PERFORMANCE INDICATORS: A LOGIC OF ESCALATION?

Christopher Pollitt, Stephen Harrison, George Dowswell, Roland Bal and Sonja Jerak-Zuiderent

A paper for presentation at the European Group for Public Administration Conference, Erasmus University Rotterdam, September 2008

Institutional details (correspondence to)

Christopher Pollitt
BOF/ZAP Research Professor of Public Management
Public Management Institute
Katholieke Universiteit Leuven

Christopher.Pollitt@soc.kuleuven.be

For CV, see www.publicmanagement.be (go to ‘staff’)

(egpa08cadspaper/27-06-2008)
PERFORMANCE INDICATORS: A LOGIC OF ESCALATION?

ABSTRACT

The Netherlands and England are near neighbours that share many fundamental health care challenges (ageing societies, rapidly changing medical technologies, rising demand, constrained health and welfare budgets, etc). The relevant policy communities have also usually been well aware of what is going on in the other country. Nevertheless, for two decades the development (or non-development) of their performance indicator systems was strikingly different. England went overboard for national healthcare PIs, and the Netherlands virtually shunned them.

A now–standard explanation for difference is that England possesses what is, by Western European standards, an unusually centralized, majoritarian and adversarial political system, which frees it from many of the constraints and compromises which most continental, consociational governments must engage with. Our examination of the evidence of performance indicator development since the early 1980s in part confirms this image, but also indicates that, in parallel, there are other structures and processes at work. These additional processes have produced both similarities and further differences between the two systems – not least the ‘late’ appearance of a Dutch national system of healthcare PIs from 2003. Our analysis thus supports the ‘new institutionalism’ in so far as it shows that broad institutional patterns do indeed have a significant influence on policy development. But it also suggests that there is a need for at least two complementary strands of analysis and explanation. First, while the institutional patterns may help us understand why things remain on the same, or different, paths, they do not work so well as explanations for the periodic departures from those paths. These ‘punctuations’ to ‘business as usual’ require more specific attention. Also, second, we argue that there are technical and political factors which are endogenous to PI systems, and which, once these systems are in place, appear to exert a logic of their own.

We therefore conclude that, in order to explain the existence and characteristics of these (and other?) PI systems, a broad institutional analysis needs to be complemented both by particular attention to major changes in course and by a more detailed consideration of the internal logics (and external political effects) of particular policy technologies. To some extent PI systems have their own ways of life, as can be seen in both England and the Netherlands.
PERFORMANCE INDICATORS: A LOGIC OF ESCALATION?

Research design and methods

The main aim of our research project was to give a comparative account of the trajectories and patterns of development and redevelopment and usage of nationally-established PIs for health care in the English NHS and the Dutch health care systems.

The study comprised two longitudinal national-level case studies, one for each country, covering a period beginning in the early 1980s and going through to 2007. Our comparison was intended to take analytical advantage of a number of Anglo-Dutch similarities and differences. Both countries have similar levels of public expenditure on healthcare, similarly highly-developed hospitals, technologies, and health professions, and similar health care demand pressures. The policy-academic communities of the two countries have had a good deal of interaction in relation to health policy matters such as ‘rationing’ criteria, ‘evidence-based’ medicine, quasi-markets, and health services research more generally. There are, however, important contrasts in respect of population (the Dutch population, while less than one third of England’s, is more densely-packed) and health system organisation. On the latter point, England has a tax-funded single payer for over 80% of health care expenditure and (for practical purposes) government ownership of hospitals. The Netherlands has multiple payers and private providers in a mainly social insurance arrangement, with less than 5% of direct government funding. At a more general level, although both countries have ‘modernising’ governments with relatively high degrees of public transparency, they exhibit contrasting types of democratic systems, respectively ‘Westminster’ (majoritarian) and ‘consensus’ (with coalition governments and numerous corporatist arrangements) in Lijphart’s (1999) classification. This is potentially important because attempts to link such characteristics to trajectories of public management reform have suggested that the UK tends towards rapid, centrally-directed reforms whilst The Netherlands exhibits slower and more diverse patterns of change (Pollitt and Bouckaert 2004). This fitted the histories of healthcare PI development as we knew them before the project: England had first developed them in the early 1980s, some twenty years before The Netherlands. (In order to avoid the complex and subtle differences between the National Health Services of the four countries within the UK, we confined the study to England.)

