

## Comparing how to Compare: An evaluation of alternative performance measurement systems in the Field of Social Care

Clarkson, P., Challis, D., Davies, S., Donnelly, M., Beech, R., & Hirano, T. (2010). Comparing how to Compare: An evaluation of alternative performance measurement systems in the Field of Social Care. *Evaluation*, 20(1), 59-79. DOI: 10.1177/1356389009350020

**Published in:**  
Evaluation

### Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

#### General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

#### Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [openaccess@qub.ac.uk](mailto:openaccess@qub.ac.uk).

# Evaluation

<http://evi.sagepub.com>

---

## **Comparing How to Compare: An Evaluation of Alternative Performance Measurement Systems in the Field of Social Care**

Paul Clarkson, David Challis, Sue Davies, Michael Donnelly, Roger Beech and Takayuki Hirano

*Evaluation* 2010; 16; 59

DOI: 10.1177/1356389009350020

The online version of this article can be found at:  
<http://evi.sagepub.com/cgi/content/abstract/16/1/59>

---

Published by:



<http://www.sagepublications.com>

On behalf of:



The Tavistock Institute

**Additional services and information for *Evaluation* can be found at:**

**Email Alerts:** <http://evi.sagepub.com/cgi/alerts>

**Subscriptions:** <http://evi.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.co.uk/journalsPermissions.nav>

**Citations** <http://evi.sagepub.com/cgi/content/refs/16/1/59>



# Comparing How to Compare: An Evaluation of Alternative Performance Measurement Systems in the Field of Social Care

Evaluation  
16(1) 59–79

© The Author(s) 2010

Reprints and permission: <http://www.sagepub.co.uk/journalsPermission.nav>

DOI: 10.1177/1356389009350020

<http://evi.sagepub.com>



**Paul Clarkson, David Challis, Sue Davies**

University of Manchester, UK

**Michael Donnelly**

Queen's University, Belfast, Northern Ireland

**Roger Beech**

Keele University, UK

**Takayuki Hirano**

Nihon Fukushi University, Japan

## Abstract

Comparative performance evaluation has taken different forms depending upon the purposes of performance monitoring and the types of measures available. This paper investigates the different performance measurement systems in place in the social care setting, in particular for older people receiving community care services. In England, earlier systems to assist performance management within organizations have been eclipsed by national systems of regulation with top-down implementation of standards and measures. In Northern Ireland, by contrast, organizations have been compared descriptively without the use of national targets. Internationally, in Japan, organizations arranging similar services have had more local information available collected in a bottom-up fashion with greater employment of service user-level data. These differences in performance evaluation are located within an analytical framework permitting comparisons of system design and the use of measures. Conclusions are drawn concerning the breadth of evidence available for successfully monitoring service provision in this setting.

## Keywords

institutional perspective; local management; performance evaluation; social care

---

## Corresponding author:

Paul Clarkson, PSSRU, University of Manchester, First Floor, Dover Street Building, Oxford Road, Manchester M13 9PL, UK.  
Email: [paul.clarkson@manchester.ac.uk](mailto:paul.clarkson@manchester.ac.uk)

## Introduction

The growth of 'evaluative systems' in government and public sector organizations has been a subject of debate (Leeuw and Furubo, 2008). One evaluation system of note is the performance monitoring or performance measurement system, which many suggest shows both complementarities and differences with evaluation as broadly conceived (Davies, 1999; Nielsen and Ejler, 2008). In this paper we seek to describe and to critically assess the different performance measurement systems in existence in the field of social care, where public welfare services, primarily the responsibility of local authorities, are subject to different monitoring arrangements in different countries, both of the United Kingdom (UK) and more widely. These different approaches are rooted in different regimes for monitoring performance and for the reporting of information. In other words, these approaches are built around contrasting social systems, which cannot be divorced from their approaches to evaluative activity as a whole (Brignall and Modell, 2000; Leeuw and Furubo, 2008). These different systems are seen to reflect the different purposes of performance monitoring and also generate differences in the types of measures available.

Our comparison of the different approaches used to monitor performance in the social care setting draws on material and data from a study under the UK Economic and Social Research Council's Public Services Programme (Clarkson et al., 2007). The study primarily employed a national survey in two countries of the UK, England and Northern Ireland, intended to form an overview of the different arrangements in place for monitoring social care services from the perspectives of two constituents: 'performance leads' – information officers and local authority personnel responsible for the design of performance measurement systems and data collection within their agencies – and senior managers – responsible for the day-to-day running of services to older people, who form the bulk of recipients of social care services.<sup>1</sup> We take the example of these two UK countries whose social care systems, broadly speaking, provide similar types of services and which are similarly financed, but which have developed quite different mechanisms for monitoring and evaluating performance. The value in comparing these two countries is that they provide a focus for a 'natural experiment' in the construction of performance monitoring systems across two devolved administrations of the UK, with the opportunity for relating the conduct of the systems to a range of institutional, political and cultural factors. These arrangements within the UK are contrasted with those in Japan, a country where social care provision is differently financed and which offers a contrasting approach in terms of the use of performance information but which, nevertheless, shares certain similarities in terms of the kinds of social care services provided and the role of 'street-level' professionals ('care managers') with responsibility for coordinating services. In Japan, comparable organizations (local municipalities) are charged with monitoring the services available through the provision of public long-term care insurance (Campbell and Ikegami, 2003). Data from the performance measurement system in Japan were collected through the software program adopted by over 60 percent of municipalities in that country, engaged in analysing the implementation of the insurance system locally (Community Care Policy Network, 2001; Institute of Health and Economic Programmes, 2002).

We begin with a brief overview of social care services and the systems prevalent in the three countries to first of all set the context for our comparison. We then offer some examples of the kinds of evaluative activity undertaken under the rubric of performance measurement of social care within these countries. It will be seen that the three countries considered here exhibit very different systems regarding performance evaluation and that these systems are changing over time. These different systems also show diverse linkages with other evaluation systems, such as those of inspection or oversight and monitoring and evaluation (Leeuw and Furubo, 2008). Finally, we use

an analytical framework, or 'logic model' to compare the different systems in use, concentrating on aspects of system design and the use of measures. On this basis we draw conclusions concerning the breadth of evidence available for monitoring social care services, according to a conception of performance evaluation as being embedded within complex institutional arrangements that largely define the scope and use of measures.

## Overview of performance measurement systems

It has been argued that performance measurement and evaluation are complementary but distinct forms of activity (Nielsen and Ejler, 2008). In a relatively early paper on performance evaluation, Arvidsson (1986) highlighted the pressures for public services to more thoroughly measure their performance but also drew attention to a lack of technical capacity for so doing. His account distinguished performance evaluation from other evaluative activities that may occur in government (such as policy analysis) according to several aspects: *object* (the performance of programmes or undertakings are measured rather than people, patients, or clients per se); *time perspective* (past or present performance is analysed); *criterion* (performance is measured against the accomplishment of objectives, which are multidimensional, rather than purely 'procedural' concerns); and *evaluator's position* (the purposes behind performance evaluation are linked to the interests of different stakeholders and so this type of evaluation will tend to vary according to the role of different evaluators advocating it). Performance measurement is thus concerned with comparing (programmes, units) against objectives or results and Arvidsson's account brings to light the fact that the 'dynamics of performance evaluation' (1986: 627) will vary with organizational context.

