.

Monitoring the many dimensions of performance of the mental health system can help to facilitate better use of the resources allocated to mental health. However, these performance measurement systems face what may be unique challenges – defining the social and cultural boundaries of what constitutes poor mental health; difficulties in making diagnoses; and ensuring that there is a clear understanding of the different elements of service provision. For instance, outpatient care is very different to community care yet is sometimes used as an indicator of the implementation of the latter.

Issues of human rights and dignity are of paramount importance given that the mental health system uses involuntary detention and treatment in some circumstances. It is increasingly recognized that mental health system impacts on health outcomes cannot be assessed by looking at changes in mental and physical health status alone. Other key outcomes include individuals' ability to live independently and, particularly, to return to employment. Additional measurement difficulties are created by poor quality data and by shifting boundaries between health, social care and other sectors, e.g. where vocational rehabilitation services may be provided.

To a large extent, progress in performance assessment to date has depended on political agendas and the differing national priorities accorded to mental health. The majority of developments have been initiated in the United States but there are different examples of how this is being driven forward across the globe. Often initiatives are undertaken at regional level. For example, the Australian government developed national standards reflecting a number of important dimensions of performance for the national mental health strategy initiated in 1992. These nationally agreed indicators have since been monitored in different ways across the different states and territories (Andrews & Peters 1994; Rosen et al. 1989).

Significant developments in performance assessment initiatives are in place or due to be implemented in some parts of Europe, notably in England, Iceland, Ireland, the Netherlands, Norway and Scotland. However, these are exceptions rather than the rule. Wahlbeck's (2006) recent survey of twenty-five European Union countries noted that data

on suicide rates and the number of psychiatric beds were readily available but other data were scarce. The report concluded that 'clearly, there is a need for Member States to develop their mental health monitoring systems'. A survey of eighteen OECD countries suggests that much information that would be useful to performance assessment is already available (e.g. in Denmark or Sweden) but is not used as part of a performance assessment process (Garcia-Armesto et al. 2008).

Drives towards performance monitoring have often been initiated through a desire to inform programmes and systems or to reduce expenditure rather than to inform treatment decisions for individual service users. This means that some systems may have been designed to use data that meet the needs of policy-makers or system managers rather than clinical staff. Of course, it is essential to provide information to inform policy-making. However, there may be an adverse impact on implementation if clinicians perceive the process of performance measurement as a threat or a paper-filling exercise, with no clinical value. This challenge was acknowledged in the development of a new benchmarking system for mental health in Scotland. This stressed the need to set up an expert implementation group charged with working with local health bodies and other stakeholders to develop and agree on the dimensions of the system to be measured in order to help facilitate uptake. The recommendations also emphasize the need for stakeholders to work together to align costs with service definitions and functions (Donnelly 2008).

We have highlighted the challenge posed by the need not only to develop effective information and data systems that make use of administrative data but also (ideally) to use integrated data systems with information on measurable and appropriate indicators across the different dimensions of performance. Initiatives to develop and make use of such indicators can be identified. Both the OECD and the European Union have recognized the importance of mental health performance indicators and are developing plans to monitor aspects of mental health in member countries, although these policy drives are still in their infancy. The OECD HCQI project identified a number of measures for international benchmarking of the quality of mental health care (Hermann et al. 2004a; Hermann et al. 2006). Actual benchmarking has been delayed because of the difficulties in ensuring common definitions of services across countries (Garcia-Armesto et al. 2008).

The MINDFUL (Mental Health Information and Determinants for the European Level) project also put forward a plan for a comprehensive mental health information system to cover not only mental health problems but also positive mental health, mental health promotion and the prevention of mental disorders (Lavikainen et al. 2006). Supported by the European Commission, this project has been revising mental health indicators that appear in the European Community Health Indicators list in order to support the development of the proposed European Health Survey System.

At a European level, WHO has relied on self report by countries to publish some basic data on the structure of mental health systems within the region. However, some of the major variations in the availability and balance of services in this report can be attributed to difficulties in obtaining comprehensive data and in how different countries (despite the provision of guidance) defined different services and types of mental health and related professionals (Petrea & Muijen 2008).

At a global level, WHO has developed the Assessment Instrument for Mental Health Systems (WHO-AIMS) to collect essential information on the mental health system of a country or region. Both a brief and a long-form instrument are provided to collect a broad range of data in a common format across countries, primarily low- and middle-income. A number of European countries are participating including Portugal, Greece and Ukraine (WHO 2005a). WHO has also published two editions of an atlas on adult mental health. This contains brief basic information on the structure of mental health systems on a country by country basis, including the development of new policies, funding for mental health and the level of resources available (WHO 2005a). This information has many limitations and gaps but has increased awareness of disparities in coverage for mental health across Europe and elsewhere. Atlases on child mental health and people with learning difficulties are also available.

Policy-makers face another key challenge – it is not sufficient to improve access to information on the services provided within the health-care system alone. The greater focus on the promotion of mental well-being in health policy-making across Europe and beyond also implies the need to develop initiatives that promote and maintain this. Developments in indicators for well-being are still in their infancy. We have already noted that the boundaries between health and social care vary considerably across countries (McDaid et al. 2007).

Clearly, performance frameworks that can integrate data from health and social care and provide a coherent set of performance measures have considerable advantages (Clarkson & Challis 2002). However, key services and supports may also be provided entirely outside of this system. For instance, interventions may be delivered by education services within a school setting and employment services may focus on helping and supporting individuals with mental health problems to be fully integrated into the workplace. Such developments will become critical as policy-makers increasingly embrace the language of service-user empowerment and choice. They are also necessary for adequate assessment of new mechanisms for funding mental health services. This includes the direct allocation of budgets to service users which in theory allows them to purchase services that best meet their needs – within health, social care or other sectors.

Finally, from a policy-making perspective, institutional arrangements need to be in place to promote participation in any system of performance assessment. Our analysis indicates that improvements in system performance can be encouraged by mechanisms such as collaboration with multiple stakeholders in system design; financial incentives; routine data collection; and feedback to providers. Some emerging evidence suggests that performance assessment may also help to improve individual health outcomes but much more evaluation and analysis is required. There should be careful consideration of how information arising from performance assessment systems can best be used to help facilitate change in both policy and practice.

References

Allison, DB. Mentore, JL. Heo, M. Chandler, LP. Cappelleri, JC. Infante, MC. Weiden, PJ (1999). 'Antipsychotic-induced weight gain: a comprehensive research synthesis.' *American Journal of Psychiatry,* 156(11): 1686–96.

Andrews, G. Peters, L (1994). Measurement of consumer outcome in mental health. In: CRUFAD (ed.). *A report to the National Mental Health Information Strategy Committee.* Sydney: CRUFAD.

Appleby, L (2004). *The national service framework for mental health – five years on.* London: Department of Health.

Ashaye, O. Livingston, G. Orrell, M (2003). 'Does standardized needs assessment improve the outcome of psychiatric day hospital care for

older people? A randomized controlled trial.' *Aging & Mental Health,* 7(3): 195–199.

Babiker, IE. Thorne, P (1993). 'Do psychiatric patients know what is good for them?' *Journal of the Royal Society of Medicine,* 86(1): 28–30.

Berk, M. Dodd, S. Henry, M (2006). 'The effect of macroeconomic variables on suicide.' *Psychological Medicine,* 36(2): 181–189.

Berndt, ER (2004). 'Changes in the costs of treating mental health disorders: an overview of recent research findings.' *Pharmacoeconomics,* 22(2 Suppl. 2): 37–50.

Berndt, ER. Bir, A. Busch, SH. Frank, RG. Normand, SL (2002). 'The medical treatment of depression, 1991–1996: productive inefficiency, expected outcome variations, and price indexes.' *Journal of Health Economics,* 21(3): 373–396.

Berndt, ER. Busch, SH. Frank, RG (1998). *Price indexes for acute phase treatment of depression.* Cambridge, MA: National Bureau of Economic Research (NBER Working Paper no. 6799).

Berndt, ER. Busch, SH. Frank, RG (2001). Treatment price indexes for acute phase major depression. In: Cutler, DM. Berndt, ER (eds.). *Medical care output and productivity.* Chicago: University of Chicago Press.

Berndt, ER. Busch, AB. Frank, RG. Normand, SL (2006). 'Real output in mental health care during the 1990s.' *Forum for Health Economics & Policy,* 9(1): 1–17.

Bero, LA. Grilli, R. Grimshaw, JM. Harvey, E. Oxman, AD. Thomson, MA (1998). 'Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. The Cochrane Effective Practice and Organization of Care Review Group.' *British Medical Journal,* 317(7156): 465–468.

Bower, P. Sibbald, B (1999). 'On-site mental health workers in primary care: effects on professional practice.' *Cochrane Database of Systematic Reviews,* Issue 4.

Brooks, R (2000). 'The reliability and validity of the Health of the Nation Outcome Scales: validation in relation to patient derived measures.' *Australian and New Zealand Journal of Psychiatry,* 34(3): 504–511.

Burns, T. Catty, J. Becker, T. Drake, RE. Fioritti, A. Knapp, M. Lauber, C. Rossler, W. Tomov, T. van Busschbach, J. White, S. Wiersma, D (2007). 'The effectiveness of supported employment for people with severe mental illness: a randomised controlled trial.' *Lancet,* 370(9593): 1146–1152.

Callaly, T. Hallebone, EL (2001). 'Introducing the routine use of outcomes measurement to mental health services.' *Australian Health Review,* 24(1): 43–50.

Callaly, T. Arya, D. Minas, H (2005). 'Quality, risk management and governance in mental health: an overview.' *Australasian Psychiatry,* 13(1): 16–20.

Callaly, T. Coombs, T. Berk, M (2003). 'Routine outcome measurement by mental health-care providers.' *Lancet,* 361(9363): 1137–1138.

Callaly, T. Hyland, M. Coombs, T. Trauer, T (2006). 'Routine outcome measurement in public mental health: results of a clinician survey.' *Australian Health Review,* 30(2): 164–173.

Callan, A. Littlewood, R (1998). 'Patient satisfaction: ethnic origin or explanatory model?' *International Journal of Social Psychiatry,* 44(1): 1–11.

Chisholm, D. Sanderson, K. Ayuso-Mateos, JL. Saxena, S (2004). 'Reducing the global burden of depression: population-level analysis of intervention cost-effectiveness in 14 world regions.' *British Journal of Psychiatry,* 184: 393–403.

Chow, JC. Jaffee, K. Snowden, L (2003). 'Racial/ethnic disparities in the use of mental health services in poverty areas.' *American Journal of Public Health,* 93(5): 792–797.

Clarkson, P. Challis, D (2002). 'Developing performance indicators for mental health care.' *Journal of Mental Health,* 11(3): 281–294.

Corrigan, PW. Wassel, A (2008). 'Understanding and influencing the stigma of mental illness.' *Journal of Psychosocial Nursing and Mental Health Services,* 46(1): 42–48.

Demyttenaere, K. Jaspers, L (2008). 'Review: Bupropion and SSRI-induced side effects.' *Journal of Psychopharmacology,* 22(7): 792–804.

Department of Health (1992). *The health of the nation: a strategy for England*. London: HMSO.

Department of Health (1999). *National service framework for mental health: modern standards and service models*. London: Department of Health.

Dickey, B. Sederer, LI (2001). *Improving mental health care: commitment to quality*. Washington, DC: American Psychiatric Publishing, Inc.

Donnelly, RR (2008). *Mental Health Project final report: National Benchmarking Project. Report 2*. Edinburgh: The Scottish Government.

Dow, MG. Boaz, TL. Thornton, D (2001). 'Risk adjustment of Florida mental health outcomes data: concepts, methods, and results.' *Journal of Behavioral Health Services & Research,* 28(3): 258–272.

Druss, BG. Rosenheck, RA Stolar, M (1999). 'Patient satisfaction and administrative measures as indicators of the quality of mental health care.' *Psychiatric Services,* 50(8): 1053–1058.

Durbin, J. Lin, E. Layne, C. Teed, M (2007). 'Is readmission a valid indicator of the quality of inpatient psychiatric care?' *Journal of Behavioral Health Services and Research,* 34(2): 137–150.

Ettner, SL. Frank, RG. McGuire, TG. Newhouse, JP. Notman, EH (1998). 'Risk adjustment of mental health and substance abuse payments.' *Inquiry,* 35(2): 223–239.

Evers, S. Salvador-Carulla, L. Halsteinli, V. McDaid, D. The MHEEN Group (2007) 'Implementing mental health economic evaluation evidence: building a bridge between theory and practice.' *Journal of Mental Health,* 16(2): 223–241.

Evers, SM. Van Wijk, AS. Ament, AJ (1997). 'Economic evaluation of mental health care interventions. A review.' *Health Economics,* 6(2): 161–177.

Faria, JG (1997). 'Quality development in mental health care in Europe. Recent contributions by WHO.' *Epidemiologia e Psichiatria Sociale,* 6(Suppl. 1): 211–215.

