

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/314950390>

Ranking Dutch hospitals.

Technical Report · March 2013

DOI: 10.13140/RG.2.2.12743.70561

CITATIONS

0

READS

391

3 authors:



Julia Topp

Weiterbildungszentrum FH St. Gallen

14 PUBLICATIONS 51 CITATIONS

[SEE PROFILE](#)



Iris Wallenburg

Erasmus University Rotterdam

39 PUBLICATIONS 219 CITATIONS

[SEE PROFILE](#)



Roland Bal

Erasmus University Rotterdam

195 PUBLICATIONS 1,836 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



PhD project: Modeling Practices in Science Policy [View project](#)



Patient safety [View project](#)

Erasmus



iBMG Working Paper W2013.02

Institute of Health Policy & Management

The performativity of rankings

On the organizational effects of hospital league tables

Julia Quartz
Iris Wallenburg
Roland Bal

The performativity of rankings

On the organizational effects of hospital league tables

Date of publication

November 2013

Authors

Julia Quartz
Iris Wallenburg
Roland Bal

Contact

Erasmus University Rotterdam
Institute of Health Policy & Management
Tel. +31 10 408 8555
research@bmg.eur.nl
www.bmg.eur.nl

Chapter 1 Introduction

Since the early 2000s, several rankings of Dutch hospitals are published. The most well known are the rankings of *Algemeen Dagblad* and *Elsevier*. These rankings, based on data from performance indicators from the Healthcare Inspectorate, data from the Centre for Client Experiences in Health care and on peer review, attract much media attention within the Dutch hospital sector. Although it has been demonstrated that the validity of these rankings is dubious (Pons, Lingsma, and Bal 2009) —which is a more general finding related to composite measures (Jacobs, Goddard, & Smith, 2005)— there is at least anecdotal evidence that they influence hospital policies to a considerable extent (van der Aar, 2008; Huisman, 2008; Pons, 2009). These paradoxical findings, while in need of further corroboration, call for research into organizational responses to rankings.

This is especially the case as transparency has become a dominant value in health policy, with both the Ministry of Health, its main policy advisers, such as the Council for Health and Healthcare (RVZ), and the Healthcare Inspectorate putting an increasing emphasis on transparency as a mechanism to ensure competition between healthcare providers. Despite much effort put into creating transparency, however, results so far have been disappointing. For example, the national audit office of the Netherlands recently found that the effort of the Ministry of Health in setting performance indicators has not resulted yet in information that could be used by third parties to choose, procure or oversee care on the basis of comparative health information (Rekenkamer, 2013). A study on the performance indicators published by the agency for transparent care (*Zichtbare Zorg*) in the Netherlands found that the validity of indicators is questionable (Kringos et al., 2012), arguing that more work needs to be done to standardize measurement and creating better indicators. More investments in information systems in hospitals would be needed to create better, more useable data. The same study also concluded that, despite the lack of valid comparisons, hospitals had been investing much in the collection of data.

How, to what extent and in what ways the call for and investments in transparency is affecting healthcare organizations remains largely unknown. Within the literature, both performance enhancing effects and ‘perverse effects’, such as gaming and tunnel vision, are recognized (Bevan and Hood 2006). For example, Jerak-Zuiderent and Bal (2011) showed in a qualitative study how the hospital indicator for diabetes worked against integrated diabetes care in one Dutch hospital. Further research on such organizational responses to transparency is urgently needed. Such research would need to focus on what drives healthcare providers in their reactions to transparency, what effects different ranking systems and underlying performance indicators have on healthcare organizations, and how and under what circumstances transparency for improvement might actually work.

Rankings are currently hotly debated in the Dutch health care sector, including the hospital sector. Much is expected from an increasing transparency of the performance of Dutch health care as performance data is expected to enable patient choice and healthcare procurement and to contribute to competition between healthcare providers. We however have very little understanding of organizational responses to rankings. Whereas much energy is devoted to creating ‘better’ indicators and in looking at strategies to enable patient choice, insight in such organizational responses would be necessary to create an understanding if and how rankings might be used to improve the quality of care within organizations.

Although there is now some indication that rankings indeed do influence the performance of public service organizations (with evidence mainly stemming from the educational sector, see e.g. (Koning & van der Wiel, 2010)) we have very little insight on whether this is the case for healthcare organizations as well, and what the possible ‘perverse’ or ‘unintended’ effects of such rankings are on healthcare

organizations and policies. Research on university rankings has for example shown the growing distrust between universities as a result of ranking systems, as well as a process of 'tunnel vision', i.e. focusing on improvements on aspects measured in rankings while neglecting other aspects of quality (Sauder & Espeland, 2009). Others have argued that performance measurement might lead to 'cherry picking' or 'gaming the system' (Bevan & Hood, 2006). With the growing importance of rankings of healthcare providers in the Netherlands (e.g. their current uptake by health insurers), further insight into the organizational effects of rankings is in dire need.

Theoretically, rankings can have an effect both through the 'choice' and the 'reputational' routes. That is: rankings can affect organizations because of customer behaviour or because they affect the status of the organisation, thereby mainly creating an internal incentive to do better. The research on choice thus far has by and large found that patients hardly use comparative information on health providers. Rather, the research that has been done has argued that rankings have an effect on organizational performance mainly through the 'reputational' route (Berwick, 2002; Fung, Lim, Mattke, Damberg, & Shelleke, 2008; Hibbard, Stockard, & Tusler, 2005). That is, even without rankings having an effect on consumer (or payer) choice, they are expected to affect organizational performance. According to Power, rankings as well as other instruments aimed at external accountability and transparency have given rise to a new area in organizational risk management, targeted at 'reputational risk' (Power, 2007). Rankings, and the individual indicators that they are made up of, have in this way become a new object for organizational policies. How this new 'reputation management' impacts on work floor levels is still largely unknown, however, as research has until now mainly focused at the organizational (meso) level. Based on for example Foucauldian understandings of 'discipline' these studies have conceptualized the impact of rankings on work floor levels as one of 'tight coupling', that is, the direct translation of rankings from organizational to work floor levels (e.g. Sauder & Espeland 2009). Given the complexity of health care systems and organizations it is unlikely that such tight couplings will be dominant. In this research project, we therefore also researched the relation between organizational and work floor levels, targeted at a further conceptualization of organizational responses to rankings.

Moreover, much of current research on rankings has assumed that organizational effects of rankings are similar across organizations and affect different parts of organizations in the same way. Given that health care organizations find themselves in different competitive environments, and that this might even differ within the organization, we might however hypothesize that health care organizations can react to rankings in different ways, depending on the competitive environment in which they operate. Health care organizations in more competitive environment are then expected to react to rankings in more direct ways. In those organizations 'tight coupling' might be a more expected strategy, whereas in hospitals within less competitive environments 'loose coupling' would be expected.

Given the current state of research, much work still remains to be done in conceptualizing organizational responses to rankings (Espeland & Stevens, 2008) in order to get a better grip on pathways through which rankings affect health care organizations. One way of getting to such conceptual clarifications is through detailed qualitative research, leading to an analysis that is grounded in the practical experiences of the people working at different levels within healthcare organizations (Strauss & Corbin, 1998). This study builds on such an approach by performing an in depth qualitative analysis of three Dutch hospitals and how they are affected by rankings.

The main goal of this research is to get a better understanding of the possible impacts of rankings on Dutch hospitals and to clarify the pathways through which such impacts can be understood. The grounded analysis of impact will on the one hand lead to a conceptual framework that can be used in subsequent research aimed at quantifying ranking effects, and on the other to comparative research (both between sectors and countries) aimed at understanding system influences on these effects.

Furthermore, the study aims at providing a contribution to the policy debate on rankings in the Netherlands by specifying the types of effects rankings have on Dutch hospitals. Such effects might include changing relations between hospital management and clinical staff, changing relations between hospitals and their wider environments, the growth of 'reputation' as a new management object for hospitals, as well as desired or undesired effects on hospital performance. Given the increasing emphasis on and use of rankings, e.g. by insurers as well as public media, such an input seems highly relevant.

The following research questions will guide the research:

1. What structural and policy responses to rankings have been made at the organizational level in the Dutch hospital sector?
2. What different strategies have hospitals followed in their organizational response to rankings?
3. To what extent and how do these structural and policy responses affect actual care delivery?

Research Design

Not much research has been done as of yet into organizational responses to rankings. The little evidence there is, is moreover focused on the university sector. Michael Power and colleagues have for example shown the great impact of international rankings of business schools on university policies (Power, Scheytt, Soin, & Sahlin, 2009) whereas Espeland and Sauder have come to similar conclusions regarding law schools (Espeland & Sauder, 2007; Sauder & Espeland, 2009). Both studies show that universities, in response to ranking systems, have set up special organizational units and have devised specific policy measures, pointing to 'reputational risk' and to the normalizing and disciplining powers associated with rankings as explanations for organizational change. For the hospital sector, somewhat similar claims have been made for the effects of benchmarks (Triantafyllou, 2007), but this work has been mainly theoretical rather than empirical. Also, studies on the organizational effects of performance indicators point in this direction (Jerak-Zuiderent and Bal 2011).

Moreover, the empirical work that has been done so far mainly focuses on either the role of specific indicators on hospital organizations, including the relations between work floor and higher management levels, or have looked at the board level of organizations only. This study departs from this in two ways. First, by looking at rankings, the study aims for the analysis of so-called composite performance indicators, which combine many underlying specific indicators (e.g. patient experiences, outcome, process and structure indicators) in the hospital sector. Secondly, we aim to study this from a whole-organization perspective, so including work floor levels. Rather than assuming that board-level reactions to rankings discipline the organization and cause 'tight coupling' at work floor levels, the study aims at asking if and how board-level reactions affect lower organizational levels.

The proposed study makes a first step in this analysis by studying three Dutch hospitals in depth in an explorative, qualitative way. Such a qualitative analysis seems warranted given both the lack of current knowledge about the ways in which rankings affect hospital performance and the lack of conceptual frameworks to study this relation. For example, how do rankings relate to the increasingly networked

character of Dutch (hospital) care? How do they influence trust relations between hospitals, as e.g. Espeland and Sauder point at a growing distrust amongst American law schools, who accuse each other of gaming the numbers, arguing that it is always the others that game the system (Espeland and Sauder 2007). Can the concepts of 'tight' versus 'loose coupling' form a basis for describing the effects rankings have on Dutch hospitals and how do these mechanisms vary between hospitals? Is there indeed a growing emphasis in the hospital sector on 'reputation management' as analysed by Michael Power and colleagues regarding business schools, suggesting that rankings may help to carve out a new object of risk management (Hilgartner, 1992)? How does this reputation management get shaped and in what way does it affect clinical processes? In what way, if so, are rankings used for internal quality improvement? Do rankings lead to 'tunnel vision', that is do they make hospitals focus in those aspects of their performance that are used for rankings while neglecting others? Building on the work done by e.g. Power Espeland and Sauder, this study therefore proposes to make a grounded analysis (Strauss and Corbin, 1998) of reactions to rankings in three Dutch hospitals. The results of the study will form the input in making a conceptual model that can be used in follow-up research to study the extent to which found relations do indeed occur or for comparative (cross country or cross sectoral) analyses.

In order to answer our main research questions, we conducted a qualitative, ethnographic study in three Dutch hospitals. Hospital selection was done by looking at similar size hospitals but in different competitive environments, because of the expectation that the level of competition hospitals find themselves will influence the ways rankings affect hospitals, with more competitive regions showing higher levels of tight coupling.¹ In total we approached four hospitals for the study of which one refrained from participating as this hospital was going through an accreditation process that took most of its time. The studied hospitals were similar in size, with 551 (hospital A), 673 (hospital B), and 709 beds respectively but differed as to their geographic locations and especially their competitive environment. Although we did not quantitatively measure competitive environments, based on the amount of hospitals in a range of 30 kilometres, hospital A was in the strongest competitive environment, and hospital C in the least competitive environment. All studied hospitals were in a merger process at the start of the study, or announced a merger during the study period, a theme we will get back to as performance indicators, and especially indicators on volume of care were one of the drivers of the mergers.

Within each of the hospitals, interviews were held with quality managers, communication staff, information department staff, medical specialists, nurses, executive directors and, where possible, members of the Board of trustees. Where possible, external stakeholders like patient organizations, GP-associations, nursing homes, the account holder of the Healthcare inspectorate and the main negotiator for insurers were also interviewed, or meetings with these external stakeholders were observed during the study. The amount of interviews in the hospitals ranged from 14 (hospital A), 19 (hospital B), and 17 (hospital C) and ranged from 30 minutes to 2 hours. Also, we observed relevant meetings (e.g. quality & safety and other committee meetings, meetings with outsider stakeholders such as insurers and the healthcare inspectorate), registration work (e.g. registration of patient data at clinical wards, activities of coders and information and/or communication managers), and, where possible, clinical work (in hospital A, observing clinical work was not permissible). Observation time for each of the hospitals was 12 days (hospital A), 8 days (hospital B), 10 days (hospital C). All interviews were digitally recorded; most of them were transcribed. Observations were written down during breaks or soon after

¹ Note that we did not select hospitals on the basis of performance on the rankings, as these are highly volatile and don't compare between different ranking systems.

meetings and were then written in observation reports. Interviews and observations were done in a period of about three months in each of the hospitals.

The case studies were aimed at getting detailed insights of the internal (including impact on administrative and clinical processes) and external organizational responses and institutional pressures to ranking systems. The cases were set up so to first get a general idea of the functioning of the hospital in terms of quality policies, position of the information, communication and quality departments, use of performance data in relations between management and clinical staff, as well as in its relations with outside stakeholders such as insurers and the healthcare inspectorate through interviewing key actors in the hospitals. In this phase we also read relevant documents such as quality policies and overall strategy documents, the hospital websites, and specific information related to performance indicators and how they were used in the hospitals. Secondly, we focused on administrative processes concerning indicator data, following indicators from work floor registration up till the use of indicator data in internal and external policies. In each hospital we decided to follow two indicators to get a clear picture of this process; with one indicator held constant over all of the hospitals. This last indicator was the indicator on delirium that was introduced recently by the healthcare inspectorate and is also part of the indicators that should be measured for the Safety Management System (*Veiligheidsmanagement Systeem*, or VMS) that each hospital in the Netherlands has to comply with as of 1 January 2013. This indicator was chosen first because it was relatively new, so we would be able to observe the building of administrative and performance systems within each of the hospitals. Second, the delirium indicator cuts across many different disciplines and wards, thus allowing for us to observe negotiations between actors within the hospitals on setting up the administrative and performance systems. In each of the hospitals we also asked questions about volume indicators as these were of particular importance towards external strategies (e.g. mergers, negotiations with insurers) as became clear early on in the study. In each of the hospitals we then discussed our initial finding with a member of the board and with the quality manager related to indicators, in order to feed back our findings and get additional input.

Based on our data collection we wrote detailed, 'thick' descriptions (Geertz, 1973) for each of the hospitals. These included an analysis of:

- changes in the organizational structure of the hospital over the last five years: types of departments, background of employees (e.g. have more communication experts been hired and what are their roles in the organization);
- the way rankings or underlying performance indicators have been used by the hospital in designing internal or external (quality) policies;
- quality projects that have been done by the hospital over the last five years and how they relate to rankings or underlying performance indicators;
- the way that rankings are used in external and internal communication of the hospital;
- the way rankings are used in communications with specific stakeholders (e.g. GPs, insurers, patient organizations; nursing homes);
- the way that rankings or underlying performance indicators are used in discussing quality issues between higher management levels and clinical and nursing staff;
- the way that rankings or underlying performance indicators have been used by clinical departments in designing or implementing (quality) policies;
- development of information infrastructure for quality of care information (e.g. electronic health records) and if and how these have been targeted at collecting data for rankings.

These thick descriptions then formed input for the cross-case analysis presented in this report.

In this study also interviews were held with the main producers of rankings to get an understanding of the ways in which the diverse rankings are made and how they have changed over time. For this part of the study, preparatory work has already been done in the form of a bachelors thesis in which the development of ranking systems in the Netherlands was analyzed (Pons, 2009; Pons, Lingsma and Bal, 2009) and the present study will build on this analysis. Since then, the Elsevier ranking has changed its ranking system, making use of indicators of 'Zichtbare Zorg' (literally: transparent care), because of criticism of its earlier use of a peer review system. Policy actors at e.g. the Healthcare inspectorate, the Ministry of VWS, the NZa and the RVZ will also be interviewed to get a clearer understanding of the ways in which rankings are thought to function in the changing healthcare system. For example, the current focus on volume indicators by some insurers ties closely into policies of both the Ministry of Health and the Healthcare Inspectorate to set volume standards; the difference between rankings has however been seen as a problem as health consumers are thought to get conflicting information (RVZ, 2007) and the use of the basic set of indicators of the Healthcare Inspectorate by the *Algemeen Dagblad* has in the past been seen as a misuse of indicators meant to function as risk assessment tool for the Inspectorate (Berg et al., 2005). Policy reactions to rankings are therefore not straightforward and might influence the ways in which hospitals react on them.

The set-up of the report is as follows. In the next chapter we explore what discourses are used in the hospitals we studied concerning rankings. How important do hospitals and the professionals working in those hospitals regard rankings to be for their work? This chapter will show that whereas on the one hand rankings are thought to be of no interest, they nevertheless are of high value to the practices of the hospital. Chapter 3 will then go on to discuss what happens in the process of data collection and use—what work do hospitals do to collect indicator and ranking data and how is this data used, both internally and externally. This chapter also discusses changing and new positions and functions in the hospital and the role of information technologies. Chapter 4 will then go on to discuss some of the dysfunctional consequences of rankings and the ways in which the hospitals try to deal with those consequences, while chapter 5 discusses the ways in which rankings are affecting the governance of hospitals as well as the ways in which hospitals try to manage their external environments. In chapter 6 we then answer our research questions and discuss the consequences of our findings for policy and further research.

Chapter 2 How important are league tables? On the ambivalences of being ranked

How important is it for the hospitals we researched to perform well on the rankings? Are they actually confronted with the consequences of performing well or not? Does it matter to them if they rise or fall on the rankings in consecutive years? And do they watch how other hospitals are doing on the rankings, and if so, who do they see as their main peers or competitors? These are some of the questions we asked to different people working in the hospitals we studied—executive director's, quality managers, communication managers and professionals—to get a feel for the importance the hospitals placed on the rankings. Our main hypotheses guiding this part of the research were that if rankings were seen to be of consequence to the functioning of the hospital they would also influence how hospitals effort to restructure themselves and the care that they give to perform in better ways; and that hospitals in more competitive environments would watch the rankings with greater care than hospitals in less competitive environments would. We will answer to these questions and highlight how ambivalent the importance of league tables actually is.

Our findings indicate that in the early days of rankings scoring high on the rankings became an issue of importance to all of our case study hospitals. The executive director of hospital C for example recalls that while he initially ignored rankings, he later realized that they became more important after his quality manager realized how important rankings are for hospital image and how helpful they could be to actually improve quality of care, too (executive director, hospital C, 12.11.12). In the early days, also all case study hospitals followed the strategic goal to be one of the top 25 hospitals in the most prominent Dutch league tables.

Recently, the importance of rankings in hospital changes and hospital A for example abandoned its goal to be one of the top 15 hospitals in the country. One reason is that the hospital does not find its' positioning in the ranking to be of much consequence to external actors, such as patients. Also, respondents in hospital B and C feel, the relevance of rankings decreases. A division director in hospital C argues:

I don't do much with rankings, particularly as public reactions towards rankings decrease steadily. Increasingly rankings grow towards each other and become less distinctive, particularly when it comes to quality and safety indicator performance.' According to him 'the question of how media influences on patient choice is much more interesting'. However, he then continues after a short pause, 'I actually do use rankings to check whether something is going astray, whether there are really good or really bad indicators.' If scores are too low he talks to the respective professionals (division director, hospital C, 29.11.12).²

Overall, then, our findings seem to indicate that neither of the hospitals think that rankings are of much importance for clinical work these days. Yet, statements about how rankings become increasingly irrelevant are highly ambivalent. For example, the manager from hospital A who was pointing towards the limited usefulness of rankings continues her argument in stating: "We still want to end high. When we dropped from the top 25 to place 60, that wasn't liked much" (interview quality manager, hospital A, 12.11.12). Rankings, our empirical material shows, generate ambivalent reactions.

We observed meetings with the healthcare inspectorate and with insurance companies. In neither of those conversations the position of the hospital on the league tables was addressed (hospital C, 16.12.12; 16.01.13; hospital A 17.10.12; 24.10.12). However, individual indicators that underlie such rankings are discussed

² All quotes have been translated from the Dutch by the authors.

with the healthcare inspectorate extensively (observation meeting IGZ yearly meeting, hospital C, 16.01.13, hospital A 17.10.12). Indicator performance also increasingly matters in negotiations with insurers, “after indicator-based payment dependent much on the insurer in previous years, and for 90 % it was about finance, quality and indicators only were an afterthought (‘sluitpost’).” (division director, hospital C, 06.12.12). Likewise, one of our respondents from hospital A argued when we asked him whether quality indicators:

“[Our main insurer] each year has a list of our scores on the ZiZo and IGZ indicators, as well as on the quality information they ask themselves. And they feed those back to us. And, well, on some we do good, on others bad and on still others average. For some of the areas where we are doing good we can become a ‘quality hospital’, meaning that you can grow in those areas. For example, last year we could grow in hip and knee replacements. That sounds great, but in practice you cannot steer on this. For one thing, the negotiations with the insurer are usually only finished halfway through the year. And you cannot easily change the referral behaviour of GPs. Moreover, each of the insurers we deal with does it a bit different so you may be allowed to grow in this area from one insurer but not from the other and that really doesn’t help much either (controller, hospital A, 24.11.12).

It also seems that hospitals do not ascribe much value to rankings when it comes to patients and their mechanisms to choose for health care providers. When being asked about whether patients look into the rankings for guidance in choosing a hospital, the communication manager of hospital A responds:

I don’t think that patients use them. Actually, this is probably the group that least uses them. They might look at zorgkaartnederland.nl [a rating site that uses patient feedback and scoring of health providers] but not at the [AD or Elsevier] rankings. (...) We have also asked Motivaction [a marketing consultant] to research for what reasons patient choose our hospital, and then you see that, on top of my head, 75% of patients choose us because their GP told them to come here. Proximity is also very important (communication manager, hospital A, 26.11.12).

The communications manager of hospital C adds to that argument:

Rankings? ... Well, the media is keen on it. And this is exactly what I try to manage. Elsevier, CQ index, benchmarks, patient experiences... Reputation management is extremely relevant. I try to react to how these lists emerge. (...) For example, I report to [cooperation of health care insurers] where they go wrong in their analysis. I also report such failure to the health care insurers. Plus, [my colleague] discusses the future development of patient-related indicators within national forums (marketing and communications manager, hospital C, 06.12.12).

Over and over again, we also encountered professionals that – mostly in interviews – articulated a lack of interest in rankings. For example, the head of the medical staff in hospital B argues in an interview that rankings have no relevance for his work at all. Asking whether these would influence management meetings when scores are bad was negated.

On the first sight, then, it appears as if patients, payers, regulators and hospitals don't consider rankings to be of much importance. There are many reasons why different actors in the hospital are critical about rankings and their influence on the hospital, on

which we will further elaborate below. These narratives of insignificance of rankings are, however, only one dimension of a more complex process of interaction between rankings and care processes. To say that rankings do not matter at all is not true. While often invisible in interviews or formal reports managers and medical communities alike discuss rankings results in team meetings, in personal communications, etc. An excerpt from management meeting notes in hospital B is insightful:

Excerpt from notes from a management team meeting on indicators and rankings, hospital B, 17.01.13

[Medical specialist] notes (...) that rankings are not the only measure for quality; quality can still be good internally. [Executive director] relies that indicators do increasingly converge. The goal is not to be at the top, but it is a positive side effect. It is generally known that the 'lists' are sometimes just filled in such a way that it turns out positive; what matters is your integrity in this. The unit management had to fill in the scores. Indicators are important, but the care process is the most important and in the end that pays off in the rankings. You cannot keep hiding behind the remark that your care is good and the ranking is no use. The end goal is not accountability, but improving your processes.

In the same hospital, we ask a care group manager who indicated in a previous interview that rankings had no meaning for his work about the relevance of rankings for his work a second time. This time he answers:

Well, if we score badly in [the most prominent national rankings] ...that is absolutely no topic of discussion, neither in the policy committee, the meeting of the medical managers and the care group managers, nor amongst medical staff meetings. In the tri-lateral meetings between medical managers, ward managers and care group managers. Thus, you cannot say it matters ... well it might pass by and we laugh about it, but then we return to important business (care group manager, hospital B, 02.05.13).