Since that time, initiatives under the New Public Management reforms have brought about a range of performance measurement systems, directed towards different audiences, with varied aims and uses of information (for an account within European local governments, see Bloomfield, 2006). Hood (2007) describes the variants of these systems under three headings: management 'intelligence' (whereby information is collected concerning performance but is not interpreted in any fixed way, with no explicit grading of units); target systems (where performance is measured against a specified standard); and ranking systems (where performance of different units is explicitly compared). It has been argued that in the UK, and particularly England, management intelligence has largely been superseded by target and ranking systems in the public sector in a range of settings, particularly over the past 10 years. In other countries, performance evaluation has taken different forms, often involving hybrids of these three applications, stressing some to a greater extent than others. The particular characteristics of societies, their governance and culture, have been offered as partial explanations for the rise of different forms of performance evaluation in different countries. Different settings – such as education or central government – have also stressed some applications to the exclusion of others (Hood, 2007; Hood et al., 2004).

Social care, in contrast to health care, is a relative newcomer to performance measurement. Services, largely dealing with personal and domestic care to vulnerable populations, are often located within local government but the precise configuration of these and their relationship to acute health care differs across jurisdictions. A major investment, in terms of finance and personnel, within social care lies in the care of older people, reflecting ageing populations (Timonen, 2005). The development of social care services to older people and the associated budgetary and consumer-driven pressures to account for expenditure and performance explain the rise of performance evaluation and quality monitoring in this setting (Wiener et al., 2007).

Performance measurement in the settings to be described is principally concerned with comparison. As Arvidsson (1986) makes clear, it is the relative success of objectives that is compared, often

across units of provision. Progress on achievement is compared with a standard, which denotes when performance is considered acceptable or when there is a need to consider remedial action, although the relative formality of such standards differs between countries. In aged care, administrative efficiency (containing costs), maintaining care in community-based settings rather than institutions, and monitoring responsiveness to consumers are central issues of concern. Each of the countries has responded to these issues in different ways and, in the following sections, we investigate *how* comparison takes place to monitor performance. We first describe the systems in each country, concentrating on the aims of performance measurement, who data are provided for and their form. Second, we present some concrete examples of practice, which make it possible to discover the links between performance and other forms of evaluative activity existing in the three countries. We then compare 'how to compare' by locating the performance measures employed by the three systems within a framework, determined by both the production process for social care and the levels at which performance data are designed and analysed.

## England

Social care in England, organized through local authorities, has shown marked variation as local agencies were primarily left to monitor their own arrangements, admittedly with enabling legislation from the centre, until a government White Paper in 1998 (Dept of Health, 1998) set out a tighter regulatory regime with the associated use of targets and rankings of performance (Challis et al., 2006a). This system was a reaction to the perceived problems around area variation and the lack of national standards: the so-called 'post-code lottery' in the delivery of services (House of Commons Health Committee, 2005). Before that time, social care performance evaluation was primarily conceived of as a system to monitor inefficiency and waste and developments to monitor services through collections of performance indicators were vigorously opposed by front-line social workers who perceived such measures as blunt tools for monitoring service delivery (Llewelyn, 1996). Before the reforms of 1998, social care performance information in England was, on the whole, orientated towards 'management information', whereby data were used to describe the operation of different units within authorities and to raise questions about policy directions (Barnes and Miller, 1988; Miller, 1986). However, different authorities used measures in different ways, with varying degrees of sophistication, and these were not standardized across the country. Inter-authority comparisons were made on a national basis (e.g. Audit Commission, 1996; Dept of the Environment, 1981) but these, again, functioned primarily as management 'intelligence' in that local authorities were not compared against explicit national standards nor were they ranked against each other. These kinds of data essentially functioned as 'vital statistics' on the operation of social care throughout the country or in the context of similar local authorities (Miller, 1986).

The performance regime that developed in England after 1998 moved towards a system of explicit *rankings* of performance whereby composite ('star') ratings and a general set of national ('Performance Assessment Framework') indicators were used to compare social services provision across all local authority organizations. The aim behind this was to ensure that these organizations 'drive up their standards to match those of the best' and 'secure continuous improvements in performance' (Dept of Health, 1998: paras 7.3 and 7.4). This aim was operationalized with reference to broad objectives such as to maintain older people in community settings and for services to be responsive to individual needs and preferences as well as to maintain cost efficiency (Henwood and Wistow, 1999). *Targets* were also often set, such as those related to improvements in efficiency (Dept of Health, 2003a). Often average standards were used from which to compare authorities (usually the median value on a particular indicator), using aggregated data at the local authority

level. This comparison, however, was not merely for descriptive purposes but was coupled with incentives for local social care organizations to improve through the treatment of ratings as rewards and punishments; central government had the ultimate sanction to take responsibilities away from 'poorly performing' authorities (Huber, 1999). Thus, comparison was explicitly tied to the regulation of local authorities by central government and data were used for accountability purposes, to oversee the conduct of local agencies and to provide information for the public (Milburn, 2001).

The response to these central government-imposed measures and oversight has been one of reduced morale on the part of social workers and managers (Burnham, 2006). Social workers, acting as 'care managers' to assess and plan the care delivered, are principally those who generate the activity from which are derived the data used for performance review, but most monitoring takes place at the managerial level (Challis et al., 2006a). There was some concern from those subject to this regime that the measures employed represented a fairly crude means of judging achievement. One of our survey respondents replied that:

Nationally set indicators need to avoid ambiguity, data [are] flawed due to interpretation; authorities are held to account against inaccurate benchmarking. Systems [are] not designed to work with front-line professionals [and] after-the-event recording creates mistakes and omissions. (English manager)

More recently, however, this system is changing. In line with changes in political values concerning how to measure performance in this sector, a new reduced set of measures has instead been advanced as necessary to monitor the work of social services (Dept for Communities and Local Government, 2007). This will involve the cooperation of local authorities rather than their being subject to competitive ratings from central government, although it is too early to tell what, if any, standards will be set.

### *Northern Ireland*

Social care in Northern Ireland has, in the content of practice, largely followed the lines of that in England, for example the key professionals assessing and planning care locally are, similarly, care managers (Challis et al., 2006b). However, the form of organization is different. In Northern Ireland, social care and health care are integrated at the organizational level through Health and Social Services Trusts (in England, by contrast, responsibility for social care and health lies with local and central government respectively). These organizations, of which there are now five (formerly there were 11), have a high degree of autonomy in relation to higher authorities to which they are accountable – Health and Social Services Boards and the Department of Health, Social Services and Public Safety (DHSSPS). Integration might be expected to offer a more flexible and efficient working environment which would lead to improved performance in social care, through, for example, the smoother transfer of patients from hospital to support services in the community. However, few research evaluations have tested out whether these organizational arrangements have brought about benefits compared with England,<sup>2</sup> although a recent review has concluded that these arrangements have varied across trusts with consequent variations in performance (Appleby, 2005). There have been further concerns that Northern Ireland is lagging behind England in its health and social care performance, particularly in waiting times for treatment and delays in discharges from hospital, and one reason cited is the relative absence of incentives for local agencies to improve through national targets, as exist in England. Existing health and social services performance management is viewed as lacking 'appropriate performance structures, information and clear and effective incentives – rewards and sanctions – at individual, local and national organizational levels to encourage innovation and change' (Appleby, 2005: 162). This review

noted that accountability arrangements were unclear and uncertainty as to which part of the system was responsible for poor performance.