Fearon, P. Kirkbride, JB. Morgan, C. Dazzan, P. Morgan, K. Lloyd, T. Hutchinson, G. Tarrant, J. Fung, WL. Holloway, J. Mallett, R. Harrison, G. Leff, J. Jones, PB. Murray, RM (2006). 'Incidence of schizophrenia and other psychoses in ethnic minority groups: results from the MRC AESOP Study.' *Psychological Medicine,* 36(11): 1541–1550.

Fleischhacker, WW. Cetkovich-Bakmas, M. De Hert, M. Hennekens, CH. Lambert, M. Leucht, S. Maj, M. McIntyre, RS. Naber, D. Newcomer, JW. Olfson, M. Osby, U. Sartorius, N. Lieberman, JA (2008). 'Comorbid somatic illnesses in patients with severe mental disorders: clinical, policy, and research challenges.' *Journal of Clinical Psychiatry,* 69(4): 514–519.

Fonagy, P. Matthews, R. Pilling, S (2004). *The Mental Health Outcomes Measurement Initiative: report from the Chair of the Outcomes Reference Group.* London: National Collaborating Centre for Mental Health.

Ford, D (2006). 'Counting hoops: measuring the quality of mental health services in the US.' *The Mental Health Review*, 11:15–20 (http://www.commonwealthfund.org/Content/Fellowships/Harkness-Fellowships/2004-2005-Fellows/Dominic-Ford--U-K.aspx).

Frank, RG. Berndt, ER. Busch, AB. Lehman, AF (2004). 'Quality-constant "prices" for the ongoing treatment of schizophrenia: an exploratory study.' *Quarterly Review of Economics and Finance,* 44(3): 390–409.

Frank, RG. Berndt, ER. Busch, SH (1998). *Price indexes for the treatment of depression.* Cambridge, MA: National Bureau of Economic Research (NBER Working Paper no. W6417).

Frank, RG. Busch, SH. Berndt, ER (1998a). 'Measuring prices and quantities of treatment for depression.' *American Economic Review,* 88(2): 106–111.

Frank, RG. McGuire, TG. Normand, SL (2006). *Cost-offsets of new medications for treatment of schizophrenia.* Cambridge, MA: National Bureau of Economic Research (NBER Working Paper no. W12643).

Garcia-Armesto, S. Medeiros, H. Wei, L (2008). *Information availability for measuring and comparing quality of mental health care across OECD countries.* Paris: OECD.

Garcia, R. Joseph, T. Turk, J. Basu, R (2002). 'A comparison of parent and therapist ratings of outcome in a child mental health clinic.' *Child and Adolescent Mental Health,* 7(4): 168–172.

Gartlehner, G. Thieda, P. Hansen, RA. Gaynes, BN. Deveaugh-Geiss, A. Krebs, EE. Lohr, KN (2008). 'Comparative risk for harms of second-generation antidepressants : a systematic review and meta-analysis.' *Drug Safety,* 31(10): 851–865.

Geddis, PW (1988). 'Health service performance indicators.' *Ulster Medical Journal,* 57(2): 121–128.

Gilbody, S. Whitty, P (2002). 'Improving the delivery and organisation of mental health services: beyond the conventional randomised controlled trial.' *British Journal of Psychiatry,* 180: 13–18.

Gilbody, S. Whitty, P. Grimshaw, J. Thomas, R (2003). 'Educational and organizational interventions to improve the management of depression in primary care: a systematic review.' *Journal of the American Medical Association,* 289(23): 3145–3151.

Gilbody, SM. House, AO. Sheldon, TA (2001). 'Routinely administered questionnaires for depression and anxiety: systematic review.' *British Medical Journal,* 322(7283): 406–409.

Glover, G (1995). 'Mental health informatics and the rhythm of community care.' *British Medical Journal,* 311(7012): 1038–1039.

Glover, G. Farmer, R. Preston, D (1990). 'Indicators of mental hospital bed use.' *Health Trends,* 22(3): 111–115.

Greenberg, GA. Rosenheck, R (2005). 'Using the GAF as a national mental health outcome measure in the Department of Veterans Affairs.' *Psychiatric Services,* 56(4): 420–426.

Greenwood, N. Key, A. Burns, T. Bristow, M. Sedgwick, P (1999). 'Satisfaction with in-patient psychiatric services. Relationship to patient and treatment factors.' *British Journal of Psychiatry,* 174: 159–163.

Grol, R (2001). 'Successes and failures in the implementation of evidence-based guidelines for clinical practice.' *Medical Care,* 39(8 Suppl. 2): II46–54.

Gutierrez-Recacha, P. Chisholm, D. Haro, JM. Salvador-Carulla, L. Ayuso-Mateos, JL (2006). 'Cost-effectiveness of different clinical interventions for reducing the burden of schizophrenia in Spain.' *Acta Psychiatrica Scandanavica. Supplementum*, 432: 29–38.

Halsteinli, V (2008). Personal communication. November 2008.

Halsteinli, V. Ose, SO. Torvik, H. Hagen, TP (2006). 'Allocation of labour to somatic and psychiatric specialist care – the effects of earmarked grants.' *Health Policy,* 78(2–3): 115–127.

Hansson, L (1989). 'Patient satisfaction with in-hospital psychiatric care: a study of a 1-year population of patients hospitalized in a sectorized care organization.' *European Archives of Psychiatry & Neurological Sciences,* 239(2): 93–100.

Harris, EC. Barraclough, B (1998). 'Excess mortality of mental disorder.' *British Journal of Psychiatry,* 173: 11–53.

Haywood, TW. Kravitz, HM. Grossman, LS. Cavanaugh, JL. Davis, JM. Lewis, DA (1995). 'Predicting the "revolving door" phenomenon among patients with schizophrenic, schizoaffective, and affective disorders.' *American Journal of Psychiatry,* 152(6): 856–861.

Health Research Board (2008). *WISDOM.* Dublin (http://www.hrb.ie/health-information-in-house-research/mental-health/information-systems/wisdom/).

Health Scotland (2006). *Establishing national mental health and well-being indicators for Scotland*. Edinburgh.

Healthcare Commission (2007). *Survey of users of mental health services 2007*. London.

Hendryx, MS. Dyck, DG. Srebnik, D (1999). 'Risk-adjusted outcome models for public mental health outpatient programs.' *Health Services Research,* 34(1 Pt 1): 171–195.

Hendryx, MS. Teague, GB (2001). 'Comparing alternative risk-adjustment models.' *Journal of Behavioral Health Services & Research*, 28(3): 247–257.

Hermann, R. Mattke, S. & Members of the OECD Mental Health Care Panel (2004). *Selecting indicators for the quality of mental health care at the health systems level in OECD countries*. Paris: Directorate for Employment, Labour and Social Affairs (OECD Health Technical Papers no. 17).

Hermann, RC. Mattke, S. Somekh, D. Silfverhielm, H. Goldner, E. Glover, G. Pirkis, J. Mainz, J. Chan, JA (2006). 'Quality indicators for international benchmarking of mental health care.' *International Journal for Quality in Health Care,* 18(Suppl. 1): 31–38.

Hermann, RC. Palmer, H. Leff, S. Shwartz, M. Provost, S. Chan, J. Chiu, WT. Lagodmos, G (2004b). 'Achieving consensus across diverse

stakeholders on quality measures for mental healthcare.' *Medical Care*, 42(12): 1246–1253.

Horton, R (2007). 'Launching a new movement for mental health.' *Lancet*, 370(9590): 806.

Huxley, P. Evans, S (2002). 'Quality of life routine outcomes measurement: lessons from experience in the USA and the UK.' *Epidemiologia e Psichiatria Sociale*, 11(3): 192–197.

Jenkins, R. Glover, G (1997). 'The importance of service level measures for mental health policy.' *Epidemiologia e Psichiatria Sociale*, 6(Suppl. 1): 229–237.

Jenkins, R. Strathdee, G. Carr, S. Rawaf, S (2000). Equity in mental health. In: Rawaf, S. Orton, P (eds.). *Current issues in health care – health improvement programmes*. London: Royal Society of Medicine.

Kalman, TP (1983). 'An overview of patient satisfaction with psychiatric treatment.' *Hospital & Community Psychiatry*, 34(1): 48–54.

Kelleher, MJ. Chambers, D. Corcoran, P. Williamson, E. Keeley, HS (1998). 'Religious sanctions and rates of suicide worldwide.' *Crisis*, 19(2): 78–86.

Kisely, S. Campbell, LA. Crossman, D. Campbell, J (2008). 'Routine measurement of mental health service outcomes: Health of the Nation Outcome Scales in Nova Scotia.' *Psychiatric Bulletin*, 32: 248–250.

Knapp, M. McDaid, D. Amaddeo, F. Constantopoulos, A. Oliveira, M. Salvador-Carulla, L. Zechmeister, I. Mheen Group (2007). 'Financing mental health care in Europe.' *Journal of Mental Health*, 16(2): 167–180.

Kramer, TL. Owen, RR. Wilson, C. Thrush, CR (2003). 'Relationship between self-report and clinician-rated impairment in depressed outpatients.' *Community Mental Health Journal*, 39(4): 299–307.

Lavikainen, J. Fryers, T. Lehtinen, V (eds.) (2006). *Improving mental health information in Europe*. Helsinki: STAKES.

Leavey, G. King, M. Cole, E. Hoar, A. Johnson-Sabine, E (1997). 'First-onset psychotic illness: patients' and relatives' satisfaction with services.' *British Journal of Psychiatry*, 170: 53–57.

Lebow, J (1982). 'Consumer satisfaction with mental health treatment.' *Psychological Bulletin*, 91(2): 244–259.

Lien, L (2002). 'Are readmission rates influenced by how psychiatric services are organized?' *Nordic Journal of Psychiatry*, 56(1): 23–28.

Ling, DYC. Busch, AB. Frank, RG (2004). *Treatment price indexes for bipolar disorder*. Harvard: Harvard Medical School.

Lyons, JS. McGovern, MP (1989). 'Use of mental health services by dually diagnosed patients.' *Hospital & Community Psychiatry*, 40(10): 1067–1069.

Lyons, JS. O'Mahoney, MT. Miller, SI. Neme, J. Kabat, J. Miller, F (1997). 'Predicting readmission to the psychiatric hospital in a managed care environment: implications for quality indicators.' *American Journal of Psychiatry,* 154(3): 337–340.

Manderscheid, RW (2006). 'Some thoughts on the relationships between evidence based practices, practice based evidence, outcomes, and performance measures.' *Administration and Policy in Mental Health,* 33(6): 646–647.

Mann, JJ. Apter, A. Bertolote, J. Beautrais, A. Currier, D. Haas, A. Hegerl, U. Lonnqvist, J. Malone, K. Marusic, A. Mehlum, L. Patton, G. Phillips, M. Rutz, W. Rihmer, Z. Schmidtke, A. Shaffer, D. Silverman, M. Takahashi, Y. Varnik, A. Wasserman, D. Yip, P. Hendin, H (2005). 'Suicide prevention strategies: a systematic review.' *Journal of the American Medical Association,* 294(16): 2064–2074.

McCrone, P (1995). 'Predicting mental health service use: diagnosis based systems and alternatives.' *Journal of Mental Health,* 4(1): 31–40.

McDaid, D. Oliveira, M. Jurczak, K. Knapp, M. & the Mheen Group (2007). 'Moving beyond the mental health care system: an exploration of the interfaces between health and non-health sectors.' *Journal of Mental Health,* 16(2): 181–194.

McFarland, BH. George, RA. Goldman, W. Pollack, DA. McCulloch, J. Penner, S. Angell, RH (1998). 'Population-based guidelines for performance measurement: a preliminary report.' *Harvard Review of Psychiatry,* 6(1): 23–37.

Meehan, T. Bergen, H. Stedman, T (2002). 'Monitoring consumer satisfaction with inpatient service delivery: the Inpatient Evaluation of Service Questionnaire.' *Australian & New Zealand Journal of Psychiatry,* 36(6): 807–11.

Meehan, T. McCombes, S. Hatzipetrou, L. Catchpoole, R (2006). 'Introduction of routine outcome measures: staff reactions and issues for consideration.' *Journal of Psychiatric & Mental Health Nursing,* 13(5): 581–587.

Ministry of Health (2006). *National Mental Health Information Strategy: implementation plan 2006.* Wellington: New Zealand Ministry of Health.

Mohan, R. McCrone, P. Szmukler, G. Micali, N. Afuwape, S. Thornicroft, G (2006). 'Ethnic differences in mental health service use among patients with psychotic disorders.' *Social Psychiatry and Psychiatric Epidemiology,* 41(10): 771–776.

Ness-Abramof, R. Apovian, CM (2005). 'Drug-induced weight gain.' *Drugs of Today*, 41(8): 547–555.

Newcomer, JW (2005). 'Second-generation (atypical) antipsychotics and metabolic effects: a comprehensive literature review.' *CNS Drugs,* 19(Suppl. 1): 1–93.

Ohio Department of Mental Health (2007). Ohio Mental Health Consumer Outcomes Initiative (http://www.mh.state.oh.us/oper/outcomes/outcomes.index.html).

