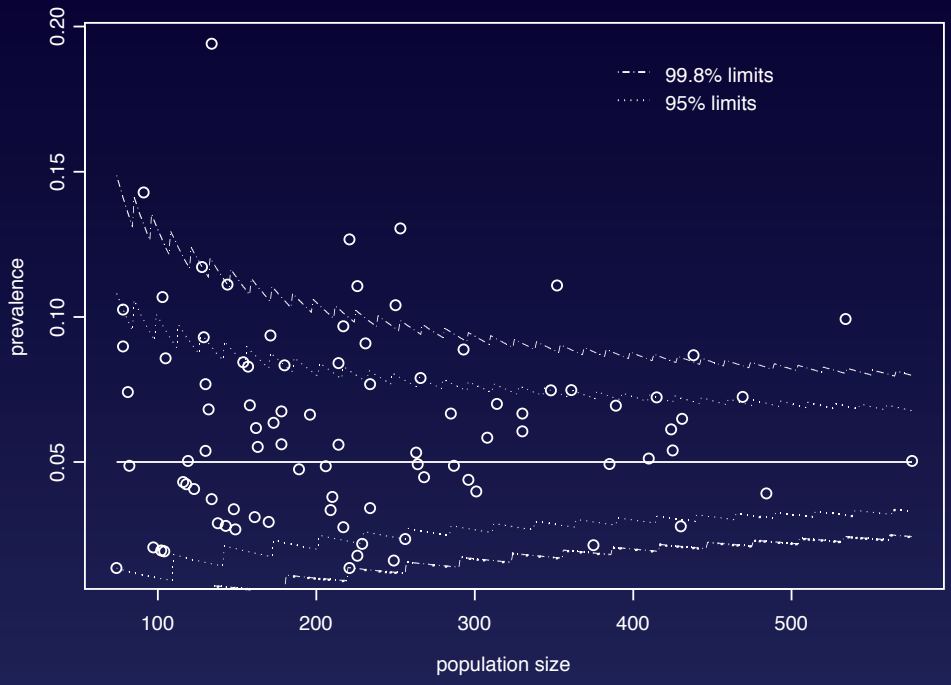
INDICATORS FOR QUALITY OF HOSPITAL CARE: BEYOND THE NUMBERS



Anne-Margreet van Dishoeck

Indicators for quality of hospital care: beyond the numbers

Anne-Margreet van Dishoeck

Indicators for quality of hospital care:
beyond the numbers

Indicatoren voor de kwaliteit
van zorg in ziekenhuizen; meer dan cijfers

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam op gezag van de rector magnificus

Prof.dr. H.A.P. Pols
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 11 november 2015 om 11:30 uur

door

Adriana Margaretha van Dishoeck
Geboren te ’s-Gravenhage

Erasmus University Rotterdam

Erasmus

PROMOTIECOMMISSIE

Promotoren	Prof.dr. J.P. Mackenbach Prof.dr. E.W. Steyerberg
Overige leden	Prof.dr. J.W. Deckers Prof.dr. J. Kievit Prof.dr. P.B.M. Robben

CONTENTS

Voor Pappie, omdat ik nu begrijp wat je me toen vertelde.

Chapter **1**

General Introduction

In the last decades the attention on quality and safety in health care has increased enormously. The role of transparency in quality of care is becoming ever more important and hospitals face increasing demands with regards to performance. Information on quality of delivered care enables various stakeholders to compare hospitals and to improve the quality of care.

HOSPITAL PERFORMANCE AND MEASUREMENT

The Dutch Quality Act (1996) provides a framework in which health care providers are given responsibility to be transparent on their outcomes of delivered care. The essence of the Quality Act is that care providers are held responsible for providing good and affordable care and improvement of the quality of care. Transparency is to be achieved through the communication of information on performance. The information on the quality of delivered care should further enable patients to make an informed decision when choosing professionals or heath care institutions. Furthermore, it supports value-based purchasing for insurers. The Health Care Inspectorate supervises and evaluates the quality and safety in health care institutions partly aided by this quality information. An overview of the different stakeholders is summarized in table 1.

Stakeholders	Transparency in quality of care
Governments	Public health, policy making, International comparisons
Health care Inspectorate	Supervision and patient safety; identifying poor performance among healthcare providers
Health insurances	Purchase; high value low costs
Patients	Supporting patient choice
Health care institution management	Policy making within the care institution
Professional	Monitoring, benchmarking and improving quality of care

Table 1; Stakeholders and their quest for transparency in quality of care

These developments are creating a growing focus on hospital performance in which hospitals face increasing demands with regard to the quality, transparency and accountability of delivered care.