The performance regime in Northern Ireland, similar to England, involves reporting of information at a national level through the 'Key Indicators' publications produced since 1998. However, there is no explicit ranking of units, as has existed in England, and no composite measures used to construct 'league tables' of performance. Instead, the aim of this system is to provide 'detailed comparative information, which affords Boards and Trusts the opportunity to view their Board/Trust in relation to others, [to] raise questions in relation to expenditure, the level and complexity of service provision and the variations that exist across Trusts' (Northern Ireland Statistics and Research Agency, 2003: 21). Data are therefore used to compare across units but not to offer any assessment as to which are performing poorly; that is, descriptively as 'intelligence'. The regime's purpose is to promote debate, stimulate research and influence decision-making (Northern Ireland Statistics and Research Agency, 2003: 23). Thus, the Executive and local agencies are more involved in a mutual exchange of information rather than, as in England, an adversarial relationship with a lack of trust in the ability of local authorities to self-manage.

Recently, however, changes have been initiated in response to the perceived problems with performance, particularly in comparison with England. A major reorganization implemented new structures from April 2009 and there will be a sharper focus on performance measurement, in particular the meeting of targets (DHSSPS, 2008).

## Japan

Japan represents an interesting case study in terms of performance evaluation. As a country in which the drive towards formalized measures of quality and implicit control mechanisms in the business sector were stressed throughout the 1980s, through such pioneers as Ouchi (1981) and Deming (1982), it may be expected that the public services too would have been infused with a performance culture. However, in contrast to England, Japan has not had a tradition of central government monitoring and performance measurement in the public sector has arrived relatively late. Both prior to and since the Government Policy Evaluation Act (GPEA), similar to the Government Performance Results Act (GPRA) in the United States, of June 2001 (Ministry of Public Management, 2001), local government (municipalities) has taken the lead in developing measurement systems. It is an open question, however, whether municipalities have sufficiently developed the appropriate tools with which to monitor performance and development work is ongoing. As Japan has experienced the most rapid growth in the older population in the world, it has faced mounting financial and social pressures to reform its long-term care provision for older people (Campbell, 1992). Therefore, monitoring of performance is crucial to evaluating recent changes there.

Since 1 April 2000 Japan has introduced a mandatory system of Long-term Care Insurance (LTCI) (*Kaigo Hoken*) to provide services for frail older people who require care (Campbell and Ikegami, 2003; Ikegami, 1997). Half of the costs of this system are financed through general taxation and the other half through insurance premiums. Eligibility for benefits under LTCI is determined by a computerized 'needs assessment' of functional and cognitive status (Ikegami, 2007; Morris et al., 1997), with this classified into six dependency levels. Each of these levels determines the maximum amount of benefits available (the 'benefit limitation' amount) for older people and their families to purchase long-term care services, which include geriatric units in hospitals, nursing and home care. Care managers operate to coordinate and plan the delivery of these services and have an important role in monitoring the quality of services on behalf of individual clients, but do not have a formal role in the performance measurement system (Wiener et al., 2007). Local municipalities (cities, towns and



villages) are the insurers and are responsible for setting premiums, overseeing the provision of services and managing the finances within a framework set by the Ministry of Health, Labour and Welfare. There is a 10 percent co-payment by older users at the point of service provision, with the remaining cost covered half by premiums and half by taxation with deductions or waivers for those on low incomes (Ikegami, 1997). As local government units (municipalities) are the insurers for LTCI, a need emerged to monitor the administration and efficient running of this system at the local level.

A survey of local governments' utilization of performance measurement in Japan by Tanaka (2006) highlighted the diversity in the foci of measurement and use of measures. Managers utilizing performance measures in local government were broadly satisfied with the design of their systems, but less so with aspects such as the precise measures and targets employed. The difficulties experienced by local government with the implementation of performance measures to measure progress of the insurance system have increased the role of third-party researchers in development work. These, in contrast to government regulators, appear to have been more important in Japan than in other countries (Tanaka and Ono, 2002; Wiener et al., 2007). One example of this is a research team at Nihon Fukushi University, Nagoya, convened at the inception of LTCI. Funded initially by a central government grant, the team planned a framework around which to devise indicators for monitoring LTCI focusing on input, output and outcome indicators. In contrast to England, the primary demand for a performance monitoring system came from the municipalities. Local managers responsible for administration required a system that would enable them to evaluate the operation of LTCI in their areas in comparison with other municipalities. Therefore, the role of the research team was to support the municipalities in developing comparative analyses, not to evaluate them in terms of league tables or 'star ratings' as has been the predominant concern of government in England (Adab et al., 2002; Dept of Health, 2003b). The performance system subsequently developed thus formed part of a management control system to ask relevant questions concerning the operation of LTCI and to assess whether the system was performing as planned. In June 2001 the research team developed a software programme for the analysis of performance under LTCI and during 2002 the programme was delivered to all municipalities along with a manual to assist operation (Community Care Policy Network, 2001). This application software was a tool enabling the original electronic data generated by the LTCI system to be translated into data reports including some performance indicators. Due to the nature of the insurance-based system, data are available at the user or client level and can be analysed according to dependency, including measures of functional, cognitive and sensory status.

More recently, in 2004, the team delivered the software to municipalities to enable comparative and time-series analyses to take place. Planners in each municipality automatically receive the database concerning users of LTCI services and a monthly monitoring report containing the analysis of selected indicators. This report is used as feedback to municipalities in order to facilitate their evaluation, priority setting and planning. A number of pressing problems remain to be tackled in terms of the operation of LTCI. The number of beneficiaries has increased, reflected in expenditure patterns and some budget deficits.<sup>3</sup> The system was subject to review in 2005 and monitoring is more recently directed towards analysing expenditure and trends in service receipt by dependency so as to potentially avoid unsustainable demand (Tsutsui and Muramatsu, 2007).

## **Performance measurement of social care: Examples of practice**

Some examples of uses of measures and the types of analyses sanctioned as part of the systems prevalent in the three countries are now considered. It needs to be stressed that these systems were those in use at the time our survey data were collected (throughout 2008) but that they are changing in line with shifting perspectives on the need for performance information from different stakeholders.

## England

At a national level in England there have been around 20 indicators monitoring services to older people, with these concentrating on the achievement of objectives related to central government priorities: care at home rather than in institutions (e.g. delays in care between hospitals and the community); efficiency; and customer responsiveness (e.g. timeliness in the delivery of assessments and services) (Commission for Social Care Inspection, 2007). Authorities have been compared against national standards so as to judge achievement. Analysis of performance against these standards has been relatively simple, concentrating on authorities' distribution on the indicators and ratings against the average or against a defined direction denoting good performance (such as a low figure for the number of residential admissions). Composite indicators have been used as league tables testifying to which authorities are judged to be 'excellent', 'good' or 'adequate'. From such measurement, incentives have been used to direct the performance of authorities with 'naming and shaming' for poor performers and 'earned autonomy' for those performing well. Therefore, in terms of the inventory offered by Leeuw and Furubo (2008), the English system falls into the category of 'performance monitoring' but has been a system that is highly centralized and thus shows links with 'inspection and oversight' regimes.

## Northern Ireland

Indicators in Northern Ireland, published annually, are also aggregated at a national level with authorities as the unit of analysis. The measures concentrate on services and 'care packages' (the content of the services each user receives) provided across Trusts. There are broad indications of the direction in which services are moving, such as increases in expenditure or reductions in the numbers entering residential care. This information is used comparatively but without any judgements being made between Trusts. Instead, the performance measures are used to facilitate the monitoring and review of individual local services. This system is, again, a 'performance monitoring' one (Leeuw and Furubo, 2008) but without the strong oversight and competitive characteristics prevalent in England.