Osborn, DP. Levy, G. Nazareth, I. Petersen, I. Islam, A. King, MB (2007). 'Relative risk of cardiovascular and cancer mortality in people with severe mental illness from the United Kingdom's General Practice Research Database.' *Archives of General Psychiatry,* 64(2): 242–249.

Parker, C (2007). Developing mental health policy: a human rights perspective. In: Knapp, M. McDaid, D. Mossialos, E. Thornicroft, G (eds.). *Mental health policy and practice across Europe.* Buckingham: Open University Press.

Perreault, M. Rogers, WL. Leichner, P. Sabourin, S (1996). 'Patients' requests and satisfaction with services in an outpatient psychiatric setting.' *Psychiatric Services,* 47(3): 287–292.

Petrea, I. Muijen, M (2008). *Policies and practices for mental health in Europe – meeting the challenges.* Copenhagen: WHO.

Potts, Y. Gillies, M. Wood, S (2001). 'Lack of mental well-being in 15-year-olds: an undisclosed iceberg?' *Family Practice,* 18(1): 95–100.

Priebe, S. Badesconyi, A. Fioritti, A. Hansson, L. Kilian, R. Torres-Gonzales, F. Turner, T. Wiersma, D (2005). 'Reinstitutionalisation in mental health care: comparison of data on service provision from six European countries.' *British Medial Journal,* 330(7483): 123–126.

Priebe, S. McCabe, R. Bullenkamp, J. Hansson, L. Rossler, W. Torres-Gonzales, F. Wiersma, D (2002). 'The impact of routine outcome measurement on treatment processes in community mental health care: approach and methods of the MECCA study.' *Epidemiologia e Psichiatria Sociale,* 11(3): 198–205.

Rea, CA. Rea, DM (2002). 'Managing performance and performance management: information strategy and service user involvement.' *Journal of Management in Medicine,* 16(1): 78–93.

Renvoize, E. Clayden, D (1990). 'Can the suicide rate be used as a performance indicator in mental illness?' *Health Trends,* 22(1): 16–20.

Rose, D. Lucas, J (2007). The user and survivor movement in Europe. In: Knapp, M. McDaid, D. Mossialos, E. Thornicroft, G (eds.). *Mental health policy and practice across Europe.* Buckingham: Open University Press.

Rose, D. Thornicroft, G. Slade, M (2006). 'Who decides what evidence is? Developing a multiple perspectives paradigm in mental health.' *Acta Psychiatrica Scandinavica,* 113(Suppl. 429): 109–114.

Rosen, A. Miller, V. Parker, G (1989). 'Standards of care for area mental health services.' *Australian and New Zealand Journal of Psychiatry,* 23(3): 379–395.

Rosenheck, R. Fontana, A (1999). 'Changing patterns of care for war-related post-traumatic stress disorder at Department of Veterans Affairs medical centers: the use of performance data to guide program development.' *Military Medicine,* 164(11): 795–802.

Sayce, L. Curran, C (2007). Tackling social exclusion across Europe. In: Knapp, M. McDaid, D. Mossialos, E. Thornicroft, G (eds.). *Mental health policy and practice across Europe.* Buckingham: Open University Press.

Schacht, LM. Hines, H (2003). 'Recent applications of risk adjustment for performance measures used by state inpatient psychiatric facilities.' *Harvard Review of Psychiatry,* 11(4): 220–224.

Schomerus, G. Angermeyer, MC (2008). 'Stigma and its impact on help-seeking for mental disorders: what do we know?' *Epidemiologia e Psichiatria Sociale,* 17(1): 31–37.

Slade, M (2002). 'What outcomes to measure in routine mental health services, and how to assess them: a systematic review.' *Australian & New Zealand Journal of Psychiatry,* 36(6): 743–753.

Slade, M. McCrone, P. Kuipers, E. Leese, M. Cahill, S. Parabiaghi, A. Priebe, S. Thornicroft, G (2006). 'Use of standardised outcome measures in adult mental health services: randomised controlled trial.' *British Journal of Psychiatry,* 189: 330–336.

Social Exclusion Unit (2004). *Mental health and social exclusion.* London: Office of the Deputy Prime Minister.

Stiegler, M. Rummel, C. Wahlbeck, K. Kissling, W. Leucht, S (2005). 'European psychiatric treatment guidelines: is the glass half full or half empty?' *European Psychiatry,* 20(8): 554–558.

Taipale, V (2001). 'Mental health and quality of mental health care.' *Medicine & Law,* 20(4): 531–542.

Tapsell, R. Mellsop, G (2007). 'The contributions of culture and ethnicity to New Zealand mental health research findings.' *International Journal of Social Psychiatry,* 53(4): 317–324.

Tennant, R. Joseph, S. Stewart-Brown, S (2007). 'The Affectometer 2: a measure of positive mental health in UK populations.' *Quality of Life Research,* 16(4): 687–695.

Thornicroft, G (2007). 'Most people with mental illness are not treated.' *Lancet,* 370(9590): 807–808.

Thornicroft, G (2008). 'Stigma and discrimination limit access to mental health care.' *Epidemiologia e Psichiatria Sociale,* 17(1): 14–19.

Wahlbeck, K (2006). Mental health in EU health monitoring systems. In: Lavikainen, J. Fryers, T. Lehtinen, V (eds.). *Improving mental health information in Europe.* Helsinki: STAKES.

Wang, PS. Aguilar-Gaxiola, S. Alonso, J. Angermeyer, MC. Borges, G. Bromet, EJ. Bruffaerts, R. De Girolamo, G. De Graaf, R. Gureje, O. Haro, JM. Karam, EG. Kessler, RC. Kovess, V. Lane, MC. Lee, S. Levinson, D. Ono, Y. Petukhova, M. Posada-Villa, J. Seedat, S. Wells, JE (2007). 'Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys.' *Lancet,* 370(9590): 841–850.

Warner, JP. Singhal, S. Dutta, A (1994). 'How well do we assess patients' satisfaction?' *Medical Audit News,* 4(6): 157–158.

Watters, C (2007). The mental health care of asylum seekers and refugees. In: Knapp, M. McDaid, D. Mossialos, E. Thornicroft, G (eds.). *Mental health policy and practice across Europe.* Buckingham: Open University Press.

Watters, C. Ingleby, D (2004). 'Locations of care: meeting the mental health and social care needs of refugees in Europe.' *International Journal of Law and Psychiatry,* 27(6): 549–570.

Weinmann, S. Koesters, M. Becker, T (2007). 'Effects of implementation of psychiatric guidelines on provider performance and patient outcome: systematic review.' *Acta Psychiatrica Scandinavica,* 115(6): 420–433.

Westert, GP. Verkleij, H (eds.) (2006). *Dutch health care performance report.* Bilthoven: National Institute for Public Health and the Environment.

WHO Regional Office for Europe (2003). *Mental health in the WHO European region.* Copenhagen & Vienna (Fact sheet EURO 03/03) WHO Regional Office for Europe.

WHO Regional Office for Europe (2005). *Mental health action plan for Europe. Facing the challenges, building solutions.* Copenhagen: WHO Regional Office for Europe.

WHO (2005). *Mental health atlas 2005.* Geneva: World Health Organization.

WHO (2005a). *The World Health Organization assessment instrument for mental health systems (WHO-AIMS).* Geneva: World Health Organization.

Wilkinson, DG (1982). 'The suicide rate in schizophrenia.' *British Journal of Psychiatry,* 140(2): 138–141.

Wing, JK. Curtis, RH. Beevor, AS (1996). *HoNOS: Health of the Nation Outcome Scales: report on research and development July 1993-December 1995.* London: Royal College of Psychiatrists.

Wittchen, HU. Jacobi, F (2005). 'Size and burden of mental disorders in Europe – a critical review and appraisal of 27 studies.' *European Neuropsychopharmacology,* 15(4): 357–376.

4.4 | *Long-term care quality monitoring using the interRAI common clinical assessment language*

VINCENT MOR, HARRIET FINNE-SOVERI,
JOHN P. HIRDES, RUEDI GILGEN,
JEAN-NOËL DUPASQUIER

Introduction

Residential care has been the mainstay of long-term care delivery systems in industrialized countries for decades. However, changes in acute care financing; individuals' preferences for remaining in the community; and the ageing of the elderly population mean that individuals with increasing frailty and impairments occupy these long-term care facilities. Most long-term care systems have evolved idiosyncratically as countries have faced different demographic imperatives and responded to different regulatory and medical-care systems. The need to characterize the needs of the population of long-term care users and the types and quality of services they receive has come to the forefront as the acuity of long-term care facilities has increased and as countries attempt to rebalance these budgets in order to provide more community support.

This chapter describes the development of a comprehensive clinical and functional assessment instrument – the nursing home Resident Assessment Instrument (RAI), more commonly known as the Minimum Data Set (MDS). This was designed in the United States on the basis that the proper provision of the complex care needed by frail older persons is predicated upon a comprehensive clinical assessment and it is the absence of such that underlies deficient quality of care. Originally intended as a clinical care planning tool, this minimum set of clinical and demographic data on all nursing home residents has been adapted as a vehicle for determining payment levels and to monitor the quality of care.

Several European countries have adopted the RAI within their long-term care systems. Similar applications are in place, either by governmental mandate or on a voluntary basis, in Canada and several European countries such as Switzerland and Finland. Various provinces in other countries are currently considering adopting this approach. The long-term care sector shares many of the conceptual and technical difficulties that health policy-makers face when attempting to compare quality performance in hospitals or medical groups. However, long-term care facilities are also individuals' homes and therefore the adequacy of the living experience must be addressed by understanding quality of life, not just quality of care, issues.

In this chapter we document how the RAI-MDS has been transformed into an assessment based data system that serves multiple research and applied policy functions, ranging from casemix reimbursement to outcomes measurement and quality performance monitoring. Since all industrialized countries are facing rising ageing populations and are therefore grappling with how to develop and/or modify their long-term care systems, there is substantial international interest in the development of the RAI for clinical assessment, educational purposes and for policy applications. The second half of this chapter focuses on the use of RAI data for benchmarking nursing home quality via public reporting and quality improvement efforts in the United States, Canada, Finland and Switzerland.

Origin of the RAI in the United States

Complaints about the quality of nursing home care began soon after Medicare began reimbursing for post-hospital nursing home care and Medicaid began paying for long-term nursing home care in 1966. Scandals about the quality of care in nursing homes have occurred periodically and prompted the formation of a new investigatory commission, the promulgation of new regulations, or both (Davis 1991). In 1984, the Institute of Medicine initiated a study of the quality of care in nursing homes, led by Sidney Katz.

Recommendations from the committee's report *Improving the Quality of Care in Nursing Homes* (Institute of Medicine 1986) were translated almost entirely into the 1987 Nursing Home Reform Act of the Omnibus Budget Reconciliation Act. One of the key recommenda-

tions was to mandate a comprehensive assessment that would provide a uniform basis for establishing a nursing home resident's care plan. This was based on the observation that the lack of training and education among direct line nursing home staff meant that they were unable to identify patient needs. It was thought that a systematic assessment would structure the clinical information necessary for care planning and form the basis for a common lexicon for describing patients and their needs. Like the ICD, the MDS for nursing home resident assessment was designed to become a common language of functional impairment and disability for long-term care (Mor 2004).

The MDS was designed by a consortium of academic medical centres under contract from the Health Care Financing Administration (now the Centers for Medicare and Medicaid Services – CMS). Hundreds of experts representing the academic disciplines and professional organizations serving geriatrics, psychiatry, nursing, physical and occupational therapies, nutrition, social work and resident rights advocates participated in the design and testing of the instrument between 1989 and 1991. The goal was to create an instrument that captures the basic information needed to determine whether patients have various common geriatric problems and to develop a care plan that considers individuals' co-morbidities as well as their strengths and residual capacities. The domains of problems to be included in the assessment were specified in the 1987 Nursing Home Reform Act (Hawes et al. 1997; Hawes et al. 1997).

Version 1.0 of the RAI-MDS was implemented in all nursing homes in the United States in 1991. As a 'condition of participation' in the Medicare or Medicaid programmes, nursing homes had to complete the assessments for all residents regardless of their payer source. Thus, the population of all nursing home residents was represented in the data in all certified facilities. Assessments were required upon admission; re-admission; when the resident experienced a significant change in condition; and quarterly following the initial admission assessment. In 1999 an updated version (RAI 2.0) of the instrument was implemented along with a mandate that all facilities must computerize all assessments and submit them to CMS (Morris et al. 1997) With the adoption of a subset of RAI items as a measure of casemix acuity for casemix reimbursement, Medicare post-acute hospital nursing home admissions had to be assessed more frequently in the weeks follow-

ing admission in order to determine the level of reimbursement based upon residents' assessed acuity.

Various studies have evaluated the impact of the Nursing Home Reform Act. One focused on understanding the impact of introducing the RAI-MDS in nursing homes in the United States, based upon a longitudinal study of 250 randomly selected facilities in 10 states with all data collected by independent research nurses. The investigators found that processes of care in several areas (restraint use and pressure ulcer prevention services) improved between the period prior to and after the implementation of the RAI (Hawes et al. 1997). Using MDS-based measures of cognitive function and ADL and mobility as outcome measurement scales, residents were found to be less likely to decline functionally and less likely to be hospitalized than they were before the Omnibus Budget Reform Act (Fries et al. 1997; Phillips et al. 1997; Mor et al. 1997) This study revealed that, when used by trained research staff, RAI-MDS has the capacity to identify specific care process problems and to measure changes in functional status.