Historical perspective

In 1863, Florence Nightingale was the first to suggest the recording of hospital performance in what she called "Hospital Statistics"; "These methods, if generally used, would enable us to ascertain the mortality in different hospitals, as well as from different diseases and injuries at the same and at different ages, the relative frequency of different diseases and injuries among the classes which enter hospitals in different countries, and in different districts of the same country. They would enable us to ascertain how much of each year of life is wasted by illness,—what diseases and ages press most heavily on the resources of particular hospitals." [1] "The truth thus ascertained would enable us to save life and suffering, and to improve the treatment and management of the sick and maimed poor." Hereby she expressed a great confidence in the statistical possibilities of measuring quality of care. It would take until the late twentieth century for hospital statistics to be applied on a large scale, now known as performance indicators or quality indicators.

Definitions of quality of care

Quality of care is a broad concept for which different definitions exist. The Dutch Quality Act defines quality of care as "care of a high standard being efficient, effective, patient oriented and matching patients' real needs", concentrating on the patient.[2] The American Institute of Medicine formulated quality of care as "the degree to which health services increase the likelihood of desired health outcomes for individuals and populations and are consistent with current professional knowledge", aiming at the professionals in health care.[3, 4] The World Health Organisation defines quality of care as being effective, efficient, accessible, patient-centred, equitable and safe [5]. Hospitals usually focus on the domains: effectiveness, efficiency and safety to address the quality of care.

Measuring quality of care

Aiming at measuring quality, Donabedian described care as a function of three components: structure, process and outcome.[6, 7] The evaluation of structure consists in the appraisal of the instrumentalities of care and their subsequent organisation, equipment, manpower and financing. Donabedian stated: "If these meet certain specifications, it is likely that good care follows".[7] Process data refers to what is actually done during contact between patient and health care professional. Process indicators assess what was done during this contact and how well it was done.[8-10] For example, prescribing statins for patients with cardiovascular disease. An optimal process leads to the best outcome for the individual patient. The outcome indicator

captures the effect of care processes on the health and wellbeing of patients. For example, the occurrence of pressure ulcers during hospital admission (nosocomial pressure ulcers), but also the degree of the patient's satisfaction with care. Data capturing structure, process, or outcome, are known as performance indicators.

A performance indicator, being a retrospective measurable element of practice for which there is evidence or consensus, provides insight in quality of care.[11, 12] As from the early 1980s, an increased interest in measuring hospital performances, has led to the development of a growing number of performance indicators. These measures address different areas of health care and are constructed in different sets of indicators. In the Netherlands, these sets are used by several assessors for different aim, focus and internal or external accountability (table 2).

Indicator set	Aim	Primary focus	Accountability
Inspectorate; basic indicator sets for health care organisations	Supervision	All health care institutions	External
Hospital Standardised Mortality Ratio	Supervision	Hospitals	External
Safety Management System (VMS)	Patient safety	Hospitals	Internal/external
Patient Reported Outcome Measures (PROMs)	Patients experience and satisfaction	Patients	External
Dutch Institute for Clinical Auditing (DICA); disease specific national databases	Benchmark	Medical specialists, and other health care providers	Internal
National Institute for Public Health and Environment (RIVM); PREZIES network database of health care related infections	Benchmark	Hospitals and medical specialists, and other health care providers	Internal

Table 2; Sets of indicators monitoring quality of health care organisations in the Netherlands

Nowadays performance indicators are used for public accountability, for external assessment, for supervision, for purchase, for supporting patient choice, for internal management control and for internally driven quality improvement.

CHALLENGES TO THE MEASUREMENT OF HOSPITAL PERFORMANCE

Professionals using performance indicators are faced with a number of challenges. These challenges are often related to the difficulties of measuring something as abstract as quality of care. Aspects such as definition of concepts, quality of the data, risk of gaming, role of disturbing factors, and the intended use for quality improvement, influence the reliability and validity of performance indicators.

Definition of concepts

Except for structure indicators, performance indicators are usually measured with a numerator (the occurrence of a particular outcome or process) and a denominator (the population for whom the outcome or process is relevant). For instance, the readmission rate in heart failure patients is calculated as the total number of patients readmitted within a period of three months after hospital discharge (numerator) divided by the total number of patients who were admitted for heart failure (denominator). Each of these elements must be defined unambiguously. For instance, what is exactly meant with "re-admission"? Every readmission in this time period or re-admission for heart failure or unplanned re-admission? So, the term readmission is far from unambiguous. Defining concepts in a clear and explicit way has been addressed in the past decades and needs ongoing attention and adjustments.[9, 13, 14].