## Japan

Indicators derived from data concerning the LTCI system are arranged at two levels: a 'simple benchmarking set' containing indicators designed from routinely collected data and a 'comprehensive benchmarking set', which includes data from additional surveys. As routine data are constructed for the purpose of the insurer (containing data on persons insured, eligibility and service use) they are relatively easy to collect, although they do not provide a comprehensive comparison of outcomes for older participants. The comprehensive dataset is therefore more detailed and provides outcome indicators in terms of both objective factors (changes in care level; remaining at home or entering institutions) and subjective factors using detailed health and functional data (changes in carer burden, morale and depression, for example). These indicators, allied to routine data for the same subjects, are available within ten municipalities in Aichi prefecture (the south-east region from Nagoya), covering a sample of approximately 6,000 older people with disabilities.

Several discrete data collections and evaluations have also been prepared from the data on LTCI. The research team have conducted a national survey to determine the content and effectiveness of satisfaction surveys of older users carried out by municipalities. The team have also conducted a community survey in 10 municipalities (the Aichi Gerontological Evaluation Study) which used standardized scales

of depression, morale and carer burden as outcome measures on a sample of disabled elderly people ( $n = 11,000$ ) eligible for LTCI services (Kondo, 2003). These data were linked to official data on eligibility, the contents of care plans, costs of care provided under LTCI and circumstances of death. This dataset enables comparisons to be made concerning disparities in performance between municipalities using subjective outcome measures in one prefecture (second tier of government, group of municipalities).

Analysis of these indicators has taken the form of comparisons across insurer (municipalities), across care package, and in terms of costs at each care level. Relatively sophisticated regression analyses have also been undertaken to test the influence of various factors on cost (per recipient) and on the relative 'balance' between home and residential care.

The Japanese indicators, and their analysis, therefore form part of a 'monitoring and evaluation' system (Leeuw and Furubo, 2008) that aims to evaluate the consequences of the programme with its close links to a university institution. This performance system is also allied with experimental (or 'quasi-experimental') evaluation and 'evidence-based policy'; in other words, it is concerned with 'what works'. This system is therefore quite different from the other UK countries in that it shares many commonalities with other forms of evaluation.

## Comparing how to compare: An evaluation of the different systems

Looking across these systems, it is evident that different aims are ascribed to performance evaluation in the three countries. In England the primary aim is one of control of subordinate agencies by central government (or its arms-length inspection agency);<sup>4</sup> in Northern Ireland, the aim is one of describing the complexity of outputs of the system so that local Trusts can compare their provision with others and plan for service provision locally. In Japan, the system is geared towards monitoring of the long-term care insurance scheme by municipalities and thus provides more detailed data at the local level. In comparing these systems we draw a similar conclusion to Behn (2003) that these different purposes generate different types of measures, although there are similarities between them in the types of domain-specific measures chosen. Table 1 lists some of the main differences between the cases analysed, in terms of the performance systems of the three countries; concentrating on aims, context, the data available; issues considered; and any perceived problems.

We now attempt to evaluate the different systems in terms of their design and use of measures. To accomplish this, a production model is used as a starting point to determine whether all relevant aspects of performance are covered (Arvidsson, 1986). Such 'logic models' (Schacter, 2002) are used to frame the design and planning of performance systems in that they dictate what measures need to be employed to permit comprehensive evaluation in different settings. Therefore, such models are the first step from which to design and collect relevant indicators. Models are not always used, particularly when performance is measured by a regulator as in the UK examples. However, we use a model here in a normative manner, to compare the systems in terms of which aspects should be included to offer a breadth of evidence concerning performance in a particular setting. The Performance Indicator Analytical Framework (PIAF) clarifies the domains of evaluation, from which indicators should be developed, and the relationships between them (Challis et al., 2006a ; Clarkson and Challis, 2003). Here, each domain, following the production process characterizing social care, is defined as:

Need – 'who needs care' (e.g. number of older people living alone);

Contextual Factors – 'aims to be achieved' (e.g. national/local objectives);

**Table 1.** The three 'cases' of performance measurement systems in social care

	England	Northern Ireland	Japan
Aim	To regulate and direct performance of local agencies by central government	To develop understanding of complex system	Evaluation of local provision of mandatory insurance system
Organizational context	Centralized accountability relationships between national and local governments (150)	Accountability relationships of mutual learning and development between Executive and local Trusts (5)	Local control by municipalities (1,787 in July 2008) within a framework set by central government
Form of data	Aggregate authority-wide data for explicit comparison ('ranking'); very few local data particularly on characteristics and dependency of users	Aggregate authority-wide data for descriptive comparison ('intelligence'); very few local data particularly on characteristics and dependency of users	Individual-level data for purposes of management control ('intelligence'), including that characterizing dependency of users
Performance issues	Cost efficiency; timeliness; balance between home and residential care	Cost efficiency; range of outputs; balance between home and residential care	Cost efficiency; balance between home and residential care
Perceived problems	Lack of trust in rankings; gaming and manipulation of data by local agencies	Lack of incentives and unclear lines of accountability for improved performance	Rising costs of insurance scheme; appropriate analysis of data by local municipalities

Supply – 'what is available' (e.g. number of home care hours);

Practice Process – 'what care managers do' (e.g. number of ongoing cases);

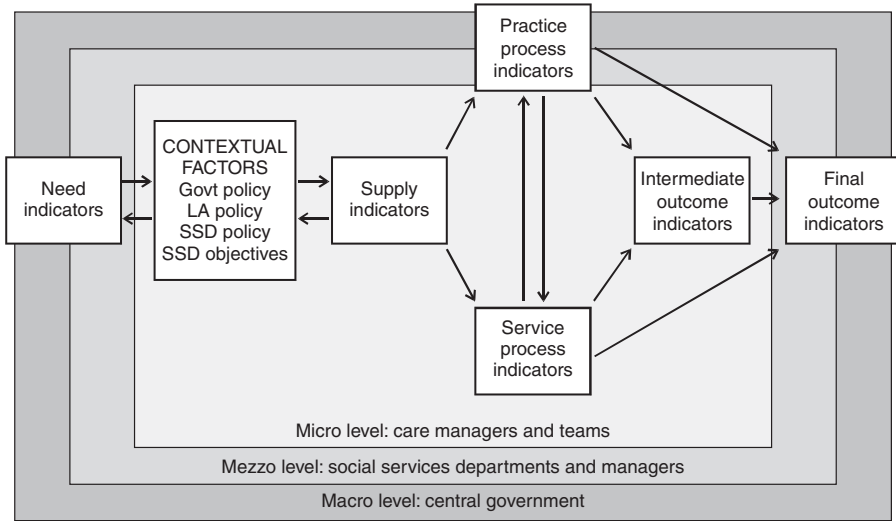
Service Process – 'how services are organized' (e.g. costs of care packages);

Intermediate Outcome – 'what is provided' (e.g. number of residential admissions);

Final Outcome – 'with what effect' (e.g. number of users satisfied with service).

We combine this model with three levels of analysis at which data can be aggregated: the *macro* level of central government; the *mezzo* level of social services departments (or their equivalents) and managers; and the *micro* level of care managers (professional staff – common to the three countries – who screen, assess, plan care and provide ongoing contact and review (Applebaum and White, 2000)) and teams. Together, this framework (Figure 1) helps to see whether all appropriate aspects of social care are considered in each of the three cases.

In Table 2 details on the availability of performance data for each of the six domains of the PIAF model (excluding the contextual factors domain already described) are combined with the three levels of analysis. Information to populate this table was drawn from national sources in the three countries combined with our survey data from England and Northern Ireland and material from the software program analysing the Japanese performance data. For these sources, the most up-to-date information available, at the time of writing, was used.



**Figure 1.** The Performance Indicator Analytical Framework at different levels of analysis

Sources: Clarkson and Challis (2003); Challis et al. (2006a).