Reliability and validity of the MDS

The MDS was tested repeatedly for inter-rater reliability among trained nurse assessors in large and small, for-profit and voluntary nursing homes throughout the country. These tests revealed high average levels of reliability as measured by kappa. The MDS was implemented nationally in late 1990; a modified version was designed and retested in 1995 and found to have improved reliability (Hawes et al. 1995; Mor et al. 2003; Morris et al. 1990)

Subsequent epidemiological and health services research studies using data from several states that used computerized versions of MDS found considerable evidence for construct validity. For example, Gambassi et al. (Gambassi et al. 1998) linked MDS assessment records with the Medicare hospital discharge claim that immediately preceded the MDS nursing home assessment. They found that the positive predictive value of an MDS-based diagnosis of a chronic condition affecting function or treatment exceeded 0.7 when compared to the hospital claim discharge diagnosis. In addition, in comparisons between drugs taken by residents and their MDS-based diagnoses they observed high levels of correspondence between the diagnosis and the appropriate

class of drug for its treatment. Subsequent analyses of patients with diagnoses ranging from Parkinson's disease to congestive heart failure revealed similar positive associations (Bernabei et al. 1998; Bernabei et al. 1999; Gambassi et al. 1998) Finally, a series of analyses examining the relationship between the presence of selected diagnoses and functional and cognitive status found that each of these measures strongly predicted mortality in the expected direction (Gambassi et al. 1999; Gambassi, Lapane et al. 1999).

The discriminant validity of the MDS was also established by a series of smaller studies that compared summary indices derived from selected MDS data elements. Morris and colleagues created the Cognitive Performance Scale (CPS) by crosswalking variables in the MDS with the mini-mental state examination administered by research staff (Hartmaier, Sloane et al. 1995; Morris, Fries et al. 1994). They (and others) found the CPS to be strongly correlated to clinical and research tools assessing cognition and to a diagnosis of Alzheimer's disease and subsequent mortality (Gambassi et al. 1999; Gruber-Baldini et al. 2000). Various forms of ADL indices have been constructed using MDS variables characterizing patients' mobility; self-care performance; and the amount of assistance required to perform those tasks. Morris reported that both hierarchical and additive versions of the ADL scale were found to be strongly related to staff time – residents with more ADL impairment receiving more assistance (Morris et al. 1999) Other multi-item summary indices based upon the MDS assessment have been developed for domains such as pain; distressed mood and behavioural disturbances; and social engagement (Mor et al. 1995; Frederiksen, Tariot et al. 1996; Fries et al. 2001). Each of these manifested discriminant validity, clearly differentiating patients with different diagnoses, levels of functioning and nursing care needs.

Policy applications of the RAI

The RAI was designed as an assessment tool to facilitate care planning for nursing home residents but it was not long before the assessment data were being applied to very different functions ranging from reimbursement to quality monitoring. The precedent for this multifaceted use of clinical assessment data was established in the original studies

that tested their utility and validity for research purposes since part of the evaluation required the creation of summary indices of residents' outcomes. Indeed, as described below, much of the work on applying the assessment data for reimbursement purposes was performed contemporaneously.

Casemix reimbursement

Casemix reimbursement came to long-term care in the 1980s in states such as New York, which was intent upon controlling its nursing home costs in the Medicaid programme. This was initially based on the Resource Utilization Group (RUG) system, a mandated, uniform data collection tool that classified patients largely by functional status (Fries and Cooney 1985). During the 1990s, many other states began adopting a prospective reimbursement system based on casemix (Feng, Grabowski et al. 2006). This trend was greatly accelerated by the universal availability of the MDS and by revision of the RUG system to incorporate new data elements that captured the characteristics of the more clinically complex patients entering nursing homes in increasing numbers. RUG was revised under the federally funded Nursing Home Case-Mix and Quality Demonstration project to include the far richer and clinically more complex data elements contained in the MDS. Thus, RUG-III was created for application to the Medicaid and Medicare patients in facilities from six states that participated in the demonstration project (Fries et al. 1994) Although not without controversy, the Medicare programme adopted the RUG classification system and applied it to a per diem payment for Medicare-reimbursed skilled nursing facility stays (Davis et al. 1998; Matherlee 1999).

It is interesting that virtually all evaluations of the impact of introducing casemix reimbursement at both federal and state level have relied upon the MDS data. Numerous researchers have merged nursing home level data on staffing levels with resident level data from the MDS. The resulting hierarchical and longitudinal data have been used to test the effect of introducing casemix reimbursement on staffing levels and skill mix; and the average acuity of residents and the outcomes they experience, for both Medicare and Medicaid beneficiaries (Feng et al. 2006; Konetzka et al. 2006; Konetzka et al. 2004; Wodchis et al. 2004).

Creating quality indicators to monitor provider performance

Researchers have frequently proposed and used measures of nursing home quality but generally only for a small number or select groups of facilities. Until recently, most such measures were based upon aggregate data reported by the home as part of the federal requirement for survey and certification (Zinn 1994). Many early studies of the determinants of quality of care in nursing homes produced contradictory findings because they used facility-level data that could not be risk-adjusted for differences in casemix (Davis 1991).

The availability of clinically relevant, universal, uniform and computerized data on all nursing home residents raised the possibility of using this information to improve care quality. Several approaches were suggested. The MDS data were thought to have utility in directly guiding efforts to improve the quality of care in a single nursing home (Popejoy et al. 2000; Zimmerman 2003). Several states instituted the use of MDS-based indicators of nursing home quality as part of the Case Mix Reimbursement and Quality Demonstration (Reilly et al. 2007). As with most efforts designed to improve health-care quality, this offered multifaceted incentives and targets. First, government regulators anticipated that the creation of indicators of nursing homes' quality performance would guide and systematize existing regulatory oversight processes that had been characterized as idiosyncratic. Secondly, more enlightened facility administrators felt that such information could facilitate their own existing quality improvement activities. Finally, advocates for nursing home residents thought that making this information available would create greater transparency to guide consumers' choice of a long-term care facility.

Initially, few nursing facilities across the country had the sophistication to use the MDS for institutional planning, staff loading or outcome monitoring but now many are actively using the MDS for one or more of these functions. Some states, particularly those that began statewide computerization of their MDS data before the CMS mandate in June 1998, began rudimentary efforts to report aggregated quality indicators from a variety of different MDS domains (Castle & Lowe 2005) These efforts were designed to make facilities aware of the potential uses of the MDS and to allow comparisons between their quality of care and the state-wide averages.

As part of the Nursing Home Case-Mix and Quality Demonstration, Wisconsin's Center for Health Systems Research and Analysis (CHSRA) was charged with developing an array of readily useable facility and resident quality indicators based upon computerized data from the resident assessment instrument (Zimmerman 2003; Reilly et al. 2007). Numerous versions of these proposed indicators were reviewed by various clinical and industry panels for appropriateness, meaningfulness and their potential for attributing problems to the care provided in the facility. Indicators included the prevalence of pressure ulcers; prevalence of use of anti-psychotics; and the incidence of late loss ADL. The CHSRA team created algorithms to identify individual residents and aggregate them to the level of the facility and then designed reports to help facilities and state inspectors to use this information to isolate problem areas.

Various other efforts were undertaken to develop and test quality indicators focused on quality of life issues such as mood or well-being. As it was easier to gain expert consensus on the meaning of clinically pertinent quality indicators, far fewer broader quality of life measures have been developed and promulgated (Castle et al. 2007; Mukamel et al. 2007). Furthermore, psychosocial measures included in the RAI-MDS have been shown to have poorer inter-rater reliability and suffer from ascertainment bias – under-identification of pain, mood and behaviour problems, for example (Bates-Jensen et al. 2004; Simmons et al. 2004; Roy & Mor 2005; Wu et al. 2005). Additionally, the MDS contains information on distressed mood and even involvement in social activities but does not capture patients' preferences or satisfaction. However, a separate 'industry' has arisen to produce resident satisfaction surveys in the United States over the last decade and these are increasingly available in facilities across the country (Lowe et al. 2003; Castle 2006; Straker et al. 2007).

In the late 1990s, CMS expanded their commitment to use quality indicators in their efforts to improve nursing home quality (Clauser & Fries 1992; Harris & Clauser 2002). The first objective was to improve and expand upon extant clinically relevant quality indicators based upon the universally available MDS information (Berg, Mor et al. 2002). The second objective was to develop measures that were fully responsive to the quality of life concerns of long-term care facility residents, such as food quality and preferences, autonomy and percep-

tion of treatment with respect (Kane 2003). These updated measures of quality performance were intended to meet the information needs of four distinct audiences: providers, regulators, purchasers and consumers. The first two groups had had some experience of interpreting and working with the MDS-based quality indicators developed under the Nursing Home Case-Mix and Quality Demonstration. However, the involvement of purchasers and consumers meant introducing some level of public reporting of the information. Public reporting has presented challenges to both the National Committee for Quality Assurance (in the managed care plan realm) and to the Joint Commission, which has been struggling with hospitals on this issue. Data can be misinterpreted or tell only part of a story and providers and insurers are uncomfortable that data are made available to a public who may not understand the meaning of the performance measures. This reluctance has frequently resulted in disagreements about the precise definitions and construction of the performance measures, particularly whether and how to risk adjust the data (Sangl et al. 2005; Zinn et al. 2005; Castle et al. 2007; Gerteis et al. 2007; Phillips et al. 2007).

CMS quality measures cover both long- and short-stay nursing home residents, with a numerator and denominator defined for each measure. Cross-sectional measures such as the proportion of residents with physical restraints are repeated quarterly as are longitudinal measures such the proportion of long-stay residents with declining physical functioning. However, longitudinal measures require the residents to have two measures and ignore censoring due to death or discharge. Rules on reporting are based on the number of patients for whom a measure can be calculated. However, the result can be quite volatile even when there are at least twenty or thirty patients (Mor 2005; Sangl, Saliba et al. 2005). For example, it is not uncommon for the measure of the proportion of patients declining through late onset ADL impairments to be well over 30% in one quarter and well under 20% in another, shifting the providers' quality ranking from near the top to near the bottom (Mor 2004). Statistically, less than 25% of the variation in a quality measure reflecting one quarter's performance can be explained by that of the next quarter. Even more importantly, the correlation between clinical quality measures (e.g. rate of functional decline, pressure ulcer prevalence) is less than .05, meaning that providers doing well in one area may not be doing well in another (Mor et al. 2003; Baier et al. 2005; Mor 2005). Consumers, families and advo-

cates who use this information to choose a provider do so because they believe that the past will be a good predictor of the future. When quality measures are volatile they will not be good predictors of future performance – nor will they guarantee that good performance in one area means good performance in another.

Comparisons of data quality problems in relation to the prevalence or incidence of selected quality indicators revealed that almost half of the observed inter-state differences are due to systematic coding differences in the assessment items that make up the quality measures (Wu et al. 2005a). This is consistent with several small studies that compared nursing home providers' performance in areas such as pain management or incontinence care. The authors found substantial inter-facility and inter-state differences in the prevalence of clinical conditions that seemed unrelated to differences in the patients studied (Schnelle et al. 2003; Simmons et al. 2004).

In spite of concerns about the validity of data and consumers' use of publicly reported quality information, in 2002 the CMS began posting aggregated quality measures on their Nursing Home Compare web site (see below). This had previously contained information about staffing levels and the results of annual inspections of facilities (Castle & Lowe 2005; Castle et al. 2007; Mukamel et al. 2007). The resulting publicly reported data are now promulgated widely throughout the Internet. Many companies repackage the information in a more user-friendly format to help consumers and their families to choose a facility and many states have gone beyond CMS by adding selected information about facilities (Castle et al. 2005). At present CMS is initiating a demonstration project that pays nursing homes extra. These bonuses are based on performance on the publicly reported quality measures for reductions in acute hospitalizations and associated costs which are presumed to accompany improvements in quality (Rahman 2006).

Use of RAI for quality monitoring and benchmarking: international examples

Nursing Home Compare in the United States

CMS initiated a six-state pilot project in April 2002 in which facility-specific, MDS-based quality measures were promulgated for every Medicare/Medicaid certified nursing facility in each state. Applied to

both long and short-stay post-acute patients, the quality measures included items such as pressure ulcer prevalence, restraint use, mobility improvement, pain and ADL decline. Advertisements presenting the rankings of area nursing homes were taken out in every major newspaper in every community in these six states.

Most nursing homes in the state were ranked and data on all measures for all facilities were included on the Nursing Home Compare web site (http://www.medicare.gov/NHCompare/home.asp). These data are readily accessible to consumers and advocates who may be seeking a facility. Having indicated a chosen geographical location, any number of facilities can be selected by various characteristics such as size, ownership or specialized services. This generates printable reports that compare the selected facilities in terms of staffing levels, inspection results and quality measures. Fig. 4.4.I provides an example of a comparison of one of the RAI-MDS based quality measures in several facilities in the state of Rhode Island. The comparative report includes information on the national and state average of the measure in order to provide context for the performance of each facility.