Our main focus was on PIs in the acute hospital sector (where most political priority had been placed). The primary aim was to trace the ‘story’ of the invention, development and usage of PIs over time. For the Netherlands, we also sought data on the health care performance and management policies pursued in the period when PIs were not on the policy agenda. We began by constructing, from secondary sources, an outline chronological account for each case before proceeding to primary data collection. This enabled us to focus on potential ‘critical junctures’ and to begin to identify appropriate informants, with whose assistance we could ‘snowball’ further respondents. We interviewed some 49 respondents for the English case study and 24 for the Dutch,
including policymakers; official health care regulators; academics and policy entrepreneurs who have contributed to the critique and/or development of PIs; specialist media health correspondents who have commented on PIs; and former senior health/hospital managers whose organisations had been the subject of PIs. Interviews were audio-recorded (with participants’ consent). We used the facility in the Atlas.ti software to analyse audio files, then producing written summaries. We also examined extensive documentary data sources, in the form of both official documents and news media reports. We used these data to construct integrated accounts for the two countries, in order to serve as the basis for theoretical analysis and to address our research questions.

One preliminary caveat would be that we were not, in our research, seeking to arrive at a judgement about the overall value of healthcare PIs, still less to make recommendations about how these should be developed in future. We describe the course of events and analyse why certain things happened (and others did not). If this account throws up instances of PIs not working well or having unintended consequences (as it inevitably does) that should not be taken as meaning that we are in some general sense ‘against’ performance measurement. [As a matter of fact, we are not]

The big picture: different patterns of institutions

The in-a-nutshell story of English and Dutch healthcare PIs is one of lengthy divergence followed by some more recent convergence. In 1983 the British government imposed a first national set of PIs on the National Health Service (NHS). This system was subsequently progressively developed (we have documented numerous shifts in the number and type of measure) and from the early 1990s it became a more and more prominent part of NHS management, along with a proliferation of regulatory institutions. It continues today. In the Netherlands, however, there was no such adoption of a national PI system, despite some similar problems (see below). Only two decades later, in 2003, was such a system put in place and, even then, it was much less complex and prescriptive than the system currently in force in England.

Although health system PI development in The Netherlands took place twenty years behind that in England, no such lag characterised other aspects of health care organisation and regulation, whose overall trajectory bears a strong similarity to developments in England. For example, expenditure concerns arose in the early 1980s, and cost reduction measures were taken from 1983 onwards. Regulated competition was proposed in 1987, introduced in a limited way between 1988 and 1994, attenuated between 1994 and 2000, and revived in 2001 (Helderman et al 2005). In the 1990s, interest in health care ‘quality’ began to be articulated and legislation in 1996 required health care providers to have quality management systems in place, a philosophy that resembles that subsequently developed in England as ‘clinical governance’. Thus PIs were an exception in the rough parallel between the two countries.

There is a rather obvious general explanation for the different English and Dutch stories. This was developed in an earlier paper (Pollitt, 2007). The core of the explanation was, first, that national differences in cultural and political systems effectively prevented the
Dutch government from imposing any centralized system of government-designed indicators, and, second, that the organization and finance of the Dutch healthcare system also made a centrally-run performance management system improbable.

The English political system is very different from the Dutch. England has a majoritarian electoral system producing one-party executives. These executives can usually discipline the legislature to their will, and thus a very high percentage of government-initiated legislation gets through. Furthermore, sub-national governments and government agencies have very weak constitutional protections – central government is normally able to impose itself on local authorities, and certainly on the National Health Service. The Netherlands, by contrast, has a proportional electoral system which usually leads to multi-party cabinets. Getting parliamentary assent for government measures is by no means as straightforward as in England. The whole Dutch system is strongly imbued with what comparativist political scientists have long termed a ‘consensualist’ culture. Lijphart’s summary tables give something of the contrast (see Table 1).

<table>
<thead>
<tr>
<th>UK</th>
<th>NETHERLANDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average effective number of political parties represented in the lower houses of parliaments following elections, 1945-96</td>
<td>2.11</td>
</tr>
<tr>
<td>Proportion of time when cabinet was single party, 1945-96</td>
<td>100%</td>
</tr>
<tr>
<td>Index of executive dominance</td>
<td>5.52</td>
</tr>
</tbody>
</table>

TABLE 1: SOME MEASURES OF POLITICAL SYSTEM DIFFERENCES (source: Lijphart, 1999, pp76-77; 110-11; 132-133)

Footnote

1 This is not the same as the number of self-defined parties. For details of measurement see Lijphart, 1999, pp74-77

2 Lijphart’s Index of Executive Dominance is a complex measurement, and its full derivation is too complex to spell out here. Essentially it is related to two different measures of the average durability of cabinets in each country, where longer duration is taken as a sign of a strong executive. For details see Lijphart, 1999, pp129-139

Thus:

‘Politically and culturally, the Dutch need to discuss new ideas at length, and seek agreement between all the major stakeholders (in our case the hospital associations, the medical profession and the government). Sharply judgemental use of indicators (for example, to pick out certain hospitals as poor performers, as happens in the UK) would be unlikely to secure an appropriate level of acceptance, in the absence of which a multi-party governing coalition would find it risky to push ahead.’ (Pollitt, 2007, p158)

Or, as one of our Dutch respondents put it:
“Naming and shaming, which has happened in England, is not happening so much here. They don’t publish everything. Also, it’s part of our Polders model, discussing a bit, talking with each other, and eventually you get the right direction” (Interview 214)

Furthermore, the specific institutional pattern of the hospital sector is rather different. The English NHS constitutes a single, centralized system, mainly financed from general (non-hypothecated) taxation, whose hospitals the government effectively owns. The Netherlands is a pluralistic system predominantly financed from hypothecated insurance payments mediated through many different companies and systems. The Dutch hospitals themselves are a mixture of diversely-owned non-profit and publicly-owned with, like the UK, a small private for-profit sector. Thus it can be argued that the greater diversity of finance in the Netherlands, including the more substantial component of private insurance, is reflected in a greater need for consensus about reforms (Pollitt, 2007, p159).

One of our Dutch interviewees memorably described the process of trying to get the various interested ‘players’ to agree on a set of indicators as being like trying to keep a bunch of frogs in a wheelbarrow (Interview 203).

This general explanation is therefore heavily structural and institutional. The Dutch Ministry of Health never felt itself to be in the position of hierarchical superiority that the Whitehall Department of Health enjoyed. It was not in the business of telling either privately-owned hospitals or the representative associations of the medical profession what to do. On the contrary, its response to fiscal problems during the 1980s was typically Dutch – open negotiations with all concerned to see if some general reform could be consensually agreed upon. Furthermore, the creation, in 1995, of a Dutch healthcare inspectorate further complicated the picture. Initially, this Inspectorate appears to have worked closely with hospital boards and to have kept some distance between itself and the Ministry. Only later, when it came under public criticism and found it necessary to draw closer to the Ministry, were Dutch PIs born. In short, the pattern of institutions in the UK permitted a top-down, centralized approach from government, the Dutch pattern did not.

When the big picture changed: 1983 and 2003

While the institutional features described may explain how the two countries remained on different courses between 1983 and 2003, they do not explain either the original English adoption of PIs in 1983 or the belated Dutch move in 2003. To understand these ‘punctuations’ we looked more closely at the particular events surrounding each case.

England, 1983 Although there had been some official concerns with hospital efficiency (measured as patient ‘throughput’ in relation to beds) in the early 1950s (Cutler 2007), and we found some evidence of regionally-based indicator development in Oxford and the West Midlands in the late 1970s, the first quantitative NHS performance regime that can be seen as both relatively comprehensive and relatively durable was ‘invented’ in 1982, coming into operation in the NHS in 1983 in the form of a package of 70 PIs, mainly derived opportunistically from existing administrative datasets, and very much
weighted towards measuring inputs and efficiency (Pollitt, 1985). The Conservative government elected in 1979 had inflation control as a major economic objective, with public expenditure reduction as one means to this end, but was also under a number of political pressures in relation to the NHS, including sustained lobbying from the Confederation of British Industry concerning the substantial growth of the NHS workforce at a time when its workload appeared to be static.

More directly, highly critical reports from the Social Services (1980) and Public Accounts (1981) Select Committees questioned the assumption of a straightforward relationship between input levels or patterns and service provision, and criticised the (then) Department of Health and Social Security for failing to exercise financial control or demand accountability from Health Authorities. The latter report also called for ‘key indicators’ to monitor the performance of Authorities, and was told by Departmental witnesses that such indicators were being planned. In the event, it was the suggestion of a relatively junior civil servant that led to the borrowing of academic work undertaken at the University of Birmingham and the association of PIs with another accountability device, the ministerial ‘regional review process’ introduced in 1982-83. The events summarised above describe the beginning of a new policy pathway.