Key: LA = local authority; SSD = social services department (or equivalent).

Table 2 shows some interesting differences between the countries in terms of their use of performance measures and the range of information available at different levels of analysis. In England, with its pronounced use of centrally prescribed measures, there is a deficit of information at the local level (both authority wide and amongst individual teams/users). However, there was some variability in authorities' arrangements for collecting performance information. Overall, only 41 percent of authorities had locally developed performance indicators (defined as those linked to objectives set out in local service or business plans or to assist managers in running the service). More specifically, between 13 and 21 percent of authorities had indicators available locally characterizing each domain of our model. Need indicators were most generally available, followed closely by intermediate outcome indicators. However, these locally developed indicators were limited compared to those required at a national level. Indicators aggregated at the level of teams or areas within authorities, or at an individual user level, were also limited. Whilst only half of authorities had information on the number of cases per team or individual worker, for example, only 13 percent of authorities routinely used data on the spend per case. In contrast to this, a range of information was available at a national level with Practice Process and Intermediate Outcome domains being well represented.

In Northern Ireland, the number of units is smaller and, whilst local measures were available these were largely at an early stage of development. A greater part of the time of performance leads, on the whole, was spent in preparing information for statutory (national) reporting. However, these data, appropriately disaggregated, were also used for internal performance monitoring within Trusts. Two Trusts, for example, had data available on the 'spend per case' and 'case load size', indicative of a more localized approach to measurement.

In Japan, the striking difference is in the use of individual-level measures. Because, unlike in England, data are collected routinely in electronic form at the local level, this enables local

Table 2. The range of measures in the three systems<sup>5</sup>

Levels of analysis/domain	England	Northern Ireland	Japan
Macro level (central government) Need	x	Prop. of population over 65/75; socio-economic indicators	*
Supply	Cost of intensive social care; unit costs of home care; availability of single rooms in residential care	Residential places in homes (by type); Total cost per capita	*
Practice Process	Prop. of people receiving statement of needs and how they are to be met, prop. receiving a review, delayed transfers of care from hospital, acceptable waiting times for assessment/care package, ethnicity of those receiving assessment/services	No. of care packages in year	*
Service Process	Prop. of equipment delivered within 7 days	Respite admissions to residential care; occupancy rates; no. of day care/home care/meals users	*
Intermediate outcome	Intensive home care as prop. of total care; households receiving intensive home care; older people helped to live at home; older people receiving direct payments; carers receiving a carers' service; admissions to residential care	No. of admissions to residential care	*
Final outcome Mezzo level (authority) Need	User satisfaction	x	*
Supply	* (21%) No. of older people living alone * (13%) Hours of home care available	Prop. of population over 65/75; benefit claimants No. of places in homes by sector; expenditure on older people	Prop. of population over 65/75 Total cost of services/nursing homes per capita; No. of institutional care beds

**Table 2.** (Continued)

Levels of analysis/domain	England	Northern Ireland	Japan
Practice Process	* (17%) No. of assessments leading to care package	No. of care packages commenced/ongoing	Prop. of high dependency users; Costs of care management
Service Process	* (19%) Unit cost of home care; aggregated spend	Respite admissions to residential care; occupancy rates; no. of day care/home care/meals users; costs of care packages* (40%)	Costs of high dependency users; Care use against benefit limits; No. of users as prop. of insured (by dependency); Prop. of users by dependency receiving care packages (composed of no. of services)
Intermediate outcome	* (19%) No. of users of home care; admissions to residential care	No. of admissions to residential care	No. users helped to live at home as prop. of all users (by dependency)
Final outcome	* (15%) User satisfaction	* (20%) user satisfaction	User satisfaction; carer burden
Micro level (user/team)			
Need	* (20%) No. in each eligibility band	* (40%) No. in each eligibility band	Age; dependency status
Supply	* (28%) Hours of home care per case	* ('occasionally') hours of home care per case	Expenditure on service packages; No. of beds available
Practice Process	* (50%) No. of cases per team	* (20%) No. of cases per team	Expenditure on care management
Service Process	* (13%) Spend per case	* (40%) Spend per case	No. of services used; Benefit limit levels
Intermediate outcome	* (57%) Admissions to residential care	* (60%) Admissions to residential care	Services received (type, duration)
Final outcome	* (15%) User satisfaction	* (20%) user satisfaction	User satisfaction; carer burden

Sources: English Commission for Social Care Inspection (2007); Northern Ireland Statistics and Research Agency (2007); Japan Ministry of Health, Labour and Welfare (2002); Survey of performance measurement approaches in England and Northern Ireland; Data from Japanese LTCI system.

\* Limited availability (% of organizations); × Not available

government units to monitor the effects of LTCI and assists in planning and comparing activity between authorities. The LTCI scheme necessitates data being held at the individual user level (through the electronic transmission of bills from service providers to insurers), which permits aggregation of the data in a form suitable for analysing the effects of such factors as regional characteristics or client conditions upon the services older people receive. It has been argued that this has led to the 'informatization' of the Japanese care system (Matsuda and Yamamoto, 2001), resulting in data becoming generally available with which to analyse performance locally. In contrast, there are few measures routinely available at a national level, other than those collated by the Ministry of Health, Labour and Welfare, which are, in general, average figures across the country relating to aspects such as the numbers of people requiring support at each care level, the numbers using each service, expenditures and the ratio between the benefits paid and the benefit limitation amount of each care level (see Ministry of Health, Labour and Welfare, 2002). Japanese analysis relies on the concept of 'continuous improvement' (or *kaizen*) and 'benchmarking' (*dantotsu*) (Camp, 1989) against the practices of the 'best' municipalities, in terms of those who have successfully managed their LTCI arrangements. Thus, indicators such as 'benefits paid against benefit limitation' are key ones for actively managing the conduct of the scheme locally against the practices of other areas.

## Conclusions

The data collected here, on the workings of different performance measurement systems for social care in the different countries, testify to a notion of performance evaluation as rooted within institutional arrangements that largely define the scope and use of measures. Performance evaluation, in the words of March and Sutton (1997: 698) is a 'socially constructed evaluation of organizational effectiveness'. As such, although often presented as a neutral, technical approach to appraising services, it is, in fact, part of the political environment (Jowett and Rothwell, 1988). In this respect, certain broad features of this environment have been argued to at least partly explain the emergence of such divergent approaches to performance in different countries (Hood et al., 2004). One feature is the nature of the political system in operation (Lijphart, 1999). As social care operates within the ambit of local government, one obvious dimension of interest is Lijphart's (1999) 'division of power' one; that is, the relationship between central and other tiers of government. For example, the particularly centralized nature of the UK (particularly English) government has tended to produce policies which are centrally driven, whilst in Japan a decentralized system (especially in welfare policies) has necessitated that local government units have a considerable degree of autonomy and essentially control their own performance, although they do depend on national government finance (Preston, 2000). Northern Ireland remains a distinctive case in this analysis; formally part of the UK but now controlled by its own Executive, it is less centralized than England and is unique in terms of the UK in that its health and personal social services are integrated at the organizational level. Moreover, in recent times, one issue ('green-orange' politics) has dominated and the major concern of government decision-makers has been security rather than health and social care or other important issues of relevance to the lives of citizens. This 'democratic deficit' (Wilford et al., 2003) has perhaps led to a fairly stable performance evaluation system, without the, sometimes widespread, changes characterizing that in England. In addition, as a relatively smaller country, there is less distance between the 'overseer' and 'agents' in Northern Ireland, with a 'village life' community characterizing relations between them (where few targets and high level PIs operate as in England and where the Executive and local agencies are more involved in an exchange of information) (Hood et al., 1999). These differences have, to an extent, framed the environments in which these performance systems have been imbedded.