Thus, our research indicates that rankings actually do matter in hospitals, and that performance in rankings is part of managerial agendas. However, as the quote of the care group manager in hospital B indicates, such (often informal) conversations are usually accompanied by elaborate argumentations as to why rankings are problematic. Observations from hospital A are exemplary in this respect.

The day after the results of the AD league table is published, rankings were discussed at a meeting of the steering committee on indicators, and our field notes indicate how criticism we have encountered in all our case study hospitals takes off: 'The rankings are meaningless'; 'The criteria are opaque and change all the time'; 'The one year you are in the top 10 and the next year you are way below'; and 'Patients do not understand this, they don't know where they have to be.' The committee chair refers to a specific hospital: 'They were as dead, and they are now in the top 3', indicating this to be impossible if the rankings were really displaying quality of care (observation, hospital A, 12.11.12).

Overall, rankings are usually downplayed in their relevance to improve quality and safety of care. Yet, despite the general critical attitude towards rankings, these also matter to the hospitals. Rankings give feedback on the way the hospital is performing relatively to others, and professionals can be sensitized through rankings - especially

if respective results relate to their professional status. We will therefore firstly elaborate on the main problematic issues respondents see with regard to rankings before we delve into empirical evidence that demonstrates the relevance of rankings thereafter.

The difficulties with rankings

In all hospitals, the AD and Elsevier rankings are studied—sometimes quite intensely by the information and communication managers as in hospital A and C. They are mainly used as comparative tool to see how the hospital was doing on specific underlying indicators vis-à-vis other hospitals, often using colour codes to express where the hospital did well, average or poor. Results are usually also spread and discussed in different meetings, including those of the board and the board of trustees of the hospital. Thus, all three the hospitals try to learn from rankings.

Yet, for various reasons, respondents feel that performance measurement through rankings and performance of primary care processes don't always match. We will elaborate on this criticism in the following paragraphs. We start with criticism on the *design of rankings*, and focus on the *unpredictable composition* of rankings across years, the *questionable validity of particular indicators*, and *data collection methods* to be precise. We begin with the latter, and an observation from hospital C the day after the Elsevier ranking is published in this respect:

Today, the 2012 Elsevier ranking is out. The case study hospital scores somewhere in the middle range. I walk through the 'management corridor' and see how the division director's secretary copies the list results. I asked the quality manager what he does with the ranking. "Elsevier? ...Not so much..." He explains that the division director went to the pre-publication congress, which allowed him to get some hunch of the results ahead of publication. Apart from that, the quality manager argues, one congratulates ones staff if the result is ok. He feels the Elsevier list is more of an "opinion research". In contrast, he would take results from the AD 100 list much more seriously. If results are bad here, he increases pressure internally, particularly with doctors (observation, hospital C, 01.11.12).

Comparable evidence is collected from hospital A. We asked a member of the quality staff of hospital A which of the rankings she acknowledges most. She answers that the AD list was taken more seriously because 'it is older' and 'Elsevier takes information from all kinds of places: waiting times from internet, IGZ, ZIZO...' (observation, hospital A, 29.9.12) The marketing manager of hospital B adds to that point and argues that particularly the Elsevier ranking was perceived as less trusted particularly due to its peer-review methods, where fellow professionals would evaluate other hospitals in previous years (communications manager, hospital B, 21.05.13). Overall the AD ranking, which exclusively draws on data from the health care inspectorate, is trusted more amongst professionals and managers (also see Pons, Lingsma & Bal, 2008). Rankings are thus attributed with different degrees of credibility, depending on how trustworthy the data collection process is considered. Beyond that, also individual indicators and their ability to evaluate performance in hospitals are questioned. The head of the quality committee in hospital B stars the interview with saying:

Respondent: Well, I find indicators and the sense they make a difficult topic.

Interviewer: *Why?*

Respondent: Well, because if I bring a new medicine to the market or a new treatment method, then I have to show that it helps. And that did never happen in the case of indicators. That is zero ...there is zero evidence that it

actually helps. It is not ... proven that it helps. (...) I was a real opponent of indicators in care, because I didn't see how one could measure care with numbers. That is difficult. You can also not say that someone is this and that healthy and this and that sick with such a score. You can do something with such scores, but that is then one score in a whole clinical picture (head quality committee, hospital B, 28.05.2013).

The microbiologist questions the scientific evidence underlying particular indicators. He questions the validity of particular indicators to measure performance in health care. The empirical analysis now shows that it is not always the rankings themselves that are central to discussions, but the indicators that underlie such rankings.

With regard to design of rankings, respondents from all hospitals argue that rankings are sometimes generating surprising positions in rankings. For example, one of our case study hospitals was ranked as best Dutch hospital in one year and fell down approximately 50 positions in the next year. A respondent from hospital A explains:

'The tricky issue is that you don't know where they [rankings] pay attention to. This is why you score differently every year. And sometimes that paints a quite faulty picture' (observation, hospital A, 25. 09.12).

The marketing manager of hospital C made an analysis of all rankings with regard to information about patient satisfaction (CQ, Elsevier, AD) and tried to put these next to local evaluation tools (complaints, mirror conversations, contract with society, patient satisfaction survey). But an analysis of all rankings across years did not generate coherent results - neither amongst the dissimilar rankings and rankings within one ranking across years, nor in relation to local evaluation insights (marketing manager, hospital C, 06.12.12).

Respondents are critical of the rankings themselves, arguing that they are 'lotteries' that are unpredictable, if only because the indicators that are used in creating league tables change every year, are different between ranking systems or are weighed differently each time. Respondents criticize the relevance of rankings 1) by questioning the scientific evidence that legitimizes the use of a particular indicator, 2) the overall relevance of particular indicators to measure performance in the primary care process, and 3) the volatile and thus unpredictable nature of ranking designs. Earlier research has already shown that both the comparison of rankings over time as well as the comparison between ranking systems lacks coherence: hospitals tend to jojo through consecutive versions of rankings and scoring well on the one rankings is no predictor at all for scores on the other (Giard, 2005; Pons, Lingsma, & Bal, 2009). Rankings thus are considered as unpredictable tools to actually measure performance in hospitals, as we will further elaborate next.

A care manager in hospital B argues that the hospital 'has to perform a constant balancing act', where one aims to end high in the rankings (best 25%) while one feels that steering on basis of indicators is a difficult instrument to actually improve quality. He refers to how rankings can be used to actually *improve performance*. We focus on the issue of *multiplicity* and the *limited 'newness'* of rankings-based governance information here. We start with the former.

The patient service office manager in hospital C recalls that in the beginning the Elsevier list was important as it 'generated shock amongst professionals about how much performance can be improved.' However, she argues, rankings also generate a 'multitude of indicators', to which it is 'difficult to relate to over the years.' For example, with regard to patient experience, AD,

Elsevier, CQ, IGZ and health care insurers all demand slightly dissimilar information. She argues that this is consequential for her work, as she tries to react to the external demands, and tries to take up changing indicator sets into the local patient experience survey in order to be able to report and improve the respective indicators. She argues that she therefore 'keeps on searching' what exactly she has to include in the local patient survey. Due to the multiplicity and volatility of external performance demands, she argues, this search is increasingly difficult. Meanwhile, she feels that one almost exclusively uses the lists to 'generate a sense of urgency,' while actively steering on quality improvement is difficult. (head patient service office, hospital C, 06.12.12)

Like the manager of the patient service office, respondents repeatedly argued that the multiplicity of information generates incoherence more than insight into quality improvement work. For example, the marketing manager in hospital C argues:

The patient experience survey helps to steer wards. The insurers send a consumer quality index with some 200 questions in addition. Then another 100 questions from the insurers, and I do also handle 60 questions locally...and this then is only about patient experience. We have to stop this. We have to search for what is relevant. (...) It is about time that we get a clear set of indicators. I also can't sell the many slightly different measurements anymore (marketing manager, hospital C, 06.12.12).

Increasingly, rankings draw on comparable yet slightly dissimilar indicators. This means that hospitals have to deal with a multiplicity of (often only slightly different) indicators that are collected for dissimilar agents. Hospitals are only to a limited degree able to synchronize such multiplicity of information and hence are only to a limited degree able to steer performance improvement processes on basis of such indicators.

Simultaneously, rankings generate only limited amounts of new information that would help to steer performance in hospitals. Again, observations in our case study hospitals are insightful again:

Today is 'the day after' the publication of the AD league table. The hospital has gone up in the ranking considerably. (...) The quality manager looks [results up and checks how] hospital has done in comparison to surrounding hospitals and what criteria were used to make the ranking. There are no great surprises: the hospital 'is not performing well on these aspects', especially pressure ulcers and readmissions. 'We already knew that' (observation, hospital A, 01.10.12).

In a comparable fashion a respondent in hospital C remarks when being asked about the impact of the most recent AD league table: 'We know since year and day on which indicators we score red [colour coding system indicating bad performance in hospital C]. And those remain red. Our reports are always about data collection, but seldom about improvement' (senior quality staff, hospital C, 12.11.12).

Respondents from all case study hospitals remark that while results in rankings might be surprising due to the particular composition of the indicators selected for the respective ranking, performance in underlying indicators is actually not surprising at all. Hospitals, and particularly those respondents who do compliance management and quality control, are usually well aware of which indicators score well or badly

locally. Thus rankings do not generate new information and rather reiterate indicator-based performance information.

Knowing which indicators are badly scored upon does not yet help to improve primary care processes either:

We know that we don't score well on particular indicators. But what exactly is that? What does that contain? What exactly do we not do well? And what could we do instead? And do we want to do that? Not only management but also professionals... If we don't know what we can improve, then nothing will happen (head patient service office, hospital C, 23.10.12).

Likewise executive director of the hospital B argues: 'By looking at indicators alone, we will not get better. We have to understand what is actually happening' (executive director, hospital B, 02.05.13).

Respondents argue that quantitative, indicator-based performance information is not always an appropriate means to steer improvement processes that happen in complex primary care scenarios, as the executive director highlights. Next, we therefore focus on the relationship of complexity of health care and indicator governance and elaborate why respondents feel that indicators (and thus rankings) are reductionist tools to represent organizational realities that are messy and complex. We start with the *level of abstraction* that rankings generate before we elaborate on the difficult question of how *complex health care practices can be measurable*. The executive director of hospital B argues:

It is a really high degree of abstraction that underlies [such rankings] It is about 'Is it a good hospital', because the picture that rankings paint is always on the level of a whole hospital. Mostly, it deals with whole hospitals, which is nonsense, because something like 'the hospital' does rarely exist. It might well be that the department for neurology is doing well and the department for dermatology is mediocre, that can happen. The external world pays attention to [rankings] and sort of gives a all over certificate in saying something like 'It is a good hospital.' This is what matters for the external world. Drawing a picture, image, all these sort of things (executive director, hospital B, 02.05.13).

The executive director's criticism is about the generalizability of ranking results. He argues that all performance measurement relates to smaller entities, such as particular units or individual professionals. Respondents argue that what is often perceived as a good or a bad hospital in rankings does actually not exist. Yet, rankings tend to impute universality when positioning hospitals in a hospital-wide score table.

Criticism moreover relates to how rankings interact with practices in health care that are more complex than performance indicators. A microbiologist in hospital B argues:

Indicator thinking ... makes that you look for measurable things. The biggest mistakes in health care, however...are made in the realm of non-measurable things. The biggest mistakes in healthcare are, if you ask me, always individuals that make diagnostic mistakes. Like: 'You have pain in your knee, I put in an implant.' While the correct question must be 'You have pain in your knee, are you able to walk on it 5 more years? Then I will wait with the implant.' It is much more difficult to research what goes wrong on this terrain for the Inspectorate. But whether the washing runs on machine 129 or 130

degree heat doesn't matter. Bacteria are already dead 10.000 times. That is not interesting, but you can very well link numbers to such questions. With the indicator-based thinking you look for such things ...and you will say...the machine is not well regulated and you have to take care of this.

Interviewer: I get your point, what is the consequence?

Respondent: The consequence is that you fool yourself, about what you are looking at and where you put your time and energy into (interview head lab & head quality committee, hospital B, 28.05.13).

The respondent explains that the very problem of indicator-based performance measurement is its focus on measurable agendas. Health care, however, according to our respondent is only to a limited degree measurable. In consequence, he argues, indicator-based governance would always focus on measurable agendas while the most difficult agendas with regard to performance in the domain of quality and safety are related to tacit expertise of professionals, such as diagnostic interpretations. Thus, according to our respondent, complexity of medical practice does not fit indicator-based health care governance. The argument of our respondent is a representative criticism on rankings that above all medical professionals have expressed in interviews.

In sum, respondents feel that rankings and performance of primary care processes don't always match. Rankings are mainly criticized for faulty design, the inability to actually improve performance, and the difficult relationship with the complex nature of health care. This criticism, if considered exclusively, would offer sufficient explanatory power to our initial observation that rankings 'don't do much' (quality manager, hospital C) in organizations. We will therefore now move to the benefits that respondents ascribe to rankings and thereafter elaborate on how we can understand the ambivalence that rankings generate in terms of *pragmatic compliance* in our intermediary conclusion.

The benefits of rankings

As mentioned above, if being asked about the relevance of rankings, managers and professionals alike would, accompanied by a variety of reasons, argue that rankings have only limited impact on hospital organization and work in the primary processes. However, particularly managers and quality staff – if being asked – come up with beneficial aspects of rankings. A conversation with the executive director of hospital C, where we asked him how he evaluates the relevance of rankings for quality improvement, is exemplary in this regard:

He argues that he tries to use the indicators that underlie rankings for internal quality management. Yet, for him the local quarterly report that also builds on external performance indicators was a much better steering instrument than rankings, because it allowed for short-cyclical feedback. He then reasons that particularly the AD list is a good means to compare the hospital to other Dutch hospitals. He pauses and continues: 'Rankings are quite important when it comes to the medical staff. They consider the AD list important, and are quite irritated by low rankings in that list. Therefore, I do use the AD list to stimulate professionals. I use it to start conversation with doctors' (executive director, hospital C, 12.11.12).

For the executive director, rankings indirectly have the capacity to stimulate performance improvement, as they *enable comparison* with other hospitals and serve a means for *negotiations with professionals*. In the following we will elaborate on benefits of rankings in more detail, starting with the latter.

As we had argued above, when being asked about the impact of rankings, medical professionals tend to negate that rankings have any impact on their work. Mostly such narratives are accompanied by a variety of arguments about the validity of both individual indicators and collection methods or the argument that rankings purely relate to image and reputation of hospitals, and thus not to medical work but rather to managerial practice. However our field notes demonstrate that rankings actually are important means when it comes to professional group performance. A quality manager from hospital C argues:

Rankings are surprisingly relevant. They give more guidance than we thought they could years ago. (...) Professionals can be struck by particular results of a ranking, the AD in particular. This puts you on your place in comparison to colleagues. And this has directly to do with professional esteem. It generates energy to improve. (...) The AD list makes everyone react. During the year, indicator collection is much work. (...) Yet, no one gets too excited about it. Then, Sunday night, when the AD list comes out, the executive director calls and asks about particular things, explanations, about what has to improve. Doctors start discussing results on Monday then, the medical executives discuss it on Tuesday, put it on their agenda. Then blood streams through the organization... (quality manager, hospital C, 07.12.12).

Interviews with medical professionals, as for example performed in hospital B and C, generate comparable results. Professionals discuss rankings results informally in their team meetings. For example the head of the medical staff from hospital B argues that rankings are, formally speaking, not relevant to his work. Informally speaking however, they impact as results are discussed in team meetings and the policy committee of the hospital. In such meetings not rankings overall but the specific indicators that have relevance to respective specialization are discussed. Also, while rankings are by professionals largely conceptualized as a managerial practice and thus irrelevant for medical practice, the relevance of indicators (and particularly those developed by medical communities) is recognized (head medical staff, hospital B, 02.05.13).

Thus, professionals consider rankings important when it comes to medical group reputation. Managers again are aware of the effect that above all the AD ranking has on a group's standing and strategically use rankings as a means to negotiate performance with professionals. Rankings then are a strategic means for executive managers to shape performance improvement within organizations.

A second aspect in favour of rankings that respondents mention is of strategic kind, too. A conversation with the executive director of hospital B, which is located in the highly-competitive area, is insightful here:

Transparency does everything to competition and reputation. [Our hospital] is quite vulnerable in this respect. We have to fight against two academic centres which patients have big expectations about. A taxi driver said recently that he would go to [neighbour hospital] for small things, but for really important matters, he would go to [the academic centres]. [Another hospital in the region] had something royal to it, as Beatrix is the patron and the supervisory boards were crowded with high ranks. [Yet another hospital] has a classic touch and is quite well-known, [and another one] is quite busy with marketing. [Our hospital] is often overlooked in this context. External marketing is increasingly relevant (executive director hospital A, 22.06.12).

Likewise, hospital C, which is located in the least competitive environment does generate a comparable policy and the quality manager recalls that his

Our first reaction to rankings was 'We don't work for the indicator.' This was not tenable after all. It is unavoidable that when you are throughout the year evaluated on basis of percentages then you have to steer on basis of that (quality manager, hospital C, 07.12.12.).

Thus, the ambition to react to rankings seems to be independent of regional settings, and respondents from all hospitals argue that executive directors set the aim not to 'end up' in the lower parts of rankings (communication officer, hospital A, 26.11.12). Benchmarking information and mirror information is considered an important steering instrument for the executive director in the highly competitive area - particularly with indicators where hospitals perform badly as compared to others (executive director, hospital A, 06.11.11).

However, competitive advantage due to particular outcomes in rankings is clearly most prominent in hospital A, which

is located in such a [competitive] environment, where one cannot say 'again a ranking from an insurance company...I won't participate.' Then you are one of the only ones – apart from if you agree with each other not to do it – but who stops (information manager, hospital A, 18.09.12).

Competition is less relevant in hospital B and C, even though both hospitals are currently undergoing merger processes with (regional) neighbour hospitals. Respondents in hospital B even argue that patient flows into the hospitals are still stable, and that therefore rankings have of course to do with reputation but less so with patients choosing for other hospitals instead. Reputation management, however, is an important aspect for all respondents in all regions, as we will further elaborate in chapter 3.

Intermediary conclusion: dealing with ambivalence

Our empirical analysis highlights that rankings are widely criticized by professionals and managers alike. Some of these criticisms are related to the design of rankings, others to the claim that rankings do not generate relevant performance information. Again others stress the mismatch between organizational complexity and linearity of indicator-based measurements. Criticism of rankings, then, to its largest extent is about the 'architecture' of rankings and the difficult relationship with organizational complexity.

However, hospitals don't ignore rankings. Quite to the contrary, rankings are taken very serious as they influence the increasingly important reputation of the hospital. Rankings, in the first place seem to be ambivalent creatures then. Why paying attention to rankings while heavily criticizing them? Our analysis approaches this ambivalence in demonstrating that these narratives of insignificance of rankings are only the front stage to a more complex process of interaction between rankings and care processes. To say that rankings did not matter at all is not true. At some moments we found that they did very much matter. Every front stage narrative has a backstage, too (Goffman, 1990). At this backstage, we encountered professionals that are wary about their group reputation and managers who use rankings to negotiate performance indicators (and other quality tools) with professionals. This glance at the backstage, then, offers a more nuanced conclusion: rankings *do* matter. They serve as strategic means to position a hospital in its respective (competitive) environment and enable managers to negotiate and shape performance

improvement agendas with professionals internally, while respondents perceive rankings as only limitedly able to steer performance improvement.

That leaves us with the question of how hospitals then deal with the multiplicity of demands, the limited steering information that comes from rankings, and the design problems that respondents perceive to be problematic? Do they simply ignore rankings? Or do hospitals select for particular rankings in which they would want to score well? In short, how do hospitals deal with the ambivalence that rankings generate?

The argument of the head of the quality committee in hospital B, who himself is trained as medical doctor and who heads the laboratory of the hospital, is insightful in this respect. He argues that indicators and rankings are only useful for actual performance management in his lab to a very limited extent. We ask him whether he then pays attention to indicators (and even rankings) in his lab. He sighs and answers:

What we indeed do...No, the word is as it is, I would say. I do have a laboratory here and once in two years I am visited by the [accreditation organ]. If I don't perform well, my certificate will be invalid, my accreditation will be invalid. If I follow their processes, I get my certificate. Even if I find matters ... lets say not really important for the quality of the laboratory? Then I resolve them nevertheless, because I want the certificate and I know that if I try to fight against what they find relevant the bureaucratic way ... in 99 of 100 cases I am quicker in resolving the issue (head laboratory, hospital B, 23.05.13).

The head of the laboratory thus complies with demands of external parties even he does not consider all demands useful for improvement of quality and safety in his lab. And he complies so for a simple reason: not complying with indicators would cost him more effort than simply fixing what he sometimes considers as irrelevant for good lab practice. Likewise, the reactions of the executive directors show how the ambivalence towards rankings and indicators and the way of dealing with it: pragmatic compliance. Pragmatic compliance is central to understand how hospitals currently deal with the ambivalence that rankings generate.

Chapter 3 Co-constructing rankings and hospital organizations

In the previous chapter we have seen how rankings call for ambivalent reactions in hospital organizations. On the one hand, rankings are said to be of little practical consequence to the hospital and its workers. Rankings, as is argued, do hardly affect external relations of the hospital and lack validity to be of consequence for hospital policies. Criticizing rankings is a standard rhetoric we encountered in interviews and meetings. On the other, however, rankings are seen to be highly consequential, affecting the reputation of the hospital and individual professionals and managers. Below we will for example highlight how rankings change the ways in which quality management is organized and the communication between managers and professionals. In a way, this ambivalence can be explained by analyzing this in terms of front- and backstage discursive processes: in the front stage, rankings are being criticized because of their lacking validity, but in the back stage, they nevertheless are seen to influence what happens in the hospital.

In the conclusion of our previous chapter we have already indicated that hospitals deal with the ambivalence that rankings generate by help of pragmatic compliance. In this chapter we leave the field of discourse and deepen our look at the practice in the hospitals. What is actually done in the different settings of the hospitals to organize for rankings? We decided to focus on four specific areas: the organization of administrative processes, changes in the structure of the hospital organization, including the positioning of quality and communication departments, the use of performance review and the organisation of learning processes. Below we analyse our data on these four fields respectively.

Investments in form: administrative work

All hospitals we studied went through great changes in terms of the organization of administrative processes. Collecting information for indicators is work. Information has to be collected for hundreds of indicators, stemming from insurers, patient and professional organizations, the healthcare inspectorate, *zichtbare zorg* (ZiZo, literally 'transparent care'), and the safety management system. What is euphemistically called an '*uitvraag*' (an information demand by an external party) actually entails sometimes many months of work for the quality and information departments in collecting information from different sources in the hospital, let alone the amount of work that goes into administrating that information in the first place. Administrative work is collective work and is spread across the whole hospital organization: from clinical departments to the central organization of the hospital, including the board; from peripheral, largely invisible work done by medical secretaries and coders, to highly visible work done at the communications department. It is also heterogeneous work in that it comprises activities of both human actors and machines: information and communication technologies abound and all hospitals had high hopes that ICT systems would ease the pain of the administrative work—and as a consequence invested huge amounts of money in building electronic patient records and data warehouses.

One way of describing the administrative work done is to just follow the information collection from the bottom up. Data collection starts with professionals registering information on their patients. This is no easy task as all kinds of information have to be collected. Apart from information guiding the treatment process, health professionals have to collect data on all kinds of scores necessary for the collecting of data for performance indicators. Nurses for example have to do risk assessments for pressure ulcers, delirium and malnourishment, and have to regularly check whether their patients are in pain. But not only do they have to collect the information, they must also make sure the information is registered in their patient systems—

usually some form of electronic health record. Professionals are notoriously bad at registration work and incomplete records have been an issue for a very long time. When for example Marc Berg and co-workers in the mid 1990s observed health professionals working with electronic health records they noted that registration was often seen as a nuisance as it distracted attention from the care process—the primary concern for health professionals (Berg, Goorman, Harterink, & Plass, 1998). As a consequence, patient records were often incomplete.

The hospitals we studied all had installed different methods to make sure registration of care was actually done. These included building indicators in the electronic patient record, disciplining professionals by publishing information on registration, and ‘policing’ professionals to make sure registration was actually done. Hospital A and C went through quite some changes in making registration work, and hospital B was in the process of focusing more on registration, as a reaction to unexpected low positions in the rankings. An observation in hospital A helps in illustrating how EPRs are used to make registration happen.

