Feasibility first: Developing public performance indicators on patient safety and clinical effectiveness for Dutch hospitals

Marc Berg a,*, Yvonne Meijerink a, Marit Gras a, Anne Goossensen a, Wim Schellekens b, Jan Haeck c, Marjon Kallewaard d, Herre Kingma c

a Institute of Health Policy and Management, Erasmus University Medical Centre, Rotterdam, The Netherlands
b Dutch Health Care Improvement Institute (CBO), Utrecht, The Netherlands
c Dutch Health Care Inspectorate, Den Haag, The Netherlands
d Institute for Public health and Environment (RIVM), Bilthoven, The Netherlands

1. Introduction

Performance indicators are seen to be a promising answer to the demands for increased transparency, accountability and quality within health care. Rather than ensuring the public function of provider organizations or health plans through detailed rules and regulations, performance indicators monitor the level of performance of such organizations (percentage of children vaccinated; waiting lists below an upper limit). It is up to the latter just how to achieve these performance levels: what matters and has to be monitored is that they do. Bureaucratic interference with and regulation of health care organizations, in other words, would be reduced in favor of a more ‘businesslike’ approach based on performance contracts. Concurrently, these indicators would give professionals insight in the results of
their work processes: helping them to spot problems, redesign work processes, and measure the continuous improvement thus set in motion. They are, then, only held accountable for that which they are motivated and equipped to produce: high quality professional work [1–3].

In addition, performance indicators can help payers and clients make an informed judgment about providers: the client becomes a critical ‘consumer’, and payers may closely monitor their money’s worth. On the basis of a balanced set of indicators (on the clinical effectiveness, the efficiency, the safety and the patient-centeredness of the care, for example), rankings of care plans and delivery organizations can be made [8]. Finally, publishing performance information would further stimulate quality improvement processes: by turning ‘the heat on’ or ‘creating a burning platform’, the natural organizational resistance to change can be overcome [2,10–12].

The hope to simultaneously attain increased transparency, accountability, quality, professional motivation, deregulation and consumer emancipation has created an excited wave of performance indicators programs in Western health care [13,14]. Yet many are concerned about the risks of publicly ‘shaming’ professionals or institutions – especially when their ‘poor performance’ may not even be attributable to them [3,15]. Whether a certain indicator (say, vaccination rate of a population) indeed reflects the quality of care is endlessly debatable. Low vaccination rates may be due to population characteristics, for example. In addition to these validity problems, performance management may have perverse effects such as the tendency to ‘game the numbers’ or dump difficult cases 

1 These developments are not unique to health care: they are one widespread manifestation within the public sector of the drive to incorporate more ‘business like’ models of governance (see e.g. [4–6]). In addition to individual organizations and regional arrangements, indicators can also be focused on individual departments within organizations, individual professionals and even individual professionals [7].

2 In Europe, the UK is leading this development with its NHS Performance Indicators initiatives, see e.g. http://www.bbc.co.uk/iplayer/hi/panorama/3013062.stm, accessed February 7, 2005) and [9]. Many performance indicator initiatives are up and running in the US. The majority of these have so far not been obligatory, but that is changing rapidly (see e.g. www.iha.org, http://www.jcaho.org/pms, http://www.cms.hhs.gov/medicare/press/release.asp?Counter=1343, all websites accessed February 7, 2005). (and/or ‘cream off’ the easy ones), so as to manipulate scores. Also, researchers have pointed out that the hunt for good scores increases bureaucracy and induces ‘tunnel vision’ and suboptimization (because of excessive focus on improving isolated scores). Finally, performance indicator programs may erode solidarity between care institutions or stifle innovative change, thus actually hindering the diffusion of best practices out of fear of lowering one’s scores. Although anecdotes and predictions as to these effects abound, it is still too early to form a definite judgment [3,4,10,13,14,16].

Against the background of these discussions, the Dutch Health Care Inspectorate decided early 2003 to develop a set of Hospital Performance Indicators, covering patient-safety and clinical effectiveness. The Dutch health care system is a complex mixture of private, mostly not-for-profit health care organizations within a health care system that is closely regulated by government. Of its financing, 80% is public (compulsory sickness funds, social insurance programmes and general taxation). The remainder consists of private insurance schemes and direct payments. In recent years, the Dutch Government aims at reducing its own role in the overall management of the health care system by increasing ‘market incentives’ and stimulating the insurer’s role in steering the health care providers they contract. One core, public responsibility the government holds on to is the supervision and monitoring of the quality of the care delivered by both public and private providers. The Dutch Health Care Inspectorate, an autonomous section of the Ministry of Health, Welfare and Sport, is responsible for this task.

Until now, the Dutch Inspectorate uses a system of surveys and both random and incident-triggered inspection visits to health care providers. By developing performance indicators that would be both obligatory and public, the Dutch Health Care Inspectorate would be provided with data on the actual performance on

3 Newspapers and magazines carry many of these stories. The BBC – Panorama documentary ‘Fiddling the Figures’ (broadcasted June 29, 2003) see http://news.bbc.co.uk/1/hi/programmes/panorama/3013062.stm, accessed February 7, 2005) criticizing the UK focus on ‘targets’ for hospitals, in particular, telling.

patient safety and clinical effectiveness of all Dutch hospitals. This would meet three simultaneous goals:

- It would allow a first ‘screening’ of the care delivered in individual care providers, which would enhance the effectiveness and efficiency of the Inspectorate’s monitoring of the quality of this care.
- By being the first nation-wide initiative of its kind in the country, it would enhance the transparency of the hospital sector, which is an important government strategy in the light of its desire to introduce market elements.
- Through the public reporting of hospital performance, individual hospitals would be stimulated to improve their scores (the ‘burning platform’ effect mentioned above).