The Netherlands, 2003 Although some centralisation and integration had taken place in relation health regulation, with the present Dutch Healthcare Inspectorate (IGZ) established in 1995 by amalgamation of previous regulators, its approach was still very much rooted in consensus and mutual trust between inspectors and the Boards of the institutions that they assessed. Inspections were generally brief and based primarily on discussions with the Board rather than observation or discussions with professional staff. Moreover, hospitals often failed to implement the above-mentioned requirements to introduce quality management systems. However, in 1999 the Dutch National Audit Office (Algemene Rekenkamer) produced a highly critical report on IGZ, concluding that its work did not provide a sound basis on which the Minister of Health could be assured of the quality of the country’s health care. In particular, IGZ was insufficiently focused on actual quality of care, unquestioning about vaguely-expressed professional standards, did not undertake risk analysis as a means of prioritising its work, and was not rigorous in following up whether its recommendations had been implemented. Although the Minister of Health publicly defended IGZ, from this point in time the organisation repositioned itself so as to be closer to government and less close to hospitals, paid greater attention to following up its recommendations, and sought to develop an approach to risk analysis. It was in this latter context that IGZ began in about 2002 to develop PIs, which were first published in 2003 (and annually since). The original PI set contained some 30 items, including input, process and outcome indicators, and the intention was that these should be relatively stable, with few annual changes in the indicators themselves. They were presented as a tool for risk analysis, and were to be publicly available through hospitals’ own websites, which would also be allowed to state any relevant qualifications, such as atypical casemix, to the figures.

At more or less the same time, the Dutch government once more began to plan for increased marketization of healthcare (hospitals being pressed to compete for business
from the health insurance organizations). During the internal debate over this policy many of the policy elite recognized that if quality was to be maintained or improved within a more cost-conscious market, far greater quality transparency would be essential. PIs were very much part of this agenda.

As in England, the creation of PIs at this point in time represented a break from the previous policy path. However, this does not mean that the Netherlands had suddenly taken on the characteristics of a centralized, majoritarian state. There were some important differences with England in 1983 (Pollitt, 2007, pp151-154). Most obviously, the intervening 20 years had seen PIs become common in many countries’ healthcare systems – Dutch supporters of PIs could argue that they were not dangerously experimenting with something new but rather seeking to establish a system based on learning from extended experience elsewhere, and adapted to the particular Dutch context. Furthermore, the Dutch PIs were said to be focused on patient safety and clinical effectiveness, not on the efficiency goals which had been to the fore in England in 1983. And the Dutch healthcare inspectorate had made agreement with the medical specialists’ association and with the Dutch hospitals association prerequisites for the introduction of the system. Finally, as mentioned above, hospitals retained a measure of control over how the data was presented and explained on their websites: “The hospitals themselves deliver the indicators – they have to put them on the internet. Every now and then the Inspectorate takes a sample to see if the indicators are correct or not” (Interview 217, Dutch healthcare politician)

Overall, then, it was a ‘softer’ (but also a more public) launch than the NHS had experienced in 1983.

Performance indicators on the move: a logic of escalation?

Thus far we have reviewed two types of explanation for two different sets of phenomena. First we examined the institutional underpinnings of the long period of difference between the two countries, a period in which England embraced PIs and the Dutch shunned them. Then, second, we noted that more particular explanations were needed of the ‘punctuations’ – the moments at which there was a change of course and a new system was brought in. We found these explanations in terms of the particular concatenation of political and organizational forces when political elites faced severe criticism and ‘needed to do something’ – in the English case in 1981-3 and in the Dutch case in 1999-2003.

However, the story does not end here. In our interviews and documentation we found considerable evidence to indicate that, once a quantitative PI system is in place, there is a certain dynamic or logic to the way it is likely to develop. Here, despite the short period during which the Dutch system has been in operation, we believe we can see the beginnings of a similar dynamic to that exhibited during the longer history of PIs in England. Boldly characterized, this dynamic runs as follows:

1. An initial tendency for a few simple measures to become a more comprehensive package. “I think that there should be about 1000 areas where every organization should be trying to achieve something. But I do think it is correct that out of that
1000 or so areas you should pick a few which are the big ones” (Interview 120: senior health service manager and auditor). “To assess risk, I’ve given up the idea that there are 10 indicators that do it all…So I end up with two and a half thousand indicators – I look for triangulation and patterns in the results” (Interview 135: healthcare regulator). “The service woke up to the notion of managing performance and therefore having some measurables…Having a limited number of measurables and targets. That was important to the new generation of general managers and accountants, and it quickly moved from input/output to input and impact – effect. That journey is still continuing” (Interview 143: top NHS manager, describing developments in the late 1980s and early 1990s). A number of our Dutch interviewees also recognized that the limited set of measures with which the Dutch system began would also have to expand (Interviews 202, 203, 204, 217). “They may not be intricate enough” said one leading reformer (Interview 217). Another observed that “I think the movement is still going on and all other parties, patients, are asking for more data. More than only the set of indicators that is [that] the hospitals are obliged to give” (Interview 204). A third said that the current set of PIs gave “a very small, it is a very rough impression of what the hospitals really perform. But it is a beginning” (Interview 203).