The nurse shows me the electronic patient record (EPR), pointing at the activities that need to be done with the patient today: blood sugar measurement in the morning and the afternoon, general controls every three hours (tension, temperature, saturation, pulse), delirium score (twice), and pain score (also twice). Measurement for the risk of falls and pressure ulcers should be done twice a week, at pre-defined days. The nurse then clicks to the ‘order list’, made by the nurses themselves, that contains the same measurement tasks but now ordered in the day as a sort of checklist. If an activity is over time it will turn yellow and the screen will flash ‘over time’ (observation hospital A, 23.10.12).

Electronic patient records are a much used way of making sure registration work gets done; they not only make it easier to do the registration, they also enable ordering registration work during the day and can also have built in alerts to show that registrations are due—or over time. In that way, EPRs order the day for care work, in this case especially for nurses. All hospitals also found however, that electronic systems cannot do this work on their own. As a consequence, additional interventions were necessary

Some respondents argued that existing problems with electronic patient filing and indicator collection could eventually be resolved at one point in time when a coherent ICT system for the whole hospital would exist (quality staff, hospital B, 25.05.13). Such a comprehensive system would allow hospitals to obtain information that is relevant for rankings and indicator collection at large by ‘just by one click.’ However all hospitals we studied have several ICT systems in place - like planning systems of the operating theatre, complication registration systems, lab systems, pathology systems, etcetera. Also other sources, like DRG registrations that doctors fill in for reimbursement purposes or other types of data sources like clinical registries are important sources for rankings. Thus, it seems overly optimistic to effectively construct ICT systems that offer performance indicator data ‘by just one click’. As a date warehouse administrator in hospital C puts it:

‘Often, the same indicator is registered a number of times with dissimilar results. (...) In consequence, this generates 4 times the same data; and often this data differs depending on the file. There is too much of the same registration. We work with too many sub- systems.’ He goes on to explain that it would be almost impossible to merge all data in one particular ICT system, particularly as different professional groups use dissimilar ICT systems with different demands and also with different needs with regard to data secrecy.

Therefore, he manually merges all sources (data warehouse administrator, hospital C, 19.12.12).

ICT systems cater to different demands of different professional groups, and a full convergence of ICT systems in one source appears to be difficult at the current time. Another of the problems with EPRs is that the hospital is dependent on the supplier of the information system to build in the indicators in the system. This takes time and money, and suppliers may have other priorities than changing the information systems. Because of the high costs of EPRs, most hospitals have moreover formed consortia with other hospitals to buy and implement them. As a consequence, they are also dependent on the other hospitals to see that changes are made, and the hospitals may have different priorities again. At the same time, indicators are volatile and unpredictable as we had shown in the previous chapter. So it may happen that when they are finally built into the system, the indicator has changed. Consequently, only relatively stable indicators or indicators that the hospital wants to collect anyway are therefore built into the system. As a quality manager in hospital A expressed:

The starting point is to keep the 'registration burden' with doctors as low as possible. As much as possible, we want to get the information out of electronic systems. But, with building in indicators into the systems, you count on the stability of indicator sets and practice turns out to be different. Especially the IGZ indicators change regularly. This makes it hard to keep the registration burden low. And that in turn makes it hard to keep the clinical department on board (quality manager, hospital A, 04.09.12).

Similar obstacles were found in the other hospitals. Parts of the performance data in hospital C is collected in the centrally accessible data warehouse Business Intelligence (BI), which is an ICT application that can combine various indicator-related data from different data sources. It also allows for the monitoring of indicators on short-cyclical modes (monthly), and generates visually appealing graphs. This enables short-cyclical monitoring of key performance tasks. The system can be individually adjusted. The use of this system, however, is limited as it demands a corresponding ICT system that functions on the basis of structured and formatted fields. Therefore, so far it accommodates approximately half of the hospital's indicators; i.e. those which are stable across time (e.g. postoperative wound infections, or POWI), which are prestigious (e.g. access time) and which are crucial for the functioning of the hospital (e.g. some of the IGZ indicators). Depending on account settings, managers can access the system and use it for monitoring. Yet, care group managers don't regularly use BI information to steer towards QI on the wards they are responsible for (care group manager, hospital C, 07.12.12). A similar real-time monitoring system is present in hospital B, yet this system disposes of only a fraction of indicators, namely 6 care-related indicators such as pressure ulcers, that however are checked by care group managers on a regular basis (care group manager, hospital B, 28.05.13).

What's worse, EPRs can even make it more difficult to get registrations done. In a meeting of the HSMR committee in hospital A for example, mention was made that many discharge letters were missing. This had been a repeated observation and the chair of the committee sighed that he had asked the hospital board many times to deal with the issue, but it still wasn't in order. One of the medical specialists present mentioned that since the EPR was introduced controlling the process of making discharge letters: "now you don't see over filled pigeon holes any more that make you say to the assistants 'you should dictate your letters'. We have to come up with a solution for that." (observation HSMR meeting, hospital A, 17.09.12). Paper files may have disadvantages, they sometimes also have some things to say for them; a stack of paper gives information, only by being a stack, and can point at patients having

long histories in the hospital (Berg, Goorman, Harterink, & Plass, 1998) or residents not doing their registration work in time.

Information technologies thus definitely have advantages to get registrations done—and especially to collect data throughout the hospital, although as we shall see later on what works still needs to be done there—but have to be ‘made to work’. For this reason, the hospitals have developed other types of interventions to stimulate registration. Hospital A for example started using a benchmark for the registrations in nursing care and used this to compare performance on ratings between wards. The benchmark was an initiative from the hospital board after it was discovered that the hospital was doing comparatively badly on the nursing indicators in the *AD* ranking. The hospital felt that it could do better by placing registration more up front.

What we saw a couple of years ago is that we scored badly on the indicators in general and the rankings in particular, on the indicators that deal with nursing. (...) We then decided to focus on registrations first and look at outcomes later, and meanwhile, within a year and a half, we have gone up to 80% of registration of the indicators. 100% is unattainable. (...) We have made registration really visible, on the intranet, every month again, to compare clinical units and have made a competition out of it. So nurses became aware that ‘yes, we really have to make visible what we are doing’ (board member, hospital A, 06.11.12).

On all wards in the hospital the registration scores on the nursing indicators would be displayed on the walls of coffee or other meeting rooms. Such visual displays of performance can be found in similar rooms in all hospitals—only what they focus on might differ. Registration however is a topic in all hospitals. And visualizing performance is seen to be a way to stimulate it and often, like in hospital A, coupled to some sort of competition; in hospital A the ward that this month did best on registration would get cake on Monday morning and have their photo on the intranet with the hospital board member bringing in the cake. So despite the rhetoric of disdain for rankings, hospitals in fact use them themselves in order to stimulate—in this case—registration.

Even then, competition and visualization are not enough in themselves. Nurses for example can—and do—ignore them and we would hear nurses say that they actually never looked at the numbers (but many others did). So, in addition to visualization, other interventions were used to get nurses to do the registration work. Team leaders in hospital A and C for example defined themselves as the policemen of indicators. They are the ones that have to assure the nurses actually do the registration. In hospital C, such responsibilities are clearly set in an organizational chart indicating which team leader is responsible for which indicator. Indicator compliance without control is often felt like something nurses cannot relate to; there is no conversation about indicators.

In hospital A, team leaders employ different methods to get the nurses to register. On their patient lists on the EPR the team leaders can see which scores have been registered for which patients. “We might say in the morning that registration has been low and that in these and these rooms patients have to be registered; if it is not done by the end of the day we make the nurses do it still, without overtime.” Not all patients have to be registered though: “that patient is here for a sleep test; he doesn’t have to be measured or delirium” (observation, 23 October 2012). This indicates that there might still be some leeway in registration work—but also that not all care work fits registration work, a point we come back to below.

Apart from team leaders, also board member pay attention to nurses’ registration efforts and act as policing agents. During ‘ward rounds’ for example in

hospital A and C nursing registration is checked and discussed. An observation from hospital A:

When I do a safety round, like last week at [name ward], well, I always carry this [showing the nursing registration benchmark] and they often scare. And I would ask them 'why aren't you doing pain?' And there might be good reasons why they are not doing it, like at paediatrics we can't measure pain well with young children, or at the intensive care unit, we can't measure pain when patients are sedated. And I would ask them, well, how they think about how to solve this, because we really want to have the pain scores three times as day (executive director hospital A, 06.11.12).

Policing nurses, either by team leaders or board members in a ward round, in combination with the building of the EPR and the visualization of performance, is all part of an effort of getting registrations done.

For some indicators, however, additional measures are needed. This is particularly true for new indicators, where nurses still have to learn how to do registrations. One example is the delirium indicator, measuring the risk for delirium with patients older than 70 years. All hospitals have specialized nurses (so-called 'aandachtshouders') looking into delirium and trying to get more attention for delirium throughout the hospital. Also, sometimes other care workers, like dieticians can fulfil this role (in this case, related to the malnourishment indicator). Hospital A for example also had a training program for malnourishment after it scored badly on the malnourishment indicator in the *Elsevier* ranking.

The hospital has many malnourished patients and we pay a lot of attention to them. But what was important for the *Elsevier* ranking was the evaluation at the 4th day in the hospital, which did not happen much due to a lack of staff. The *Elsevier* ranking was a trigger to start doing this. The dieticians needed extra staff for this, but the hospital could not afford that at the time, so we decided to train some of the food assistants for the measurements. This also worked for the good food assistants, to give them more responsibility (quality manager, hospital A, 04.09.12).

Administrative work needs interventions in the care process itself. Administration, it is often argued, may not affect care processes as such: if only administration changes, care could still be done in a similar way. However, this is misleading and indicators do lead to changes in care, for example to the standardization of care processes. Standardisation can be both a direct and an indirect consequence of performance indicators. Especially process indicators tend to lead to standardization of care. Take for example the indicator used for stroke—the 'door to needle' time, pointing at the time between the patient entering the hospital and the start of treatment. As neurologists in hospital A note in a paper they wrote for a professional journal: "The *door-to-needle* time has been used in recent years as an important performance indicator for the quality of care for patients with an acute cerebral infarction." They then go on to argue that multidisciplinary collaboration is crucial to get to good outcomes on this indicator, pointing at the changes in personnel that has to be taken into account, due to handovers but also due to changing assistants. Because of this, "it is crucial that everyone involved in the care process knows his or her part and takes responsibility. Having good process descriptions therefore is essential."

Such direct relations between performance indicators and standardisation processes of care were also found in relation to breast cancer care in hospital A. When the hospital did not get the 'pink ribbon' from the patient association for breast cancer, the organizational manager of the ward—who was new to the hospital at the time—was surprised and started an investigation. She found out that almost 80% of

patients weren't operated on in the standard time because the planner for the operating theatre was not involved from the start. Planning therefore only took place after the pre-operative screening; from then on everything went well, but the standard started counting after the decision to operate, rather than after pre-operative screening. And 99% of patients who went for pre-operative screening were actually operated on. So changing the process meant the hospital would again be well within the norm (interview organization manager, hospital A, 18.10.12).

Also indirectly, performance indicators led to standardizations of care processes, for example by developing care pathways, structuring care processes including the implementation of forms to measure progress and complications, as the case of post-surgical wound infections and access times in hospital C demonstrates. Administrating makes the position of professionals as well as supporting services more visible as well as more formal, stimulating an organization wide approach to care.

Registration work, then is not simply administrative work that health professionals can do but needs many added interventions to make registration happen—investments in electronic patient records, in performance management, in policing, in training, in bringing in new types of professionals and in standardizing care processes. This juxtaposition, or layering of interventions we saw in all hospitals we studied, although the forms they took, and the indicators they focused on might differ. Such 'investments in forms' (Thévenot, 1984) were necessary ingredients to making measurement happen—and to perform well on rankings.

Registration difficulties

Even then, though, the investments sometimes didn't work out and we encountered many difficulties with registration due to several reasons. For example, registration work might not comply with the work flow at specific wards. In hospital C for example, patients admitted to the emergency department had to be screened on delirium risk in order to comply with the indicator that says that such screenings have to be done in the first 4 hours in the hospital. This generates problems, as we found in two observations of the work in hospital C.

I ask the nurse, who also functions as a senior quality advisor, why she thinks organization implements DOSS/delirium scoring at the emergency department (ED). She tells me that the indicator demands to score patients within 24 hours. On the ED patients are seen by doctors within 1 hour, and where it is therefore easy to fulfil the norm. This is different at the Acute Ward (where all non-emergency patients are submitted to). Here, first contact with doctors can take a whole day. So, in order to commit to the indicator, the doctors on the emergency unit carry the responsibility to score for delirium. She stops and says: "It is impractical as it is not relevant for the ED." Thus, indicators at times do generate frictions with local work logics (discussion with nurse from Acute Ward, hospital C on delirium indicator during lunch break in the coffee room, 12.12.12).

The emergency department specialist I follow knows about most of the indicators that are prescribed, he also knows where to click in his EPR, yet they doesn't comply. He knows that he has to fill in the delirium score, yet doesn't do that. And while he basically agrees that indicator work is relevant for the improvement of health care, he feels that it is "counterintuitive" to his work. He explains that his work practice is organized around very different parameters: he sees his patients only shortly, and in that period of time he is focused on speed intervention and triage work most of the time (e.g. choosing

for the most vulnerable patient to receive most resources). Indicator compliance is not on his priority list (observation emergency department, hospital C, 23.01.13).

Similar problems occurred with short stay patients in hospital A, where nurses often did not make checklists in the EPR as they thought it was not worth the trouble:

The nurse opens the EPR to do some registration work. She says that registrations with neurosurgical patients don't get done as well as with neurological patients. 'We have to watch this better.' She argues that the neurosurgical patients get hospitalized for a (to) short period of time. That doesn't invite the nurses to make a list of orders for the measurements (observation at ward, hospital A, 23.1012).

Problems could also have to do with the indicators themselves, for example if different rankings or indicator systems used slightly different definitions of indicators; especially if these differences also had an effect on work logics in the hospital. This occurred for example on the indicator on the appraisals of individual medical specialists (the so-called 'individueel functioneren medische specialist' or IFMS indicator).

The information manager points at the IFMS indicator. The *AD* ranking awards a maximum of points to this indicator if 75% of the medical specialists are appraised each year. The hospital does not live up to that standard. He argues that this is also not necessary if you live up to the IGZ indicator that states that such appraisals need to be done every two years. It might be that the hospital decides to increase the frequency of the appraisals, but that would actually be an unnecessary burden, he argues. The hospital is very much focused on individual (dis)functioning and safety. 'If somebody isn't functioning well I don't wait for the next yearly appraisal, but will call him in and discuss this in order to improve his functioning' (observation hospital A, 01.10.12).

Volume indicators, which are increasingly seen as important, especially by insurers, also create much confusion in the hospitals as they go against the logic of work in and between hospitals. Whereas volume indicators work from the assumption that care is located in one particular place, this is often not the case, as hospitals might have regional agreements and professional (and sometimes whole operating teams) might work in different hospitals. Hospital B might serve as an illustration here.

Due to the upcoming merger with another hospital in the region, specialties are shared. For example, colon cancer surgical procedures are distributed amongst both locations. As hospital B is a regional trauma centre, complex patients that demand open surgery are operated in hospital B, while less complex patients are treated by help of colonoscopy in the other regional hospital. And while surgeons work in both hospitals and work as one team, the volume norms in both hospitals were not sufficient, as it was not indicated that surgical teams commute from one hospital to the other. (observations hospital B, 02.05.13 and 23.04.13)

Volume indicators here work against the organisational logic of health care in that they ignore the mobility of health care facilities between hospitals. In doing so, they force hospitals to change their policies (see more on this later in the report; see also (Zuiderent-Jerak, Kool, & Rademakers, 2012)), as well as generating difficulties for registrations.

Also changing medical language impacted on registration work. To be able to make this visible, we had to get into the cellars of the hospitals to visit the medical coders. Until recently a highly invisible group, medical coders have become a more pronounced and visible group in recent years (Jerak-Zuiderent & Bal, 2011). Coders are mediators between medical records and national registrations, in that they translate medical records into the categories and language of, say, the International Classification of Diseases (ICD). This is difficult and highly specialized work as it means being able to understand what is in the record—and thus have the clinical knowledge to read those—and translate them into appropriate codings. This translation or mediation process is crucial to the indicators as it for example gives information about the amount of patients with specific conditions (including ‘co-morbidities’) being treated in the hospitals. It is also highly interpretative and many mistakes are possible. In the Netherlands, especially the debate on the hospital mortality ratio has made this clear (Kalkman et al., 2013). In our study, we found similar issues, for example in hospital B:

The coder in hospital B feels that there is way too little training. She feels that education for medical coder is not sufficient in order to do demanding registration work. She and her fellow coders are educated as medical coders and also follow additional trainings, mainly at KIWA Prismant and other specialized institutions. The last training she received was about how to use ICD 10 (instead of ICD9). Coding, however, is an increasingly complex task. For example, the ICD 9 had three codes for a cerebral infarct, but ICD 10 has nine. Hospital B therefore split coding work and each coder has particular areas of attention and responsibility for particular specialties. She is responsible for cardiovascular diseases and explains that for example recognizing what caused a cerebral infarct is necessary for proper coding work that does not only constantly code “9” – which is “other categories” (observation hospital B, 25.06.13).

Coding errors, respondents argue, happen frequently in the hospital due to many reasons, such as typos, forgotten registration, wrong coding, or technical failure, as a coder in hospital C argued (interview data warehouse administrator, hospital C, 17.01.13). In the case of surgery, for example, treatment codes are attributed by secretaries the day after surgery. There are no separate coding teams doing that work. This is a well-known source of registration mistakes. In the case of hip fracture, for example, he found out later that secretaries used different codes than the administrator prescribed for hip fracture. In consequence, a large quantity of hip fractures did not appear in the registration system, which again generated a problem with regard to the volume of treated patients eventually. Therefore, the coder spends much time with ‘putting results next to each other’, checking dissimilar systems, and trying to find out where coding mistakes happen. He does not only do this when things go obviously wrong as in the case of hip fractures, but he always tries to ‘think along the numbers’. He argues that much of data collection work is controlling based on experience and ‘horse sense’.

The additional work in translating medical records into indicator information is then another possible source of misfit with what is happening in the hospital organisation—and thus another issue to address in governing administrative work.

A last reason why registration work creates difficulties is because certain indicators might work against the governance logics of hospitals. Most hospitals have been reformed in recent years by creating ‘result based units’ in which wards (or larger units) function as self sustaining departments within the hospital. This then generates

problems with indicators that go against this compartmentalized logic of the organisations. As we observed in hospital C:

Hospital C recently joined a national pilot on Patient Related Outcome Measures (PROMS). The aim of this collaboration is to “bring together soft and hard indicators.” The manager of the patient service office describes the project as a difficult endeavour, because the hospital still entertains much of a ward-based structure of responsibility, organized in polyclinic and clinical care. Also, she feels it is difficult to evaluate patient experience, where “people don’t see beyond their nose” and where “parochialism” makes it difficult to pinpoint who in the care process is responsible for good or bad experiences of patients. PROMS, she then seems to imply, doesn’t fit the organizational structure yet (manager patient service office, hospital C, 01.11.12).

In summary then, we can say that while all hospitals worked hard to get registrations going—and were largely successful in doing so—registration work also encountered many problems in getting to the data that were needed to report on indicators. In this section, we have analyzed how hospitals tried to deal with these difficulties by focussing on the ways in which health professionals were disciplined in doing administrative work. A layering of interventions—that we referred to as ‘investments in forms’ after the French sociologist Laurent Thévenot—was seen to be used in all hospitals, albeit with slight differences between them. This did not overcome all difficulties however, as indicators sometimes conflicted with the flow of work of clinical work, with the mobility of care between wards and hospitals, with changing clinical language, and with the governance structure of hospital organisations. In the next chapter we will analyse how these changes and resulting difficulties played a role in organizational changes and collective sense-making in the hospitals. In the remains of this chapter we rather move upwards in the organisation, to see what happens with the data itself once it is registered.

Aggregating and processing data

Data collection and processing is a collective effort in the hospitals and many different departments are connected to the processes involved. In the hospitals we studied, the main actors in collecting the data—apart from clinical staff as discussed in the previous section—were quality and information managers, but controllers, and communication staff would also be connected in some way to the data collection process. Controllers, for example, are important because performance data is at some points, say the use of DRGs, connected to financial information,³ and communication staff would be involved in both internal and external communication on registration work and/or performance data. Usually, in the hospitals we studied, some four to five people were directly connected to data collection and processes; for two or three it would be an almost fulltime job—taking care of the data collection processes for a whole range of indicator and ranking systems.

As already pointed at, electronic systems like EPRs play a large role in the data collection process, but still data that needs to be collected for the indicators comes from a very diverse set of systems. The hospitals had built data warehouses in order to combine data from different systems, but still, as one of our respondents in hospital A said: “there is still much hand work attached” (information manager,

³ We have not seen explicit connections between data on quality and costs being combined in the performance indicator and rankings processes we have observed; however, cost-benefit ratios, for example in the form of business cases, were used in the hospitals to decide on investments and performance data could be used for those purposes. Cf. (Bal & Zuiderent-Jerak, 2011; Zuiderent-Jerak, 2009).

hospital A, 18.09.12). This is the case for example if different information systems used come to different data on say the number of patients with a certain diagnosis in the hospital, which often happens. In this case, systems have to be checked against each other to establish which number is the right one.

One way of checking the number is to back-and-forth data between the quality and information department and the clinical wards. Getting to the 'right' number is a process of triangulation where different data sources—including human ones—need to be compared. A quote from hospital A is exemplary for all case study hospitals in this respect:

The process is...we get a list of questions from an external source and then we first start looking at what we can get out of the systems, say out of the DRG system, or that we would have to ask the medical specialists for that information, but before we send anything internally we will first check what information we have and fill in all the questions as far as we can retrieve that. And once we have done that we will send the whole package to the medical specialist or care manager concerned, for them to control and approval, with a deadline and we will contact them to ask if this is going to work for them or they need further assistance. Well, and then you get it back and by then it is approved so to speak, so we can then use it for sending to the external party. (...) Sometimes they will say 'this number is wrong' and then we consult with them and I also then usually would have a look at other registrations whether they match" (information manager, hospital A, 18.07.12).

Getting to the 'right' data means taking up a relational epistemology (Verran, 2013) in which back and forting of data between information and quality departments, clinical departments and different information systems in the hospital lead to a settlement on the 'right' data. This relationality also extends in time and place, with the information manager looking at values that were derived in previous years as well as to national data in order to 'verify' the data for this year:

...I will look into the year concerned, but also at previous years to compare...what kinds of numbers are we talking about, does it look normal, is there a progression or not and then I would also look at the national data so that I can mirror... (information manager hospital A, 18.07.12).

Even with all those controls, mistakes do happen, so more—sometimes external controls are necessary and get done. The data for the IGZ and ZiZo indicators for example has to be uploaded to a central website. This is a manual process and happens by the medical secretaries. Because there are so many data, mistakes occur and therefore controls have to be done. "For example, we had recorded that children with Crohn's disease come to the hospital 3 to 4 times a month, but this should be per year, and in stead of the mean waiting time we had filled in the maximum waiting time; that makes your scores worse" (Quality manager, hospital A, 04.09.13). Like for other hospitals, data for hospital A are externally 'cleaned' by MediQuest—a commercial bureau active in the performance indicator market. Data for the *Elsevier* ranking are collected by SIRM—another commercial consultant—and also there controls of the data occur: "In the data that we work with we had an open MRI, but there are only a few of those in the country and actually we use the one of [a neighbouring hospital], so we corrected that" (ibid.).

To govern the process of data collection and processes, all hospitals have formed steering groups or similar structures, of which medical staff, quality and information managers, controllers and clinical managers are members. An indicator steering group in hospital B discusses the 'bottlenecks' and 'crazy things' that emerge from

the collection of mandatory indicators, such as IGZ and ZIZO. The head of this group is herself a medical doctor and she then addresses her colleagues in cases the steering group identifies responsibilities and problems related to indicators. This group also arranged for an analysis of recent rankings results. The bad results of last year's rankings intrigued the management to perform a disease-specific analysis (medical doctor, member steering group indicators, hospital B, 21.05.13).

Also in hospital A a steering group has been set up that has as its task to "develop a vision on indicators and increase the support for indicator in the hospital." Specific goals of the steering group relate to minimizing the registration burden of clinical staff, to monitor national developments and to "start a movement in which external indicators will be used as internal steering information in the hospital (i.e. a dashboard)."