In this paper, we will first describe core discussions on indicator development and use that were crucial to the emergence of our approach. Subsequently, we will elucidate the Inspectorate’s vision: rapidly producing a feasible set of indicators, obligatory for all hospitals, that could and would be used in such a way that the mentioned aims would be fulfilled, while maximally preventing ‘side effects’ such as misinterpretations, defensive or perverse reactions.5

In realizing this vision, we argue, we bridged some of the classic distinctions between ‘internal’ and ‘external’ indicators, and questioned the seemingly self-evident call for exhaustive validity in discussions on external indicators.

2. Internal and external performance indicators

‘Performance indicators’ do not constitute a precisely defined category. Broadly speaking, by using this term one makes clear that one is primarily interested in whether (rather than how) the organization at stake achieves goals. These goals can be defined both by and for the organization [5]. The classic distinction of ‘structure’, ‘process’ and ‘outcome’ indicators partly overlaps with the category of ‘performance indicators’. Obviously, an organization’s ‘outcomes’ (say the 30-day mortality in stroke care) are part and parcel of that organization’s ‘performance’. Yet often, process indicators (such as length of waiting lists) or structure indicators (such as presence of internal, ‘blame-free’ complication registrations) are clear cut ‘proxies’ for outcomes. Alternatively, one can be interested in structural features of an organization such as the presence of disease-specific, integrated care arrangements, for example, to make sense of achieved clinical outcomes with respect to these patients.

Several authors have emphasized the importance of distinguishing the internal and external use of indicators [13,17]. Internal indicators are used by health care providers to monitor and improve the outcomes of their care processes. Professionals and managers can use these data to investigate where potential problems lie, and how they may be approached. On the basis of such analyses, care processes may be redesigned, and the indicators can then be used to monitor the consequences of these improvement attempts [18]. External indicators, on the other hand, are used by governments, patient organizations and payers to assess the quality of care of a health care provider, and to compare that quality to the performance of other health care providers.

The differences in purpose result in relevant differences in characteristics of the indicators involved. Failure to take these differences between internal and external indicators into account, according to these authors, results in indicators that are unfit for their task [18,19]. For internal purposes, indicators need to be relevant for the managers and professionals involved: they have to be specific to the care process at stake, aimed at its particular peculiarities and problems, and sufficiently detailed to capture the impacts of (planned or unplanned) changes to that process [17]. To make comparisons between organizations, however, this level of detail and specificity is irrelevant. Such comparisons require a more overall view of the performance of (groups of) professionals or whole institutions: what are the overall outcomes, average waiting times, global costs? Consumers and payers seem to be mainly interested in global patient satisfaction figures, and overall quality indicators such as reputation and accreditation [10,20]. In their turn, such coarse-grained indicators provide too little grasp to build improvement projects on.

5 The Inspectorate asked the Institute for Health Policy and Management, Erasmus University Rotterdam to help execute the project. In addition, the Institute for Public Health and Environment (RIVM) and the Dutch Health Care Improvement Institute (CBO) were vital participants and resources.
External indicators, according to these authors, require exhaustive validation. Comparisons have to be fair and real: relevant differences in patient mix, for example, may not be missed. The best surgeon will often have the highest mortality figures, since he will operate on the most difficult cases. A general practitioner in a deprived neighborhood may have very low vaccination rates because of the low compliance of patients, or because of the fact that his ‘acute’ workload is so high that there is simply no time for preventive activities. Finally, exhaustive validation is required to prevent perverse effects such as ‘creaming and dumping’ (selecting those patients who will guarantee high scores, and denying service to ‘difficult’ patients to similarly improve the indicator score) [3].

Exhaustive validation, then, means correcting for all the possible relevant differences between the care contexts compared. This requires large amounts of precisely registered and comprehensive data, including all the potentially relevant differentiating variables such as age, sex, gender, ethnicity, socio-economic status, severity and range of complaints, stage of the disease, diagnostic and therapeutic activities undertaken (including the reasons for doing so), and so forth [21–23]. This ‘scientific soundness’ comes at a high cost (literally speaking): the required number of data and the registration-effort approach the levels required in a clinical trial [17,24].

For internal use, such levels of validation are not required. First of all, this usually concerns outcome or process measurements over time within organizations, where many variables remain constant (and thus need not be corrected for). In addition, the same professionals that produce internal indicators subsequently interpret them. To a large extent, they are optimally equipped to ‘filter’ irrelevant bias or noise out of the data: they know each other’s patients, for example, and are aware of potential differences in case mix, context of care-delivery, and so forth [25]. Their ‘ownership’ and the direct transparency of the data (since no statistical risk adjustments and so forth have been applied) enhances the professionals’ motivation. Small data samples usually suffice; the pragmatics of the individual improvement process at stake here overrides the requirement for universal usability of the numbers [18].

This way of juxtaposing ‘internal’ versus ‘external’ indicators portrays them as two separate categories.

As Berwick argued in 1998, ‘measuring for improvement is not measurement for judgment’ [26]. In this dichotomy, internal indicators, jointly working on ‘improvement’, are about trust – in each other’s intentions, in the common purpose of improving the patient’s care, and in the conscientious treatment of ‘soft’ numbers. External indicators, on the other hand, are about formal relations of accountability: mechanical rankings of meticulously processed numbers, whose ‘hardness’ is to form the objective arbiter of ‘good’ versus ‘bad’ performance (cf. [27]).

As we will see further, however, there are some deep problems with especially the way external indicators are portrayed in this analysis. We will now describe what the starting points were for the project described here. In the discussion we will come back to the internal-external distinction, to the issue of validity, and argue for a more fruitful understanding of these different uses for and types of indicators.