Our findings suggest that these different institutional arrangements have generated different types of measures at different levels of aggregation. Their different modes of analysis also indicate divergent links to other forms of evaluative activity, conceived more broadly. By use of our analytical framework, conclusions may be drawn as to the breadth of evidence available in the three systems for successfully monitoring the range of service provision in social care.

For example, in certain of our cases, notably England, performance evaluation has been allied to competitive oversight by central government. This is a particularly centralized form of performance measurement in comparison to some European countries (Bloomfield, 2006). Such a system does not allow a broader assessment of organizations' processes and outputs as it is aimed primarily at regulating the behaviour of local agencies. It thereby necessitates the presentation of quite simple descriptive comparisons to permit easy interpretation by those subject to performance evaluation. Thus, links to other evaluative activities, such as more formal statistical testing and multivariate analysis, are limited and the system can be criticized on this basis (see Clarkson and Challis, 2006; Goldstein and Spiegelhalter, 1996). Encouraging local government to measure its own performance may suffer in such a regime since those producing the data are subject to 'measure pressure' to instead orientate performance measurement towards public reporting to higher bodies (Van Thiel and Leeuw, 2002).

In other cases, such as Japan, the system shows many commonalities with other forms of evaluation, such as quasi-experiments and evidence-based policy (Leeuw and Furubo, 2008). In that country, performance data are primarily employed to generate knowledge about the operation of the system, in other words to build up management 'intelligence' (Hood, 2007). Information from the performance system is largely viewed by those generating it as serving the 'Plan-Do-Check-Act' process of management quality control made famous by the original pioneers of this approach (Deming, 1982; Shewhart, 1931). There is greater use of locally based data in the Japanese system, which can track the care received by individual users, so permitting more sophisticated formal analysis, traditionally the preserve of research investigations in this setting (see e.g. Challis et al., 2004; Clarkson et al., 2006). This 'bottom-up' approach has arisen precisely because the responsibility for variations between areas in implementation is the legitimate purview of the municipalities as insurers and national monitoring is, by contrast, relatively undeveloped. In particular, the locally administered information relating to a common system of insurance makes possible the generation of user-level data in a similar format, facilitating the development of a shared perspective of effective performance between similar municipalities, thereby permitting local benchmarking (Camp, 1989). The English system, by contrast, is limited to broad indicators aggregated at the local authority level. Such a shared approach as exists in Japan is hampered in England by the lack of appropriate common information within local authorities, other than those items of information prescribed from the centre. Thus we find, in England, that many authorities had not designed their own local indicators and these did not cover the full range of domains expected from comparison against the production model (the PIAF) used here.

In Northern Ireland also there is a limited range of performance data available nationally. Those indicators relating to outputs (such as admissions to residential care) and service processes predominate. Currently, there are few bespoke local indicators designed to capture management practice with effort directed towards reporting data for the returns requested by the DHSSPS. The performance monitoring that takes place at a national level involves more sharing of information and is also less competitive than the system in England. However, broad changes are taking place and our survey gives a clear impression that tighter performance monitoring (in terms of national standards and targets) is to receive much more attention than before and that indicators and their measurement will be more sophisticated, or at least more comprehensive:

Very early days ... [the] current performance system is fledgling but attempts to provide a view of performance from multiple perspectives ... currently we are not flowing performance information on a regular basis to operational managers. [However,] the system is well defined and useful with meaningful information ... the focus over the next four months will be the development of Directorate specific performance systems ... this will lead to the development of key performance indicators responsive to local need. (Northern Ireland managers)

The main conclusion to be drawn is therefore that the nature of the performance measurement system can either constrain or enhance relationships with other types of evaluative activity. Where the system is embedded within particular accountability relationships, essentially to motivate and direct organizations' actions (such as those in England), this may hamper efforts at using information to serve other ends (Benjamin, 2008). In essence, such systems create tensions for evaluators, and those attempting to manage by use of performance data, in that their ability to employ data reflecting local concerns is superseded by the need to respond to the requirements of those to whom they are answerable (Day and Klein, 1987). While the behavioural consequences of this are not well documented in the social care setting, other research in the English health service has brought to light the ways in which practice is constrained to fit with the performance measures rather than examining other objectives, which may be relevant locally or from the perspective of those using the service (Mannion et al., 2005). By contrast, where accountability is focused more on encouraging mutual understanding of a complex system (as in Northern Ireland), there is less need for explicit comparisons and ratings between organizations. Instead, indices are presented that permit a broad overview of (primarily) outputs in order that the direction of services against objectives can be discerned. Alternatively, where the system is part of an evaluation of a relatively standardized programme, which is administered locally – as in Japan – then analysis necessary to understand and monitor the programme is augmented by other forms of evaluation. Explicating the differences between these systems and their different links to other forms of evaluative activity is important in gauging what happens when systems change. This is likely in the three cases studied here. Investigating the range of performance information prevalent in divergent systems in other sectors also holds much promise in unravelling how performance measurement links with evaluation conceived more generally.

### Acknowledgements

The study to which this work relates was funded by the UK Economic and Social Research Council's Public Services Programme (grant: RES-166-25-0062). Data from the Long-term Care Insurance System in Japan were collected with the assistance of a grant from the Nihon Fukushi University, 21st Century Centre of Excellence (COE) Programme.

### Notes

1. The national survey was of the 149 local authorities with social services responsibilities in England (Isles of Scilly excepted) for which 117 authorities replied with data from at least one of the groups of respondents (79% response rate); the survey in Northern Ireland was of all five Health and Social Services Trusts which achieved a 100% response rate.
2. Some research evaluations have compared England and Northern Ireland in terms of their care provision. These evaluations, however, have been either small-scale exploratory studies of the views of stakeholders (Heenan and Burrell, 2006) or studies investigating patterns of team working (Reilly et al., 2003; Challis et al., 2006b). Evaluations at the organizational level using performance data to gauge the effects of system differences are limited, the review by Appleby (2005) being one recent exception. One broad conclusion

- has been that integrated structures, although conducive to inter-professional working, do not necessarily guarantee it.
3. For example, in 2004 expenditure on long-term care insurance beneficiaries amounted to approximately US\$42 billion (Wiener et al., 2007). Expenditure has increased steadily and by October 2006 there were over 4 million beneficiaries with these increasing at a faster rate than the increase in the older population. However, the growth in the number of beneficiaries is more pronounced at lower levels of need, suggesting that more older people with only limited dependency have been prompted to apply for the assessment as knowledge of the system has become more widely disseminated.
  4. Strictly speaking, the local authority organizations being regulated are not 'subordinate' to central government but are locally elected bodies accountable to local citizens. However, the functions they perform are subject to guidance and eventual sanctions from the centre. The degree of oversight of English local authorities by central government (and its agencies) has increased dramatically since the 19th century and the tension between central government influence and local authority discretion forms the backcloth for considering changes in the types of performance measurement system employed. On this point, in general, see Hood et al. (1999) and for social care see Challis et al. (2006a).
  5. Details of the measures provided here are those used in the formal analysis and monitoring of performance in each of the three countries. Other data and indicators are available but are not employed in regulatory contexts or in terms of comparing the activities of authorities or the programme of LTCI in Japan. In England, for example, a wider set of 'key' indicators is available for authorities, users of services and researchers (available at: [www.drfooster.co.uk/localGovernment/kigs.asp](http://www.drfooster.co.uk/localGovernment/kigs.asp)). In Northern Ireland, other data are available presenting the extent of service provision in the Community Statistics publications (e.g. Northern Ireland Statistics and Research Agency, 2006). In Japan, utilization and expenditure data for LTCI are available through the Ministry of Health, Labour and Welfare (see: [www.wam.go.jp/wamappl/bb05Kaig.nsf/aCategoryList?OpenAgent&CT=20&MT=060&ST=01](http://www.wam.go.jp/wamappl/bb05Kaig.nsf/aCategoryList?OpenAgent&CT=20&MT=060&ST=01)), although these are not available in English.