As noted above, consumers and their advocates are not the only users of these data. State inspectors of nursing home quality use the information on quality measure performance to guide their inspections, focusing on those aspects of the care process in which the facility appears to perform most poorly. Additionally, the Quality Improvement Organizations (QIOs), contracted by CMS to help facilities to institute quality improvement programmes, generally focus on improving those aspects identified as problematic in the publicly reported quality measurements. Finally, both CMS and some states are experimenting with pay-for-performance programmes that pay bonuses to high-performing facilities, based on the quality measures and selected structural factors such as staffing levels (Rahman 2006; Arling et al. 2007).

To date the impact of public reporting of nursing homes' perform ance is poorly understood. A recent survey of administrators suggests that most providers are keenly aware of how they compare to their local competition or peers; those who see their performance as sub-par report having instituted quality improvement programmes (Mukamel et al. 2007). Another recent study revealed that fewer than half of all consumers correctly interpreted the meaning of the bar graphs on Nursing Home Compare (see Fig. 4.4.I), suggesting that the quality
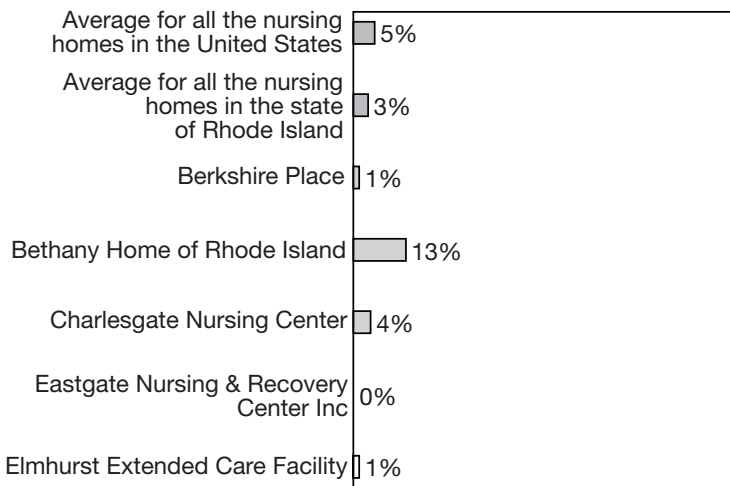
Fig. 4.4.1 Percentage of long-stay residents who were physically restrained

Source: http://www.medicare.gov/NHCompare/Include/DataSection/Questions/ProximitySearch.asp

information has relatively low utility to the end user (Gerteis et al. 2007). Similar results have been observed from efforts to inform consumers about the quality of insurers, hospitals and even physicians (McGee et al. 1999; Sofaer & Firminger 2005).

Benchmarking initiatives involving interRAI data in Canada

Multiple organizations undertake efforts to improve the quality of health care in Canada. The Canadian Institute for Health Information (CIHI) is an independent agency that houses and reports on data related to health expenditures, health services, health human resources and population health. It provides national reports on a range of health indicators for a variety of sectors including acute care, continuing care, home care, rehabilitation and mental health. The Canadian Council on Health Services Accreditation (CCHSA) works at the organizational level to evaluate and identify opportunities to improve quality in health care. Its accreditation standards require performance indicators to be used within internal quality improvement efforts but the organization does not produce provincial or national comparative

reports based on those indicators. Following a national commission on the future of health care, the Health Council of Canada was founded to promote public accountability and transparency. Its reports have focused on progress related to federal and provincial governments' commitments in the 2004 ten-year plan for health system renewal. Since the establishment of this national agency, a number of provincial governments have created parallel agencies to perform similar functions at their level.

There is widespread implementation of interRAI instruments in Canada. For example, the nursing home Resident Assessment Instrument 2.0 (RAI 2.0) was first mandated as the standard assessment instrument for all patients in Ontario's Complex Continuing Care (CCC) hospitals/units in 1996. Seven other provinces/territories have since undertaken to implement the instrument and CIHI established the Continuing Care Reporting System (CCRS) to serve as the national data warehouse for RAI 2.0 data. As noted in the summary, versions of the RAI assessment instrument appropriate for home care and other populations with disabilities are also being implemented in multiple Canadian provinces.

The Ontario Hospital Report initiative was the first large scale effort to report on the quality of care using interRAI data in Canada. Data from CCC hospitals/units were used as part of a scorecard that aims to report on clinical utilization and outcomes; patient and family satisfaction; financial performance; and system integration and change. Of particular interest here, quality indicators developed by Morris and colleagues are used to benchmark hospital performance in thirteen areas including depression, communication decline, falls, pain, pressure ulcers and physical restraint use (http://www.cms.hhs. gov/NursingHomeQualityInits/Downloads/NHQISnapshot.pdf). The reports include provincial level distributions, regional rates and hospital-specific performance on individual quality indicators. The financial quadrant of the report uses resource utilization groups to provide a casemix adjustment for benchmarking the direct costs of care per weighted day (Fries et al. 1994). The system integration and change quadrant examines trends toward improved care through evidence based practice; use of information technology; integration of care; and use of the RAI 2.0 to inform clinical practice. All reports from this initiative are publicly available through the research collaborative (www. hospitalreport.ca). Fig. 4.4.2 provides an example of a report compar-
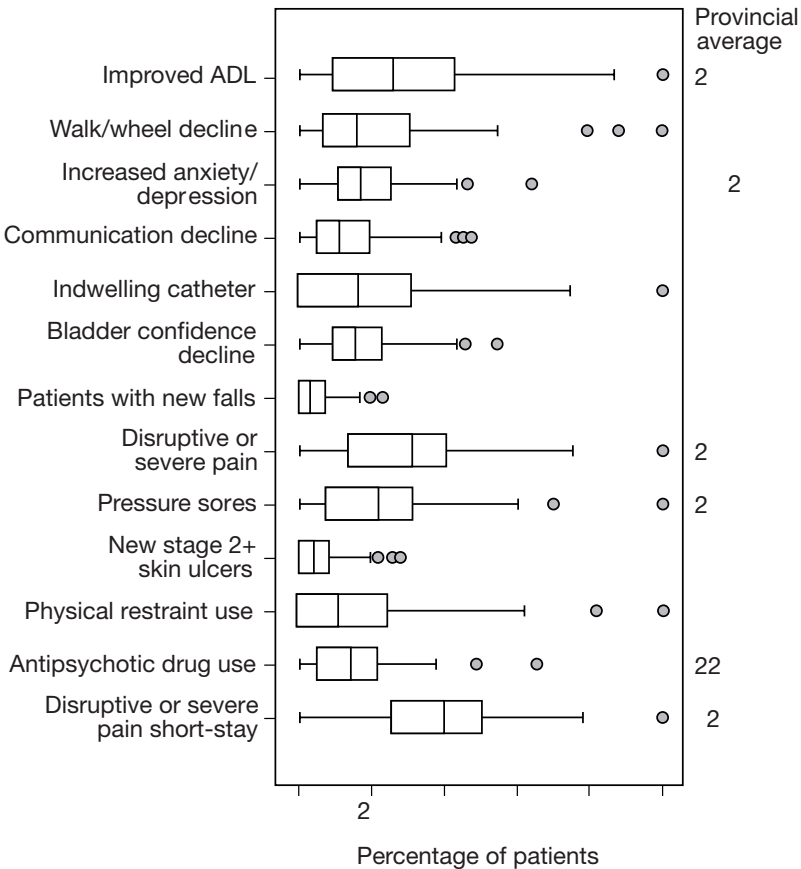
Fig. 4.4.2 Distribution of provincial indicator results

Source: Continuing Care Reporting Systems 2005–2006, CIHI.

ing CCC providers on a number of different quality measures. Box plots for each measure indicate the median facility score and the distribution of providers that are outside the range of most providers on each measure. Fig. 4.4.3 compares a number of providers on a given quality measure (new pressure ulcers) and indicate the provincial average of all other CCCs in much the same way as the CMS Nursing Home Compare report.

There was some initial concern about how public reporting would impact on hospital performance but such transparency is now accepted as common practice in Ontario hospitals. Long-term care facilities

New stage 2 or greater
skin ulcers by LHIN
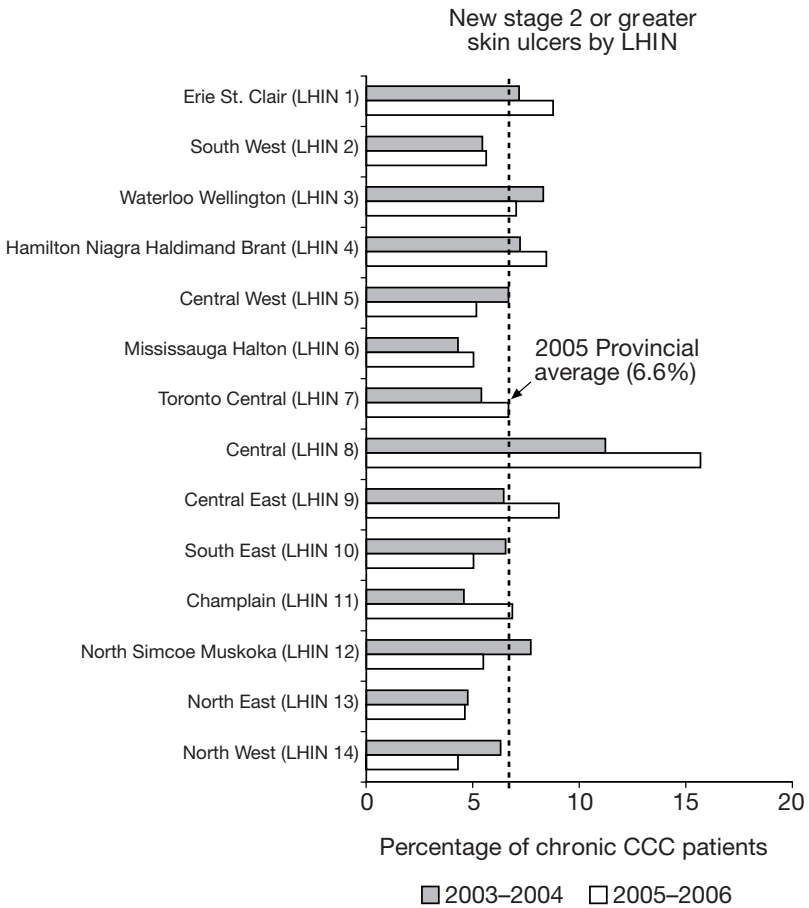
Fig. 4.4.3 New stage 2 or greater skin ulcers by LHIN

Source: Continuing Care Reporting Systems 2003–2004 and 2005–2006, CIHI.

have not yet fully implemented the RAI 2.0 and so it has not been possible to produce equivalent reports for that sector. Recent high-profile media coverage of several instances of poor care in nursing homes has increased demands for improved quality in that sector. Indeed, there is now general agreement on the need for increased accountability and transparency in all continuing care settings. However, the issue of risk adjustment has been a source of some concern, given the great heterogeneity of CCCs and nursing homes. For example, Ontario's CCC hospitals/units serve a considerably more clinically complex,

post acute population than is typical of nursing home residents in that (or other) province(s). Early quality indicators based on the RAI 2.0 included some resident level risk adjusters but these are acknowledged to be inadequate to control for the substantial facility-level differences in the populations served. A CIHI-funded research initiative is exploring the use of direct adjustment methods to control more adequately for these differences without over-adjusting the indicator. A report on this new approach is expected by mid 2008.

Comparing performance of nursing facilities in Finland

In Finland, long-term care for older individuals has traditionally been divided into two categories: (i) hospital based long-term care delivered in health centres; and (ii) residential homes (nursing homes). The population aged 65 or over will increase by nearly 75% between now and 2030. However, particularly the proportion of the oldest old; the number of long-term beds; and the proportion of the elderly population living in them have been decreasing during the past ten years.[1]

The National Institute for Health and Welfare (STAKES) is a research and analysis unit that functions immediately under the Ministry of Social Affairs and Health. Its responsibilities include enhancement of best practices in the care of older persons in addition to collecting data and maintaining national registers on this field. However, it has no controlling or regulatory power. The counties are responsible for overseeing and supervising nursing homes but regular visits for these purposes are practically nonexistent. Also, data about conditions and the nature of the population served were sporadic and lacking information about performance until the RAI-benchmarking project was launched.

RAI benchmarking project in long-term care
STAKES and collaborating organizations in the RAI benchmarking project launched RAI activities as a pilot study in 2000. Project aims included implementing the RAI assessment system in Finnish long-term care facilities; educating facility staff and management in RAI assessment technology; developing performance measures to monitor efficiency and quality of care; creating software for facility manage-

[1] STAKES: http://www.stakes.fi/EN/Aiheet/olderpeople/statistics.htm?KwPath=S tatistics&TextSize=medium accessed Dec. 26, 2007.

ment to monitor web-based reports; and creating a forum for ongoing educational and best practice dissemination. Participation has been voluntary but facilities that committed to participation were required to assess every resident.