3. Indicators for the inspectorate: starting points

In addition to the aims laid out in the Introduction, the Dutch Inspectorate set two important prerequisites at the start of the project. First of all, the cooperation of hospitals and professionals had to be gained. The Dutch Hospital Association (NVZ), the Association of Academic Hospitals (VAZ) and the Dutch Medical Specialists Association (Orde) were critical about the idea of hospital-specific, public performance indicators. They argued that defensive reactions would be inevitable, and that manipulation of numbers was to be expected. Given the Dutch traditions of governing-by-consensus, and the direct dependence of the project’s success of the cooperation of hospitals and specialists, this critical position was to be taken very seriously. The Inspectorate set out to develop a vision on how to develop and use these indicators in such a way that there would be no need to be overly defensive, and that would prevent perverse reactions.

Second, the Inspectorate did not want to start a large-scale project, which would only result in practically usable indicators after several years of fundamental groundwork. Current guidelines for the construction of (external), ‘evidence-based’ indicators emphasize the need to first scientifically establish where quality
improvement is possible and feasible. Subsequently, indicators are prioritized and selected on the basis of literature reviews as to their relevance, potentially influencing factors, and so forth. After that, evidence-based measure specifications have to be developed, including specifications on whom to include and exclude, how to adjust for case mix, and how to register and analyze the data [21–23]. This process would be both lengthy and costly, while the Inspectorate wanted results within 1 year, and had limited resources.

From the literature on indicators and the experiences of indicator projects elsewhere (see the previous paragraph), however, we had deduced four important lessons. First of all, the more ‘valid’ the indicator has to be (in terms of comparability between institutions, for example), the more work it is to construct and report on them. For those creating the indicators, it implies that precise risk adjustment schemes have to be developed and all potential confounders have to be taken into account. For those having to report on the indicators, higher validity comes with high demands on the precision and extensiveness of data registration (see the grey line in Fig. 1) [21–23,28,29].

Second, because of the required use of increasingly complex statistical procedures to compensate for case-mix and other potentially confounding variables, there is a point after which the aim to increase validity and thus comparability in fact decreases the transparency of the indicator. That is to say, the statistical processing required to make numbers (more) comparable decreases the direct meaningfulness of the numbers – especially to those professionals and managers who are directly responsible for their production [13].

Third, the more you try to make indicators ‘really’ comparable, the more your claims are challenged, and the more defensive reactions you will get from those having to report on them. The more ‘comparability’ is stated as an achievable goal of an indicator project, in other words, the more meticulously those critical of the indicator scheme (for whatever reason – fear, interest, principle) will attempt to challenge that aim. The advanced case mix adjustment techniques and precise specifications of inclusion and exclusion, for example, can always be contested. There is simply never enough unequivocal evidence available to avert such discussions, for one. Also, there may always be reasons for confounding that were not foreseen when the case mix adjustments were made [21].

Fourth, and closely related, the more direct and serious the consequences of having high or low scores, the more manipulations and perverse reactions you may expect. It is no coincidence that the UK ‘star system’ invites creative uses of data ‘massage’. Since high star rankings yield financial and managerial benefits, and low rankings may immediately result in similar punishments, there will be high pressure to ‘perform’ – in whatever way. We do not here want to argue that systems as used in the UK have an overall negative impact – preliminary data coming out of the first evaluations seem to show that the positive impacts are far more prominent [30]. Yet it is obvious that systems that are less punitive, and/or treat indicators less as a direct instrument for rewards or penalties, will suffer less from such perverse effects.

All these insights led our team to take a second look at the stated aims and prerequisites of the project. As stated above, we were to:

- create a ‘screening’ instrument for the Inspectorate of the quality of care delivered in individual care providers;
- enhance the transparency of the hospital sector;
- stimulate individual hospitals to improve their scores through the ‘burning platform’ effect; while:
- keeping the hospitals and professionals on board;
- doing all this in a short timeframe.

This did not imply, we decided, that we required an objective ‘ranking’ of institutions, or a meticulously crafted set of indicators that would show at a glance...
what hospital was ‘good’ or ‘bad’. What we wanted to achieve, rather, was to start a process in which quality improvement initiatives would be stimulated, ‘accountability’ would be seen as a normal feature of professional work, and perverse effects would be minimized. This led us to the following starting points for the construction and use of the Dutch Hospital Performance Indicators.

3.1. The indicators do not constitute a ‘cockpit’ for the Inspectorate, but lead to requests for (more) explanation, additional investigation and/or the start of a dialogue

The indicator set does not function as a set of control panels for the Dutch Inspectorate, from which the quality of the individual hospitals can be simply deduced. A high score on an indicator such as ‘heart failure readmissions’ (or, for that matter, a very low score on the indicator ‘pressure ulcers’) is a signal that additional questions and/or investigations are required. It is a first screening that forms the basis for the next phase of more in-depth investigation. Only when this additional step demonstrates a real failure to deliver the required quality of care (or to properly investigate this quality), sanctions may be put in place.

3.2. The indicators should stimulate internal quality improvement within hospitals

Continuous quality improvement by hospitals and professionals themselves is the only way the quality chasm can ultimately be bridged [1, 31]. The indicator set, then, is tuned so as to stimulate such projects. Wherever best practices are being implemented, the indicator set sets an indicator so that those hospitals having implemented this best practice (through, for example, a Breakthrough project6) are rewarded with a high score. Conversely, those hospitals that have not yet implemented this best practice are given an additional incentive to do so as well. In addition, national quality initiatives that have already been developed by professionals themselves (such as the national surgical complication registry) are not taken over by the Inspectorate. After all, the more hospitals and professionals are themselves responsible for assessing and assuring the quality of the care they deliver, the better. In such instances, the participation in such initiatives is scored – keeping the development of an adequate, national complication registry itself out of the hands of the ‘external’ Inspectorate [35].