Governing performance

So far, we have seen how hospitals have invested in administrative work and in the 'relational epistemology' necessary for the collection and processing of data. In this section we will rather look at what happens after this process is completed: what actually happens in the hospital with the data that has been submitted for external accountability? How have the hospitals we studied organized for giving shape to the 'movement in which external indicators will be used as internal steering information in the hospital', to repeat one of the goals of the steering committee in hospital A. Put into different terms: how do the hospitals govern performance?

Overall, we found that all hospitals use the performance data for internal steering. However, they do it in different ways, relating to the ways in which they have organized quality policies more generally. Hospital B, for example, has a tradition of lean management, in which decentralised control over quality is emphasized and in which performance data, though important, is not meant for comparison across services but to measure improvement over time. However, recent bad results in a ranking led to an analysis that led to:

That we give renewed attention to registering and weighing outcomes and more actively use the data for quality improvement. Also, we use the data in the fall and spring meetings with the medical units, and as a follow up we use them in the contracts we have with the medical staff (executive director, hospital B, 02.05.13).

Overall, rankings results in hospital B are not consistently analyzed or benchmarked. The responsible care group manager argues she benchmarks results with the neighbour hospital with which a merger is underway. She feeds back results into the bilateral meetings amongst manager and professionals. Also, she goes directly into teams and asks professionals as to why particular scores are so bad (observation hospital B, 21.05.13).

Hospital B has a steering system, where care group manager together with medical directors steer departments. The organization is also characterized as 'informal' by many respondents, where "much information flows on corridors and only to limited extend on paper" (care group manager, hospital B, 02.05.13). Respondents also describe this governance system as decentralized, where managers and professionals share leadership and where professionals enjoy much freedom with regard to indicator governance. Some examples: 1) the executive directors allow for internal negotiations about the relevance and evidence of particular indicators amongst professionals. 2) Patient experience has to be collected on every ward once a year, yet the wards can choose collection methods individually. 3) Professionals are allowed to deviate from indicators when they can reason deviance medically.

However, if indicators are relevant to perform well on particular rankings these are filled in even though professionals consider the indicator 'nonsense'. For example, the cataract indicator generated much discussion with regard to existing evidence. But as the indicator is important for ZiZo it is filled in nevertheless. This again points towards what we have described as pragmatic compliance in the previous chapter. 4) Professionals are allowed to prioritize improvement activities on basis of indicators they consider relevant.

In hospital B, quality and safety compliance is governed according to what the executive director calls a 'fishing net approach' (02.05.13). Here, the executive director tries to bring together 'signals' from the organization that are of relevance for the quality and safety agenda, such as calamities, disciplinary cases ('tuchtzaken'), and complaints. Also, individual findings from the medical management, the nursing council, the safe incident reporting committee, are used by the executive director to "generate a coherent picture about quality and safety" in the hospital. Quality issues are also discussed bilaterally in the contract negotiations that happen twice a year amongst care group management and professionals. However, the secretary of the executive director—who plays an important coordinating role towards quality & safety policies throughout the hospital—reports, this "does not yet lead to quality improvement, but rather helps the organization to get a bit of an overview on what happens with regard to quality and safety" (interview, 23.04.13).

A set of 6 care-related indicators is monitored in hospital B in real time in the electronic patient file. A three-monthly management report for quality and safety monitors twelve selected indicators (including pressure ulcers, malnutrition, pain, delirium, and safe incident reporting). Yet, it is not a comprehensive monitoring tool as to how registration and compliance with indicators is performed in the hospital. Progress is not monitored, also, there is no 'ample colour system' that would indicate performance for particular indicators. The secretary of the executive director feels that so far, there was not much focus on the collection and monitoring of indicators, particularly as staff was resistant. The care group manager who has quality and safety in her portfolio argues that this only gives her very limited ability to steer on basis of indicators. But, she also argues, not all quality work was related to indicators, and the hospital's 'fishing net culture' again helped her to get her fingers on important issues nevertheless (care group manager, hospital B, 02.05.13).

Hospital A showed many similarities with hospital B, but was further along the path of using the performance indicator for benchmarking purposes. Indicator and ranking data were much discussed in the hospital between the managerial and clinical departments, as we have already seen in relation to the nursing indicators that were set as a priority on the basis of ranking scores. The quality and information manager mostly related to the performance indicator processes make detailed analyses of the relative performance of the hospital that then get send throughout the hospital.

The quality manager shows me the Excel sheets. Long lists with data, including explanations for the comparisons. Colours indicate points of attention. She shows me the ZiZo mirror reports. "We work with percentages: the bottom 25%, the middle 50% and the top 25%. Then you have to look at what is good or bad; if we score badly we make it red and give that to the hospital board, if we do well we make it green," she explains. "We list the things where we are scoring bad, and where we score well too: you also have to give a positive message!" Score-sheets also go to the ward managers who discuss them at the wards. Also the board of trustees, who have a quality & safety commission that meets 4 to 5 times a year, is informed. At one of their meetings, the performance indicators will be discussed as well as the policies taken for improvement of the scores (quality manager, hospital A, 04.09.12).

Also in hospital A, indicators are discussed in the meetings between the board and the clinical wards.

In hospital C, the external focus on market-based health care and its accompanying transparency technologies lead to that the hospital underwent a restructuring process when the hospital tried to generate a better link with the market and the capacity to react [to changes] more quickly. In hospital C, the department for quality and safety was one of the first quality and safety departments nationally, and is comparably well staffed. Due to the increasing national demand to comply with indicators and other performance tools, well-trained staff for indicator compliance increased. Also, the hospital restructured its quality department whose staff has been decoupled from quality improvement work and does (officially speaking) not support improvement work in the primary process anymore. Thus, ward-based quality improvement work is not in the task package of the quality department anymore. In that way the department differs from many other hospital quality departments in the country, including hospitals A and B, where quality improvement and quality assurance were more integrated. The changed structure of the department for quality in hospital C extrapolates the relevance, which the hospital ascribes to the management of quality and safety, and to the compliance management with indicators in particular.

The external focus on market-based health care and its accompanying transparency technologies in particular also generated a new 'type' of quality manager, whose task is (at least formally) more oriented towards compliance management. However, hospitals consider the role of quality-related indicator management to be broader. A meeting with the healthcare inspectorate, in which a new supervisory system is piloted, is exemplary here:

After the inspector introduces herself, her team, and the aim of the pilot, the quality manager explains why and how the hospital splits quality control from quality improvement work. She listens and then asks the quality manager whether he 'is the compliance manager of the hospital.' He becomes silent. The executive director explains that the quality manager 'translates between the inside and outside world', and that it is important to take apart the role of compliance management and improvement. The quality manager now also jumps in and argues that the hospital underwent a process of 'clearance of roles and tasks,' while he himself was 'more than a compliance officer.' He argues that he 'constantly was searching for discourse within the hospital and outside the hospital to unite these two worlds.' (observation hospital C, 16.12.11).

The increasing focus on indicator-based governance generated a new 'type' of quality manager, whose task is (at least formally) more oriented towards compliance management. However, this also creates ambivalence, as quality improvement work still is valued.

Performance management in the hospital

In hospital C, compliance with national quality indicators is one out of three parts of the hospital's quality management system and thus receives much attention. Indicator performance management entails the monitoring of external performance indicators (IGZ indicators, ZiZo indicators), HSMR, VMS themes, access time, nursing scoring compliance results, and the local patient satisfaction survey. With regard to national indicators, the hospital has spelled out ambitions (e.g. be amongst the top 15% performing hospitals for IGZ indicators and top 25% for ZiZo indicators). Indicator performance management does not contain information about rankings, and this was a conscious choice. Both the executive director and the quality manager felt

that rankings are too volatile and with too much of a time slack in order to actually steer quality improvement.

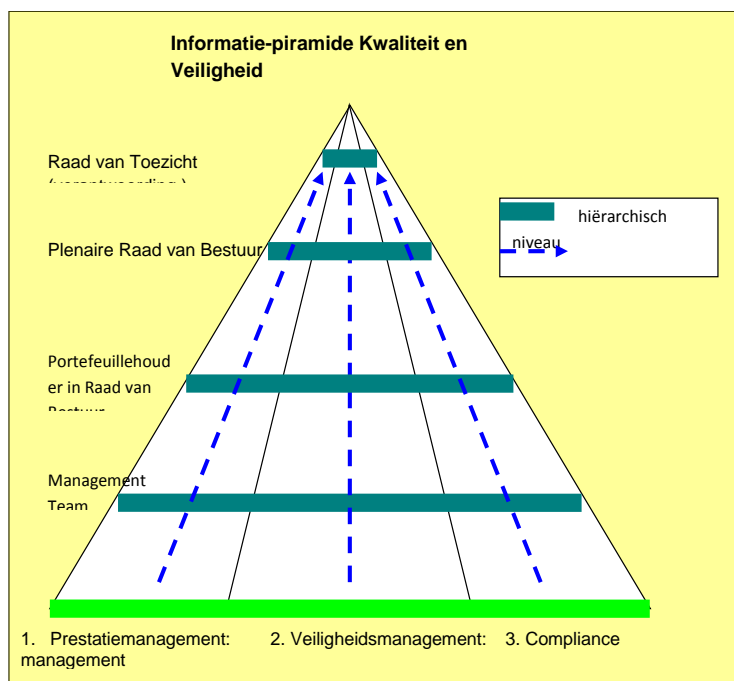


Figure 1: organization of quality and safety in hospital C

Compliance with indicators is basically monitored in an excel sheet that offers:

- 1) a benchmark to the previous year and the national average
- 2) internal trends (arrow up, down, explanation)
- 3) an explanation section for bad performance and internal trends
- 4) a visual coding system that indicates the overall status of compliance

Next to indicators, risks and compliance are monitored in comparable excel sheets, and together these form the three sections of the quarterly quality report. The report builds on the plan-do-study-act cycle. The department for quality and safety collates all data, that derives from the primary process and which has to be delivered by professionals and managers. It reviews, summarizes and interprets the outcomes, and the preliminary analysis is shared with ward managers, who react to results before the official quarterly report is discussed in the hospital management team by all middle managers and executive directors. The executive director approves of the report and is responsible to address professionals with regard to improvement. In that way one "hopes to generate quality improvement on basis of indicators." (EG, 12.11.12) The report delivers different forms of aggregated information for middle managers, the RvB, and the RvT respectively. As shown in figure 1, the process through which quality data is collected and used is depicted in a hierarchical fashion, showing the strong centralization of quality policies in the hospital.

In hospital C, in response to the AD result, the quality manager also published a letter on the intranet. The letter was accompanied by figures that illustrates how particular treatments scored in comparison to what would deliver 100% of the rankings score. Next to that, the department for quality and safety also generated a more detailed analysis that evaluated the respective indicator sets according to: maximum score & hospital score, other regional hospitals, explanation, score according to own measurement (see figure 2).

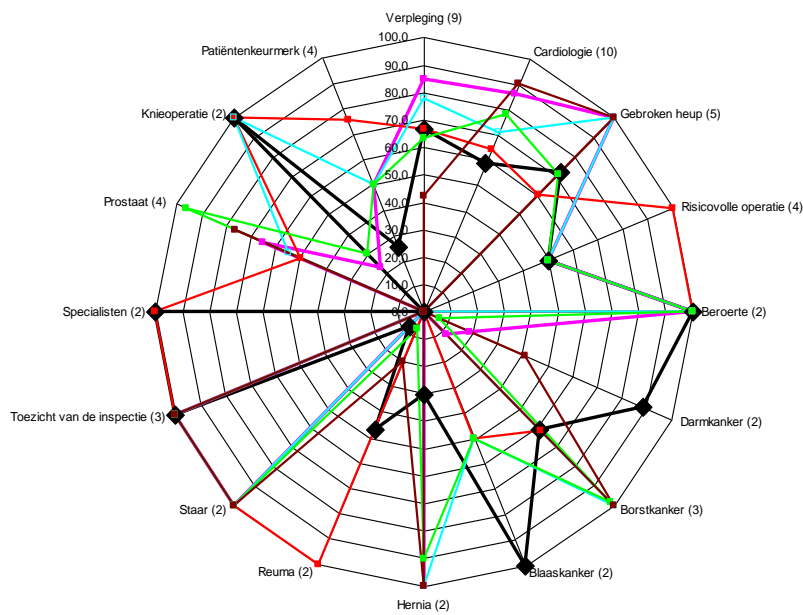


Figure 2: Regional comparison of treatment-specific indicator sets

Although in all hospitals performance measurement was an important topic and relative performance of the hospital in relation to national and/or regional data was used as an input to quality work, there were large differences between the hospitals, with hospital C standing out as the most centralized performance governance. Whereas hospitals A and B combined quality assurance and quality improvement, this has been separated in hospital C; likewise, data intensity in hospital C was much higher than in the other hospitals. Hospital A was somewhat in the middle position. Contrary to our expectations, competitive environment does not play a role here, as hospital C is actually situated in the *least* competitive environments. Rather two other factors seem important to explain the differences between the hospitals. First, history matters (Pollitt, 2008). This became particularly clear in hospital B, which had a history of lean management that actually went against central control and planning. Second, in all hospitals, bad scores on rankings seemed to have sparked more centralisation. Especially in hospitals A and B this seems to have been the case, making the hospitals deviate from more decentralised and informal approaches towards a stronger focus on performance information.

Bringing in the marketers

In the previous section we started to have a look at the ways in which performance data was used internally in the hospitals, an analysis that will be extended in the next chapters. Here we want to focus rather on the external use of data, more particularly how and in what ways rankings or underlying performance indicators are used in the hospitals to build a reputation for the hospitals. As we saw in the introduction, previous research into rankings has shown that one of their effects—and also one of the central mechanisms through which rankings have an effect on organisations—is that they create reputation as a new ‘risk object’ that needs to be managed by organisations (Power, Scheytt, Soijn, & Sahlin, 2009). In this respect, it could be expected that marketing becomes a new and important function within hospital

organisations, replacing the earlier focus of communication departments on patient information. To some extent, we indeed saw such a movement happening within all hospitals, but again differences were found, with hospital C standing out with a more pro-active marketing strategy than the other hospitals.

In hospital B, the manager of the communications office stresses that he is not a marketer, as he and his crew would not invest into analyzing potential patient flows and operating on basis of that. According to him, patient flows are still sufficient for the hospital, despite all transparency technologies. Nevertheless, he argues, his work increased in manifold ways in the previous years, and that he “has to keep a hundred balls in the air all the time.” He therefore prioritizes his work. Rankings, in this context, are secondary to his work, he argues (communication manager, hospital B, 21.05.13). The executive director of hospital B confirms this, after asking what communication strategy the hospital takes when it has performed badly on the rankings:

We communicate them internally what we will do and externally we communicate only in a re-active manner, when we are asked, that we are startled and that it is an impulse to improve. What we won't do attacking the ranking by 'it doesn't say anything' or 'it is worthless', that kind of nuance of the ranking we will not do. Internally we won't do that either, although we might nuance it a bit more. We would rather use it in a way that says 'well, what our external environment thinks of us is increasingly important', but what matters most is that we can show our patients that we are actually performing well (executive director, hospital B, 02.05.13).

Rankings are also not the leading factor for the communication managers in hospital A. Particularly the volatility of rankings make them 'risky' in using them for communication purposes, the communication managers argue.

...we try to be a bit neutral, as it is hardly comparable (...) I can understand that hospitals that are on top of the list want to shout it from the rooftops, but there is a risk, I think. You might be last next year and what are you going to do then? (...) Once you start using the league tables for your marketing, you can be hurting yourself. So I would not use them that much, at least externally. We rather see it as a danger (communication manager, hospital A, 26.11.12).

Although they will make postings about the rankings, they will always cloud this in nuancing expressions. Nevertheless, rankings are closely watched and the hospital will make actions on them, for example when they are wrong, thus trying to guard the reputation of the hospital.

The quality manager refers to an article by the Consumer Association that had made a comparison of hernia treatments on the basis of ZiZo data. The hospital was scoring badly and the quality manager had investigated this. It appeared that the presence of a physician assistant was weighed heavily in the ranking. When the ranking was published, the hospital did have a physician assistant, but not in the year on which the data were based. The neurologists were unhappy with this and so the hospital contacted the Consumer Association in order to verify the data; in a subsequent publication, this was corrected (quality manager, hospital A, 04.09.12).

In hospital C, new organizational structures were initiated as a direct reaction to the increasing public interest in quality and safety work, which is partly caused by transparency technologies such as rankings and national indicators. After the

department for communication was expanded and transformed into the department for marketing and communication, also a new marketing and communications manager was hired. She has a history of marketing in public institutions and is responsible for the hospital's overall reputation management, and thus also for the communication of ranking results to the staff and the outside world alike. The communications manager also briefs the executive director for meetings with politicians about how to 'sell' the current standing in rankings such as Elsevier and AD (communication manager, hospital C, 06.12.12).

In hospital C, the marketing manager often knows rankings results before they are published, usually through informal contacts. She then prepares her communication strategy in advance. If the ranking demonstrates an improvement trend, she publicly increases the relevance of the rankings for quality and safety. She also proactively communicates results to journalists then. If results demonstrate a deteriorating trend, she focuses on what exactly is decreasing, pays attention to the internal analysis, diverts attention from the rankings to other things, emphasizes the time slack of data analysis and result publication, and finds explanations as to why the ranking was so bad. She downplays bad results, and features good results.

The communications manager in hospital C differentiates what she perceives as main audiences of dissimilar rankings and how she reacts to that. For example, the AD and Elsevier are usually important for journalists, so if the hospital would drop there, she would invest mostly in discussing the results with journalists. On the contrary, the CQ index received much more attention from insurance companies, which intrigues her to discuss these results with insurance companies in the first place. Interestingly, patients are rather mentioned as main audience of less prestigious rankings, such as the 'gastvrijheidssterren' of a daily newspaper, *de Telegraaf*.

Communication managers in other hospitals might also be aware of the different audiences of rankings. The communications manager in hospital B for example argued that the Elsevier ranking was only read in the Rotterdam/The Hague area and related this to the dissimilar population structure.

In line with a more general trend in Dutch hospitals (Adams, 2011; Groenen, 2013) social media is seen as particularly important by the communication departments in the hospitals we studied. In hospital C, the marketing and communications department for example focuses on how the hospital is represented in social media. One of her staff members continuously monitors such pages and reacts immediately (e.g. tweet: 'long waiting time 1st policlinic visit' → retweet: 'that's annoying, how do you explain this?'). "The best advertisement for the hospital is a well-disposed twittering patient", the communication manager argues (communication manager, hospital C, 06.12.12). Also in hospital B, the communication manager we spoke to feels that particularly social media needs attention. For this reason he drafted a 'social media plan' where 0.5 fte staff shall be used to monitor and communicate with social media platforms (communication manager, hospital B, 21.05.13).

Whilst the communication managers were quite aware of the rankings and how they could (or could not) influence outside perceptions of the hospital, they also immediately combined this with an internal focus, arguing that rankings are actually there to learn from, so emphasizing quality improvement rather than analysis and communication. As the communication manager in hospital C termed it:

I always hear that we score badly; we always have to do more analyses. But we don't need more analyses. We have to act. More research is always safe and my closet is filled with analyses. But we have to get into action. We know at which points we are not scoring well. We know the backgrounds. And we

want to improve, but instead of action we rather have more analyses (marketing and communications manager, hospital C, 23.20.12).

Learning

As our data shows, and as we have seen already in many examples in this chapter, learning did take place. Analyzing indicator scores regularly lead to changes in processes of care in all hospitals we studied. Especially those care processes that were measured in the rankings were the focus of attention; like in all hospitals the nursing indicators. As the quality manager in hospital A noted, the hospital gave a lot more attention to pressure ulcers as an effect of performance indicators; while before the indicators were introduced the problem was recognized, but there was way too little support for change.

Scoring badly on rankings creates support for change and for learning. In hospital C, a range of quality programs/projects emerged as reactions to bad indicator scores, and for example the culture program emerged as a reaction to an Inspectorate visit that claimed that the hospital had no culture of safety and collaboration in place.

From setting goals to realising them

- Changes are not proclaimed, they are attained.
- There are methods for effective implementation: daily discussions for example. They need to be applied more.
- You don't jump higher when you look at the score-board.
- Motivate, visualize, organize, try, practice; these actually help you improve
- We communicate much about goals, now we focus on how to attain them.
- We will make classrooms, workshops, seminars, and a network of like-minded
- We learn a lot from conferences with celebrities.
- Now the options will be less open-ended.

From plan to action

Give medical specialists access to the learning spaces, including e-learning on many subjects. Let coordinators guide.

Source: quality plan hospital C

Also, training sessions were organized in some hospitals to learn from indicators. Hospital C's elaborate three-monthly management analysis for example showed that indicator compliance lacked behind consistently over the last years. On basis of that finding, the quality manager induced a team-based learning session on how to better improve on basis of indicators (see excerpt summary meeting in the box on the previous page).

In hospital B similar sessions were organized. In a management meeting we observed for example, one of the care group managers gave a presentation on performance indicators, starting with a six points plan 'from measuring to improvement' taken from a lecture by a professor on quality improvement. The presentation informs the meeting what actions the hospital will take to work on performance measurement and management.

Outside expertise was often brought in—e.g. in the form of slides or videos—but likewise some of the hospitals also used visits to other, well-performing hospitals, to learn how to improve. With regard to the CQ index, hospital C for example visited well-performing hospitals in order to learn from their experiences.

Concluding notes

In this chapter we have focused on how rankings affect hospital organisations, focussing first on administrative processes and then going on analyzing structures and processes of making use of the collected data, both internally and externally. Rankings, and the performance indicators underlying them, were shown to affect hospital organisations to a great extent, despite the finding in the previous chapter that the rhetoric surrounding rankings is often one of disdain. Rankings lead to huge 'investments in forms' – including the building and implementation of electronic health records and other types of information systems, the disciplining of professionals to actually register and the standardisation of care processes to make such registrations possible. They also lead to new work and sometimes changing roles for quality and information managers, introducing a compliance function to quality management—and sometimes, like in hospital C—even to a separation of quality control and quality improvement. They lead to a renewed emphasis on learning, stimulating the use of performance indicator data to improve on care processes, and in this way 'internalising' the performance indicators data collected for 'external' purposes into the hospital. And they lead to new marketing strategies. Especially scoring badly on rankings was seen to be a stimulus for hospitals to start working on these internalisation processes.

We also saw differences in the ways in which the three hospitals we studied reacted on rankings, with hospital C taking up a more centralised approach, focussing attention on quality assurance, whereas hospitals A and B still used more decentralized approaches. However, rankings do seem to perform a centralising tendency, making the hospital as a whole into a more governable entity and thus stimulating managerial action. Interestingly the differences in approach could not be reduced to the competitive context of the hospitals, but seemed rather to reflect the histories of the hospitals in which previous choices have created a pathway for further development. Nevertheless, such pathways were influenced by the rankings leading to new types of governance.

In the next chapter, we continue our study of the ways in which rankings affect care processes in the hospitals, before we turn our gaze more explicitly to hospital governance.

Chapter 4 Quantification work and its reactions

The data collection process, which forms the fundament for performance technologies, is not a straightforward process that generates unambiguous and comparable data across institutions, as we have already seen in the previous chapter. In that chapter we focused on the activities and interventions that were employed to get to performance information nonetheless—and on how this data then was used upstream in the performance management process. In this chapter we return to some of the underlying tensions in this process. Where do the tensions lie in indicator measurement? We will show that those underlying tensions lead to many distortions of the data, making the data an often untrustworthy source. We will also explore possible unintended consequences that arise from this – both in terms of data managing and in terms of changes in care processes. We were also interested in the reactions to those distortions and unintended consequences: did alternative practices of accounting for performance arise?

Professionals, “rightfully tend to question indicator-related compliance demands,” as a senior staff member, who is responsible for performance control in hospital C, remarked (12.11.12). In this chapter we explore the background to this. First, we summarize some of the reasons that bring us to argue that data collection is an interpretive, situated practice. Second, we show some of the unintended consequences of performance indicators and last we discuss some of the reactions in the hospital to what is felt to be an overtly confident use of performance data by outside actors.

Negotiations over data

Indicator collection work begins with the definition of the indicator. Therefore, the process of deciding what data has to be collected and where is crucial. A division director in hospital C explains that particularly the IGZ delivers vague definitions for its indicators, while ZiZo indicators are much more precise in their inclusion/exclusion criteria (06.12.12). Care insurers (apart from Achmea) often only deliver one page with indicators. This implies that particularly the indicators of insurers and the IGZ can be interpreted in different ways.

For example, the IGZ indicator ‘ventilated patients’ does not make any clear reference to the group of patients to be included, i.e. should one only count ICU patients or include all patients? Obviously, including all acute patients would dramatically improve the score, a practice which the case study hospital doesn't do. Vaguely formulated indicator sets generate leeway for strategic behaviour; inclusion and exclusion decisions are one way of strategically influencing on performance results. Another example is door-to-needle time for stroke patients.