2. A tendency for the uses of PIs to move from voluntary to mandatory, and drift from the formative increasingly into the summative. Commenting on the NHS national PIs of 1986 one long-standing Department of Health official recalled “No big impact – out of date and not backed by performance management” (Interview 116). “You approach a time when a new deal is needed to freshen up relationships. ‘We really are hands off this time’, but then, of course, ministers or No 10 come along and we end up setting targets for matrons or something. Ministers are schizophrenic in that they’ll agree in principle to hands off, and then half an hour later they’ll say ‘what we need to do is…’” (Interview 131: civil servant). A long-standing Department of Health official described the sudden shift under Secretary of State Alan Milburn in the late 1990s as follows: “I don’t remember any major debates on star ratings, but it was a culmination of this approach of going from indicators to measures, to incentives and targets, which we seemed to move through incredibly quickly, given the problems we’d had with keeping those distinctions earlier on” (Interview 118). A long-standing Dutch medical academic described the history up to 2002 thus: “So you have the movement from guidelines [which] are not being implemented, to developing indicators based on the guidelines, and [then] the IGZ inspection asking for responsibility to develop performance indicators’ (Interview 221). Another Dutch respondent explained at length how, while perhaps 20% of hospitals would voluntarily get involved with developing better practices through an intrinsic motivation to improve, the other 80% would need “external incentives”, which could be money but was most likely to be the desire to preserve reputation (Interview 209: experienced medical manager and inspector). Once transparency rules put “the basic facts of your performance on the table” hospital managers will act because they don’t want to look bad in the press and in front of insurers and patients’ organizations.
3. As part of 2, indicators become associated with targets or quantified standards that organizations are expected to achieve. Thus, for example, the Dutch recently set a standard that, for hospitals to continue undertake a particular type of complex gullet surgery, they needed to be dealing with a minimum of 10 cases per year. Research looking at hospitals that appeared to be just above that threshold found considerable problems with the available data which purported to show that the necessary throughput had in fact been achieved (Smolders et al, 2008). Quite a few of our Dutch respondents argued that, at the very least, minimum standards for patient safety needed to be set, and hospitals that could not meet these would see ‘their’ patients being cared for elsewhere.

4. As PIs become more summative, and as targets are increasingly linked to incentives and/or penalties, so the pressures on staff tend to produce more gaming and even cheating (Bevan and Hood, 2006; Meyer and Gupta, 1994; Pitches et al, 2003; Smith, 1996). “I think that there is more pressure to fiddle once you start putting targets and managers’ salaries” (Interview 101: a long-standing PI consultant and researcher). “If you have a blame culture people will cheat or do whatever they need to do to meet targets or get around them” (Interview 141: healthcare academic). “There are some that have promoted perverse behaviour” (Interview 122). “The overwhelming majority of Chief Execs in the NHS live in fear.” (Interview 126: healthcare regulator). “In theory, they [the PIs] may become less summative. But given the macho-political thing, I would be surprised if some of that basket of performance metrics doesn’t end up being used summatively” (Interview 109: an experienced healthcare manager commenting on the moves since about 2005 to reduce the number of hard targets). A leading Dutch healthcare politician readily acknowledged that, though essential, they brought certain problems: “They can of course be manipulated. I am sure they are [laughs]. That’s why the Inspectorate should inspect them” (Interview 217). “These institutions have their own interests in what they tell you – if they can manipulate their information, they will do it” (Interview 220: Dutch medical academic)

5. Simple single indicators are a) subject to standardization in order to achieve more valid comparisons and b) amalgamated into complex, composite indices. Both make PIs more difficult for non-experts to understand. “I didn’t want to aggregate these [individual standards] – I wanted them produced like [school results] – you’re good at history and crap at maths. I lost that debate – ‘the great British public can’t cope’” (Interview 135: healthcare regulator). “You start getting clinical indicators – and all the same problems that you had with process indicators ‘ its not my data/how can you compare me with him’ and all of that. It is even more complicated – the health process, and I worked on it for forty years – is very complicated” (Interview 140: senior NHS manager and Department of Health official). In the Netherlands a bureau for transparency in healthcare, set up in 2007, now has a working group focusing on standardizing and controlling hospital statistical reporting. One proposal is that hospitals should hire accountants to audit their PI data. The difficulties of just offering ‘simple’ data
was well-summarised by one of our Dutch respondents: “Well, you have to give the general public everything, I think, but you have to be clever in at what level of information you are giving, raw data doesn’t bring us anywhere. It just confuses this discussion” (Interview 203: regulator).