The Inspectorate, then, does not produce a national ranking of hospitals. Nor does it give ‘stars’ or produce hospital ‘report cards’. Much more important than the ability to objectively make comparisons between hospitals is the ability to trace improvements over time within hospitals. This is not to say that comparisons between hospitals are or cannot be made. The Inspectorate does so internally to fine-tune their subsequent queries for information and investigations. Also, some of the process- or structure-indicators are rather straightforward, so that rankings are possible on those individual items. Finally, hospitals and professionals compare themselves with their direct competitors. Since these actors are highly aware of the specific conditions in which they and these competitors have come to the results listed, they will be able to attach meaning also to uncorrected outcome figures.

3.3. The results are made public by the hospitals themselves

The public nature of the individual hospital scores was a highly contentious issue. As elsewhere, some of the most successful, national quality improvement projects (surgical complication registration, wound infection registration, Breakthrough projects, and so forth) are carefully designed to keep the identity of those supplying the data secret. As said above, the movement towards openness about outcomes is a very precarious one, and professionals are (often rightly so) concerned that ‘raw’ data may be misconstrued when ‘out in the open’. ‘Trust’ is crucial for true reform to flourish on the shop floor.7 Individual hospitals can get information on how they ‘perform’ in respect to their peers, but these ‘peers’ are usually made anonymous.

6 The Dutch Health Care Improvement Institute (CBO) has been running ‘Breakthrough’ projects for several years now [32]. The ‘Breakthrough’ concept was developed by the US Institute of Healthcare Improvement. It is a collaborative improvement model in which multiple health care organizations (or parts thereof) work in parallel to bring existing ‘best practices’ into their own work environments [33, 34].

7 See also the current focus on methods such as ‘blame-free reporting’ to fight the high mistake-rates in health care. See e.g. [31].
In addition, the data, aggregated or not, remain hidden from the public.

The Inspectorate, however, is a government agency, and is therefore not able to keep the data it gathers out of the public domain.\(^8\) In addition, the Inspectorate wanted to use their indicators to enhance the transparency of the public sector. To emphasize the importance of the internal use of these indicators, however, the Inspectorate now asks hospitals to report their individual results themselves, hospital by hospital. Each hospital can choose its preferred way: through a website, a separate report, integration in yearly reports already produced, and so forth. The only conditions are that the requested information is complete, on time, and that the Inspectorate should be informed about where the data can be found.

To further emphasize that the data are from (and largely for) the hospital itself, the Inspectorate has made a deliberate choice to only use information directly from the hospitals itself. No use is made of national registries or patient surveys that are not ‘owned’ by the individual hospitals themselves.

Having the individual hospital publish its own figures, in addition, also allows it to add explanatory remarks to indicators. A hospital that has a high readmission rate for heart failure patients, for example, can explain why the categories of heart failure patients they treat will inevitably yield such a high readmission rate. Alternatively, it can admit that the score is higher than they aim for, and elaborate how the hospital will be ameliorating this situation. In this way, thus, it is not the Inspectorate that has to demonstrate why a score can be compared between hospitals – it is the individual hospital that is free to choose to argue why their score is not comparable.

3.4. Feasibility first: the indicators are but a first step in an ongoing process

Registering and analyzing the indicators has to be feasible for both hospitals and Inspectorate. It was crucial, therefore, to aim for a limited set of indicators that wherever possible could draw upon existing registries or databases within hospitals. We stated as a central starting point that additional registration demands is only acceptable if this additional work is a prerequisite for the ability to deliver adequate quality of care. Not all hospitals, for example, register the prevalence of pressure ulcers in their wards. Yet it is widely accepted that a proper pressure ulcer prevention policy requires a proper registration to prosper.

The first version of the Hospital Performance Indicators set, then, is by no means a finished product. It would be an illusion to think that the quality of hospital care could be adequately captured in this set of some 30 indicators. Much work will still need to be done with regards to the further fine-tuning of indicators, the search for additional indicators covering relevant care domains that are now missing, and so forth. This work will be greatly aided by the feedback that is now being received from hospitals actually working with this first set, and from our own experiences in trying to make sense from the results reported.

Importantly, however, the aim is not an ever-extending set of indicators. It is imperative that the indicators remain practically feasible. It is crucial, also, that the refining of the indicators does not lead to a move beyond the ‘optimal’ validity level indicated in Fig. 1 by the yellow column: the point where the registration work, the transparency and the indicator’s validity are in an optimal balance.

4. Results

In March 2003, a small group of researchers and Inspectors started with the project, of whom only two junior researchers were available most of their time.\(^9\) After having established the here described starting points, the Inspectorate’s aim and planned approach was presented to a selection of stakeholders (specialists, hospital managers, their organizations, the Ministry, and so forth) during a so-called ‘expert meeting’ in April.

During and after the expert meeting, intense contacts with the Dutch Hospital Association (NVZ), the Association of Academic Hospitals (VAZ) and the...

---

\(^{8}\) According to Dutch law, information held by Government and government-related bodies are in principle open to public access (‘Wet Openbaarheid Bestuur’, or ‘WOB’).

\(^{9}\) Exact time investment varied, but on the average 1.5 for junior researcher time was available throughout the project. The two Inspectors, two RIVM senior researchers and two IBMG senior researchers spent a few hours until one day per week on the project.
Dutch Medical Specialists Association (Orde) led to a ‘joint venture’ in which it was agreed to ensure a complementary rather than competing strategy in performance indicator development. These organizations were starting to develop indicators for their own purposes, and the Inspectorate announced that it would gladly incorporate their suggested patient safety and clinical effectiveness indicators within its own set. In addition, these organizations would construct additional patient-centeredness and efficiency indicators. All indicators were to be jointly piloted and presented.