The performance measures adopted were based upon the models available in 2001. The nursing home casemix index had been validated in Finland in 1995, as had several RAI-MDS based summary outcome scales (Bjorkgren et al. 1999; Morris et al. 1999). (Burrows et al. 2000; Fries et al. 2001) The only nursing home quality indicators internationally tested at that time were those created by the University of Wisconsin-Madison and therefore that form was adopted (Zimmerman 2003). There were twenty-six indicators with set thresholds – twenty-two prevalence based; four incidence based. Five of the indicators were also risk adjusted (stratified by risk status).

The RAI benchmarking project established continuous feedback between the facilities and STAKES. A copy of the RAI assessment data is sent to STAKES biannually for benchmarking and research purposes. Within a month STAKES produces web-based, password-protected benchmarking results together with individual reports for each of the wards in the facility. STAKES organizes biannual two-day seminars in order to educate facility managers and clinical leaders and to facilitate sharing of best practices among providers. Over the eight years of the project, the number of voluntarily participating facilities increased from 41 in health centres and 43 in residential homes (overall 84) to 110 in health centres and 261 in residential homes (overall 371). The number of semi-annual RAI-assessments conducted increased from 2300 to 9000 and the number of nurses participating in semi-annual training seminars increased from 100 to 1000.

In order to highlight the comparisons possible with the benchmarking data, we have drawn upon examples that include only those communities in which every long-term care facility uses RAI. The performance measures embedded in the RAI assessments can first inform management of changes in the mix of residents' acuity levels. Fig. 4.4.4 reveals differences in means of the casemix and proportions (%) of light-care residents in four small or medium size towns. Light-care residents are independent in the personal activities of daily living and have minimal cognitive impairment. Presumably, health-care resources should be allocated accordingly.
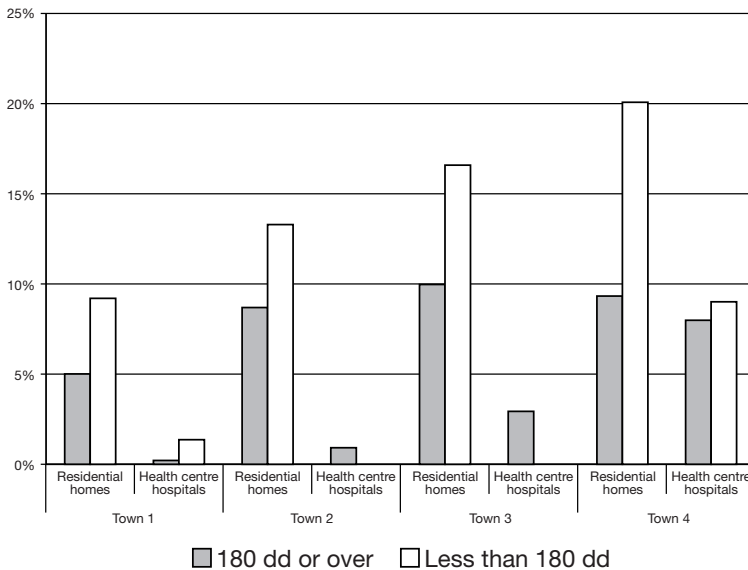
Fig. 4.4.4 Light-care residents (%) by type of long-term care facility in four medium or small sized towns, 2006

Source: Noro 2005

Town 4 has a smaller difference between casemix in the two different types of care (health centre hospitals, residential homes) than other towns. In addition, Town 4's intake of light-care residents is considerably higher than in the peer towns. These data indicate potential inefficiency in the case management processes designed to sustain older persons in their own homes in Town 4, where the eligibility criteria for long-term care settings are worth revisiting. Conversely, Town 1 has the lowest overall prevalence of light-care residents but there is also a small proportion of newly admitted light-care residents in health-care centres.

In order to benchmark quality of care, the facilities are first encouraged to ensure that peer groups are selected correctly, e.g. they have reasonably similar acuity levels in terms of cognitive and physical functioning. Fig. 4.4.5 shows a comparison of casemix index, staffing ratios and the prevalence of grade 2-4 pressure ulcers in four residential homes belonging to same organization. This shows some vari-
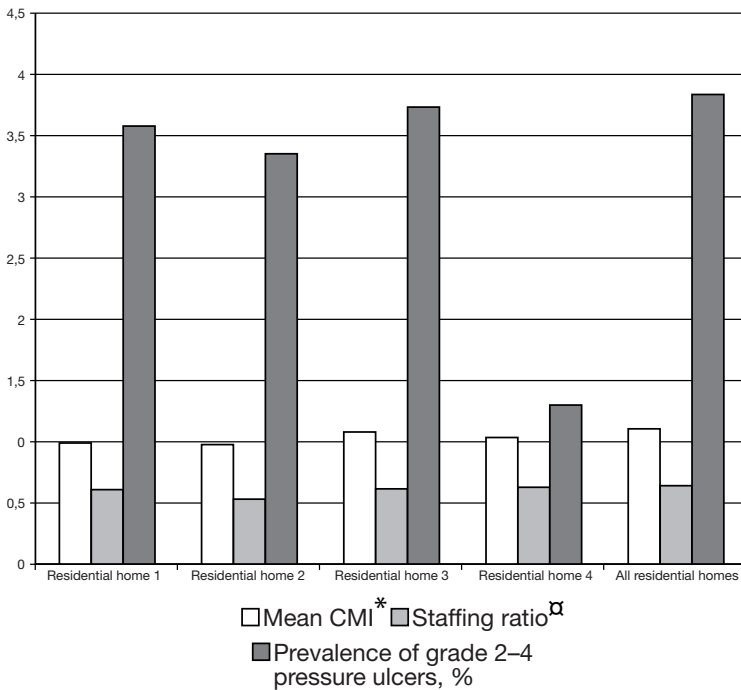
Fig. 4.4.5 Mean case-mix index, staffing ratio, and prevalence of grade 2–4 pressure ulcers (%) in four residential homes compared to their peers in 2006

* CMI: casemix index, range 0.42–2.52

¤ Staffing ratio: number of licensed and practical nurses in the ward per number of residents

Overall number of assessed residents in the residential home numbers 1–4 is 885

Source: Noro 2005

ability in the prevalence of grade 2-4 pressure ulcers but comparable casemix and staffing ratios across the four facilities.

Benchmarking in intra-facility management

Intra-facility comparisons between wards follow the same guidelines as inter-facility comparisons. However, individual wards may have special profiles based upon management decisions such as concentrating ambulating persons with dementia and behavioural problems in some wards and relatively independent residents with mental illness in

others. In these cases the wards are encouraged to compare themselves with ward-specific peer groups calculated by STAKES. This grouping of wards according to the severity of the casemix index and percentage of residents with dementia produces fourteen categories of ward, regardless of the type of facility. Every ward receives the suggested peer grouping values independently. Identification of the appropriate peers helps to create networks between similar units, to set reasonable goals for the units and to enable systematic work to reach them. Facilities are encouraged to identify target areas for which particular improvement can be expected and to set specific goals for each of the performance measures. One successful effort substantially reduced the use of psychotropic drugs but it is a challenge to hit a moving target when all residential homes improve their performance (Noro 2005).

In summary, RAI benchmarking was implemented successfully in Finland in 2000. Apart from measures for psychotropic medications, nursing rehabilitation and new pressure ulcers, the overall level of performance measures has remained relatively stable. However, looking only at those facilities involved in the project since 2001, eight of the twenty-six quality indicators have remained stable; four show deteriorated quality of care and fourteen have improved. This suggests that monitoring performance measures on a regular basis is a valuable tool for nursing managers in long-term care facilities. The observed changes in care patterns have occurred as a consequence of strong management actions within the facilities. These actions have not resulted from external pressures such as sanctions or changes in legislation or requirements. It is also evident that the changes have occurred only where both leaders and staff have used the measures for multiple purposes.

Nursing home performance measurement in Swiss cantons: Q-Sys approach

Since the late 1990s seven cantons in the German-speaking part of Switzerland have adopted the RAI. This serves as the basis for health sector reimbursements to facilities and for measuring nursing home quality as part of a broader voluntary adoption of the RAI in all facilities in participating cantons. By 2006 over 300 facilities in 7 cantons serving over 20 000 residents were participating in the RAI residents' assessment, facility payment and quality improvement system operated

by a company called Q-Sys AG, led by geriatricians and software engineers (http://www.rai.ch/). Instituted primarily as a care planning tool with substantial educational content for skilled and unskilled staff in Swiss nursing homes, the RAI-MDS has been used for both financing and quality monitoring and improvement efforts. The long-term care funding agencies in each canton have accepted the RAI based RUG-III casemix reimbursement financing model, a system that has been validated in many other countries (Ikegami et al. 1994; Hirdes et al. 1996; Ljunggren and Brandt 1996; Jorgensen et al. 1997; Carpenter et al. 2003).

Much of the movement towards the adoption of the RAI in selected Swiss cantons is attributable to the Health Insurance Law revised in 1994. This altered the basis for payment of nursing homes to produce a more uniform system of coverage for long-term care in all Swiss Cantons. The regulations required a geriatric assessment using a standardized instrument for all residents of nursing homes who wished to be reimbursed under the new long-term care financing law. Furthermore, nursing home providers were obliged to undertake some form of quality assurance and improvement programme in order to continue receiving reimbursements. A health information services company devoted to processing RAI data and producing the reports and data that nursing home providers need to generate quality reports was founded in 1999. Q-Sys AG receives RAI assessment data from all participating nursing facilities in the seven Swiss cantons which have adopted this approach to reimbursement and quality monitoring. The report produced for each provider summarizes their performance on twenty-four different quality indicators first developed by Zimmerman and his colleagues (Zimmerman 2003).

Many different presentations of performance are generated in the form of reports to each provider and to the consortium of providers in each canton. Fig. 4.4.6 provides an example of the variable performance among providers in eight different areas, displaying intra- and inter-cantonal differences in the distribution of the proportion of residents receiving psychotropic medications in the absence of a psychiatric diagnosis. These data cover 2006 but similar reports are generated semi-annually. Other reports made available to all providers within a canton and to specific providers demonstrate changes in the prevalence of the quality indicators in participating nursing homes over a four- or five-year period. Most recently, Q-Sys investigators
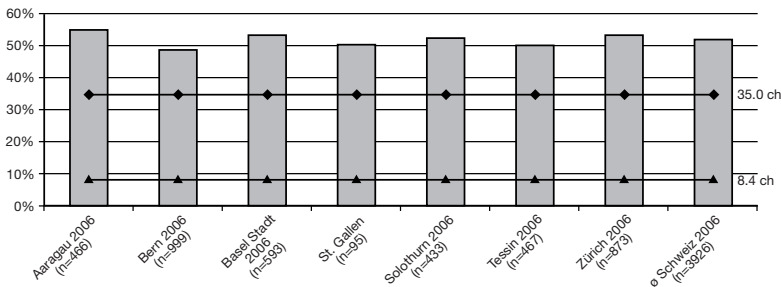
Fig. 4.4.6 Inter- and intra-canton comparisons of psychotropic drug use

Source: Q-Sys AG web site (http://www.rai.ch/)

and other European colleagues collaborated to produce some cross-national comparisons of these longitudinal data. These are intended to engage providers and cantons in a wider understanding of quality improvement by providing an opportunity to view their activities in a broader international context.

Summary and implications

The availability of uniform clinical data on nursing home residents' characteristics makes it viable to create quality performance measures for multiple purposes. Like uniform discharge abstracts for hospitals in the United States, the availability of the RAI-MDS on all facilities in selected geographical regions makes it possible to compare providers' performance on important parameters relevant to quality of care. There are still numerous conceptual and technical problems associated with interpreting differences among providers on the quality performance measures. However, the examples from the United States, Canada, Finland and Switzerland clearly reveal that the impetus for quality improvement is greatly stimulated by comparative data.

Provider quality performance measures can be used as a management tool to identify areas for quality improvement. This is reinforced when providers come together as a consortium to share best practices in quality improvement strategies and track performance changes, as in Finland. Performance measures can also be used to assist governmental or non-governmental inspectors charged with ensuring that providers meet minimum standards in order to retain certification

for reimbursement, as in the United States. Public reporting of performance measures can help consumers and their advocates to select high-quality facilities that provide the types of services they require. Finally, governmental or insurance entities charged with reimbursing long-term care providers can use performance measures as a basis for bonuses for high quality or to adjust payment levels in accordance with the quality of care provided (Rahman 2006; Grabowski 2007; Kane et al. 2007).

Policy challenges

Numerous policy challenges arise when common assessment systems are introduced to evaluate residential care facilities' quality in a country. First and foremost, should these systems be mandatory or voluntary? A related policy challenge is whether it is viable to use data intended for clinical use for policy applications such as casemix reimbursement and public reporting of quality performance. Finally, if the data are to be used to drive quality improvement through public reporting of results that influences consumers' choices, there needs to be an understanding of the policy implications if consumers are not able to interpret or use publicly reported quality data to make such choices.