When does the clock count with regard to the indicator ‘door to needle time’? When the patient arrives to the emergency room, at the point in time that the neurologist enters the room? What if the patient is registered for the wrong diagnosis? What if information in the pre-hospital trajectory is missing, and the patient not recognized as a CVA case? Where does door-to-needle begin then? ... You can get hundred answers to one indicator (division director, hospital C, 06.12.12).

The director makes clear that data collection depends on local interpretations. In hospital C, quality staff for this reason communicates with medical managers and problem owners about how data should be collected (way of collecting data, codes, etc.).

Because of these interpretations, our respondents argued that it is easy to fiddle with the data. A surgeon in hospital A for example talked about the time to

operation for breast cancer patients, that, according to the criteria for the 'pink ribbon' of the breast cancer association has to be within five weeks.

Where you get into trouble, I noticed... in 2011 we were way over the standard and we had say 20 patients that were not operated in time. But with 11 it was at their own request. But you can't register that (...) and if you bring these data into the open, than the public can no longer see that... Because, well these patients were all very satisfied as they had been operated on in the time set by themselves, the first week after they returned from holidays, but yes... (...) And the ZiZo data, you fill that yourself, so you can choose to leave these patients out, that is possible. (...) It is still the butcher that adopts his own meat. If you want to fiddle with the data, nobody is going to check on them. And, well, that just happens. It's not fraud, but if you are filling your data and you see that somebody went to Mexico for three weeks, well you have to really honest. So I think many people will score 'nicely on time' (surgeon hospital A, 18.02.13).

Many similar examples are in our data. A quality manager in hospital B for example noted that post-operative wound infections are highly interpretive as infections might have many causes, and need not be related to the operation. "You can say that 'well, this wound leaks a bit and there is some redness, but I don't think it is a post-operative wound infection as the patient is also bleeding from his nose and he has a cold.'" (28.05.13) He goes on to say: "We had a meeting recently with medical managers and it was said 'well, we are doing bad on these readmissions, but [other hospital in the region] is not reporting any of these readmissions so they must be fiddling the data.'" Likewise, the information manager at hospital A noted:

"I know at the start, well, you can search at main diagnosis to include a number of patients, but you can also search the side diagnoses, and then you get to much higher volumes. But not all hospitals were doing that, so you could see a lot of differences in the data. A nice example is the myocard infarction indicator of the healthcare inspectorate. If you look at it from a distance it is a really nice indicator, myocard infarction is a clear diagnosis. But if you look more precisely there are a bunch of in- and exclusion criteria that they have formulated and this necessitates you to combine different information sources and do quite some interpretation" (information manager hospital A, 18.09.12).

So numbers are never just numbers but need interpretations and decisions. Yet, this finding is slightly different from what has been written about dysfunctional reactions towards performance measurement, such as tunnel vision (exclusive focus on external accountability to neglect of others) and gaming (wilful manipulation of accounting data). In line with recent ethnographic research that hints at that data collection is in itself a challenging practice, where "what counted as relevant for reporting reflected localized interpretations" (Dixon-Woods, Leslie, Bion, & Tarrant, 2012), our material also proposes that data collection work allows for multiple, locally different interpretations, and that this is not necessarily a process of conscious adaptation that leads towards strategic benefits for the respective person, group or organization. As data is always tied to specific local processes—like patients going on a holiday—decisions need to be made on how to represent them, and these decisions are bound to vary between practices.

'We don't cheat!'

Often, interviews reached a point where respondents argued that indicator-based governance is difficult and questionable, as 'cheating' was an easy thing to do. However, respondents usually would argue that it is other hospitals that would engage into such practices of cheating. For example, a medical specialist in hospital A argues: "I think we try to do it neatly but you know and suspect that some hospitals are less so" (12.12.12). Likewise, the executive in hospital C asks whether his "colleague executive directors" from other hospitals are honest with data collection. He explains that he refuses to base data on random sampling, as other hospitals do. He insists to present data that builds on all available data, which usually generates less beneficial numbers. The reason for doing this is that he wanted a realistic picture about how quality of care was about in his hospital. In consequence, he argues, the hospital was not always scoring as well as it could. "I prefer less good indicator results. This gives me insight into whether the procedures are ok. It better helps to steer internally."

Cheating for the indicators is thus a sort of cheating on yourself as this would disable learning processes. The executive director of hospital B reasons in a similar fashion when we talk to him about pressure ulcer scores:

The next thing we asked was, do we want to improve? The answer is yes, we do want that, but that is not a simple question because: to what price? You can say, well pressure ulcers don't interest us, but the number has to go down. You can then start looking at the point prevalence and take a nice and dry day and you do that twice a year and sure you will get a much better score than when you do continuous monitoring like we do. (...) The question is how do you measure. And are you measuring for the newspaper or to get an impulse to improve? (...) We emphatically went for improvement; our medical staff has confirmed that as well (executive director hospital B, 02.05.13).

So, whereas the ease of 'fiddling the numbers' was recognized by different actors in all hospitals we studied, and there was quite some suspicion that other hospitals might cheat to get to better scores, hospitals thought such fiddling to be not only unethical but also beside the point. Rather than 'scoring well for the newspapers', our respondents emphasised the learning aspects of indicators and rankings—as these enabled them to reflect on their care processes and find ways to improve on them.

Intentional gaming the numbers

This doesn't mean however that 'gaming' does not occur and we have found many examples of occasions where health professionals and other actors for various reasons 'fiddled' with the numbers. One reason why this happened for example was because professionals sometimes found it hard to do all the measurements, but their systems were nonetheless asking to fill them. So they sometimes 'worked around' the system tick-boxes. Take for example our observations of a nurse in hospital A:

Back at her computer, the nurse argues that there is an increasing stress on registration. She fills in the scores. She also gives the pain scores, saying 'I haven't asked him, he does seem to have some pain, so I will give him a two or a three [on a scale of 10].' For another patient she puts in a zero. She explains that she doesn't always ask if the patient doesn't indicate to be suffering from pain and just walks around (like this patient does). Also she sometimes adjusts the score if the patient indicates a pain level that she doesn't believe to be true: "they might say they have an 8 but then they go and have cigarette or they don't want pain killers. Well, then I just give them a

lower score. You use your clinical intuition.’ She also fills the DOSS score [for delirium] without having answered the three underlying questions. ‘He seemed to be reacting adequately’, she says (observation at ward, hospital A, 23 October 2012).

As professionals have to a lot of registration work, tick boxing occurs and is a way for them to get the work done and focus on their patients, rather than on the measuring process all the time. In doing this they ‘work around’ the system, but giving numbers without actually checking on their accuracy.

The excerpt from the observation notes also makes clear that what is the accurate number is not self evident. What the ‘actual’ pain score is for a specific patient seems to be related to several things: what does the patient indicate (when asked), how does he or she behave (e.g. walk around over the ward and outside, lie in bed; talk or not, etc), and how do these two relate. Nurses use their ‘clinical gaze’ to get to ‘accurate’ numbers. Also previous behaviour of the patient may be relevant (has this patient indicated high pain before) or the cultural background of a patient. Pain measurement is thus part of a social relationship between a nurse and a patient – and judging whether patients can be trusted as a source for giving their pain scores is part of getting to the ‘right’ numbers.

Working around the system was something that we saw in all hospitals. This did not only relate to social relations with patients or giving preference to care work, but was also sometimes related to the workings of the ICT systems in use, as an observation at the acute ward in hospital C shows:

The nurse wants to perform a DOSS score. Yet, before he can do that the system forces him to perform another score related to elderly patients in the EPD first. He does that, asking himself questions like: “hm...how red is the sternum...lets say x points...” Thereafter, he finally can do the DOSS indeed. (Nurse pauses, looks at me) He laughs and explains that his main problem with the scores is that he doesn’t have time. At moments where his time allows, he fills in scores regularly as he does not question the relevance. Often, however, he scores while performing other care tasks with the patient and then fills in the score later (observation acute ward hospital C, 12.12.12).

In this case, the EPR forces the nurse to submit other scores before entering the DOSS, but as he doesn’t have the time to actually get back to the patient to measure, he just fills in a score he thinks is about right and then goes on with the process. Another observation in hospital C confirms that these are not isolated instances:

The nurse makes a ward round. She is alone now, her colleague left for a late lunch. She registers vital parameters (temperature, pulse, RR, O2saturation, alertness) on paper charts that hang on yellow boards at the foot of the patient beds and on the mobile computer. She argues that during ward rounds doctors can better see what happens with regard to vital functions if these are visible on a paper curve. Double work, she sighs. She also tries to do all the scores she has to do, yet she is interrupted constantly. At one point, she quickly wants to enter the parameters of the patient in bed [number] in the EPR, which she measured a minute before on a little piece of paper she carries. She wants to do that before a new patient arrives. Yet, the EPR generates a pop-up that tells her that she first has to score the pressure ulcer risk of the patient before she can enter the vital function parameters. The system is blocked and she can only continue working after she scores the patient with regard to pressure ulcers. She sighs and enters some random data, which she thinks could be appropriate. “It doesn’t work out otherwise, I have to do some guessing work here.” Yet, because the patient just left to the

toilet, and because she has no time to score all patients three times a day, she enters data that she thinks can be appropriate (observation acute ward, hospital C 12.12.12).

At times, nurses '*work around*' formal requirements and thus make up scores to prevent disruption of work. Here, guessing scores is a reaction to very 'mundane' problems of nursing work: the second nurse is out for lunch, the new patient is about to arrive and demands attention, a small time slot emerges, and the system demands her to perform a particular score immediately. She synchronizes demand and available resource by guessing. This is not congruent to what literature describes as 'gaming,' and what would imply the wilful and strategic manipulation of data for the sake of better results. Here, a nurse takes a *situated decision* that helps her to make her work effective and to synchronize the mundane (and sometimes inflicting and messy) demands of health care work. The wilful manipulation of scores in our case is not of the strategic kind but an effort to keep work processes going both, in times when bottlenecks emerge *and* in mundane everyday situations.

At other times, professionals also told us that they didn't find certain scores to be relevant at all, so they again worked around the system by just giving scores, but rather as a form of 'pragmatic compliance'. An example from hospital A:

The surgeon argues that every time new measures come up, referring in particular to the stoma and prostate indicators. "The hospital board wants us to participate in that", he argues. But then goes on to say "I just give the desired scores. Taking a biops in one day is impossible, but I just indicate that we do it nonetheless. I don't spend more than five minutes on this. It is uncontrollable" (observation hospital A, 13.11.13).

Tick boxing might also occur because the amount of data to be registered is just too much. A medical specialist from hospital A told us this was also the case in some of the professional registries that have been developed over the last years, and that have become increasingly popular.

Well, I do think that these kinds of registries are good. I know from surgery that...well they have started this and their registry is so extensive; there are so many things in there because, well if you don't measure it, and then later you find 'well, actually we should have measured it'. So they have this extensive registry because they don't know. They have looked in the UK how they have done it over there, and that was a nice registry but lots of things weren't in there, but what you do need, nobody knows. So what you then end up with is a registry that takes so much time...you have to go through your records to find everything and because of that you always get like 'well, let's just fill in this, that should be about it', because if it takes so much time...'.
(medical specialist, hospital A, 12.12.12)

In this case, then, uncertainty about what measures are actually relevant induced registry developers to expand the range of measures to be filled in—a problem that may, or may not, be reduced in time when it becomes more clear what actually does matter—causing professionals to 'tick box' the registry.

Intentional gaming can also be an effect of hospitals knowing the registrations to be wrong, for example when things are not registered but are actually done because they are part of the routines of the organisation. Take an example from hospital C:

Antibiotic donation one hour prior to surgery is considered effective to prevent post-surgical wound infection. Regularly, there is no information about antibiotic donation in the files. The hospital used to register this patient *internally* as 'no data available'. In the beginning, one also decided to register such cases as 'not sufficient with regard to timing antibiotics.' "If I don't do that professionals never take care that the registration is ok. So, in the beginning we [RvB] decided to do this, this led to the registration degree improving dramatically. Internally." "And externally?" I ask. "This depends on the department of quality and safety, and management has to decide what to present. If the issue is purely about registration but you do know that it is actually done well...with some indicators you know that your score is bad but this purely relates to registration degrees or 'wrong filling', then we take samples. And if the samples do give the correct picture, you can possibly depart from what you have measured. This is a choice. For example, they do always give antibiotics in the surgical room, there is a procedure for that. But they don't register. And then you can ask yourself...The click in the system they forget...Do you evaluate them badly for this.... Externally, he recalls, he experienced that management decided to take samples and thus circumvent to include non-registered patients that would dramatically decrease the score." Internally, all data is presented. (observation hospital C, 19.12.12)

Again, this is not an example of 'classic' gaming, where organisations just make up numbers to make themselves appear better to the outside world. Given that what are known to be the routines at the operating theatre—both in terms of handling antibiotics and in registration practices—it is clear there is under-registration and reporting the 'bad' scores would seem unnecessarily harmful to the organisations *and* not representing actual post operative wound infection prevention practices. By redefining inclusion criteria, however, the hospital also nevertheless cheats, in that it doesn't report what is in the registries. In this case, the tension this brings is handled by keeping the original 'bad' scores for internal use—and use them for improvement purposes.

Similar issues occur for example when different information systems have to be combined in which different registration practices occur, which happens frequently as most registration leave leeway to professionals. All hospitals therefore report problems that arise where the definition of indicators is not in coherence with the work of the professionals. For example, with regard to gastrointestinal care in hospital C, the number of endoscopies was too low according to professionals. In consequence, systems were checked and one found that 1) some multi-morbid patients were registered for another DRG already (and thus not subsumed to endoscopy) and that 2) professionals used different DRG codes than the department for quality and safety defined previously. In consequence, a significantly lower volume of endoscopies was registered, which, if published, would have impacted on volume demands (observation, hospital C, 19.12.12). Similarly, in hospital A there was a suspicion that high mortality rates in oncology could be a consequence of registration problems; e.g. patients with cancer that died from other causes, say long infections, were registered as dying from cancer; other hospitals might be registering these as 'pneumonia' (observation hospital A, 12.12.12). Indicator registration is not straightforward and demands local ways of coping with interpretative freedom and generating synchronization.

Performance indicators and registration work, as we have shown in this section, causes paradoxes and ambivalences for hospitals. On the one hand, hospitals are well aware of the consequences of reporting performance information and of the interpretative work that this entails. While engaging in such interpretations, they want to be honest but at the same time not unnecessarily harming their reputations. While

indicator scores are seen as occasions for feedback and learning—and thus inducing honest reporting—there are, on the other hand, many situations in which compromises have to be made. This occurs both in clinical work processes, where registrations may be in conflict with the primary obligation of care professionals—seeing to their patients—but also at management levels where interpretations have to be made how to deal with missing or conflicting registrations. This is not gaming in the traditional sense of the word—which includes an element of purposeful misreporting, or plain fraud. Rather, indicator reporting forces hospitals and hospital workers, to engage in interpretative work while trying to keep to their sense of honesty.

What is striking at the same time is the general feeling we encountered that whereas their own hospitals were being honest, others must be cheating. Research on rankings in law schools in the United States has found a similar phenomenon (Sauder & Espeland, 2009). One of the effects of rankings thus seems to be the creation of distrust between organisations. If we assume that the hospitals we studied are no different than others—and we found no indications that this would be the case—such distrust has no empirical base, as all hospitals engage in interpretative work while remaining honest. Distrust may be seen then as one way of dealing with or ‘explaining away’ the volatility of rankings, but is at the same time an expression that interpretive work is necessary and will lead to differences between hospitals even if underlying ‘performance’ is similar.

Unintended consequences in care processes

The unintended consequences of rankings however do not only relate to registration and interpretation work. It also relates to care processes themselves. In this section we describe some of the examples we encountered of such unintended consequences in the hospitals we studied.

One reason for unintended consequences to occur is the performance indicators may be in conflict with each other—or that different values are at play. An example comes from an observation in hospital C:

In a meeting of the committee for delirium patients (15.11.12) the conversation quickly moves from the indicator for delirium to another indicator in the 10 themes bundle [of the Safety Management Program] that is of relevance for vulnerable elderly, which is the fall prevention indicator. The geriatrist explains that he encountered physiotherapists who delayed the release of patients because they felt that the patients encounter a risk of falling at home. He argues that it is not within the range of physiotherapists’ competences to delay release, moreover he feels it is difficult to evaluate whether patients are less well able to walk safely after hospital release. The care manager agrees and argues that in that way the fall prevention indicator increases the risk of hospitalization of elderly patients, which also increases an important outcome indicator, namely bed occupancy times. The geriatrist concludes that physiotherapists have to stick to their professional tasks, which is not to prevent the release of patients—even if they risk falling at home (observation hospital C, 15.11.12).

Indicators, the example highlights, may generate unintended conflicts for dissimilar values of quality of care, such as fall prevention, hospitalization and bed occupancy. Whereas fall prevention is a value that is certainly carried forward by the hospital, length of stay is also an important measure—not only in terms of financial consideration, but also because increased hospitalisation may induce additional risks to the patient. Sometimes, conflicting values also lead to professionals refusing to register altogether; an example being pain scores with psychiatric patients, as these

tended to anticipate on the pain scores, knowing that higher pain scores might get them more and heavier pain killers, thus stimulating substance abuse (observation, hospital A, 24 October 2012).

The example of fall prevention above also shows that indicators may cause, or become a focal point for interprofessional struggles in the hospital. Indicators always represent 'partial views'—that is, they focus on one aspect of care against many possible others. Such partial views may also be supported by specific professionals, in the case described above psychiatrists and physiotherapists. Whereas these professionals would normally not be in conflict with each other, specific indicators may point them in different directions, thus causing the need for 'coordination work' (Mol, 2003).

Such interprofessional coordination is also sometimes necessary because there is no natural 'owner' of an indicator. In hospital A for example fall prevention was struggled over between the physiotherapists and one of the neurologists. While the delirium indicator in hospital A was taken up by the recently appointed psychiatrists, this was not to say that other specialities accepted this. As the quality manager indicated, a committee was installed in which all relevant specialities were engaged, and small steps were now made in resolving the issues (observation, hospital A, 24 October 2012). In hospital C, great effort had been invested into screening for delirium. "We have given clinical lessons to health care professionals for how to register and score delirium, yet the implementation of the indicator does not work out" (care group manager, hospital C, 07.12.12). One reason for this 'implementation problem' is that delirium is not a topic on which professionals can distinguish themselves and their work. Like many cross-sectional indicators, it does not help professionals to develop their careers (geriatrist, hospital C, 09.01.13).

New risks/unsafe practice

Sometimes, performance measurements even create risks. Our observation in hospital C has generated an interesting case in this respect.

[On the oncology ward] a nurse argues that scoring lists are not always attuned to the needs of the oncology ward. She argues that standardized scoring lists even increase the risks to her patients. She takes the pressure ulcer scoring list as an example. The score functions according to risk categories such as age, weight, surgery and skin but ignores chemotherapy, which radically increases the chance of skin irritation. Thus, while chemotherapy patients might appear as low-risk patients and only have to be scored on Mondays, Wednesdays, and Fridays, oncology patients demand for constant skin supervision, the nurse argues. Following the standardized protocol could thus endanger patient safety, and the nurse concludes: 'If I followed the scores, I would lose my patients' (oncology nurse, 05.03.12).

The nurse stresses the risk of over-reliance on scoring lists, which build on indicators. This case of scoring exhibits a well-researched phenomenon of standardization: while standards are stable enough to travel across many locations, they are not situated enough to be relevant for specialized conditions, such as found on an oncology ward. Deviation from the standardized scoring system instead enables safe working practice for the oncology nurse.

Tunnel Vision

As performance indicators drive much of the work of professionals—both in terms of administrative work but also in new practices introduced to improve scores—there is a constant worry amongst professionals if they are actually working on the things that are most relevant for them, or for their patients. Registration gives feedback but also drives attention away from the patient, and with the increasing amount of measures

to be taken, this can become a problem. As a consequence, as we saw above, professionals sometimes just 'tick box' to get the registration work done, or don't register altogether. Overly relying on indicators scores, can also generate new risks, as the example from the oncology ward in hospital C showed, and the nurse had to watch other things beside the standardized pressure ulcer scores to perform good care. But ignoring registration work and standardized measurement only goes as far as it goes, as management levels in the hospital can put an increasing pressure of professionals to register, as we have seen in the previous chapter. In this case then, for example the focus on nursing indicators in hospital A can drive attention away from other types of risk or other areas of quality not measured.

The quality manager in hospital C likewise argues that some indicators might be more relevant than others for specific wards, but nevertheless need attention. In an interview, he argues for example: "At times, I addressed the broader underlying themes under the indicators. But this can lead to that the care process improves considerably, but for example your pressure ulcer percentage remains the same." He explains that if nurses improve quality based on priorities,

It can take a long period until pressure ulcers are on the agenda, for example fall prevention is much more dangerous and thus more important. But fall prevention is only on the indicator list since the last two years. Thus, taking a broad view on quality policy can mean you don't do well with regard to indicators. This is not viable. You have to focus on the specific, prescribed indicators. We now steer much more specifically to score well in indicators. (...) The negative effect is that you steer towards what someone else finds relevant who is out of our local context. Pressure ulcers are not really relevant for us. We would have better started with delirium here. (...) We primarily focus on nationally relevant indicators here, this means we have to steer towards things that are not primarily relevant to us' (quality manager, hospital C, 07.12.12)

To prevent too much tunnel vision as a consequence of indicators, the hospital also develops indicators locally, such as care pathway indicators (e.g. timing of processes, patient experience).

Volume standards

In the time when we were studying the hospitals, volume standards dominated much of the discussions on performance indicators. As we describe elsewhere in this report, volume standards (from professional associations, the healthcare inspectorate, or insurers) guided strategic discussions in the hospitals and were an important background why hospitals entered talks on mergers. Whereas concentration of care certainly is an intended effect of volume standards, it is questionable if mergers are as well. However, besides from mergers, concentration more often lead to negotiations between specific specialties on which patients would be treated in which hospitals, dividing the share of patients thus that all hospitals in the region would live up to volume indicators for specific diseases. This then often entailed only surgery, as patients would still be treated in their 'own' hospital, leading to patients moving to specific hospitals for surgery, while pre- and aftercare would still be given in the hospitals they came in first (see (Zuiderent-Jerak, Kool, & Rademakers, 2012) for similar findings).

Volume standards, however, also affected care in other ways—or were at least suspected to do so. As one of the surgeons in hospital A for example indicated:

...of course it has never been indicated that you improve quality when you do something 100 times in stead of 20, even though your intuition might indicate it does. But, well, take for example the surgeons from [neighbouring hospital],

that operate a lot in our hospital, doing some of the complex surgeries, like bladder removals. What strikes me is that the more they do them, they start doing more 'weird' things, like indications where I would think 'I would never operate on this patient.' Also, they operate much faster, but the amount of complications is not less. So I don't know if this really is improving quality (medical specialist, hospital A, 12.12.12).

Volume standards—while building on intuitive logics that practice makes improved quality—thus also lead to unintended consequences in terms of broadening the range of patients that could be operated on. This might be related to increased skills, but some of our responders also suspected that 'reaching the target' also played a role. The same surgeon from hospital A for example indicated that referrals to a regional centre for specific operations actually dropped after introduction of the volume standard, suggesting that hospitals were operating on patients they would have referred in earlier days, otherwise risking to be below the target and lose the possibility to treat these patients altogether.

Narratives and stories

Performance indicators and rankings are practices of quantification of qualities, working from a process Espeland and Sauder coined 'commensuration' (Espeland & Stevens, 1998). As became clear throughout our study, hospitals—including care professionals, quality and information managers and general managers—have to do much work to make care comparable, including solving sometimes conflicting or 'incommensurable' notions of quality of care. It was also clearly recognized though that numbers are not enough for quality improvement—or for accounting for care for that matter. Sometimes this resulted in outright resistance to quantification, but more often narratives and stories were brought in to give meaning to the data and/or to find out what underlying processes were actually at play. The executive director of hospital B for example emphasised in an interview: "By looking at indicators alone, we will not get better. We have to understand what is actually happening." Likewise, the quality committee in hospital C indicated that indicators should lead to '2nd order learning' processes, emphasising the qualitative analysis of performance data.