Meanwhile, working from the guiding principles listed above, internationally available indicators were scanned, and a broad range of experts was interviewed as to which areas would be most desirable and suitable for indicator development. In addition to our guiding principles, we used the criteria listed in Box 1 for our selection.11 (See for a more detailed discussion of this process, including the criteria and additional considerations used, Gras et al., in preparation). A second expert meeting, with a similar selection of stakeholders, took place end of June 2003. Here the first selection of indicators was presented and discussed. Refinement of these indicators, including discussion with field experts and (additional) literature reviews, continued during the summer. For every indicator, we established definitions and pinpointed how hospitals were to collect the data, using which sources. Doing so, we constantly balanced the desire to have as precise definitions and methods as possible with the requirement to not overly burden hospitals with additional registration requirements. Given the sensitivity of the latter issue, and our philosophy of starting a first step in an ongoing process, we regularly left the exact source used open, for example, asking the hospital to list the source it had used.12

In October, a pilot was undertaken, in which six hospitals volunteered to test the indicators, and attempted to deliver the requested information in a 1-month period. No serious problems were encountered, and in December 2003, the definite set of indicators was mailed to all Dutch hospitals. In total, the process had taken 10 months. By July 10th, all but one hospital had published their scores, either on the Internet or in a paper-based report. Currently, hospitals with incomplete responses have received additional reminders, and the preparation of the overall report on 2003 is under way. Simultaneously, small improvements have been made to the set (mainly in the form of some more precise definitions)

<table>
<thead>
<tr>
<th>Box 1: Criteria used for indicator selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <strong>Outcome indicator or ‘proxy’ outcome indicator</strong>: When no outcome indicators were available, process or structure indicators with a direct or proven relationship with outcomes were selected.</td>
</tr>
<tr>
<td>2. <strong>Prevalence of the issue</strong>: Indicators were preferred that focussed on issues with a high prevalence in order to obtain relevant areas for quality improvement.</td>
</tr>
<tr>
<td>3. <strong>Significant potential improvement of quality</strong>: Indicators were selected that focused on areas where variety in current quality provided was large, and the potential for improvement was significant.</td>
</tr>
<tr>
<td>4. <strong>Clear and timely connection with care activities</strong></td>
</tr>
<tr>
<td>5. <strong>Causing desirable outcomes</strong>: Indicators were selected so as to minimize obvious ‘perverse effects’ or gaming.</td>
</tr>
<tr>
<td>6. <strong>Administrative ease of implementation</strong></td>
</tr>
</tbody>
</table>

10 These purposes varied, but included a focus on creating transparency for accountability towards government, patients, payers, and so forth. At the moment the joint venture was made, all of the parties were in the beginning phases of developing indicators.

11 These criteria are commonly used; e.g. [www.qualitymeasures.ahrq.gov/resources/measureuse.aspx](http://www.qualitymeasures.ahrq.gov/resources/measureuse.aspx) (accessed February 7, 2005).

12 This could potentially also hinder trending and comparison within hospitals over time, as the hospital may end up using a data source that, in a subsequent refinement of the indicator set, would be considered inadequate. Yet hospitals that really wanted to use this first year’s data in trending could of course ensure that subsequent definitions could be internally related to these earlier ones wherever possible.
which has now been published and sent to the hospitals as the 2004 Indicator set. This same set will also be used for 2005; the first substantial improvements will be incorporated in the 2006 Indicator set, to be available for the hospitals at the end of 2005.

In the meantime, the hospital association has opened a web site that allows for comparison of hospital scores on the indicators. In addition, a large Dutch national newspaper (Algemeen Dagblad) published several front-page news items discussing the varying performances between hospitals regarding preventing pressure sores, the speed with which hip fractures were brought to surgery, and the (often far too low) volume of high risk operations in individual hospitals. In October, they published the Hospital Top 100, ranking all Dutch hospitals on the basis of the Hospital Performance Indicator set. Given the fact that data sets were often incomplete, and that inter-hospital comparison was often difficult, the newspaper had decided to base the ranking mainly on the structure and process indicators, and on the completeness of the data sets. This publication itself received much attention within the news media, both within and outside health care. Interestingly, the public reactions of the hospital and professional organizations were almost wholly positive, emphasizing the importance of this new step towards transparency while simultaneously pointing out that the data set as yet was rather incomplete and sometimes as yet ill-defined.

The 2003 set of indicators is listed in Table 1. Background information per indicator and the improved 2004 set can be found (as yet in Dutch only) on www.prestatie-indicatoren.nl.

5. Discussion

5.1. Validity or feasibility?

As in many other private and public sectors, performance indicators have entered the health care arena with high expectations attached. Embedded in a philosophy of ‘fair accountability’ (i.e., do payers or society get their money’s worth? Is the trust we invest in health care providers legitimate?), performance indicators are to increase transparency, empower the patient, while simultaneously affording professional self-regulation and -improvement.

Yet as any medication or policy instrument, performance indicators have ‘side effects’. They can lead to bureaucracy, to defensive medicine, to fiddling of figures, and, worst of all, to actions that actually worsen the quality of care on adjacent (non-measured) parameters so as to artificially score ‘well’ on the indicator that happens to be measured. According to critics, these side effects massively outweigh the potential benefit of performance indicators.

Although these critiques are only based on anecdotal evidence, this critical perception of performance indicators is widespread amongst many health care professionals and managers. For several reasons, we felt it was important to take these perceptions seriously. First of all, and very pragmatically, the Inspectorate could not risk alienating the professionals and providers they were dependent upon for the success of this new mode of working. Second, although their overall disqualification of the public use of indicators is unfounded, many of the critics’ analyses address real problems that, unless dealt with adequately, will prevent the fruitful use of public indicators.