The RAI was introduced in the United States as part of a legislative mandate designed to improve the quality of long-term care facilities, about which there was substantial consensus. Some Canadian provinces followed this example; others began with voluntary, more limited implementation only to determine that the logic of universal comparative data is so strong that mandatory implementation was required. Comparisons of the manner in which nursing home assessment systems have been implemented in North America and Europe show some interesting differences. The approach to mandating implementation in the United States and Canada is associated with public reporting uses of the information; the quasi-voluntary approach used in Finland and selected Swiss cantons is associated with a much greater focus on facility quality improvement and managerial education. It is true that quality improvement is a major focus of the performance benchmarking process introduced by both CMS and CIHI. In the United States, CMS has made a major investment in quality improvement efforts under the direction of specialized organizations in each state that work with providers to devise strategies to institute

quality improvement projects.[2] In Canada, the HRCC collaborative also undertakes continuing quality improvement projects that seek to identify strategies for performance improvement and to promulgate these as best practices among other chronic care hospitals in the consortium.[3] It is anticipated that a similar approach will be undertaken once all the nursing homes in the Ontario province have implemented performance measurement processes using the RAI-MDS.

Whether for casemix reimbursement or quality monitoring, these secondary uses of the RAI data raise questions about the validity and clinical utility of the basic resident assessment information. There was precedent in the case of hospitals' use of ICD diagnosis and procedure coding to document case mix acuity and associated payments before and after the introduction of DRGs (Hsia et al. 1988) As with RAI assessment data, hospitals had reasonably high error rates in their ICD coding, but tended to 'up-code'. To date, research in nursing homes in the United States suggests that error rates tend to be random but that systematic bias can creep readily into the process. Also, as when hospitals were paid on the basis of DRGs, clinical coding decisions became too important to leave to clinicians. Virtually all nursing home organizations now employ nurse assessors to coordinate MDS assessments, many of whom belong to rapidly growing national membership organizations that offer professional identity and education.[4] Even without responding to the incentive to up-code the MDS items, the original notion that all residents' needs would be assessed by an interdisciplinary team of professionals has fallen by the wayside as reimbursement is predicated upon the assessment information. It is not known whether this is changing the manner in which the data are used.

Research suggests that consumers find public reports of provider quality complicated and difficult to use. Also, significant unresolved technical problems may undermine the validity of direct comparisons between different providers in any one area. This raises questions about the strong push for public reporting of variations in provider quality in the United States and Canada. Certainly, the ideological rationale that underpins transparency and quality provides a strong impetus for public reporting. Nonetheless, there is increasing evidence

[2] http://www.cms.hhs.gov/QualityImprovementOrgs/01_Overview.asp#TopOfPage
[3] http://www.hospitalreport.ca/projects/QI_projects/IC5.html
[4] http://www.aanac.org/pages/membership_opp.asp.

that providers are stimulated to engage in serious quality improvement efforts precisely because their performance is open to all, including their local competition. This suggests that, despite the associated technical or conceptual problems, performance measurement that spurs providers into greater efforts to identify and improve quality problems may still have a very positive influence on long-term care in the United States and in developed economies where long-term care needs are growing rapidly. It is not clear whether providers' emphasis on quality improvement or even the validity of the underlying data might change under a regime of pay for performance (Rahman 2006).

Research needs

In all the countries that use the RAI data to develop benchmarks to which individual providers can aspire (or attempt to supercede), there is an underlying assumption that providers know how to re-organize their care processes to improve quality. It is true that the first step in quality improvement is accepting that improvements are necessary but it is far more difficult to understand which processes need to be changed and how. The provision of care in long-term care residential settings is a complex set of activities that combines medical treatments and social ministrations to enhance individuals' well-being; ensuring a safe and secure environment while allowing maximal independence in what is now the residents' home. Meeting all these needs requires innovative staff training, supervision and flexibility not normally associated with institutional care systems. The United States is introducing changes in both the physical and organizational environment in to change the culture of long-term care institutions (Rahman & Schnelle 2008). Enthusiasm for these changes appears to have outstripped the evidence for their effectiveness but it is evident that there is interest in changing institutional care to meet residents' needs more appropriately.

Research on the applicability of RAI data across providers is needed in order to better understand the implications of benchmarking for long-term care. In most countries there is considerable overlap between the needs of older people who live at home and those who enter institutions. Is it possible to develop comparable measures that are relevant to the outcomes experienced by frail older people whether they are at home or in an institution?

As noted, interRAI has developed quality indicators for home care that are in use in selected American states, Canadian provinces, Swiss cantons and Italian regions.[5] These take an approach similar to that used in the nursing home context. Initial efforts have been directed at monitoring performance measures designed to understand the sources of variation across providers in an area and to work with them to increase understanding of how to use the information for management and (ultimately) quality improvement purposes (Hawes et al. 2007). In the United States, a different assessment instrument has been mandated for all Medicare beneficiaries served by certified home health agencies. These data are used for both casemix-based reimbursement and public reporting of provider performance on a set of quality measures (Ahrens 2005). As with the MDS-RAI for nursing homes, individual agencies and consortia use these data for quality improvement (Stadt & Molare 2005; Scharpf et al. 2006).

A related research challenge with considerable policy importance is the development of measures that assess connections to the acute care setting. In the United States, almost 20% (with considerable inter-state variation) of Medicare beneficiaries entering nursing homes or even receiving home care are re-hospitalized within thirty days (Intrator et al. 2007). This may be a particular problem in that country since large differences were found in earlier comparisons of the hospitalization rates of nursing home residents in the United States and the Netherlands (Frijters et al. 1997). Nonetheless, performance measurement systems may provide incentives for facilities to discharge deteriorating residents to hospital in order to avoid reporting them. Future research will have to examine precisely how quality measures classify residents who are discharged to hospital and therefore may not contribute to the facility quality measure. It is not clear whether this is an issue in other countries but clearly the same incentives may be operating.

Conclusions

The emergence of a standardized assessment system and clinical language that is useful for educating and orienting long-term care providers has been the stimulus for standardized quality benchmarking systems in the United States and several other developed countries with rapidly ageing populations. Institutional care is always likely to be an option

[5]   http://interrai.org/applications/hcqi_table_final.pdf.

for those needing long-term care in developed countries and therefore it is important to have a means of measuring and comparing quality of care. Computerized health records facilitate performance measurement but it is possible to use a uniform assessment to characterize the needs of the population of nursing home residents without considerable investment in high tech equipment. A common assessment language helps to structure the information for subsequent reporting using simple manual summaries. Less well-developed countries – with rapidly growing ageing populations and an increasingly mobile society e.g. China – could institute an assessment system. There is considerable expertise that can be tapped to design web-based data collection and management tools appropriate to particular populations.

References

Ahrens, J (2005). 'The impact of Medicare home health policy changes on Medicare beneficiaries: part II.' *Policy Brief (Center for Home Care Policy and Research)*, (19): 1–6.

Arling, G. Lewis, T. Kane, RL. Mueller, C. Flood, S (2007). 'Improving quality assessment through multilevel modeling: the case of nursing home compare.' *Health Services Research*, 42(3 Pt 1): 1177–1199.

Baier, RR. Gifford, DR. Mor, V (2005). 'Reporting nursing home quality in Rhode Island.' *Medicine and Health, Rhode Island*, 88(6): 186–1877, 191.

Bates-Jensen, BM. Alessi, CA. Cadogan, M. Levy-Storms, L. Jorge, J. Yoshii, J. Al-Samarrai, NR. Schnelle, JF (2004). 'The Minimum Data Set bedfast quality indicator: differences among nursing homes.' *Nursing Research,* 53(4): 260–272.

Berg, K. Mor, V. Morris, J. Murphy, KM. Moore, T. Harris, Y (2002). 'Identification and evaluation of existing nursing homes quality indicators.' *Health Care Financing Review,* 23(4): 19–36.

Bernabei, R. Gambassi, G. Lapane, K. Landi, F. Gatsonis, C. Dunlop, R. Lipsitz, L. Steel, K. Mor, V (1998). 'Management of pain in elderly patients with cancer. SAGE Study Group. Systematic Assessment of Geriatric Drug Use via Epidemiology.' *Journal of the American Medical Association,* 279(23): 1877–1882.

Bernabei, R. Gambassi, G. Lapane, K. Sgadari, A. Landi, F. Gatsonis, C. Lipsitz, L. Mor, V (1999). 'Characteristics of the SAGE database: a new resource for research on outcomes in long-term care. SAGE (Systematic Assessment of Geriatric drug use via Epidemiology) Study

Group.' *Journals of Gerontology Series A: Biological Sciences and Medinal Sciences,* 54(1): M25–33.

Bjorkgren, MA. Häkkinen, U. Finne-Soveri, UH. Fries, BE (1999). 'Validity and reliability of Resource Utilization Groups (RUG-III) in Finnish long-term care facilities.' *Scandinavian Journal of Public Health,* 27(3): 228–234.

Burrows, AB. Morris, JN. Simon, S. Hirdes, J. Phillips, C (2000). 'Development of a minimum data set-based depression rating scale for use in nursing homes.' *Age and Ageing*, 29(2): 165–172.

Carpenter, I. Perry, M. Challis, D. Hope, K (2003). 'Identification of registered nursing care of residents in English nursing homes using the Minimum Data Set Resident Assessment Instrument (MDS/RAI) and Resource Utilisation Groups version III (RUG-III).' *Age and Ageing,* 32(3): 279–285.

Castle, NG (2006). 'Family members as proxies for satisfaction with nursing home care.' *Joint Commission Journal on Quality and Patient Safety*, 32(8): 452–458.

Castle, NG. Lowe TJ (2005). 'Report cards and nursing homes.' *Gerontologist,* 45(1): 48–67.

Castle, NG. Degenholtz, H. Engberg, J (2005). 'State variability in indicators of quality of care in nursing facilities.' *Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 60(9): 1173–1179.

Castle, NG. Engberg, J. Liu, D (2007). 'Have Nursing Home Compare quality measure scores changed over time in response to competition?' *Quality and Safety in Health Care*, 16(3): 185–191.

Clauser, SB. Fries BE (1992). 'Nursing home resident assessment and case-mix classification: cross-national perspectives.' *Health Care Financing Review,* 13(4): 135–155.

Davis, MA (1991). 'On nursing home quality: a review and analysis. *Medical Care Review,* 48(2): 129–166.

Davis, MA. Freeman, JW. Kirby, EC (1998). 'Nursing home performance under case-mix reimbursement: responding to heavy-care incentives and market changes.' *Health Services Research*, 33(4 Pt 1): 815–834.

Feng, Z. Grabowski, DC. Intrator, O. Mor, V (2006). 'The effect of state medicaid case-mix payment on nursing home resident acuity.' *Health Services Research,* 41(4 Pt 1): 1317–1336.

Frederiksen, K. Tariot, P. De Jonghe, E (1996). 'Minimum Data Set Plus (MDS+) scores compared with scores from five rating scales.' *Journal of the American Geriatrics Society,* 44(3): 305–309.

Fries, BE. Cooney, LM Jr. (1985). 'Resource utilization groups. A patient classification system for long-term care.' *Medical Care,* 23(2): 110–122.

Fries, BE. Hawes, C. Morris, JN. Phillips, CD. Mor, V. Park, PS (1997). 'Effect of the National Resident Assessment Instrument on selected health conditions and problems.' *Journal of the American Geriatrics Society,* 45(8): 994–1001.

Fries, BE. Schneider, DP. Foley, WJ. Gavazzi, M. Burke, R. Cornelius, E. (1994). 'Refining a case-mix measure for nursing homes: Resource Utilization Groups (RUG-III).' *Medical Care,* 32(7): 668–685.

Fries, BE. Simon, SE. Morris, JN. Flodstrom, C. Bookstein, FL (2001). 'Pain in US nursing homes: validating a pain scale for the minimum data set.' *Gerontologist,* 41(2): 173–179.

Frijters, DH. Mor, V. DuPaquier JN. Berg, K. Carpenter, GI. Ribbe, MW(1997). 'Transitions across various continuing care settings.' *Age and Ageing,* 26(Suppl. 2): 73–76.

Gambassi, G. Landi, F. Lapane, KL. Sgadari, A. Mor, V. Bernabei, R(1999). 'Predictors of mortality in patients with Alzheimer's disease living in nursing homes.' *Journal of Neurolology, Neurosurgurgery, and Psychiatry*, 67(1): 59–65.

Gambassi, G. Landi, F. Peng, L. Brostrup-Jensen, C. Calore, K. Hiris, J. Lipsitz, L. Mor, V. Bernabei, R (1998). 'Validity of diagnostic and drug data in standardized nursing home resident assessments: potential for geriatric pharmacoepidemiology. SAGE Study Group. Systematic Assessment of Geriatric drug use via Epidemiology.' *Medical Care*, 36(2): 167–179.

Gambassi, G. Lapane, K. Sgadari, A. Landi, F. Carbonin, P. Hume, A. Lipsitz, L. Mor, V. Bernabei, R (1998a). 'Prevalence, clinical correlates, and treatment of hypertension in elderly nursing home residents. SAGE (Systematic Assessment of Geriatric Drug Use via Epidemiology) Study Group.' *Archives of Internal Medicine,* 158(21): 2377–2385.