## References

- Adab, P., A. M. Rouse, M. A. Mohammed and T. Marshall (2002) 'Performance League Tables: The NHS Deserves Better', *British Medical Journal* 324: 95–8.
- Applebaum, R. and M. White (eds) (2000) *Case Management around the Globe*. San Francisco, CA: American Society on Aging.
- Appleby, J. (2005) *Independent Review of Health and Social Care Services in Northern Ireland*. London: Kings Fund.
- Arvidsson, G. (1986) 'Performance Evaluation', in F. X. Kaufman, G. Majone and V. Ostrom (eds) *Guidance Control and Evaluation in the Public Sector*, pp. 625–43. Berlin: Walter de Gruyter.
- Audit Commission (1996) *Local Authority Performance Indicators 1994/95*, vol. 1, *Education, Social Services, Libraries and Expenditure*. London: Audit Commission.
- Barnes, M. and N. Miller (1988) 'Performance Measurement in Personal Social Services', *Research, Policy and Planning* 6: 1–47.
- Behn, R. D. (2003) 'Why Measure Performance? Different Purposes Require Different Measures', *Public Administration Review* 63(5): 586–606.
- Benjamin, L. M. (2008) 'Evaluator's Role in Accountability Relationships: Measurement Technician, Capacity Builder, or Risk Manager?', *Evaluation* 14(3): 323–43.
- Bloomfield, J. (2006) *Controlling, Cajoling or Co-operating? Central Governments' Policy Approaches towards Local Government on the Issues of Performance and Cost-Effectiveness*. Birmingham: Institute of Local Government Studies, University of Birmingham, and Council of European Municipalities and Regions.
- Brignall, S. and S. Modell (2000) 'An Institutional Perspective on Performance Measurement and Management in the "New Public Sector"', *Management Accounting Research* 11: 281–306.

- Burnham, D. (2006) *Only Get Better? A Guide to Social Services Performance Measurement Processes for Frontline Staff*. Lyme Regis: Russell House Publishing.
- Camp, R. C. (1989) *Benchmarking: The Search for Industry Best Practices that Lead to Superior Performance*. Milwaukee, WI: Quality Press.
- Campbell, J. C. (1992) *How Policies Change: The Japanese Government and the Aging Society*. Princeton, NJ: Princeton University Press.
- Campbell, J. C. and N. Ikegami (2003) 'Japan's Radical Reform of Long-Term Care', *Social Policy and Administration* 37: 21–34.
- Challis, D., P. Clarkson, J. Williamson, J. Hughes, D. Venables, A. Burns and A. Weinberg (2004) 'The Value of Specialist Clinical Assessment of Older People Prior to Entry to Care Homes', *Age and Ageing* 33: 25–34.
- Challis, D., P. Clarkson and R. Warburton (2006a) *Performance Indicators in Social Care for Older People*. Aldershot: Ashgate.
- Challis, D., K. Stewart, M. Donnelly, K. Weiner and J. Hughes (2006b) 'Care Management for Older People: Does Integration Make a Difference?', *Journal of Interprofessional Care* 20(4): 335–48.
- Clarkson, P. and D. Challis (2003) 'Quality Assurance Practices in Care Management: A Perspective from the United Kingdom', *Care Management Journals* 4(3): 142–51.
- Clarkson, P. and D. Challis (2006) 'Performance Measurement in Social Care: A Comparison of Efficiency Measurement Methods', *Social Policy and Society* 5(4): 461–77.
- Clarkson, P., D. Venables, J. Hughes, A. Burns and D. Challis (2006) 'Integrated Specialist Assessment of Older People and Predictors of Care Home Admission', *Psychological Medicine* 36: 1011–21.
- Clarkson, P., S. Crompton, D. Challis, M. Donnelly and R. Beech (2007) *The Design and Use of Local Metrics to Evaluate Performance: A Comparative Analysis of Social Care Organisations*. Manchester: Personal Social Services Research Unit. URL (consulted July 2008): <http://www.pssru.ac.uk/pdf/MCpdfs/DPM160.pdf>.
- Commission for Social Care Inspection (2007) *Social Services Performance Assessment Framework Indicators Adults 2006–07*. London: Commission for Social Care Inspection.
- Community Care Policy Network (2001) *Software for Analysis of Long-Term Care Insurance Performance*. Tokyo: Community Care Policy Network.
- Davies, I. (1999) 'Evaluation and Performance Management in Government', *Evaluation* 5(2): 150–9.
- Day, P. and R. Klein (1987) *Accountabilities: Five Public Services*. London: Tavistock.
- Deming, W. E. (1982) *Out of the Crisis*. Cambridge, MA: Massachusetts Institute of Technology.
- Department for Communities and Local Government (DCLG) (2007) *The New Performance Framework for Local Authorities and Local Authority Partnerships: Single Set of National Indicators*. London: DCLG.
- Department of Health (1998) *Modernising Social Services. Promoting Independence, Improving Protection, Raising Standards*. Cm 4169. London: Stationery Office.
- Department of Health (2003a) *Social Services Performance Assessment Framework Indicators 2002–2003*. London: Dept of Health.
- Department of Health (2003b) *Social Services Performance Star Ratings*. London: Dept of Health.
- Department of Health, Social Services and Public Safety (DHSSPS) (2008) *Proposals for Health and Social Care Reform*. Belfast: DHSSPS. URL (consulted Oct. 2008): [http://www.dhsspsni.gov.uk/consultation\\_package.pdf](http://www.dhsspsni.gov.uk/consultation_package.pdf).
- Department of the Environment (1981) *Local Government Annual Reports: A Code of Practice*. London: HMSO.
- Goldstein, H. and D. J. Spiegelhalter (1996) 'League Tables and their Limitations: Statistical Issues in Comparisons of Institutional Performance', *Journal of the Royal Statistical Society, Series A* 159: 385–443.
- Heenan, D. and D. Burrell (2006) 'The Integration of Health and Social Care: The Lessons from Northern Ireland', *Social Policy and Administration* 40(1): 47–66.
- Henwood, M. and G. Wistow (1999) *With Respect to Old Age: Long Term Care – Rights and Responsibilities. Community Care and Informal Care Research*, vol. 3. London: Stationery Office.