Sometimes, qualitative techniques were also used to overcome difficulties with quantitative one. For example, hospital A had difficulty getting enough patients for the CQ-index as they served a large immigrant population. Response on patient related outcome measures (PROMs) were therefore other too low for meaningful statistical analysis. As an alternative the hospital turned to qualitative techniques like focus groups interviews. Such techniques lead to better insights in patient experiences, but are difficult to then quantify (observation hospital A, 13.11.12).

The executive director of hospital A in general felt that narrative techniques would be better than just numbers. For example, the hospital had had three mortality cases in a row concerning anticoagulation. These mortality cases had been a reason to analyze the medical files, and it was found out that all patients were on three types of anticoagulation at the same time, prescribed by different doctors and for different reasons. "And then they climb out of bed at night, fall, and die." The hospital searched the literature for a protocol, but that wasn't to be found. "Well, then we get to work to make one ourselves." (executive director hospital A, 22.06.12). Cases like this, the executive director argued, showed that interpretive analyses of underlying processes was more important than just numbers.

Concluding notes

In this chapter, we have set out to analyse some of the consequences of performance indicators and rankings by looking at some of their 'unintended

consequences'. These included 'working around' administrative systems by 'tick boxing' indicators, negotiating what values should be pursued if indicators generated conflicting strategies, and a reshuffling of patients to live up to volume standards. Indicators, so it seems, at several points generate feelings of unease in hospitals. This had to do, first, with a recognition that quantification of quality necessarily entails interpretative work, and that hospitals might do this in different ways, creating distrust between hospitals while all hospitals nevertheless pursue honesty in reporting. While this confirms on the one hand that reliability and validity of indicators is low, as has also been shown by others (Kringos et al., 2012), it also calls for reactions in the hospitals that emphasise qualitative techniques are necessary to identify underlying processes of indicator scores, or are critical of quantification work overall.

Contrary to sometimes suggested in the literature (Bevan & Hood, 2006), we did not find outright gaming processes, although there are some indications that some indicators do induce 'unintended' behaviour. Particularly volume indicators seem to be driving hospitals and specialties to strategically position themselves, and change their care practices. Processes of tunnel vision do seem to occur, but are sometimes also countered by shifting external and internal use of indicators, or by prioritizing some indicators over others.

In the next and last empirical chapter we turn our gaze towards the question concerning if and the ways in which performance indicators and rankings influence the (internal and external) governance of hospitals. To what extent do rankings affect the 'politics' of hospital organisations?

Chapter 5 How rankings are affecting hospital governance

In hospital C, a quarterly management meeting, where the quarterly report is discussed, indicator compliance is one subtheme on this agenda.

The quality manager presents the intermediary analysis of the IGZ indicators and argues that the report “deals almost exclusively with numbers, and not with improvement.” He asks how the managers think to move towards improvement. One of the division directors replies that this year, there was no capacity to deal with improvement, that registration took all available capacities. The quality manager replies: ‘So, where do we take action? Where do we improve? How do we steer on basis of the indicators? We have to show this.’ The executive director supports the quality manager and argues that so far indicators are used more as means to control compliance than as a tool to improve quality. (observation, hospital C, 12.11.12).

The meeting described above is one out of many observation moments that extrapolates the ‘gap’ which professionals and managers alike experience between quality improvement and quality infrastructure. This opens the floor for the question of how indicators help to steer quality improvement, which we discuss in the following paragraphs.

Hospital C is confronted with a paradoxical situation. While the hospital’s indicator and quality compliance system is described as ‘exemplary’ both by the Inspectorate and actors from the field, respondents repeatedly argued that this infrastructure does not help to improve quality of care all the time. Likewise the head of the patient service office in the same hospital argues, when being asked as to how she uses patient-related indicators for her work:

Not much. Actually there are too many things we have to measure only for patient-related issue: The AD, The Elsevier, the CQ, IGZ indicators... We try to incorporate important indicators such as the net promotor score in our own patient experience survey. And we use rankings to generate a sense of urgency in the hospital. It helps to discuss with division directors about: Where does it go wrong? Where do we have to go to?” (head patient service office, hospital C, 01.11.13)

The manager of the patient service office and the quality manager for example argue that they exactly know where things go wrong and that they know the hospital’s ‘pain points’ (manager patient service office, hospital C, 01.11.12). Yet, the quality manager argues, improvement does not keep step with the compliance system (quality manager, hospital C, 1.11.12). Hence, while an elaborate system for indicator control is in place, it seems to be decoupled from actual quality improvement work (Power, 1997).

Increasing decoupling between performance based governance structure and performance improvement is described as a serious issue in all our case study hospitals. One empirical questions, then, why decoupling of quality improvement and indicator governance takes place in the first place – a question on which we elaborate in the following. We start with the managerial challenge of integrating indicator-based steering mechanisms in the hospital.

Managerial challenge: the difficulty of indicator-based steering

Indicator-based governance generates managerial challenges for hospitals, which we elaborate in the following. Asking managers and directors about how indicators help

their effort to guide quality improvement on basis of indicators, it was repeatedly argued that the public exhibition of indicators (and rankings) generates a considerable time slack between publication and data collection, which is for example in the case of IGZ reports. This makes it difficult to steer on basis of public indicator publications or rankings. This challenge of indicator-based governance is recognized in all case study hospitals.

In hospital C, the department for quality and safety therefore introduced an in-between measurement. The measurement delivers an analysis that is clustered per care group and it evaluates how professionals comply with registration duties, and how care managers steer on improvement. It also compares the hospital's ambition to the achieved results. The evaluation is shared with all division managers and those responsible for particular indicators 'as a means to help managers to better steer on indicator information' (quality staff, hospital C, 01.11.12).

Next to the time slack, also the distribution of responsibilities generates challenges for indicator-based governance. Observations from a management meeting in hospital C, in which the monthly progress in implementing the safety management program is discussed, is insightful in this regard:

The group moves to medication safety and that one of the indicators—wearing gloves during medication preparation—is not followed. The care group manager in charge of that indicator argues that the doctor carries responsibility on ward that hygiene is organized. The doctor again argues that many of the hospital's labs do not have sufficient glove boxes at hand to comply with the indicator. The pharmacist starts to laugh and tells the professional: "Then I bring you gloves." The care manager interrupts promptly and replies: "No, I expect my nurses to do that." The executive director asks how such micro-issues can be resolved (research diary, 01.11.12).

The example shows how actors 'boil down' problems until they appear to be trivial micro-issues (no glove box in ward lab) without resolving them. It is illustrative of how actors tend to boil down and thus shift responsibility to others (nurses in this case).

Indicator-based governance generates managerial challenges for hospitals internally. As we show next, this internal challenge not only comprises questions of internal performance governance but also comprises questions of how to organize interaction with external actors, as the case of prioritisation demonstrates next.

Dominance of external accounting practice: on prioritisation

The executive director of hospital C, when being asked about the relevance of indicators to steer performance improvements argues that he feels that there is a "wild growth" of indicators that sometimes "emerge out of economic interest" and that hence not all are of equal relevance for him (executive director, hospital C, 12.11.12). Likewise, the executive director in hospital B argues that:

These indicators... We then discuss it with wards and ask... You know, what doesn't work out is to run through all issues on one ward. You need too much time to do that, but we do pick out a couple of issues. (...) There is a multiplicity... this also shows that it is unreal, you always have to use your brains in addition. (...) If you think this through then there is sometimes only little reason to follow it all. (...) [There is plenty of discussion about particular indicators nationally], and there is good evidence [for the indicators], but in reality and in practice it is simply not feasible (executive director, hospital B, 02.05.13).

Currently, all of our case study hospitals find it increasingly difficult to adhere to the multiplicity of external demands. A division director from hospital C argues that the overall Dutch trend for performance measurement generates an overflow of demands to collect indicator data.

It is like driving a car, there you can also only handle 7 processes at a time. The same is true for hospital governance. We put too much effort into measuring. We simply measure too much and improve too little. Numbers alone don't say anything (division director, hospital C, 29.11.12).

He argues that prioritisation of tasks is the only way out of this multiplicity so far. Respondents in all case study hospitals agree with this. Likewise, the executive director of hospital A states:

I do think that the thing we have to achieve is that we have to identify for ourselves what our hospital-wide core indicators are and on basis of what we make our own rankings and where we formulate our own goals. But we are not yet so far. We now simply adhere to all indicators and we steer where there is an anomaly (executive director, hospital A, 06.11.12).

So far, the policy of adhering to all available indicators demanded by external actors was the main strategy to deal with the multiplicity of external demands. However, as for example the head of the medical staff in hospital B indicates, the multiplicity of demands that emerge and steadily increase from external actors has to stop somehow. He therefore argues that "prioritisation is a hot topic" (head medical community, hospital B, 02.05.13). One of his colleagues, a member of the indicator steering group, already proposed to the executive director to focus on top 20 % of all indicators and no more than that (medical professional, hospital B, 21.05.13).

Prioritisation work is an important governance reply in response to the increasing amount of indicators and ranking information that from outside parties. Hospital C is an exemplary case and frontrunner in this respect. The executive director openly indicates that some indicators are more relevant than others. An example is post-operative wound infection (POWI). Taking the case of hip replacement, the executive director explains that the DRG system allows the hospital to get reimbursed for hip replacement, but not for a post-surgical infection. Complications are on the pay roll of hospitals. The financial consequences of indicators such as POWI transport these high on the agenda. Observations show that indeed POWI receives much attention in management meetings and the quarterly reports alike (executive director, hospital C, 12.11.12). Recently, the hospital has developed an internal priority setting system that prioritizes indicators which are demanded by public authorities that are of high risk for patients and that "help to resolve resource shortage." Increasingly, patient organization rankings and indicators of health care organizations are considered (division director, hospital C, 06.12.12). Likewise, hospital B handles a prioritisation system, as the executive director explains. Also, hospital B prioritises particular indicators, yet in a very different way, as the executive director explains in an interview:

We are alert for significant deviance, both upwards and downwards [in the ranking]. Thus if you are a ward manager and your pressure ulcer score is really bad, then we pick that out. (...) [Otherwise], I do prefer to stay with individual professional responsibility in place of a collective hiding exercise [i.e. behind performance demands from outside] (executive director, hospital B, 02.05.13).

The executive director argues that his main criterion for prioritisation is 'do-ability' and such indicators that are considered relevant by medical communities and that are thus able to trigger 'professional responsibility.' One of his care group managers adds in a conversation:

I always try to give priority to such indicators where I can see a direct relationship between indicator and work practice. (...) For example, the volume of repeat operations. There is a direct link between ...lets say...you can push one button and that has effect. And a repeat operation builds on indicators that are...ah...there is clear causality between behaviour and outcome. And in such cases, you have to have discussions. (...) What you also do [with regard to choosing indicators]...you look at where you can generate benefit rapidly and that leads to visible improvement (care group manager, hospital B, 02.05.13).

Thus, while hospital B's prioritisation system is less formalized, it likewise pays particular attention to the demands of governmental authorities (and increasingly to patient organizations' and insurance companies' demands). Prioritisation also knows quite informal work practices, as a respondent from hospital A highlights:

The CQ-index is quite a dubious thing if you ask me ... You have to check the website. This is a commercial business doing that. It is a sponsor of the Breast Cancer Association Netherlands. And the association did make that test mandatory...to achieve the pink ribbon. This is a really strange interaction. The first year [the hospital A] therefore refused to collaborate for this reason. And we thought: 'Leave us alone with your pink ribbon, people come [to the hospital] nevertheless. (medical specialist, hospital A, 18.02.2013)

Dissent with external demands, the quote shows, helps hospitals to cluster and thus prioritize their performance work. Prioritisation is a first step to limit external accounting demands for the sake of internal coherence and thus also performance improvement. Next we will demonstrate how also the hospital's drive to accommodate performance measurement practices locally supports hospital's drive to limit external accounting demands.

Governance beyond 'one size fits all'

Hospitals accommodate performance measurement towards local needs, as we show next. Hospital B characterizes itself as a Lean hospital, where performance improvement is individually pursued on wards without a top-down hospital wide improvement policy/strategy. Here, professionals have a strong standing and indicator governance is a negotiation-based process, as a conversation with the executive director explains:

We give primacy to our professionals. If professionals say that something is wrong, then we let them argue why it is wrong. We [i.e. executive managers] then often take over that perspective. And of course one would listen to their argument carefully. If their argument is 'I simply don't feel like it' then that is a totally different story then 'come and see, I show you how things are related to one each other' (executive director, hospital B, 02.05.13).

The hospital's strategy to prioritize professional accountability above strict performance compliance is related to the hospital's history in decentralized

governance and its experience in Lean management. Any form of compliance with performance technologies has to be adapted towards this local structure.

Hospital A offers professionals space for negotiation and adaptation as well. For example, in the case of breast cancer care the hospital was losing points because of the policy of the surgeons to aim for breast-saving operations whereas the *Elsevier* ranking is directed at the prevention of repeat operations due to remaining cancer tissue. Although the 10% of patients with remaining tissue was within the standard of the specialist association, the board of the hospital nevertheless convinced the surgeons to change their policy. "Now we are well within the range of the IGZ indicator", the executive director explains in an interview. This doesn't mean however that all processes are changed as a consequence of indicators.

Sometimes we decide that we don't want to comply (...) One example concerned the urologists where there was a story like 'you have to do this within seven days', but our process was organized differently and that was actually better for the patient, because we do all the diagnostics at one point in time, so we decided not to change (executive director, hospital A, 06.11.12).

In hospital C, on the contrary, professional-led performance negotiations is exclusively distributed into external networks. Observations from a management meeting are insightful in this respect.

Managers and doctors hotly discuss whether lowering the age from 70 to 60 makes sense, and whether screening for delirium is a useful indicator (with regard to investment and return for patient safety) after all. The executive ends the discussion and argues: 'We have to put a stop to the discussion. We lose out in important points in the AD list due to the delirium case. This simply has to happen' (executive director, observation hospital C management team, 15.11.12) .

This 'it has to happen' argument was often encountered in meetings where professionals started to question the relevance of (particularly IGZ-based) indicator sets. While this governance mode helps to end discussions amongst managers and professionals, it redistributes negotiation into informal, often profession-based circuits. This does however not imply that the hospital's rather strict focus on compliance with prioritized external performance demands prevents appropriation of external performance demands. We have already discussed the hospital's systematic prioritisation system as a means to adapt external demands towards local capacities. Another example pertains to how managers in charge for implementing performance technologies give meaning to rankings, as the following observation shows:

In response to the AD result, the quality manager published a letter on the intranet. Here, he publicly announced the bad result and where the hospital missed out on points. The letter then continues asking: 'Does that mean we are less good than we were?' He answers the question by elaborating that in comparison to the last year, the hospital did actually improve. 'With today's numbers we would, in last years' ranking, end a couple of places higher.' Yet, the letter continues 'the hospital was less fast in developing further than other hospitals' (document analysis, hospital C, 18.12.12).

The quality manager's letter stresses that rather than doing particularly bad compared to others, the hospital's speed of improvement was not high enough compared to other hospitals. He redefined the meaning of the ranking. In a conversation he explained that this was an explicit strategy to remind the staff of the

rankings' relevance and a means to keep staff motivated to invest into indicator-based improvement (personal communication quality manager, hospital C, 17.02.13).

Performance governance, then, is never a one on one translation of external demands, but seeking alignment with a hospital's overall organizational structure as much as with the meaning that is attributed to particular results in rankings. While rather top-down oriented hospitals, such as case study hospital C, organize performance governance in form of compliance, hospitals that operate on a decentralized governance mode like hospital C do rather build on negotiations (particularly with professionals) to find appropriate modes of performance governance.

Nevertheless, external demands for accounting tend to dominate organizational set-ups, as the case of hospital B demonstrates. Hospital B operates quality improvement based on Lean management methodology, which implies that there is no top-down hospital wide improvement policy/strategy (see above). Unfortunately, this approach of "undirected, organic improvement" where wards experiment on quality improvement (quality staff, hospital B, 21.05.13) generates problems with hospital-wide coherence in quality improvement work, too (care group manager, hospital B, 02.05.13). Quality indicators are external demands that have to be complied to hospital-wide, and so far indicators are not included in the Lean improvement plan yet, particularly because management feels that this would generate resistance with professionals. The bottom-up logic of Lean management in hospital B stands in contrast to top-down hospital-wide indicator governance. In hospital B, then, local governance relations will necessarily have to change from bottom-up to more centralized governance approaches, if the hospital wishes to integrate a coherent monitoring system for compliance with external indicator demands. The dominance of external accounting, then, is also represented in local governance structures. Indicators and thus also rankings change governance relations in hospitals. And while the above is an example of how external demands dominate local governance structures, our next example is about how the same pressure also generates locally adapted structures for QI.

Tin opening

Elsewhere, it is argued that performance measurement can be used as 'tin openers' to start conversation about care practices and improvement (Freeman, 2002). And indeed, we have observed such practices. For example in case study hospital C, quality improvement lacks behind, such as the implementation of the nationally demanded safety system (VMS). Therefore, and also because health care insurers expected hospital to comply by 2013 (division director, hospital C, 11.12.12), the quality manager initiated a monthly management meeting where the evolvement of each theme was discussed. Observations in these meetings highlight that one for instance deals with the technical question of whether and how to generate one document for delirium screening, while different professionals operate with the document in the acute and elective patient stream (i.e. medical specialists/nurses respectively). These meetings move "beyond registration demands" and instead focus on what actors need to improve the situation (quality manager, hospital C, 1.11.12). Here, one tries to resolve the bottlenecks which performance measurement generates beyond registration demands. The monthly management meetings are therefore an example of how qualitative structures emerge as a reaction to insufficient quantitative performance that is so far not well understood.

However, the ability of transparency technologies to foster local negotiations about performance is a difficult endeavour due to the overall dominance of external accounting. Case stud hospital C is again insightful to demonstrate that point. An excerpt from the research diary:

I participate in the monthly management meeting that wishes to increase compliance with the national 10 Themes. Amongst others, compliance problems with pre-surgical antibiotic donation (one out of four interventions in the post-surgical infection bundle and thus an important indicator to reduce post-surgical wound infection) are discussed.

Quality manager (QM) to care group manager (CGM): from when onwards can you make sure that antibiotics are given before surgical processes start?

CGM: We do register almost 100 %.

QM: I can see that, if I look into the third quality report of 2013. Come on! But when do you improve the situation? (laughs, looks to the round of managers)

Medical professional (MP) enters discussion: I am not the champion ('kartrekker'), I only monitor antibiotic policies.

(CGM doesn't answer the question and later adds: I don't want to end in interrogation ('overhoring') here. (research diary, hospital C, 01.11.12).

While the quality manager argues that indicators serve as signposts to actually improve the current quality of care (i.e. administer antibiotics), the care manager argues that compliance actually means to register (i.e. 100% registration). The example thus shows that managers have dissimilar perspectives on what indicators compliance means.

This multiplicity of meaning is consequential for the impact of indicators on quality steering: while one manager tries to steer towards quantitative results (i.e. good registration degrees), the other is concerned with qualitative agendas (i.e. how registration can be translated into improved health care practice). It opens the question of how such dissimilar logics of performance improvement are reconciled. The example of safe incident reporting (VIM) offers insightful observations in this respect. VIM is an example of qualitative accounting, where one tries to generate learning on the basis of narrative evidence in no-blame environments. However, observations in case study hospital C demonstrate that what is understood as 'good' VIM practice is considered to be of a high quantity. The logic of quantitative measurement, then, remains dominant qualitative accounting tool.

Elsewhere, it is argued that 'numbers open up spaces through which organizations can reflect their performance.' (Bal, 2012) This is sometimes difficult in our case study hospitals, as the case of safe incident reporting (VIM) exemplifies. Likewise, we have demonstrated how the dominance of external accounting practices is represented in local governance structures. We might thus so far transport the impression that indicator-based performance management does not help hospitals to govern performance improvement. However, indicators and rankings actually do steer internal policy, as the executive director of hospital B argues:

We actually don't find rankings so relevant. (...) Sometimes, we pick one [indicator] that we find relevant, because we do find that [particular indicator] relevant ourselves. And then we use the indicators support our policy (...) This is then a sort of impulse the use of het indicator (executive director, hospital B, 02.05.13).

We will therefore, next, demonstrate how indicators and rankings actually do serve as 'tin openers' and how they particularly help managers to govern performance improvement nevertheless. Particular focus is on how managers use rankings and indicators to guide internal policy for quality improvement, starting with how rankings are used to influence interaction amongst professionals and managers.

In all case study hospitals, the collaboration of managers and medical professionals, and particularly steering medical professionals towards more compliance with performance measurement, is considered as challenging. Managers

in all of our case study hospitals use external indicator demands and rankings in particular to stir negotiations about performance improvement with medical professionals. A communications manager from hospital A remarks:

We are busy with quality and safety right now, with indicators and dashboards and so. We steer very broadly on the lists, and that works out. You make things visible. You are being watched as hospital, but as capacity group and as specialist as well, and this is what the lists [rankings] do make transparent. (...) It also makes things more easily to communicate and visual, and this also helps (communication manager, hospital A, 26.11.12).

Because rankings often appear in the form of lists, rankings and tables, the scientific outlook helps correspondence with medical professionals who 'are visual and like graphs instead of text', and as they 'like the beta-feeling of numbers'. Numbers help to 'mobilize' the medical community, as they allow 'to translate complex stories into simple numbers' and thus uses the ranking scores 'towards the direction of the professionals' (division director, hospital C, 29.11.12) .

Managers use rankings to negotiate with doctors. This happens in different ways, we start with how managers use indicators to generate a need for change. The conversation with a care group manager in hospital B is representative in this respect:

And the other thing is that we of course use [rankings and indicators] as management tool to get through to medical specialists...[to get towards] particular improvement practices in the care process that have to be done. Step one, the ranking enters [the hospital]. In the following ... [For example] cardiologists score badly. In consequence I go and see the medical manager, or do sometimes even visit the whole group of specialists, and I tell them: 'Guys this is really going badly here.' Then they would tell me: 'The numbers are not correct.' Then we first look at the numbers together which they delivered ... And I tell them: 'This number was delivered, and you signed it. How come they are not correct nevertheless? What is the reason?' Then they say that the case mix is ...different. Then you check this out. Then you tell them: 'From the benchmark it seems that that is not the case [i.e. incorrect case mix]. Then you approach the core and say: 'Guys, you still score low, we took away variability, and now we have to discuss what we can do in our organization, in our work process, in our medical policy, in our care process in order to make sure that there are better outcomes next time. But then, nevertheless, the ranking is for me still an instrument in order to effect change. Rankings are not a goal in themselves (care group manager, hospital B, 02.05.13).

In hospital A and B, indicator results are negotiated with medical professionals in the yearly contract negotiations.

The relation with the executive director did change a bit [duet of indicators and rankings]. Now, for the first time we consider indicators in the yearly contract negotiations. (...) And I think for a good reason. There are points for improvement... (medical specialist, hospital A, 12.12.12).

In hospital B and C, even a small percentage of the medical professional's salary is coupled to compliance with performance measurement technologies (division director, hospital C, 29.11.12).

This is what we have accomplished together with the medical specialists. This means that we make [quality and safety related performance measurement] this part of the contracts we negotiate with specialist in 2013- 2014. To put it differently, if they don't collaborate, then we have a reason to shorten their financial benefits (executive director, hospital B, 02.05.13).

Indicators and rankings are used as a tool to negotiate with professionals. Likewise, rankings are utilized to steer relations with the supervisory board. The executive director of hospital B explains:

If we score badly somewhere along our financial indicators. (...) [We] have much manpower working at the bed, much nursing power, and this is why we score badly. This can be a reason for us to say 'Ladies and gentlemen, it may be the case [bad score], but we are doing this very consciously, because we do it for our patients. And because patients mainly meet nurses here we don't put one single bit away with that. We also discuss that with the Employee's council (*ondernemingsraad*) and the Nursing council. So we continuously use ...What you do is that you surf on the facts that offer themselves for you. (...) And a ranking is a nice example of this. Thus, if you can make use of a ranking, such a wave, that that is a nice coincidence. But if you would ask me whether we set out a policy if the [ranking] appears over a week or two, then I say: 'Sorry, I don't find it that important' (executive director, hospital B, 02.05.13).

Like in hospital B, also hospital C uses rankings as a strategic tool to negotiate particular quality-related agendas with internal parties such as professionals and supervisory committees. And while indicators and rankings often serve as strategic tin openers for negotiations within the hospital about performance management, indicators and performance improvement may also correlate unexpectedly:

The focus on stroke and thrombolytic therapy is not inspired by the indicators. Those two coincided. We wanted to improve care because we know that 'time is brain', and the period between diagnosis and treatment was too long. Thus, this is why we focused on it. We were focused on treatment. And if you do that well, then you naturally score high on indicators. But the indicator was not the goal, we wanted to improve ourselves and stick out. That was a choice, and it made sure that we have more patients now, and that we are a preferred treatment centre for stroke for care insurers now (medical specialist, hospital A, 11.10.12).