6. The ownership and use of the PIs becomes more diffuse: groups and organizations which were not part of the original ‘launch’ take up the PIs and develop them or re-interpret them for their own ends. “And there are so many indicator initiatives at the moment – instead of one, where we hoped activity would coalesce – it hasn’t” (Interview 109: an experienced health manager). In both England and the Netherlands the media and specialist consultancy groups have taken up and reshaped PIs in this way. “[…] the performance management industry has grown and become better” (Interview 107: a long-standing NHS PI expert).

7. The continuing proliferation and change of PI systems and the airing of disputatious claims and counter-claims based on rival PIs may well lead to confusion and, eventually, mistrust on the part of the lay public. Even as early as 1997, some members of the incoming New Labour administration were concerned that public trust in NHS performance statistics was fragile. Now, a decade later, we can see more clearly that he skeptical attitudes of an aggressive and statistically largely illiterate media almost certainly plays an important role in the English case (Holt, 2008, pp327-328 and 344-345). It was striking that a recent Eurobarometer survey of public trust in official statistics showed the UK as the most distrustful of the 27 EU states and the Netherlands as the least distrustful (European Commission 2007). Small-scale focus group studies indicate that many members of the English public entertain a profound distrust of government-generated hospital performance data (Magee et al, 2003). The Netherlands, however, is no paradise. At least one of our most experienced Dutch respondents forecast that increased transparency would at first lead to a decline in trust: “I think we will go through a phase next five to ten years that people will be feeling very insecure, trust is at stake [and] if people don’t trust healthcare anymore they will complain about it.” (Interview 209)

This general sense of there being an internal logic to healthcare performance measurement was perhaps best captured two quotations, the first from a London-based health expert and media person and the second from a Dutch healthcare consultant: “Pandora is out of the box…she can’t be put back. Performance data has got to get better, and we’ve got to come to terms with a more transparent world – not just hospital but team or perhaps individual performance – not just for surgeons, who are relatively easy to measure, but also physicians” (Interview 127)

“The society and the profession, both have had a tendency to improve the system. And not to stay where we started – just as a signal system” (Interview 202)

Some caveats must immediately be appended to this ‘logic model’. First, the two national stories we have traced over the period since 1980 are not as neat as the above seven steps may seem to imply. By no means all PIs get used in summative or, still less, punitive ways, even in the UK. And it is still early days for healthcare PIs in the
Netherlands. Nor does the curve lead smoothly and inexorably upwards towards more indicators and more complex indicators. There are u-turns and occasionally successful attempts to simplify this or that piece of the apparatus. Nevertheless, if we look at both national systems now, as compared with at the punctuations of 1983 (UK) and 2003 (Netherlands), it seems clear that the current systems are more extensive, open to use in a more summative spirit, more closely tied to explicit targets or standards, and more diverse than at their respective points of origin. The occasional u-turns and frequent compromises never take the system back to where it was before.

In England, the first (1983) PI package was followed by further, more extensive packages in 1985, 1989, 1993, 1998, 1999, 2000 and 2001. Concepts of ‘performance’ widened to include more attention to access and to health outcomes, and it became more difficult to identify ‘PIs’ as a distinct phenomenon as the NHS performance regime became increasingly focused on targets. ‘Efficiency targets’ had existed since 1981, but were joined in 1986 by activity targets for Regions (based on PI scores), Patient’s Charter standards in 1991, and Health of the Nation targets in 1992. Although the originators of NHS PIs had been firmly opposed to their presentation in ‘league tables’, such tables had appeared by 1994. PIs had also become public as early as 1988, with Health Authorities instructed to make them available to news media, along with commentaries on their local interpretation. There had thus occurred a progressive shift from a formative to a summative performance regime. The Labour government elected in 1997 had a manifesto commitment to treat more patients and a ‘pledge’ to cut hospital waiting lists, and had also let it be known both before and after the election that NHS managerial jobs might hang on successful delivery of such policies. This approach of ‘constructive discomfort’ (Stevens 2004) underpinned much of what followed in the performance regime, in addition to a burgeoning range of new regulatory institutions (Walshe 2003). The ‘star’ system introduced in 2001 published composite scores for NHS institutions, based on a range of indicators, and can be seen as the triumph of a summative performance regime over the formative, though both then and now there remains a ‘counterculture’ within the performance industry that firmly believes in the virtues of the latter approach. Awards of ‘stars’ were widely publicised in the news media, which in 2000 had already begun to publish league tables of hospitals based on official data. Further NHS targets came as a result of the Comprehensive Spending Reviews published in 1998, 2002, 2004 and 2007, each with associated ‘Public Service Agreements’ setting out what was to be achieved with the resulting resources. In the primary care sector, general medical practices were from 2004 onwards financially rewarded for achieving a range of process and outcome targets in relation to their patient populations (the ‘Quality and Outcomes Framework’) and from the same date patient choice of secondary care provider (‘choose and book’) was progressively introduced, supported by access to (currently rather crude) composite data about hospitals’ performance as assessed by their regulator, the Healthcare Commission. In 2007, with official support, risk-adjusted surgical mortality data was published for each individual cardiac surgeon and official aspirations expressed for the principle to be extended to surgical teams in relation to other types of surgery.