Gambassi, G. Lapane, KL. Landi, F. Sgadari, A. Mor, V. Bernabei, R. (1999a). 'Gender differences in the relation between co-morbidity and mortality of patients with Alzheimer's disease. Systematic Assessment of Geriatric drug use via Epidemiology (SAGE) Study Group.' *Neurology,* 53(3): 508–516.

Gerteis, M. Gerteis, JS. Newman, D. Koepke, C. (2007). 'Testing consumers' comprehension of quality measures using alternative reporting formats.' *Health Care Financing Review,* 28(3): 31–45.

Grabowski, DC (2007). 'Medicare and Medicaid: conflicting incentives for long-term care. *Milbank Memorial Fund Quarterly*, 85(4): 579–610.

Gruber-Baldini, AL. Zimmerman, SI. Mortimore, E. Magaziner, J (2000). 'The validity of the minimum data set in measuring the cognitive

impairment of persons admitted to nursing homes.' *Journal of the American Geriatrics Society,* 48(12): 1601–1606.

Harris, Y. Clauser SB (2002). 'Achieving improvement through nursing home quality measurement.' *Health Care Financing Review,* 23(4): 5–18.

Hartmaier, SL. Sloane, PD. Guess, HA. Koch, GG. Mitchell, CM. Phillips, CD (1995). 'Validation of the Minimum Data Set Cognitive Performance Scale: agreement with the Mini-Mental State Examination.' *Journals of Gerontology Series A: Biological Sciences and Medical Sciences,* 50(2): M128–133.

Hawes, C. Fries, BE. James, ML. Guihan, M (2007). 'Prospects and pitfalls: use of the RAI-HC assessment by the Department of Veterans Affairs for home care clients.' *Gerontologist*, 47(3): 378–387.

Hawes, C. Mor, V. Phillips, CD. Fries, BE. Morris, JN. Steele-Friedlob, E. Greene, AM. Nennstiel, M (1997). 'The OBRA-87 nursing home regulations and implementation of the Resident Assessment Instrument: effects on process quality.' *Journal of the American Geriatrics Society,* 45(8): 977–985.

Hawes, C. Morris, JN. Phillips, CD. Fries, BE. Murphy, K. Mor, V (1997a). 'Development of the nursing home Resident Assessment Instrument in the USA.' *Age and Ageing,* 26(Suppl. 2): 19–25.

Hawes, C. Morris, JN. Phillips, CD. Mor, V. Fries, BE. Nonemaker, S (1995). 'Reliability estimates for the Minimum Data Set for nursing home resident assessment and care screening (MDS).' *Gerontologist,* 35(2): 172–178.

Hirdes, JP. Botz, CA. Kozak, J. Lepp, V (1996). 'Identifying an appropriate case mix measure for chronic care: evidence from an Ontario pilot study.' *Healthcare Management Forum*, 9(1): 40–46.

Hsia, DC. Krushat, WM. Fagan AB. Tebbutt, JA. Kusserow, RP (1988). 'Accuracy of diagnostic coding for Medicare patients under the prospective-payment system.' *New England Journal of Medicine*, 318(6): 352–355.

Ikegami, N. Fries, BE. Takagi, Y. Ikeda, S. Ibe, T (1994). 'Applying RUG-III in Japanese long-term care facilities.' *Gerontologist,* 34(5): 628–639.

Institute of Medicine (1986). *Improving the quality of care in nursing homes*. Washington DC: National Academies Press.

Intrator, O. Grabowski, DC. Zinn, J. Schleinitz, M. Feng, Z. Miller, S. Mor, V (2007). 'Hospitalization of nursing home residents: the effects of states' Medicaid payment and bed-hold policies.' *Health Services Research*, 42(4): 1651–1671.

Jorgensen, LM. el Kholy, K. Damkjaer, K. Deis, A. Schroll, M (1997). ''RAI' – an international system for assessment of nursing home residents [in Danish].' *Ugeskrift for Laeger,* 159(43): 6371–6376.

Kane, RA (2003). 'Definition, measurement, and correlates of quality of life in nursing homes: toward a reasonable practice, research, and policy agenda.' *Gerontologist,* 43(Spec. No. 2): 28–36.

Kane, RL. Arling, G. Mueller, C. Held, R. Cooke, V (2007). 'A quality-based payment strategy for nursing home care in Minnesota.' *Gerontologist,* 47(1): 108–115.

Konetzka, RT. Norton, EC. Sloane, PD. Kilpatrick, KE. Stearns, SC (2006). 'Medicare prospective payment and quality of care for long-stay nursing facility residents.' *Medical Care,* 44(3): 270–276.

Konetzka, RT. Yi, D. Norton, EC. Kilpatrick, KE (2004). 'Effects of Medicare payment changes on nursing home staffing and deficiencies.' *Health Services Research*, 39(3): 463–488.

Ljunggren, G. Brandt L (1996). 'Predicting nursing home length of stay and outcome with a resource-based classification system.' *International Journal of Technology Assessment in Health Care,* 12(1): 72–79.

Lowe, TJ. Lucas, JA. Castle, NG. Robinson, JP. Crystal, S (2003). 'Consumer satisfaction in long-term care: state initiatives in nursing homes and assisted living facilities.' *Gerontologist,* 43(6): 883–896.

Matherlee, K (1999). 'Implementing the BBA: the challenge of moving Medicare post-acute services to PPS.' *Issue Brief National Health Policy Forum*, (743): 1–8.

McGee, J. Kanouse, DE. Sofaer, S. Hargraves, JL. Hoy, E. Kleimann, S (1999). 'Making survey results easy to report to consumers: how reporting needs guided survey design in CAHPS. Consumer Assessment of Health Plans Study.' *Medical Care,* 37(Suppl. 3): MS32–40.

Mor, V (2004). 'A comprehensive clinical assessment tool to inform policy and practice: applications of the minimum data set.' *Medical Care,* 42(Suppl. 4): III50–59.

Mor, V (2005). 'Improving the quality of long-term care with better information.' *Milbank Memorial Fund Quarterly,* 83(3): 333–364.

Mor, V. Angelelli, J. Jones, R. Roy, J. Moore, T. Morris, J (2003). 'Inter-rater reliability of nursing home quality indicators in the US' *BMC Health Services Research*, 3(1): 20.

Mor, V. Berg, K. Angelelli, J. Gifford, D. Morris, J. Moore, T (2003a). 'The quality of quality measurement in US nursing homes.' *Gerontologist,* 43(Spec. No. 2): 37–46.

Mor, V. Branco, K. Fleishman, J. Hawes, C. Phillips, C. Morris, J. Fries, B (1995). 'The structure of social engagement among nursing home residents.' *Journals of Gerontology Series B, Psychological Sciences and Social Sciences,* 50(1): 1–8.

Mor, V. Intrator, O. Fries, BE. Phillips, C. Teno, J. Hiris, J. Hawes, C. Morris, J (1997). 'Changes in hospitalization associated with introducing the

Resident Assessment Instrument.' *Journal of the American Geriatrics Society,* 45(8): 1002–1010.

Morris, JN. Fries, BE. Mehr, DR. Hawes, C. Phillips, C. Mor, V. Lipsitz, LA (1994). 'MDS Cognitive Performance Scale.' *Journal of Gerontology,* 49(4): 174–182.

Morris, JN. Fries, BE. Morris, SA (1999). 'Scaling ADLs within the MDS.' *Journals of Gerontology Series A, Biological Sciences and Medical Sciences,* 54(11): 546–553.

Morris, JN. Hawes, C. Fries, BE. Phillips, CD. Mor, V. Katz, S. Murphy, K. Drugovich, ML. Friedlob, AS (1990). 'Designing the national resident assessment instrument for nursing homes.' *Gerontologist,* 30(3): 293–307.

Morris, JN. Nonemaker, S. Murphy, K. Hawes, C. Fries, BE. Mor, V. Phillips, C (1997). 'A commitment to change: revision of HCFA's RAI.' *Journal of the American Geriatrics Society,* 45(8): 1011–1016.

Mukamel, DB. Spector, WD. Zinn, JS. Huang, L. Weimer, DL. Dozier, A (2007). 'Nursing homes' response to the nursing home compare report card.' *Journals of Gerontology Series B, Psychological Science and Social Science,* 62(4): 218–225.

Noro, A (2005). Vertailukehittämishanke [Benchmarking Project]. In: Noro, A. Finne-Soveri, H. Björkgren, M. Vähäkangas, P (eds.). *Ikääntyneiden laitoshoidon laatu ja tuottavuus – RAI-järjestelmä vertailukehittämisessä [Quality and efficiency of the institutional long-term care for older persons - RAI-systems in benchmarking].* Saarijärvi: STAKES.

Phillips, C. Hawes, C. Lieberman, T. Koren, MJ (2007). 'Where should Momma go? Current nursing home performance measurement strategies and a less ambitious approach.' *BMC Health Services Research,* 7(1): 93.

Phillips, CD. Morris, JN. Hawes, C. Fries, BE. Mor, V. Nennstiel, M. Iannacchione, V (1997). 'Association of the Resident Assessment Instrument (RAI) with changes in function, cognition, and psychosocial status.' *Journal of the American Geriatrics Society,* 45(8): 986–993.

Popejoy, LL. Rantz, MJ. Conn, V. Wipke-Tevis, D. Grando, VT. Porter, R (2000). 'Improving quality of care in nursing facilities. Gerontological clinical nurse specialist as research nurse consultant.' *Journal of Gerontological Nursing,* 26(4): 6–13.

Rahman, A (2006). 'The Pay for Performance demonstration program.' *Journal of the American Medical Directors Association,* 7(2): 133.

Rahman, AN. Schnelle JF (2008). 'The nursing home culture-change movement: recent past, present, and future directions for research.' *Gerontologist,* 48(2): 142–8.

Reilly, KE. Mueller, C. Zimmerman, DR (2007). 'The Centers for Medicare and Medicaid Services' Nursing Home Case-Mix and Quality Demonstration: a descriptive overview.' *Journal of Aging and Social Policy,* 19(1): 61–76.

Roy, J. Mor V (2005). 'The effect of provider-level ascertainment bias on profiling nursing homes.' *Statistics in Medicine,* 24(23): 3609–3629.

Sangl, J. Saliba, D. Gifford, DR. Hittle, DF (2005). 'Challenges in measuring nursing home and home health quality: lessons from the First National Healthcare Quality Report.' *Medical Care,* 43(Suppl. 3): I24–32.

Scharpf, TP. Colabianchi, N. Madigan, EA. Neuhauser, D. Peng, T. Feldman, PH. Bridges, JF (2006). 'Functional status decline as a measure of adverse events in home health care: an observational study.' *BMC Health Services Research,* 6: 162.

Schnelle, JF. Cadogan, MP. Yoshii, J. Al-Samarrai, NR. Osterweil, D. Bates-Jensen, BM. Simmons, SF (2003). 'The minimum data set urinary incontinence quality indicators: do they reflect differences in care processes related to incontinence?' *Medical Care,* 41(8): 909–922.

Simmons, SF. Cadogan, MP. Cabrera, GR. Al-Samarrai, NR. Jorge, JS. Levy-Storms, L. Osterweil, D. Schnelle, JF (2004). 'The minimum data set depression quality indicator: does it reflect differences in care processes?' *Gerontologist,* 44(4): 554–564.

Sofaer, S. Firminger K (2005). 'Patient perceptions of the quality of health services.' *Annual Review of Public Health,* 26: 513–559.

Stadt, J. Molare E (2005). 'Best practices: that improved patient outcomes and agency operational performance.' *Home Healthcare Nurse,* 23(9): 587–593.

Straker, JK. Ejaz, FK. McCarthy, C. Jones, JA (2007). 'Developing and testing a satisfaction survey for nursing home residents: the Ohio Experience.' *Journal of Aging and Social Policy*, 19(2): 83–105.

Wodchis, WP. Fries, BE. Hirth, RA (2004). 'The effect of Medicare's prospective payment system on discharge outcomes of skilled nursing facility residents. *Inquiry,* 41(4): 418–434.

Wu, N. Miller, SC. Lapane, K. Roy, J. Mor, V (2005). 'Impact of cognitive function on assessments of nursing home residents' pain.' *Medical Care,* 43(9): 934–939.

Wu, N. Miller, SC. Lapane, K. Roy, J. Mor, V (2005a). 'The quality of the quality indicator of pain derived from the minimum data set.' *Health Services Research,* 40(4): 1197–1216.

Zimmerman, DR (2003). 'Improving nursing home quality of care through outcomes data: the MDS quality indicators.' *International Journal of Geriatric Psychiatry,* 18(3): 250–257.

Zinn, J. Spector, W. Hsieh, L. Mukamel, DB (2005). 'Do trends in the reporting of quality measures on the nursing home compare web site differ by nursing home characteristics?' *Gerontologist,* 45(6): 720–730.

Zinn, JS (1994). 'Market competition and the quality of nursing home care.' *Journal of Health Politics, Policy and Law,* 19(3): 555–582.