In our view, many of these real problems boil down to one fundamental issue: the fundamental impossibility to found indicator systems on a rockbottom of ‘validity’. As we argued above, and as we repeatedly experienced ourselves in creating and discussing (potential) indicators, once one falls into this trap, there is no way out. Promising a ‘valid’ set of indicators, in the sense of fully representative, free of bias and confounding influences and fit for making rankings and rewards (or penalties), is setting up an impossible challenge to meet. It would take endless work, and so much data correction that the debates about the assumptions built into the statistical correction techniques would themselves explode. Since new interpretations of the data are in principle never exhausted, the noble aim of full validity is unreachable in principle.

Our objective, in contrast, was to retain the feasibility of our project in several senses of this term. The indicator construction, maintenance, processing and subsequent interpreting and reporting on the indicators had to be and remain feasible for us (the Inspectorate and the development group). Also, the understanding and
Table 1
The Dutch hospital performance indicator set consists of three subsets: (1) hospital-wide indicators, (2) indicators for the emergency ward, operation theatre and intensive care units, and (3) condition or intervention-specific indicators.

<table>
<thead>
<tr>
<th>Indicator Structure</th>
<th>Process</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital-wide</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Pressure ulcer</td>
<td>1.1 Presence of registration of Pressure Ulcer prevalence/incidence</td>
<td>1.1 Point prevalence pressure ulcers</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.2 Incidence of pressure ulcers in patients with an indication for total hip replacement</td>
</tr>
<tr>
<td>2. Blood transfusion</td>
<td>2.1 Presence of transfusion reactions registration</td>
<td>2.2.8 Transfusion Reactions</td>
</tr>
<tr>
<td>3. Medication safety</td>
<td>3.1 Availability inpatient and outpatient medication overview</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.2 Availability regional medication overview</td>
<td></td>
</tr>
<tr>
<td>4. Information technology</td>
<td>4.1 Availability of electronic data in the outpatient consultation rooms and on the hospital wards</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.2 Availability of process-supporting IT in the outpatient consultation rooms and on the hospital wards</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.3 Access to Internet and internal and external e-mail for care professionals</td>
<td></td>
</tr>
<tr>
<td>5. Wound infections</td>
<td>5.1 Presence of wound infection registration</td>
<td></td>
</tr>
<tr>
<td>6. Complication registration</td>
<td>6.1 Presence of complication registration per specialty/discipline</td>
<td></td>
</tr>
<tr>
<td>6. Risk inventory</td>
<td>6.1 Availability of clinical risk inventory</td>
<td></td>
</tr>
<tr>
<td>The emergency ward, operation theatre and intensive care units</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Post-operative pain</td>
<td>1.1 Percentage of post-operative patients having received standardized pain measurements</td>
<td>1.2S Percentage of patients whose pain score is less than 4 within the first 72h</td>
</tr>
<tr>
<td>2. Volume of high risk interventions</td>
<td>2.1 Volume of repairs of unruptured abdominal aortic aneurysm</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.2 Volume of esophageal resections for esophageal carcinoma</td>
<td></td>
</tr>
<tr>
<td>3. Laparoscopic surgery</td>
<td>3.1a Ratio of laparoscopic versus open cholecystectomy</td>
<td>3.2S Percentage of conversions from laparoscopic to open cholecystectomy</td>
</tr>
<tr>
<td></td>
<td>3.1b Ratio of laparoscopic cholecystectomy in day care versus inpatient laparoscopic cholecystectomy</td>
<td></td>
</tr>
<tr>
<td>4. Cancelled operations</td>
<td>4.1 Number of elective operations cancelled within 24h before surgery</td>
<td></td>
</tr>
<tr>
<td>5. Unplanned re-operations</td>
<td>5.1a Percentage of unplanned re-operations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.1b Top three unplanned re-operation indications</td>
<td></td>
</tr>
<tr>
<td>6. Intensive care</td>
<td>6.1 24-h availability of a registered intensivist</td>
<td>6.2 Mean and median of number of artificial respiration days per patient requiring artificial respiration</td>
</tr>
</tbody>
</table>
Table 1 (Continued)

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Structure</th>
<th>Process</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition- or intervention-specific indicators</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Pregnancy</td>
<td></td>
<td>1.1 Percentage of cesarean sections</td>
<td>4.1 In-hospital mortality</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.2 Percentage of vaginal deliveries after cesarean section</td>
<td></td>
</tr>
<tr>
<td>2. Diabetes</td>
<td>2.1 Presence of Integrated Diabetes Care service</td>
<td>2.1 Mean HbA1C value of diabetes patients</td>
<td>4.2 30 days mortality</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.2 Percentage of patients receiving an eye examination every 2 years</td>
<td>5.1 In-hospital mortality</td>
</tr>
<tr>
<td>3. Heart failure</td>
<td>3.1 Presence of Outpatient Heart Failure Clinic</td>
<td>3.2 Readmission rate for heart failure patients</td>
<td>5.2 30 days mortality</td>
</tr>
<tr>
<td>4. Acute myocard infarct</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Cerebrovascular accident</td>
<td>5.1 Presence of Stroke-service/hospital stroke unit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Hip fracture</td>
<td>6.1 Percentage of patients operated within 24 h after admission</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Total hip replacement</td>
<td>7.1 Presence of Joint Care service</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Mamma tumor</td>
<td>8.1 Presence of Outpatient Mamma Care Clinic</td>
<td>8.2 Percentage of patients receiving diagnosis within 3 days of first outpatient visit</td>
<td></td>
</tr>
<tr>
<td>9. Cataract surgery</td>
<td>9.1 Presence of cataract surgery unit and care pathway</td>
<td>9.3a Corrected post-operative vision</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.2 Presence of cataract surgery complication registry</td>
<td>9.3b Mean difference between the actual and the intended refraction after cataract surgery</td>
<td></td>
</tr>
<tr>
<td>10. Refraction surgery</td>
<td></td>
<td>10.1 Reduction in refraction</td>
<td></td>
</tr>
</tbody>
</table>

The indicators printed in red have been included by the Dutch hospital and specialist organizations. Indicators marked with an ‘S’ are merely ‘suggested’; reporting on these indicators is not obligatory.

reporting on the indicators had to be feasible to the hospitals: data had to be available or not too complicated to gather. Too much additional registration requirements would be impossible.