This 25-year history of the NHS performance regime can perhaps be summarised as follows. First, the number of PIs employed has probably been roughly static since about
1990. Because of the proliferation of organisations producing indicators that are in some sense ‘official’, and because of the various descriptions of data as ‘indicators’, ‘targets’ and ‘standards’, it is impossible to give a precise figure, though our several respondents who had been employed in this business over the whole period tended to estimate a figure of somewhere between 2,500 and 3,000. Second, it is necessary to be somewhat sceptical of official claims of reductions in targets, partly because some of these are not borne out in fact, but also because not all PIs are equally important at any point in time; NHS managers are well able to judge the ones that matter. Third, the broad performance concepts (inputs, efficiency, effectiveness, access) embraced by PIs have not changed very much since the mid-1980s, though a pattern of constant NHS reorganisations and political momentum has been associated with constant re-presentation of PI schemes. Fourth, these developments have helped to create and sustain (and be sustained by) a substantial performance ‘industry’, comprising official regulators, academic institutions and commercial companies. Finally, the length of time over which we have studied England’s NHS performance regime does reveal a fairly clear ‘logic of escalation’ over the period, along the following lines. Once PIs have been created and despite numerous official caveats, they are subsequently arranged into comparative league tables which become public. The availability of quantitative data becomes associated with target-setting, whilst regulatory institutions begin to substitute PIs for physical inspection, and individual and/or institutional sanctions (occasionally rewards) become associated with PI scores. PI scores may also become seen as a source of information upon which patients and/or primary care gatekeepers can base their choices.

In The Netherlands, although the history of PIs is much more compressed, we can discern some of the same escalation. By 2004, the news media had begun to use PIs to construct their own league tales of the ‘Top 100’ hospitals, and in 2006 the Minister of Health announced that PIs were intended to be used by patients, insurers and referring GPs to make comparison and choices between providers, using websites such as www.kiesbeter.nl. Information is also available about the prices charged by different insurers. In 2007, IGZ began work on making PIs more comparable by standardising scores, e.g. through the use of casemix adjustments. It also announced that by 2013 it would have developed some 500 further PIs in order to cover every sector of health care, and that it would also develop normative standards in areas not covered by existing standards promulgated by professional associations.

If we are correct in discerning these common dynamics, the question remains of why this should be so. Some early thoughts would include the following:

1. The move from a few indicators to many stems from the difficulty of having a complex system where some aspects are measured and others are not. If attention is paid to what is measured, other areas may be neglected, and, when this neglect comes to the attention of the authorities, one obvious response is to add the neglected elements to the measurement system. In short, distortions can develop between what is measured and what is not, and an apparent solution is to bring more into the system (for an early argument on this, see Pollitt, 1986). Of course, this eventually leads to a proliferation of PIs, and to calls for simplification (as we have frequently heard in the UK). There is therefore a cycling process here
(Pollitt, 2008, pp51-59) although evidence from England suggests that the ‘simplification’ phases of the cycle may be more cosmetic than substantive.

2. The move from formative to summative may be thought of as the result of PIs constituting a standing temptation to executive politicians and top managers. Even if the PIs were originally installed on an explicitly formative basis, they constitute a body of information which, when things go wrong, can be seized upon as a new means of control and direction. And, sooner or later, things will go wrong (indeed, as we have seen, that is one important reason why PI systems were set up in the first place).

3. The setting of targets follows closely from the reasoning at 2 above. With the global spread of ideas of performance management, the idea of making subordinate organizations steerable (Osborne and Gaebler, 1993) and accountable by setting them numeric targets has become extremely widespread (Barber, 2007; Bouckaert and Halligan, 2008; Boyne et al, 2006; Carter, Klein and Day, 1992; Cave, Kogan and Smith, 1990; De Bruijn, 2001; Ingraham et al, 2003; Pollitt, 2006a, 2006b; Smith, 1996; Talbot, 2005).

4. The reason why gaming and cheating rise as systems become more summative is obvious: there is more to lose by scoring badly against whatever targets are in force. What is not always noticed, however, is how the existence of gaming and cheating could itself become a further reason for instability in PI systems – it would be logical for regulators to change these frequently so that the cheats will not have had time to learn how to deceive (Bevan and Hood, 2006).