Improvement activity, the respondent highlights, does not always stand in relation to or react to external demands of indicator-based governance. Also, individual, strategic goals of hospitals steer improvement. Overall, internal governance of performance measurement is often a matter of choice, where appropriate indicators are linked to clinical processes or strategic goals (such as financial considerations). This means that governance of performance is never a one on one translation, but a way of how hospitals are seeking alignment of performance measurement with hospital strategies. We will next look into how collaboration with external actors is also strategically relevant for reputation management within hospitals.

Governing reputation with external stakeholders

As we had mentioned above, hospitals often perceive indicator-based performance tools as largely organized in a top-down fashion with limited effect and benefit for actual performance improvement within the organization. In correspondence to this

observation, the executive director of hospital B feels that influencing on national policy scenarios for the sake of more locally appropriate performance measurement regimes is difficult. He tells that he repeatedly tried to give feedback about what his professionals consider 'bad' indicators to the health care inspectorate – 'unfortunately without much success' (executive director, hospital B, 02.05.13). He feels that medical couple organizations could be more powerful to effect change.

However, we observed that all hospitals try to establish collaboration with national organs, such as the quality institute, hospital networks and university-based institutes to find ways to make indicator collection work more coherent locally (observation, hospital C, 26.06.13). Actually, our research shows that in the context of performance governance, hospitals are very much oriented towards the external context. For example, respondents of hospital A went to meetings and conferences to stay in tune with developments in other hospitals, and a quality staff member recalls that she for example learned in that context that 'rankings have increasingly less influence nationally' while professional registries become more important. Likewise, the executive director of hospital A, who is responsible for performance indicator management in the hospital is closely linked to national indicator-development in his position as executive manager in the Dutch Association of hospitals. Here, he uses experiences from his own hospital to discuss national developments in the area of performance indicator while he also can stay informed about most recent national developments (observation hospital A, 12.11.12).

And indeed, our research shows, there is considerable leeway for negotiation with external stakeholders, such as the inspectorate, as an excerpt from our observations shows:

The IGZ inspector hints at that there is a high rate of wound infections for hip replacements. The executive director says that orthopaedic specialists would actually do good work, and that this also shows from what was discussed so far. 'The hips are most often done by surgeons.' He admits that there was indeed a breakout of wound infections, but that this case would also show that 'the indicator movement leads to action.' According to the executive director there were a plenty of rules launched to stop further outbreak (observation hospital A, 24.10.12).

It seems as if hospital respondents can easily explain bad local scores. Similar observations were made in hospital C, where paediatricians were able to explain a bad score on malnutrition easily. The quality manager of hospital C, when being asked about how the inspectorate helped him to push quality and safety agendas in the hospital responds that the rigour of inspection would heavily depend on the personal style of negotiation that respective inspectors had (observation, hospital C, 17.01.13)

But negotiation space with the inspectorate exceeds the negotiation of local scoring results, and extends to how standards are developed nationally. A division director in hospital C argues that he 'regularly lead[s] discussions with the IGZ with regard to the theoretical appropriateness of the indicators they use. Sometimes this has effects, sometimes not' (division director, hospital C, 06.12.12). The director also negotiated indicator sets with health insurers, who often 'deliver vague indicators'. For example in the case of dialysis he defined and refined the indicators together with health care insurers and also generated checks and balances for the indicator. Also, he developed quality indicators for patient with autoimmune diseases (logistics, home care) in collaboration with health insurers. He further recalls that indicators in care can generate negative effects. According to him there is too little 'balance amongst indicators'. He feels that there is strict governance with regard to individual indicators (e.g. length of stay).

You can heavily reduce length of stay by releasing patients too early. Then you score very well, the inspectorate likes this. But if you don't ask for control indicators too, such as re-admission and complication, then you fail. It did cost me loads of energy and discussion to convince inspectors that one can never only have one indicator to measure performance. You need a second, controlling indicator. (...) After years of discussion with the IGZ' the indicator 'length of stay' is now coupled with a balancing indicators that can prevent gaming [i.e. unexpected re-admission] (division director, hospital C, 06.12.12).

In hospital C, the division director influenced how indicators should be conceptualized nationally. Comparable evidence is obtained from a division director from hospital B, where a care group manager negotiated volume indicators for colon carcinoma with the hospital's main health insurer arguing that while the volume was lower than required the expertise of surgeons was nevertheless particularly high. A medical specialist in hospital A recalled that one health insurer was asking for migraine indicators, but 'that it seemed as if half of the country did not comply with the indicator, and that the insurer thus had to give in and adapt the indicator' (medical specialist, hospital A, 11.10.12). Hospitals do influence national indicator standards—be it because standards are set too high, or because they allow for too much space to manipulate results. In sum, not only indicators influence institutions, also institutional actors effectively influence on performance measurement regimes.

On a grander scale, hospitals have been able to influence the *Elsevier* ranking considerably, by pointing at the possibilities to manipulate scores. Until 2009, the *Elsevier* ranking was based on peer review—asking GPs and medical specialists which hospitals they thought were best. Building on the insight that doctors know best which hospital to go to, the *Elsevier* used this system to calculate its ranking. However, this increasingly raised criticism from the hospitals, who thought that this system can easily be manipulated, and hired a marketing consultant to argue this (van Hamersveld & Olivier, 2009). With an increasing number of hospitals threatening to refrain from participating in the *Elsevier* ranking, the magazine changed its strategy to base its ranking on ZiZo indicators and CQ-index data thought to be more 'objective'.

New practices of accounting

The emphasis on quantitative performance measurement has repeatedly raised questions about how such numerical accounting can represent actual performance in hospitals, as we have already seen in the previous chapter. The executive director of hospital B remarks in the context of discussing the value of each qualitative accounting how the hospital is increasingly putting emphasis on qualitative practices of accounting:

You can do participant observation, shadowing...I mean, these are all elements. (...) We already have a couple of examples, such as medical students that we send out with a camera and they follow patients. Psychiatrists are far advanced there... Every Friday afternoon they discuss for one hour with their treatment teams and a couple of patients. (...) Thus, we are busy with a couple of things, but if you ask me that is still minimal and that is a really important development (executive director, hospital B, 02.05.13).

Likewise, qualitative accounting practices have been developed in hospital C. Here, the analysis of performance data delivered showed that while the hospital was

performing well according to national indicators, still patients evaluated the hospital badly in terms of patient experience. In an attempt to improve, the hospital set up a 'social contract' with stakeholder organizations in the region. The social contract comprises so called 'care guarantees', which is a (sets of) rather explicit indicators. Based on user expectations, such guarantees translated care processes into explicit promises for patients and that can be claimed and thus sustained. For example, the care guarantee for lung diseases spells out how outpatient visits are organized, how long one has to wait for treatment and results, and how privacy is dealt with. New care guarantees are signed annually in the context of the hospital's 'Contract with Society'. In sum, the development of new accounting practices such as the care guarantees evolve partly as a critique of quantitative performance measurement in our case study hospitals.

But innovative ways of accounting move beyond qualitative, patient-centred accounting strategies. One good example of alternative accounting are registries. The head of the quality council of hospital B, that focuses to a large extent on non-centralized performance governance argues:

Surgeons are difficult guys, they are a heterogeneous group. (...) Surgical procedures are standardised, surgeons often do the same kind of surgical procedures. When I was an assistant thirty years ago, I worked with a surgeon who said: 'Breast cancer I do this and that way.' A man really convinced that what he did he did well. But it was based on nothing. There was absolutely no science behind that. He did breast-saving procedures while it was very difficult to sell such choices if you ask me. He had no single proof that it worked. In the course of standardization, all surgeons now do it the same way in that they see each others numbers [registry of the Dutch Institute of Clinical Auditing, DICA] and they discuss these with each other. This means that you standardise your procedures instead of artful intervention. You simply have standard surgical procedures, and this is enormously helpful as there is one best practice and that is scientifically researched – plus you have sufficient numerical support to do that. (...) The DICA above all is for yourself and helps you to look into whether your process is ok. It is much more difficult to have third parties having an opinion about [your work]. Then it becomes much more difficult to honestly treat ... puzzles (head quality committee, hospital B, 28.05.13).

Registries such as the DICA registry for cancer treatment appear to emerge as alternatives to indicator-based and public performance measurement. Particularly the peer-based learning experience that is based on evidence from individual work processes seems inspiring here next to the fact that registries allow for standardisation of professional work practices. Also, the element of fraud and trust plays a role here, as a surgeon from hospital A explains:

'Rankings pollute and lead to 'creative bookkeeping.' The Dutch Medical Association has a complication registration, and it is paramount for the sake of truthful registration that results are not made public.' The respondent goes on to explain that indeed the professional associations steer on basis of results of such registries. 'For example, two years ago the number of reoperations for colon cancer was too high in [a neighbouring hospital]. And a group came to check how that might be, it was very harmonious. That is much nicer than being put into the newspaper. (...) The risk is that if you don't send the results to the Dutch Breast Cancer Association (DBCA) but if you have to deliver them to [a daily newspaper], then people do fraud. Then you say: 'It was the patient's guilt: one smoked, one was too bulky, one cannot be counted at all...' (surgeon, hospital A, 18.02.13)

The respondent highlights that numerical results from registries which are discussed amongst peers have a better chance to actually resolve 'puzzles' in honest interaction amongst professionals and thus prevent fraud. Registries, as they are developed and owned by the profession, according to this respondent, lead to more trustworthy comparisons, as peer pressure prevents cheating. Due to a variety of reasons, we can conclude that registries are becoming important alternative governance tools for quantitative performance measurement. They are also, however, a result of a growing pressure on transparency, and as we have seen earlier, create their own problems, for example in the amount of administrative work needed.

Concluding notes

Indicator-based governance generates managerial challenges for hospitals internally. This does not only comprise questions of internal (system-based) performance governance but also comprises questions of how to organize interaction with external actors and how to *adapt* external indicators in such a way to make them useful locally. Hospitals accommodate performance measurement towards local needs.

Respondents coherently feel that there is a 'gap' between performance measurement and improvement. Our empirical study has demonstrated that in the course of multiplicity of external performance demands, improvement strategies rest on strategic managerial choices more than overall indicator compliance. Managers for example link indicators that are considered as locally appropriate to clinical processes or strategic goals or generate standardized systems to filter most critical indicators out. Prioritisation is a first step to limit external accounting demands for the sake of internal coherence and thus also performance improvement. Non-compliance with external demands, then, sometimes helps hospitals to cluster and thus prioritise their performance work.

This implies that overall, performance governance is never a one on one translation of external demands, but a way of how hospitals are seeking alignment of performance measurement with hospital strategies and existing structures. Yet, the dominance of external accounting is represented in local governance structures, as the case of hospital B and its challenge of bottom-up Lean management demonstrates. Indicators and thus also rankings change governance relations in hospitals. But likewise do hospitals influence performance measurement regimes, as the cases of colon carcinoma and the *Elsevier* ranking demonstrate. In sum, not only indicators influence institutions, also institutional actors effectively influence both specific indicators and ranking systems.

We have also demonstrated how new accountability structures emerge as a reaction to insufficient quantitative performance and we have for example demonstrated how registries are becoming important alternative governance tools for performance measurement on the basis of public reporting, and how hospitals can create local accountability regimes.

Chapter 6 Conclusion and discussion

Hospital rankings are a new reality in the governance structure of Dutch healthcare and play into the high hopes for increasing transparency leading to better, more efficient and more patient focused healthcare. As such they are an increasingly important infrastructure for the marketization of the Dutch healthcare sector, in which the lack of transparency has been—and still is—a public mantra (see e.g. Algemene Rekenkamer, 2013). With their appeal to numerical comparison of hospital organisations, rankings have been taken up by an increasing number of organisations, including public media and patient organisations. As such, rankings have become an industry in their own right, with an increasing number of private organisations being active on the ranking market.

Rankings are also much understudied. In this project we therefore set out to study the ways in which rankings are affecting hospital organisations and the people who work in them. We did so through a qualitative, ethnographic study, researching the ways in which rankings and their underlying performance indicators influenced quality and other policies and practices in three Dutch hospitals. Based on the assumption that the competitive environment of ranking might be of influence on the ways in which hospitals might react to them (Berwick, 2002), we selected three comparatively similar sized hospitals in low, medium and high competitive environments. In all three we interviewed members of the board, communication, quality and information managers, managers of clinical wards, medical specialists, and nurses, sat in with meetings of quality commissions as well as meetings with external stakeholders like insurers and the healthcare inspectorate, and—where possible—observed clinical work practices. This approach enabled us to get a detailed understanding of the ways in which rankings were talked about and acted upon (or not), and the ways in which indicator data were collected, aggregated, transformed, used, opposed, and communicated. It also gave us an understanding how and to what extent rankings were transforming social relations within the hospitals as well as between the hospitals and their environments.

The following research questions guided our research: (1) What structural and policy responses to rankings have been made at the organizational level in the Dutch hospital sector? (2) What different strategies have hospitals followed in their organizational response to rankings? And (3) to what extent and how do these structural and policy responses affect actual care delivery?

In this final chapter we first set out to answer these questions and then discuss our findings in the light of current debates on performance measurement and management in healthcare.

Organisational responses to rankings

Rankings, as we have seen in the three hospitals we studied, induce ambivalent responses. On the one hand, when being asked about the importance of rankings, our respondent generally felt that they were not important for the ways in which the hospitals operated. The lack of validity and the volatility of rankings, as well as the lack of direct consequences in terms of patient choice or insurer commissioning of care, all lead our respondents to argue that rankings were unimportant to them. Yet, on the other hand, it was felt that rankings had to be taken seriously, as they were one of the drivers of the increasingly important reputation of the hospitals. We have demonstrated above how organizational structures, personnel, departments and ICT systems are actually adapted to fit the demands of performance measurement. Likewise, not participating in rankings was for example not seen as an option. Moreover, rankings were seen as a new way to put quality of care on the agenda and change existing social relations (e.g. between central staff and clinical wards) in the hospital. Especially falling on the rankings seemed to be an important driver for

hospitals to improve on their performance. To some extent, then, rankings led to pragmatic compliance, as just another external demand that was put on hospitals and that they had to deal with. But organisational responses went further than just pragmatically complying with external demands—they lead to significant changes in the social fabric of the hospitals.

One of the important areas where rankings and the performance indicators underlying them affected hospital organisations was in the organisation of administrative work. Administrative work, as we saw in chapter 3, entailed many investments in form, including the introduction and use of many different information technologies, the training and disciplining of clinical staff to collect and register indicator information and the standardisation of care processes to enable data collection. Hospitals set up committees and steering groups to govern the process of data collection and use, and hired quality and information managers to aggregate data and report them to external sources, like the healthcare inspectorate, ZiZo, insurers and patient organisations. External consultancies were hired to enable data cleaning, or to do some of the data collection (like CQ-index information).

Performance data was also used internally by the hospitals to improve on care processes and stimulate learning, albeit that our hospitals differed in the extent to which they used performance indicators in a systematic way to enable quality improvement. While some hospitals focussed extensively on indicator compliance and set up systems to monitor both compliance and improvement, others did so on a more ad hoc basis, mainly focussing on those areas that negatively influenced their ranking positions. Similar differences seemed to exist concerning the external use of rankings. Although for all hospitals managing their reputations was seen as a new and important task and for example invested in social media policies and practices, rankings only in some of the hospitals figured prominently on the marketing agenda, whereas others either distanced themselves from using rankings in this regard, or saw rankings (and particularly their volatility) as an external risk rather than as an opportunity for marketing. Interestingly, the competitive environment in which the hospitals found themselves did not seem to affect these different responses; rather, quality improvement strategies seemed to be path dependent, with the historical choices of hospitals, e.g. for Lean approaches, steering quality policies to a large extent. However, rankings do seem to affect those policies, mainly by stimulating more centralized, managerial approaches.

Performance indicators and rankings also seemed to induce alternative accounting practices, both qualitative and quantitative. These included an increased focus on case reports as well as different forms of patient participation. One of our study hospitals also negotiated 'care guarantees' with local stakeholders, thus creating a new accountability context in which performance could be measured and discussed. Last, especially for some surgical specialties, registries were found to be a new—and better—way of accounting for care work amongst peers; whilst not being public, these registries nonetheless opened up care practices to professional peers and stimulated standardisation of services.

Whereas most indicators we studied lead to internal restructuring and change, especially volume indicators stand out as leading to a restructuring of the hospital landscape more broadly. At board levels, volume indicators were one of the drivers of cooperation with other hospitals, often leading to mergers—and as it happened all the hospitals we studied were part of merger processes. At the medical specialist level, especially in surgery, volume indicators even more so informed cooperation across hospitals, leading to mergers of clinical groups across hospitals and the concentration of specific surgical procedures, even though there was also criticism towards the dominance of volume indicators as a driving force for 'quality of care'.

On (governance) strategy

Governing health care organizations on basis of performance measurement is a difficult task. For one, our research shows that rankings generate a 'multitude of indicators', to which it is 'difficult to relate to, even more as these change in sometimes unpredictable manners. This also implies that hospitals have to deal with a multiplicity of (often only slightly different) indicators that are collected for dissimilar agents. Hospitals are only to a limited degree able to synchronize such multiplicity of information and hence are only to a limited degree able to steer performance improvement processes on basis of such indicators. For another, rankings and their underlying performance indicators generate frictions with organizational characteristics, which again challenge the manageability of indicator-based performance in hospitals. First, rankings build on performance data from the last year respectively. They thus generate a time slack and managers have repeatedly argued that it is difficult to steer performance improvement on basis of retrospective data sets. In this context, our respondents articulate the hope that more advanced ICT systems and resulting real-time information about indicator compliance could help them to monitor and steer indicator-based performance. However, second, the robustness of ICT systems is in conflict with the volatility of performance indicators. Building ICT environments for indicator compliance is a time-consuming and thus also expensive process, and therefore hospitals tend to drastically limit the amount of indicators that are built into ICT environments. Third, performance indicators at times create confusion in hospitals as they go against the logic of work in and between hospitals, as for example the case of volume indicators for colonoscopy has demonstrated. In sum, managers face various governance challenges that have their origin in the dissimilar logics of performance indicators and organizational characteristics.

We will therefore elaborate next how hospitals strategically deal with rankings and indicators to enable performance-based governance. We start with the ambivalent responses towards rankings we already had encountered above.

As mentioned above, responses to rankings are ambivalent, if not paradoxical: while respondents tend to negate the relevance of rankings, they simultaneously invest into adapting local structures and practices towards external demands. Our analysis has approached this ambivalence in demonstrating that these narratives of insignificance of rankings are only the frontstage to a more complex process of interaction between rankings and care processes. Every frontstage narrative has a backstage, too (Goffman, 1990). A glance at the backstage offers a more nuanced conclusion: performance technologies serve as strategic tools for managers to weaken traditional hierarchies and the powerful position of doctors for (executive) and to get grip on primary care process and the work of medical professionals in particular. Rankings in their internal use increase the power for executives and serve as strategic means to enable managers to negotiate and shape performance improvement agendas with professionals.

Rankings also serve as strategic means to position a hospital in its respective (competitive) environment externally. Here, rankings are dominantly used to build a reputation for the hospitals. Previous research has already demonstrated that one of the effects of rankings is that they create reputation as a new 'risk object' that needs to be managed by organisations (Power, Scheytt, Soin, & Sahlin, 2009). The risk of bad reputation is mainly addressed by change of organizational structure and strategic communication. With regard to the latter, hospitals tend to celebrate and communicate good results to the external world while they either tend to question ranking or downplay low results. With regard to organizational texture, marketing management becomes a new and important function within all hospital organisations, with hospital C standing out with both its pro-active marketing strategy and its rapid expansion of the department for marketing and communication. In line with a more

general trend in Dutch hospitals (Adams, 2011; Groenen, 2013), also social media is seen as an important means to communicate and reason performance results. In sum, rankings are mainly used externally as image tools to maintain or improve the hospital's overall reputation.

Performance measurement shapes organizational realities. Yet, also hospitals accommodate external demands towards local context, as the case of prioritization demonstrates. The increasing multiplicity of rankings and performance urges hospitals to organize for new (governance) strategies, and our case studies demonstrate that strategic managerial choices are more common than overall indicator compliance. For example, managers link indicators that are considered as locally appropriate to clinical processes or strategic goal (such as financial considerations). In hospital B, the strategy to prioritize professional accountability above strict performance compliance is related to the hospital's history in decentralized governance and its experience in Lean management. Performance governance, then, is never a one on one translation of external demands, but seeking alignment with a hospital's overall organizational structure as much as with the meaning that is attributed to particular results in rankings. Prioritisation is a first step to limit external accounting demands for the sake of internal coherence and thus also performance improvement. Non-compliance with external demands, then, helps hospitals to cluster and thus prioritize their performance work.

In sum, hospitals strategically adapt performance measurement technologies towards local need and thus change them. Likewise do hospitals influence external performance measurement regimes, as the case of colon carcinoma demonstrates. As a result, not only do indicators and ranking systems influence hospital organisations, the relations between them are more reciprocal as institutional actors also effectively influence indicators and ranking system. In a way, then, ranking systems and hospital organizations are co-constructed, rather than one influencing the other.

On actual care strategy

In our research proposal we argued for the importance to move beyond already existing studies that focus on how rankings influence managerial practices and also study the effect of rankings on the actual care provided.

Performance measurement, our empirical analysis shows, does influence actual care strategies. We have for example shown how performance-indicators contribute to the standardization of care processes through care pathways, as the case of post-surgical wound infections and access times in hospital C demonstrates. In that way performance measurement contributes to performance improvement.

Performance measurement, however, simultaneously demands for an extra of administrative work. Data collection actually starts with professionals registering information on their patients. This is no easy task when all kind of (often only slightly) different information has to be collected. Electronic patient records are a much used way of making sure registration work gets done; they not only make it easier to do the registration, they also enable ordering registration work during the day and can also have built in alerts to show that registrations are due—or over time. Registration compliance, however, is a difficult topic in all hospitals. Policing nurses, either by team leaders or board members in a ward round, is all part of an effort of getting registrations done. In sum, while all hospitals worked hard to get registrations going, and while registration of indicator performance became an important concern for hospital managers, registration work also encountered many problems in getting to the data that were needed to report on indicators.

Performance measurement does also influence care strategies in unexpected and unintended ways. We have for example demonstrated how indicators generate unintended conflicts for dissimilar values of quality of care, as the case of fall prevention, hospitalization and bed occupancy demonstrates. Indicators, this example highlights, always represent 'partial views' and focus selected aspects of care that are supported by specific professionals. Whereas these professionals would normally not be in conflict with each other, specific indicators may point them in different directions, thus causing the need for 'coordination work' (Mol, 2003). Performance measurement might also have unintended, risky consequences. The case of the oncology nurse, which highlights the risk of over-reliance on scoring lists, exhibits a well-researched phenomenon of standardization: while standards are stable enough to travel across many locations, they are not situated enough to be relevant for specialized conditions, such as found on an oncology ward. Indicator-based governance thus generates unintended consequences, such as risky practices and interprofessional struggles.

Performance measurement, however, is not a linear process that is exclusively concerned with the implementation and measurement of performance indicators. The oncology nurse and her colleagues consciously deviate from indicator-based care standards, and non-compliance with the standardized scoring seems to enable safe working practices on the ward. The scoring system, in that sense, is adapted towards local needs. Indicators, then, are not only influencing care processes in the sense that they 'adjust' local practices towards national demands; they are similarly 'adaptive' towards demands that preoccupy organization internally. We will further elaborate on the shaping of indicators, contexts, and local practices in our discussion of performativity below.