Finally, there was a strong political sense in which the project had to remain feasible. We were constantly exploring those issues in hospital care that were ‘just ready’ to be pushed one step further. It would have been unnecessary, for example, to put effort into an issue that was already widely receiving proper attention. As an example, the issues of ‘post-operative wound infections’ or ‘surgical complications’ were not tackled, because national programs had already been set up for these topics. As mentioned above, contrary to the Inspectorate’s indicator project, these programs were based on anonymity. The Inspectorate’s direct involvement at this stage could only harm these projects.

Concurrently, in our set, hospitals are asked to report whether they participate or not – thus stimulating these projects, without meddling with their working principles. On the other hand, asking hospitals about numbers of high-risk operations per surgeon has so far been completely unheard of in the Netherlands. Yet since the correlation between the volumes of several high-risk surgical interventions and their outcomes has been proven so many times now, this was an issue that many felt overdue, and that would be very hard to object against (see e.g. [38,39,40]). Likewise, we only allowed the indicator set to ask for more, complete data registration whenever it was clear that these data would be similarly crucial to a hospital’s proper internal quality control.

In other words: our ‘feasibility first’ philosophy emphasized that our aim was to start a process of self-improvement, of increased attention to accountability...
and the legitimacy of the public’s question for transparency. It was not a ‘ranking’ we were after, nor an attempt to provide a flawless and transparent map of the hospital landscape. Such an attempt could only have invoked debates about the accuracy of map, and the injustice of some of the representational techniques selected. Rather, we opted for a deliberately rough and multi-interpretable map, leaving the hospitals free to add explanatory markings and legends, and work from there. From such a map, no simple overall rankings could be made, but many more or less interpretative comparisons between individual hospitals or scores would be possible. This strategy has proven powerful. The Inspectorate received little true opposition to the request for the indicators, and hospitals were fully responsible for the publication of their own information. When the national newspaper did rank the hospitals, public discussions mostly turned to the as yet rather incomplete data as delivered by the hospitals, and about the ranking method used by the newspaper. The Inspectorate remained outside of this particular discussion – its authority untouched by potentially endless discussions on ranking methods or validity discussions.  

5.2. Internal or external indicators? 

To elaborate this some more, we could say that we deliberately undid some of the classic distinctions between ‘internal’ and ‘external’ indicators. We asked for public presentation of a prescribed set of performance indicators, yet we allowed hospitals themselves to explain their scores. In this way, the discussion would not be focused on whether we had made the indicator ‘valid’ enough, but on the more relevant issue: whether the hospitals arguments for their relatively good or bad score made sense in the light of comparable hospital’s arguments and scores. Of course, this limited simple comparability and ranking by consumers or payers, for example. Yet this would stimulate hospitals to scrutinize the information and make comparisons with their own, relevant ‘competitors’. Having much more of the contextual ‘inside knowledge’ required to make sense of these data, they would be able to tease out relevant distinctions, and to pinpoint areas were improvement was needed, or where further self-scrutiny was called for. In addition, it would give the Inspectorate additional (not ‘unequivocal’) information about the performance of hospitals in the domains investigated, in an efficient way. Obviously, this information would only be usable when combined with other information sources – such as the experiential and factual knowledge of regional inspectors about certain hospitals, previous (individual) reports, and so forth.

We aimed, then, at creating a ‘burning platform’ effect, which has so often been pointed at as crucial for real organizational change, and at stimulating a culture of transparency, self-assessment and self-improvement. Simultaneously, we front staged the importance of interpretation of the results – by hospitals themselves, and by everyone who aimed to draw conclusions from these data.

To put this more theoretically, we explicitly undid the concept of ‘external indicators’ from some of its seemingly self-evident characteristics. As we summed up in the graph on the essential limits of ‘exhaustive validity’ when it comes to the use of indicators (Fig. 1), the stereotypical ‘external indicator’ does not exist. It is both practically and philosophically impossible to create full comparability; to correct for all the possible relevant differences between care contexts. It is possible, however, to search for the optimum in the graph; for the point before which added validity in fact decreases transparency and explodes registration work. Where exactly this optimum lies is of course itself context-dependent: the levels of registration work seen as ‘normal’ in United States hospitals, for example, are considered far out of bounds for Dutch hospitals and professionals. The point is that ‘external’ and ‘internal’ indicators are not so much two different categories of indicators as points on a continuum: ‘internal’ indicators being immediately transparent to only those within the care context itself (see the next paragraph), easy to register but impossible to use for inter-hospital comparisons. In their turn, ‘external’ indicators are indicators that are relatively better equipped for between-site comparisons because definitions and data gathering methods have been streamlined between sites as far as possible and basis case mix adjustments have been performed. External indicators are not ‘exhaustively valid’: they cannot be, are not now and never will be. Indicator programs that make such promises are doomed to fail or disappoint.
5.3. Feasibility first: a price too large to pay?

At the moment of this writing, the hospitals are preparing for the second round of reporting on the indicators. It is clear that this second round, no hospital can afford to publish incomplete sets. In addition, given all the media attention, it will be very hard for a hospital to have to report too low volumes on the selected high-risk surgeries. The move away from ‘ranking’ and towards ‘self-interpretation’ has greatly helped acceptance – both by the professional and hospital organizations, as by individual hospitals, managers and medical specialists. That we have to move towards a more public ‘accounting for’ what goes on within hospitals is an assumption nobody contests. The contestation comes as soon as concrete examples of what we will then account for are put on the table. Through our ‘feasibility first’ approach, however, the tables are turned: it is not the Inspectorate who will – endlessly – have to defend why the definition and handling of the indicator is flawless, but it is the hospital that has to argue why this ‘rough’ indicator may be poorly suited for it.