5. The proliferation of PIs noted at 1, above, produces information overkill, especially for executive politicians and the lay public. So the step to a composite index for popular consumption (such as the NHS ‘star system’) is an easy one to take – for civil servants, consultancies and the media. Unfortunately, such composites may sometimes obscure more than they reveal (Pollitt, 2009). When composite measures change it is often unclear what the contributing factors have been; good composite scores can co-exist with poor services and vice versa. Certainly aggregated measures for how good a whole hospital is (or is not) are of very limited value in guiding a specific patient to make a specific decision about his/her specific condition.

6. Once the data is in the public domain, it can be taken up and used by others – indeed, nowadays that is one of the justifications given for the existence of PI systems, that they can guide consumer choice and political judgements (Pollitt, 2006a). Management consultancies in both England and the Netherlands have re-packaged the data, added their own commentaries, and sold it. The media have published their own ‘league tables’ and have sent reporters into very low- or high-scoring hospitals to try to find out more about what is going on there. In the UK the medical profession itself has eventually found it necessary to issue its own clinical PIs – from 2007 the Society for Cardiac Surgery has published the risk-adjusted operative mortality rates of individual cardiac surgeons.

7. Finally, the proliferation of PIs, their increased linkage to significant, summative incentives and penalties, the competition between different organizations each to issue the ‘best’ PIs – all combine to produce public bafflement and, possibly, mistrust of the figures. This is a motiveless crime, in the sense that, without
anyone meaning to, the conditions of Tsoukas’s ‘tyranny of light’ have been produced (Tsoukas, 1997). There is also the important background influence that trust of governments and all their works appears to have been in decline in many countries, and healthcare statistics will inevitably get caught up in this general tide of cynicism (Fellegi, 2004)

Discussion

What we are left with might be termed a ‘nested’ explanatory model, with macro-institutional features (political systems; financing structures) explaining the broad stabilities and constraints, contingent and opportunistic behaviours explaining the ‘punctuations’, and then, innermost, the particular interactions between the technical characteristics of PIs and the strategies of executive politicians, managers, consultants and the media explaining certain patterns of dynamic change within measurement systems. This is not simply a concentric model, it is also a sequential model, in which the temporal dimension is important. One thing follows another. Political or managerial crises usher in PI systems, but subsequently the systems themselves evolve, facing politicians and managers with new choices (when does the PI system become too complex and/or burdensome, and therefore require ‘simplification’; what should be done when other organizations pick up the data and use them in ways that are uncomfortable or illegitimate from the point of view of the original system designers?). As we clearly see from the longer English experience, building up a coherent and standardized measurement system takes some years – users have to be trained and convinced, categories have to be defined, trust has to be built up, and so on. Yet other, often exogenous, developments make the likelihood of such long periods of stability quite small: each crisis or challenge tempts managers and executive politicians to adjust the system rather than just allow it to mature.

Finally, we must ask how far our findings can be applied in other contexts. We are, of course, generalizing on the basis of two national, longitudinal cases. We cannot assume that other countries will be similar, although there is some general evidence of national policies being influenced by the broad nature of the political system (e.g. Lijphart, 1999; Pollitt and Bouckaert, 2004). On this basis we could expect to find similar developments in other majoritarian, centralized states (e.g. New Zealand) and other consociational/corporatist states (e.g. Sweden, Denmark). Nor can we assume that our findings will automatically apply to all public sector performance indicator systems. We have here been examining the healthcare sector – one which poses huge measurement problems partly because it supplies such a wide variety of (often unstandardized) services, some of which may be hard to connect to measurable outcomes. Quite possibly simpler, less contentious measurement systems (e.g. for issuing licenses or completing infrastructure projects) might not behave in the ways that we have described here. On the other hand there are other complex human services which might be expected to show at least some of the symptoms we have identified here (including education and social services). Taken together, such services represent a large part of most European public sectors.
Acknowledgement

We gratefully acknowledge a grant from the UK Economic and Social Research Council’s Public Services programme, reference number RES-166-25-0051. We also acknowledge the help of the 70 plus individuals who agreed to be interviewed as part of the project, without whose memories and expertise we could not have achieved much at all.
REFERENCES

Barber, M. (2007) Instruction to deliver: Tony Blair, public services and the challenge of achieving targets, London, Politico’s


European Commission (2007) Special Eurobarometer: Europeans’ knowledge of economic indicators, Luxembourg, European Commission


Helderman JK, Schut FT, van der Grinten T, van de Ven WPMM (2005) ‘Market-
oriented health care reforms and policy learning in the Netherlands’ Journal of Health Politics, Policy and Law 30(1-2): 189-210