- Hood, C. (2007) 'Public Service Management by Numbers: Why does it Vary? Where has it Come from? What are the Gaps and the Puzzles?', *Public Money and Management* 27(2): 95–102.
- Hood, C., C. Scott, O. James, G. Jones and T. Travers (1999) *Regulation Inside Government: Waste-Watchers, Quality Police and Sleaze-Busters*. Oxford: Oxford University Press.
- Hood, C., O. James, B. G. Peters and C. Scott (eds) (2004) *Controlling Modern Government*. Cheltenham: Edward Elgar.
- House of Commons Health Committee (2005) *NHS Continuing Care: Sixth Report of Session 2004–05*, vol. 1. London: Stationery Office.
- Huber, N. (1999) "'At Risk" Local Authorities Could Lose their Social Services, Warns Hutton', *Community Care* (25 Nov.–1 Dec.): 2–3.
- Ikegami, N. (1997) 'Public Long-Term Care Insurance in Japan', *Journal of the American Medical Association* 278: 1310–14.
- Ikegami, N. (2007) 'Rationale, Design and Sustainability of LTCI in Japan – In Retrospect', *Social Policy and Society* 6: 423–34.
- Institute of Health and Economic Programmes (IHEP) (2002) *Study on Evaluation Methods of the Effectiveness of Long Term Care Insurance*. Tokyo: IHEP.
- Jowett, P. and M. Rothwell (1988) *Performance Indicators in the Public Sector*. Basingstoke: Macmillan.
- Kondo, K. (2003) *The Outline of Aichi Gerontological Evaluation Study (AGES) Project: Wave 2000 and 2003*. Nagoya: Nihon Fukushi University, Research Promotion Centre for Community Care.
- Leeuw, F. L. and J. E. Furubo (2008) 'Evaluation Systems: What are They and Why Study Them?', *Evaluation* 14(2): 157–69.
- Lijphart, A. (1999) *Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries*. New Haven and London: Yale University Press.
- Llewellyn, S. (1996) 'Performance Management in the Social Services: Its Meaning and Measurement', in I. Lapsley and F. Mitchell (eds) *Accounting and Performance Measurement: Issues in the Private and Public Sectors*, pp. 166–79. London: Paul Chapman.
- Mannion, R., H. Davies and M. Marshall (2005) 'Impact of Star Performance Ratings in English Acute Hospital Trusts', *Journal of Health Services Research and Policy* 10: 18–24.
- March, J. G. and R. I. Sutton (1997) 'Organizational Performance as a Dependent Variable', *Organization Science* 8(6): 698–706.
- Matsuda, S. and M. Yamamoto (2001) 'Long-Term Care Insurance and Integrated Care for the Aged in Japan', *International Journal of Integrated Care* 1: 1–11.
- Milburn, A. (2001) 'Speech to the Annual Social Services Conference', Harrogate, England, 19 Oct. URL (consulted July 2007): [http://www.dh.gov.uk/en/News/Speeches/Speecheslist/DH\\_4000442](http://www.dh.gov.uk/en/News/Speeches/Speecheslist/DH_4000442)
- Miller, N. (1986) 'Management Information and Performance Measurement in Personal Social Services', *Social Services Research* 4(5): 7–55.
- Ministry of Health, Labour and Welfare (2002) *Long-Term Care Insurance in Japan*. Tokyo: Ministry of Health, Labour and Welfare.
- Ministry of Public Management, Home Affairs, Posts and Telecommunications (2001) *Government Policy Evaluation Act*. Tokyo: Ministry of Public Management, Home Affairs, Posts and Telecommunications.
- Morris, J. N., B. E. Fries, K. Steel, N. Ikegami, R. Bernabei, G. I. Carpenter, R. Gilgen, J. P. Hirdes and E. Topinkova E. (1997) 'Comprehensive Clinical Assessment in Community Setting: Applicability of the MDS-HC', *Journal of American Geriatrics Society* 45(8): 1017–24.
- Nielsen, S. B. and N. Ejler (2008) 'Improving Performance? Exploring the Complementarities between Evaluation and Performance Management', *Evaluation* 14(2): 171–92.
- Northern Ireland Statistics and Research Agency (2003) *Key Indicators of Personal Social Services for Northern Ireland 2003*. Belfast: Dept of Health Social Services and Public Safety.

- Northern Ireland Statistics and Research Agency (2006) *Community Statistics 1 April 2005–31 March 2006*. Belfast: Dept of Health Social Services and Public Safety.
- Northern Ireland Statistics and Research Agency (2007) *Key Indicators of Personal Social Services for Northern Ireland 2006*. Belfast: Dept of Health Social Services and Public Safety.
- Ouchi, W. (1981) *Theory Z: How American Management can Meet the Japanese Challenge*. Reading, MA: Addison-Wesley.
- Preston, R.W. (2000) *Understanding Modern Japan: A Political Economy of Development, Culture and Global Power*. London: SAGE.
- Reilly, S., D. Challis, A. Burns and J. Hughes (2003) 'Does Integration Really Make a Difference? A Comparison of Old Age Psychiatry Services in England and Northern Ireland', *International Journal of Geriatric Psychiatry* 18: 887–93.
- Schacter, M. (2002) *Not a Tool Kit: Practitioner's Guide to Measuring the Performance of Public Programs*. Ottawa: Institute on Governance.
- Shewhart, W. A. (1931) *Economic Control of Quality of Manufactured Product*. New York: D Van Norstrand.
- Tanaka, H. (2006) *Current State of Performance Measurement Systems in Japan: Are they Properly Functioning?* Hamamatsu City: Shizuoka University.
- Tanaka, H. and T. Ono (2002) 'Performance Measurement Utilization in Japan: A Critical View', paper presented at the Annual National Conference of the American Society for Public Administration, Phoenix, AZ, 26 March.
- Timonen, V. (2005) 'Policy Paradigms and Long-Term Care: Convergence or Continuing Difference?', in P. Taylor-Gooby (ed.) *Ideas and Welfare State Reform in Western Europe*, pp. 30–53. Basingstoke: Palgrave Macmillan.
- Tsutsui, T. and N. Muramatsu (2007) 'Japan's Universal Long-Term Care System Reform of 2005: Containing Costs and Realizing a Vision', *Journal of the American Geriatrics Society* 55: 1458–63.
- Van Thiel, S. and F. Leeuw (2002) 'The Performance Paradox in the Public Sector', *Public Performance and Management Review* 25(3): 267–81.
- Wiener, J. M., J. Tilly, A. Howe, C. Doyle, A. E. Cuellar, J. Campbell and N. Ikegami (2007) *Quality Assurance for Long-Term Care: The Experiences of England, Australia, Germany and Japan*. Washington, DC: AARP.
- Wilford, R., R. Ginty, M. Roger, L. Dowds and G. Robinson (2003) 'Northern Ireland's Devolved Institutions: A Triumph of Hope over Experience?', *Regional and Federal Studies* 13(1): 31–54.

Paul Clarkson is Research Fellow at the Personal Social Services Research Unit (PSSRU), University of Manchester, UK. He is involved in work on social care performance measurement and health/social care assessment models. [email: paul.clarkson@manchester.ac.uk]

David Challis is Professor of Community Care Research and Director of PSSRU, University of Manchester and Associate Director of the National Institute for Health Research (NIHR) School for Social Care Research. He is responsible for a range of research experiments and policy evaluations in health and social care in the UK and other countries. [email: D.J.Challis@manchester.ac.uk]

Sue Davies is Research Associate at PSSRU, University of Manchester and has working experience of local authority performance measurement in England. [email: susan.crompton@manchester.ac.uk]

Michael Donnelly is Reader, Centre for Public Health and UKCRC Centre of Excellence for Public Health, Queen's University Belfast, Northern Ireland. His research includes studies of rehabilitation and recovery and the evaluation of integrated health and social care. [email: michael.donnelly@qub.ac.uk]

Roger Beech is Reader, Research Institute for Life Course Studies, Keele University, UK. His research supports the development and evaluation of services for older people with a particular focus on services for stroke care, for providing care 'closer to home' and for promoting health and wellbeing. [email: r.beech@keele.ac.uk]

Takayuki Hirano is Professor, Research Promotion Centre for Community Care, Nihon Fukushi University, Nagoya, Japan. He has been involved in evaluating the performance measurement system for long-term care insurance in Japan. [email: hirano@n-fukushi.ac.jp]