At this point, one may ask whether the redefinition of ‘external indicators’ argued for above, the loss of the ability to clearly demarcate ‘which hospital is better’, is not a price that is too large to pay. Is it not a great limitation of our approach that this remains impossible? Is ‘transparency’ not equivalent to being able to clearly choose one hospital over another, on the basis of clear-cut performance data?

We would argue that this is not the case. It is simply impossible to make such a clear-cut set of performance indicators: data may always be outdated, not relevant for the case at hand, unforeseen confounders may be at play, and so forth. Interestingly, those using public performance information know this to be true. In a study on the use of publicly published, outpatient clinic waiting times by patients, general practitioners and insurance companies, it became clear that no one used these data at face value. Everyone involved knew that the publicly published ‘waiting times’ were the result of a complicated construction process in which strategic considerations and definition problems were deeply intertwined [41]. These data, in other words, could only be used as coarse pointers that still require subsequent investigation or dialogue.

Experience shows that so far, ranking schemes and report cards do not appear to be used very much by those they are formally constructed for: consumers do not really seem to make different choices, and payers do not seem to be directly reacting to these figures by contracting different providers, for example [10,42]. We suggest that an important reason for this relative non-use is not so much a lack of sophisticated decision-making on their part. Rather, is may be these potential users’ (im- or explicit) awareness of the limitations of such seemingly objective ranking tables that explains their hesitation. On the other hand, those that are the subject of the public presentation of performance data do act strongly. Sometimes defensively, and in a dismissive way, but more often than not (simultaneously) by improving their practices vis-à-vis their peers [3,30,43].

‘Transparency’, in our view, is a relative concept. It always has to be indicated to whom something is transparent or not. Performance information that may be transparent to an insider, may be completely impossible to interpret (and thus opaque) to an outsider. Other information may be ‘transparent’ to professionals, but not to laymen. Similarly, something is never simply ‘transparent’ or not – it is always more or less so compared to an earlier or a later assessment [44].

If the project discussed here would have set out to construct a ranking of hospitals, or to construct clear-cut criteria for doing so, the project would have certainly failed. It would have lacked the required support from the field, and it would have spent all its resources making on their part. Rather, is may be these potential users’ (im- or explicit) awareness of the limitations of such seemingly objective ranking tables that explains their hesitation. On the other hand, those that are the subject of the public presentation of performance data do act strongly. Sometimes defensively, and in a dismissive way, but more often than not (simultaneously) by improving their practices vis-à-vis their peers [3,30,43].

In attempting to do so, some indicators are more useable than others. There are two ways in which it becomes progressively better possible to use indicators for direct comparisons. First of all, the more specific the care process focused on, the easier it is to make the indicators fit for comparison. Overall hospital mortality, medication error or pressure ulcer rates, for example, are much less directly meaningful than condition-specific mortality, medication error or pressure ulcer rates [28]. Second, structure and process indicators are often easier and more unequivocal to monitor than true outcome indicators [21,45]. The former indicators, then, can more easily be used in com-
We do not give up anything, then, by being realistic about the (im)possibilities of external indicators. Ranking may never be fully unequivocal or fair, and full comparability is impossible to achieve. Yet widespread variations in performance, including many practices where performance is below acceptable levels, are real. Performance indicators may not get the exact rankings right, but when properly designed they can point in the right directions. The question we have to ask ourselves, of course, is what the alternative would be. Currently, without any transparency, payment, public image and actual delivered quality are largely unrelated. With performance indicators, especially when intertwined with payment schemes that link payment to performance, we are least ensuring the convergence of these three crucial parameters [31]. Striving for the perfect system is impossible; getting lost in discussions about why external indicators are not yet ‘valid enough’ is mistaking the tool for the goal.

6. Conclusion

In this article, we described how the Dutch Inspectorate has aimed for feasibility in all its meanings: practical, political and philosophical. Doing so, it has found a course between the mutual advantages and disadvantages of internal and external indicators. More specifically, the Dutch Inspectorate has attempted to undo the classic distinction between the two, and so to optimally achieve the goals that the indicator project started out with. In a period of 10 months, a screening instrument was created that would allow the Dutch Inspectorate to assess which hospitals required additional investigation. In addition, while gaining the full cooperation of the Dutch Specialists and Hospital organizations, the project has resulted in an unprecedented insight in the safety and clinical effectiveness of the care and organizational processes in Dutch hospitals. This enhanced transparency is in turn focusing the attention of hospital boards and medical specialists on their ‘relative rankings’ vis-à-vis relevant others. Being ‘insiders’, they themselves know best how to interpret these rankings, and how to account for them. Together with attention from the press, this has already created the ‘burning platform’ that is now providing willing boards and specialists with the alibi to start thorough processes of quality improvement in their hospitals.

It will take another few years to judge whether the aims set out here will all be met. Yet the project has already started an important process of self-investigation, self-comparison, and lowering organizational tolerance for truly suboptimal care. It has been a significant external support to those managers and physicians that wanted to initiate continuous quality improvement programs, and to those arguing the importance of specialists themselves taking the lead in internal indicator development. Over the next few years, we will closely monitor the impact and the further evolution of the hospital indicator set. Since what matters is what is achieved with an indicator set, and not how exactly this tool meets certain external criteria of adequacy, we feel strengthened by the start of the process.16

References


16 The here mentioned care practices all refer to ‘best practices’ with proven improved outcomes.