Discussion

In a recent letter to Parliament, minister of health Edith Schippers reiterated the plea for transparency that has been made many times earlier in the last decade:

The healthcare sector stands for the big challenge to now really work on increasing transparency, giving insight in the comparability of care delivery. This is a joint task, in which a lot of catching up needs to be done. The case is often not to get *more* information, but *better* and especially more discriminating information; information that really gives insight in the quality of care, that is delivered timely and in a standardized way and that is collected with a direct link to the process of care delivery. (...) The performance data that comes available through this gives insight to both patient and insurer and will get subsequent meaning in the contracting of care. (Schippers, 2013)

Despite more than ten years work on and with performance indicators, and despite huge investments of all actors in healthcare, now the work on creating transparency in healthcare 'really' starts. Just a few months earlier, a report published by the Algemene Rekenkamer showed that even though 30 million Euros government money has been invested in the *ZiZo* program, transparency in healthcare has not been reached (Rekenkamer, 2013). The report also highlights that the 30 million spent only relates to governmental expenses, which overall is a huge underestimation of the actual costs of the alleged failure in creating transparency. As we have seen in this study, investments in the hospitals we studied have been quite substantial (e.g. building and implementing IT systems, training of health professionals, etc) and the time put into registration work and governing performance is high. It is easy to estimate that the total costs of registration, collecting, aggregating and acting on performance data in hospitals is extensive extends to

several 100k Euro per year—making to tens of millions of Euro per year for the hospital sector alone.⁴

In part, as we have seen in this study as well, such investments do pay off. Creating the possibilities for performance management leads to improvements in care and to improved learning. Yet, improvement largely stays confined to those areas that are covered by performance indicators, while there are also many ‘perverse’ effects. Hospitals moreover did not experience patients choosing care on the basis of performance indicators or rankings and insurers focused on cost more than quality in commissioning.

Our report, then, hints at the double-edged nature of performance measurement: while benefits in the sense of generating some insights into the performance of health care, the overall costs and the underlying challenges remain largely understudied so far, both with regard to its sociological and economic dimension. However, like the quote above illustrates, Dutch (public) debates on the value of performance measurement rather exclusively focus on how ‘better’ and ‘more discriminating’ information can be obtained and focus on the technicalities of how performance measurement can be improved for the sake of an increase of transparency in health care. Absent in the current debate are discussions of the very assumption that more transparency will generate better performance and the benefits and challenges of transparency movements after all.

On performativity

In the 2013 version of the yearly report on performance indicators, the health care inspectorate notes that:

New process- and outcome indicators almost always show a large variation in care. The example for 2011 is the indicator on obstetrics. (IGZ, 2013)

Although not spelling it out, the healthcare inspectorate seems to suggest that variation in the indicator score relate to variations in the level of quality of care between hospitals. Performance indicators, the Inspectorates assumes, are what they are: they indicate (point out or to) quality and as such represent the ‘underlying’ performance of hospitals. In this report, we have taken another course of analyses. Also, as indicators lead to a gradual standardization of care processes and administrative systems, variation of indicator scores for hospitals over time decrease. Rather than looking at what performance indicators (and their composites: rankings) ‘indicate’, we have been interested in what they ‘perform’. That is, we have been looking at the ways in which hospitals shape their practices, policies and organizational setup according to such indicators and likewise how hospitals are being shaped through the use of indicators and rankings. This focus on performance, then, allows for an alternative explanation of variations in indicator scores: rather than ‘indicating’ differences in quality, the performative perspective points at the *time* it takes for hospitals to adjust themselves to the indicator, to start measuring in ways that are similar across time and space and to adjust care practices to make them measurable. Yet, performativity as we understand it does not limit itself to the question of how hospitals adjust their practice to external demands. Also, indicators are neither stable nor always measuring the best possible practice or outcome, as the executive director in hospital B summarizes:

⁴ See also a recent blog by the CEO of the Utrecht Medical Centre, claiming that some 5 to 6 fte work in his hospital to satisfy external information demands:
<http://www.skipr.nl/actueel/id15339-indicatorenrij-kost-umc-extra-personeelsformatie.html>.

Many outcomes are not the result of careful considerations, but often the effects of compromise and politics. There is plenty of hypercorrect behaviour in the world around us where ultimately one agrees on compromise and, if you think it through carefully, there is little reason to follow everything. Plus the multiplicity ... this also shows that it is unreal, thus you also have to use your brains. I personally like to stay with individual responsibility of professionals instead of collective hiding (executive director, hospital B, 02.05.2013).

Performativity, as we understand it, also describes how hospitals *adapt* external demands, such as performance indicators, towards existing structures and dominant managerial perspectives. Performativity, then, relates to how indicators, national demands and local context shape each other and together establish local systems of performance measurement. The *logic* of performance indicators however extends beyond local contexts. For example, we saw in all hospitals that indicators and rankings emphasize centralized control in hospitals and stimulates hospitals to work on the disciplining of professionals, e.g. by focussing on adherence to administrative work. Rankings and underlying performance indicators in this sense also help shape the organisational field of hospital care, as much as rankings have shaped the organisational field of for example business schools (Wedlin 2007). An important aspect of this is that rankings suggest hospitals (and business schools alike) to be governable entities in their own right. This idea of the hospital as a governable entity is already expressed by the ranking of entire hospitals, rather than subunits like specialties. With this, rankings also strengthen already existing developments, for example in safety management, where the systemic management of risk is increasingly emphasized.

This conception of performativity also departs from a standard perspective on how indicators are implemented and used in organizations. A mainstream perspective on implementation perceives of implementation as a linear process, where rather stable products (i.e. performance tools) are embedded into organizational contexts without much deformation. Our account on performativity hints towards a different phenomenon and argues for implementation processes that are less linear and that also involve the change of the very product that is being implemented. For example, administrative work, as we have seen, requires constant interpretation of indicators against care practices. We have shown that indicator-based data collection is not a straightforward process, but rather work that demands for synchronization and correction. Such processes of data adjustment, we have argued above, do also well exceed existing theoretical accounts that wish to understand deviation as gaming or cheating. While some of the changes that are done consciously (and while some changes happen very unconsciously, too), most of such changes happen in order to attune data to what is perceived as a 'real' representation of care practices and processes. Non-compliance with 'correct' data collection, then, is also not always an act of resistance but an effort of hospitals to *situate* data within actual hospital specificities. Retrieving 'correct' data is then a practice that differs from hospital to hospital, and the very product that was once implemented (the indicator) changes depending on local context.

Such local adaptation of indicators also goes some way in explaining the lack of comparability of indicator data between hospital settings. Earlier studies seemed to suggest that lack of comparability can be explained by referring to the impreciseness of indicators, leaving room for hospitals and professionals alike to adjust them to local needs (Kringos et al., 2012). Our respondents indeed indicated that such leeway was present—and was generally found more to be the case for IGZ indicators than for ZiZo indicators, the last ones being studied by Kringos et al. However, as we have seen, *all* indicators, however defined, seemed to require adjustment because of

a lack of fit with local care practices. This finding then suggests that rather than looking for perfect definitions investments can better be directed at the ways in which performance data is used in local practices. Experiences so far with for example the use of registries, suggest that such use can lead quality improvement to a large extend.

As local adaptations are inevitable it remains to be seen whether the centralising, rationalising logic of rankings works out in the end. As one of our respondents suggested, it might be a better way to aim for increasing the responsibility of professionals, rather than aiming for a universalising logic:

Let me give you an example. A new guideline on the intensive care (IC) has just been published. If you look at the consequences of that guideline—only a small number of ICs will survive. That guideline has been made by a group mainly consisting of intensivists and what you see is that there is a high level of 'hyper correct' reasoning that it leads to a result that is totally unrealistic. And that generates a lot of discussion, really a lot. Because it might be nice to say 'well, it has to be done this way and there is a lot of evidence for that' but if it is not realisable in practice well, then I can say as a hospital director that it has to be done, but the chances of it actually being done are poor and you just close your eyes for that. I would then rather work from the position that I trust the professionals working here; starting from their professional responsibility. Not autonomy, but professional responsibility. I believe in that much more than in one-size-fits-all and control modes of governing (executive director, hospital B, 02.05.13).

Stimulating 'responsible autonomy' (Degeling, Maxwell, & Iedema, 2004) could then still make use of quantifying techniques, but would also include narrative forms of accounting. As we have seen in our study, such narrative techniques are already increasingly in focus in the hospitals that we have been studying and have developed alongside and sometimes in opposition to ranking systems. Rather than seeing them as opposite techniques however, they could be more productively combined. This would also allow for dealing with the paradoxes and ambivalences quantitative techniques now often engender in healthcare settings.

Towards a model of organizational responses to rankings

Our research has shown that despite widespread criticisms of the validity and reliability of ranking systems, rankings do have widespread influence on the organisation of hospital care. Reputation, especially in case of bad scores, is an important driver for this influence. Rankings direct attention to those areas of hospital performance that are measured through performance indicators, stimulating administrative work, the standardisation of care processes and more centralised governance approaches. Such registration work, as in the form of administration and standardization, is not straightforward work though. It demands for diverse interventions, such as constructing new patient records, training personnel and integrating new types of expertise, at different layers of the organization in order to 'work'. We observed such careful and conscious layering and diversification of interventions—to which we referred to as 'investments in forms' (Thévenot, 1984)—in all hospitals we studied in order to make measurement happen and to perform well on rankings.

However, registration work does not always imply improvement of performance in the primary care process, and all case study hospitals indicated that there is a gap between quality assurance and improvement. It is unanimously argued that while external performance measurement demands lead to an increasing

investment in form, actual improvement work is often lacking behind. Some respondents also state that performance indicators can naturally only indicate the measurable (i.e. straightforward, observable issues and linear processes), while the most crucial safety and quality gaps are to be found in complex interactions (i.e. experience-based decision making). The very nature of performance as a measurement makes that complex and resistant performance problems in hospitals stay largely untouched.

Our research has also shown that performance measurement is always candidate for change and adaptation, and we have for example demonstrated how professionals 'work around' indicators—both for mundane reasons as much as for the sake of safety. Likewise, hospitals integrate dissimilar narrative accounting tools due to various reasons such as local histories with difficulties in performance management (hospital C) or particular local expertise (hospital B). In sum, performance management is a locally multiple practice.

This said, we can extend the argument and finally elaborate as to why there is not one pathway for good performance management strategies. Our case study highlights that there are different strategies to internalize external accounting demands. Hospital C has a well-built centrally-steered performance monitoring and compliance system and improvement is to a large extent organized formally within this hierarchical top-down system. The hospital is challenged by what is often perceived as professional resistance against registration and improvement duties. Therefore, one more recently argues for a move towards more 'space' (i.e. for reflection and professional-led, locally relevant improvement agendas) in order to overcome the gap between performance measurement and improvement. Hospital C is the most different from hospital B with regard to its strategies to internalize external demands. While it also operates a (mandatory) centrally-steered performance monitoring system, this framework is very consciously much less elaborate and extensive than in hospital C. The organization follows a Lean philosophy and allows for decentralized decision making structures in the teams. Also, responsibility for improvement work largely rests in the ward-based dual management teams, where 'a thousand blossoms bloom'. However stimulating such bottom-up work is for professionals, the hospital has problems in upscaling such local experiences and implementing hospital-wide, top-down performance demands from the national outside. Hospital A is somewhere in between, showing attempts at centralisation, as we have seen for example around nursing indicators, including the disciplining of professionals in administrative work. The hospital has also invested heavily in information systems in order to facilitate both administrative work and collection of indicator information. On the other hand, much of the policies of the hospitals seem to be of a rather ad hoc nature, reacting mainly to bad scores, and a systematic policy on performance is not in place.

Our comparison is illuminating in two ways. One, it demonstrates there are different strategies to internalize external accounting demands. Hence, there is no 'one size fits all' pathway for good performance management in hospitals. In consequence, two, our comparison then also challenges linear approaches to quality and safety development. For example, Dutch hospitals are obliged to implement local safety programs according to national requirements. Here, teams have to build safety cultures by 'climbing a culture ladder', moving from ignorance to reactivity to proactive safety behavior. Underlying such programmes is one epistemological assumption: that a *systemic-managerial* perspective on safety culture, which can be established in a linear and predictable fashion and moves from denial to reaction, and proactive action is the best way to generate safe work environments in health care. Our observation clearly challenges this perspective and argues that what is an appropriate quality and safety improvement strategy depends not only on what is perceived as a mature quality system but also on local governance particularities.

There is not really a 'better' or 'worse' in how our case study hospitals organize for quality and safety.

Hence, we argue that performance management means to choose for locally fitting styles and strategies for dealing with external demands. The comparability of health care is limited. In this context, then, the challenge for hospitals is in finding appropriate fit between national demands and local context. For this end, our research and also previous research we have conducted clearly demonstrates that hospitals are in need of what we call 'reflexive spaces' (Quartz, Weggelaar-Jansen, van de Bovenkamp, & Bal, 2012). Such reflexive spaces may for instance balance hierarchical performance control and learning opportunities and focus on locally relevant problems; they might enable professionals to reveal 'blind spots' by means of inter-professional reflection on practice; they may focus on what actually goes well and might be transferred elsewhere; as much as they might establish space to unravel hidden competences and knowledge of professionals that critically contributes to the functioning of performance locally.

Literature

- Adams, S. (2011). Sourcing the crowd for health services improvement: The reflexive patient and "share-your-experience" websites. *Social Science & Medicine*, 72, 1069-1076.
- Bal, R. (2012, 7-9 June 2012). *Organizing for transparency. The ranking of Dutch hospital care*. Paper presented at the Transatlantic Conference on Transparency Research, University of Utrecht.
- Bal, R., & Zuiderent-Jerak, T. (2011). The practice of markets. Are we drinking from the same glass? *Health Economics, Policy and Law*, 6(1), 139-145.
- Berg, M., Goorman, E., Harterink, P., & Plass, S. (1998). *De nacht schreef rood: Informatisering van zorgpraktijken* (No. Studie 37). Den Haag: Rathenau Instituut.
- Berg, M., Meijerink, Y., Gras, M., Goossensen, A., Schellekens, W., Haeck, J., et al. (2005). Feasibility First: Developing Public Performance Indicators on Patient Safety and Clinical Effectiveness for Dutch Hospitals. *Health Policy*, 75(1), 59-73.
- Berwick, D., M. (2002). Public performance reports and the will for change. *Journal of the American Medical Association*, 288(12), 1523-1524.
- Bevan, G., & Hood, C. (2006). What's measured is what matters. Targets and gaming in the English public health care system. *Public Administration*, 84(3), 517-538.
- Degeling, P., Maxwell, S., & Iedema, R. (2004). Reconstructing clinical governance to maximize its developmental potential. In A. Gray & S. Harrison (Eds.), *Governing Medicine: Theory and practice* (pp. 163-179). Maidenhead: Open University Press.
- van der Aar, L. (2008). *Prestatie-indicatoren & kwaliteit van zorg*. Unpublished MSc, Erasmus Universiteit Rotterdam, Rotterdam.
- Dixon-Woods, M., Leslie, M., Bion, J., & Tarrant, C. (2012). What Counts? An Ethnographic Study of Infection Data Reported to a Patient Safety Program. *Milbank Quarterly*, 90(3), 548-591.
- Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: how public measures recreate social worlds. *American Journal of Sociology*, 113(1), 1-40.
- Espeland, W. N., & Stevens, M. L. (1998). Commensuration as a social process. *Annual review of sociology*, 24, 313-343.
- Espeland, W. N., & Stevens, M. L. (2008). A sociology of quantification. *European Journal of Sociology*, 49(3), 401-436.
- Freeman, T. (2002). Using performance indicators to improve health care quality in the public sector: A review of the literature. *Health Services Management Research*, 15, 126-137.
- Fung, C. H., Lim, Y.-W., Mattke, S., Damberg, C., & Shelleke, P. G. (2008). Systematic review: the evidence that publishing patient care performance data improves quality of care. *Annals of internal medicine*, 148(2), 111-123.
- Giard, R. W. M. (2005). Prestatie-indicatoren als maat voor kwaliteit van medische zorg: retoriek en realiteit. *Nederlands Tijdschrift voor Geneeskunde*, 149(49), 2715-2719.
- Goffman, E. (1990). *The Presentation of Self in Everyday Life* (8 ed.). London etc.: Penguin.
- Groenen, Y. (2013). *Social media en crisis in de zorg. Een onderzoek naar de toepassing van social media bij crisis in de zorg*. Unpublished MSc, Erasmus Universiteit Rotterdam, Rotterdam.
- van Hamersveld, M., & Olivier, L. (2009). *Ziekenhuiskenners Keuren Zichzelf. De Elsevier Onderzoeksaanpak Doorgelicht*. Amsterdam: MOA, Center for Marketing Intelligence & Research.

- Hibbard, J. H., Stockard, J., & Tusler, M. (2005). Hospital performance reports: Impact on quality, market share, and reputation. *Health Affairs*, 24(4), 1150-1160.
- Hilgartner, S. (1992). The social construction of risk objects. Or, how to pry open networks of risk. In J. F. Short & L. Clarke (Eds.), *Organizations, Uncertainties, and Risk* (pp. 39-53). Boulder: Westview Press.
- Huisman, G. (2008). *Werken met prestatie-indicatoren. De invloed van externe prestatie-indicatoren op de werkpraktijk van een ziekenhuis*. Erasmus Universiteit Rotterdam, Rotterdam.
- IGZ. (2013). *Het resultaat telt ziekenhuizen 2011*. Utrecht: Inspectie voor de Gezondheidszorg.
- Jacobs, R., Goddard, M., & Smith, P. C. (2005). How Robust Are Hospital Ranks Based on Composite Performance Measures? *Medical Care*, 43(12), 1177-1184.
- Jerak-Zuiderent, S., & Bal, R. (2011). Locating the worths of performance indicators. Performing transparencies and accountabilities in health care In A. Rudinow Sætnan, H. Mork Lomell & S. Hammer (Eds.), *The Mutual Construction of Statistics and Society* (pp. 224-244). London: Routledge.
- Kalkman, C., Rinkel, G., Roos, Y., Dzoljic, M., Keesman, E., & Heinen, P. (2013). Coderen van diagnoses nog veel te divers. *Medisch Contact*, 68(9), 476-478.
- Koning, P., & van der Wiel, K. (2010). *School responsiveness to quality rankings. An empirical analysis of secondary education in the Netherlands*. The Hague: CPB Netherlands Bureau for Economic Policy Analysis.
- Kringos, D. S., Anema, H. A., ten Asbroek, A. H. A., Fischer, C., Botje, D., Kievit, J., et al. (2012). *Beperkt zicht. Onderzoek naar de betrouwbaarheid, validiteit en bruikbaarheid van prestatie-indicatoren over de kwaliteit van de Nederlandse ziekenhuiszorg*. Leiden, Amsterdam, Rotterdam: Leids Universitair Medisch Centrum, Academisch Medisch Centrum & Erasmus MC.
- Mol, A. (2003). *The multiple body. Ontology in Medical Practice*. Durham: Duke University Press.
- Pollitt, C. (2008). *Time, Policy, Management. Governing with the past*. Oxford: Oxford University Press.
- Pons, H. (2009). *De rol van ziekenhuisranglijsten*. Unpublished Ba, Erasmus Universiteit Rotterdam, Rotterdam.
- Pons, H., Lingsma, H., & Bal, R. (2009). De ranglijst is een slechte raadgever. Reputatiestrijd tussen ziekenhuizen komt kwaliteit niet ten goede. *Medisch Contact*, 64(47), 1969-1971.
- Power, M. (1997). *The Audit Society. Rituals of Verification*. Oxford: Oxford University Press.
- Power, M. (2007). *Organized uncertainty. Designing a world of risk management*. Oxford: Oxford University Press.
- Power, M., Scheytt, T., Soin, K., & Sahlin, K. (2009). Reputational risk as a logic of organizing in late modernity. *Organization Studies*, 30(2 & 3), 301-324.
- Quartz, J., Weggelaar-Jansen, A., van de Bovenkamp, H., & Bal, R. (2012). *Quality and Safety in Europe by Research Country Report*. Rotterdam: iBMG.
- Rekenkamer, A. (2013). *Indicatoren voor kwaliteit in de zorg*. Den Haag: Algemene Rekenkamer.
- RVZ. (2007). *Waar staat het beste ziekenhuis?* Den Haag: Raad voor de Volksgezondheid & Zorg.
- Sauder, M., & Espeland, W. N. (2009). The discipline of rankings: tight coupling and organizational change. *American Sociological Review*, 74(1), 63-82.
- Schippers, E. (2013). *Brief aan de Tweede Kamer betreffende Uitkomstenbekostiging in de curatieve zorg*. Den Haag: Ministerie van VWS.

- Strauss, A., & Corbin, C. (1998). *Basics of qualitative research: techniques and procedures for developing grounded theory*. Thousand Oaks (CA): Sage Publications.
- Thévenot, L. (1984). Rules and implements: investment in forms. *Social Science Information*, 23(1), 1-45.
- Triantafillou, P. (2007). Benchmarking in the public sector: a critical conceptual framework. *Public Administration*, 85(3), 829-846.
- Verran, H. (2013). Numbers Performing Nature in Quantitative Valuing. *NatureCulture*, 1(2), 23-37.
- Zuiderent-Jerak, T. (2009). Competition in the wild. Emerging figurations of healthcare markets. *Social Studies of Science*, 39(5), 765-792.
- Zuiderent-Jerak, T., Kool, T., & Rademakers, J. (Eds.). (2012). *De relatie tussen volume en kwaliteit van zorg. Tijd voor een bredere benadering*. Utrecht/Nijmegen/Rotterdam: Consortium Kwaliteit van Zorg.

Appendix 1 List of codes

Primary code	Secondary code	Tertiary code
Importance of doing well on the rankings	Reactions of patients	
	Reactions of other hospitals	
	Reactions of inspectorate/regulators	
	Reactions of insurers	
	Professionals	Recruiting personnel
		Reputation amongst peers
	Regional / national relevance of ranking	
	Individual / group reputation	
	Reputation of executive director	
Reputation of the ranking	Worth attributed to the ranking	Data collection methods; weighing between indicators
		Regional relevance of ranking
		Irrelevance due to large patient volumes
		Decreasing relevance of rankings due to multiplicity
	Worths for individual indicators	
Organisational effects of indicators how indicators and organisations interact		
	Investing in reputation management	Building of marketing departments, bringing in marketers
		Other fora for building reputations (prizes etc)
		Publications on internet & social media
	New types of services	Standardisation of care work – direct and indirect
		New 'reputational' services (e.g. valet parking, new outpatient clinics)
	Prioritising domains	Introduction of 'aandachtsvelders'
		Formal/informal
	Investment in administrative processes	Compliance management function of quality managers
		Stimulating registrations – policing, benchmarks

		Information technologies; computerized (epr's etc) or paper (checklists, forms) Technical difficulties.
		Training of staff for registration
		Training of coders
	Dissimilar logics	Organizational structure
		Location-specific work logic
	Performance management in the hospital	Performance reviews
		Displaying of performance data
		Working groups
	New professionals	Marketing
		Controller
	Learning	Analysis/benchmarking
		Visit to well-performers
Quantification work/administrative work	Data collection pathway	Back and forth of indicator data. (negotiation vs 'compliance')
	Negotiations over data	Definition of indicator
	Sharing between hospitals how to measure	
	Shaping what is meant to be measured	
	Intentional gaming for benefit of score	Gaming Inclusion/exclusion criteria sampling External adaptation / internal correctness as 'aanjager'
	Intentional gaming for benefit of work flow	Immutable delegates & Work around (tick boxing etc) Localization Mundane problems
	Intentional gaming for benefit of what is perceived correct	Synchronization work Situating decision making Correction work / 'artful work' (Dixon-Woods)
	Unintentional gaming	Data collection
		Data interpretation
	Neglected indicators	Cross-sectional indicators (e.g. delirium)
	Situated resistance for safety and quality	
Unintended consequences	Conflicting values	Fall prevention case
	New risks / unsafe practice	Pressure ulcer case oncology

	Tunnel vision	Nursing indicators
		Volume standards
	Losing trust	(extended arm of inspectorate)
	Extrapolate professional conflicts	
Narratives & stories	Related to indicators	Giving meaning to data
	Reaction to indicators / against quantification	Culture programs
		Narrative accounting
		Externalizing fraude, reasoning internal failure
Governance 1: social relations and steering within the hospital	Time slack manageability	
	Central-decentral relations; new types of interaction with professionals	Compliance vs negotiations.
	External accounting outplays internal use for QI	
	Managerial logic dominance	
	Influencing internal policy for QI	Negotiation with doctors: "beta feeling for numbers" / performance-based payment
	Shifting responsibility	boiling down problems (the 'blame game')
	Multiplicity of demands	Prioritization (local history, financial interests, external powers)
	Decoupling of control and QI	
Governance 2: Strategic reputation management	influencing outside world	Careful consideration: 'can be used against us'
	New external accounting practices	Care standards
		Proactivity & transparency: Nico's case of pilot systeemtoezichts
		Negotiation relevance with insurers
		Scientification
		IGZ/VMS (length of stay)



Erasmus University Rotterdam

INSTITUTE OF HEALTH POLICY & MANAGEMENT

Visiting address

Burgemeester Oudlaan 50
3062 PA Rotterdam

Postal address

P.O. Box 1738
3000 DR Rotterdam
The Netherlands

Telephone +31 10 408 85 55
Internet www.bmg.eur.nl
E-mail research@bmg.eur.nl