Quality of the data

Measuring performance starts with accurate data collection. The underlying data of an indicator should preferably be routinely present and easily accessible. The use of administrative data is attractive because it is less demanding for health care providers than data from electronic patient files or separate stand-alone data collection.[15] Using administrative data, the data are already collected for other purposes, therefore the costs of the data collection are much lower. Clinicians often criticize the reliability of administrative data sources but the trend towards pay-for-performance has led to rigorous auditing of the data and enhanced accuracy.[16, 17] For example, recently a measurement tool was developed based on routine administrative data on hospital stay, readmission, and mortality rates (HARM), to evaluate the quality of colorectal surgery.[18] Researchers found the HARM score easy, reliable, and valid for assessing quality in colorectal surgery.[18] Despite these efforts, accuracy and completeness of clinical variables may vary or parts of the indicator may not be available in administrative data.[15, 19] If parts of indicators are registered in the clinical patient records they need to be extracted. In electronic patient files, this requires specific ICT applications. These software applications tend to be even more complicated if the data is registered in text fields. Some indicators require a separate data collection effort because they are not routinely recorded in administrative data or patient files, such as pressure ulcer prevalence. Further implementation and improvement of electronic patient files are expected to give an enormous boost to the availability of data that can be used for quality measurement purposes.[19-22]

Risk of gaming

Hospitals, as described in the quality act, are responsible for the accuracy and trustworthy origin of their own data, without any external check. This process, based on trust, makes it susceptible to confabulation when it is internally known that specific demands by governments or other influential parties, using blame and shame, are not

met. This is a known flaw in the reports delivered to governmental institutions and is described in the literature as *gaming*.[23, 24] Several English researchers stated that managing the public services with targets and using blame and shame produced this gaming response.[23, 24] In reaction to the blame and shame culture, the American Association of Health Plans recognised that a systematic approach aiming at transparency and efficiency of health care improving the quality of care that patients receive.[25] In 2008, the Australian Health and Hospital reform Commission proposed a long-term health reform plan "*Beyond the blame game*" based on accountability and benchmarking instead of ranking and blaming.[26] Gaming is a serious threat to the validity of performance indicators.

Random variation

Performance indictors aim at identifying instances of excessive variation in processes or outcomes. Part of the patient-to-patient variability can be caused by chance (random variation). Research suggests that many performance indicators are sensitive to random variation.[27] Power calculations, being a part of the study design in scientific research, has received little attention in defining and developing performance indicators.[28] As a result hereof, most differences found when comparing hospital performance can be ascribed to random variation. Random variation stays often unaddressed in reports on performance indicators. For instance, in the Netherlands, hospital performance measured using the Inspectorate's performance indicators is summarised and reported in simple bar charts (figure 1).

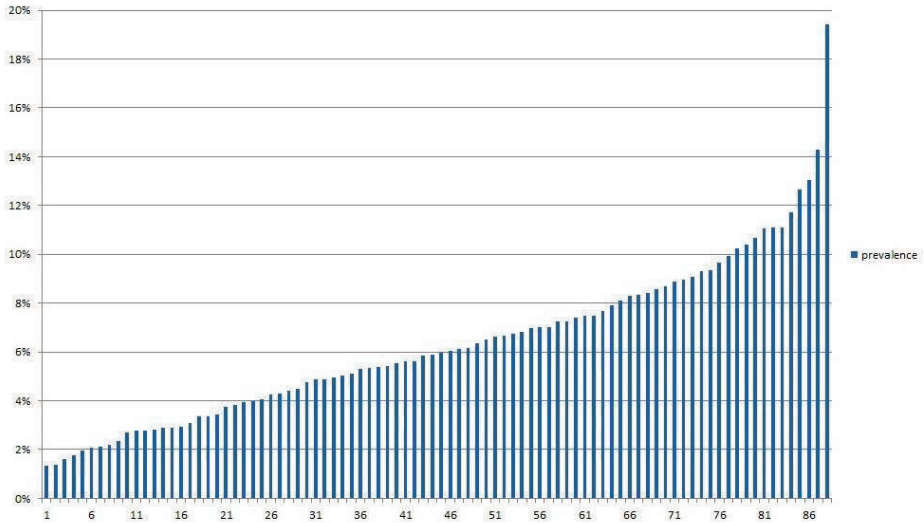


Figure 1. Pressure ulcer prevalence in 89 Dutch hospitals in 2005[29]

The x-axis represents 88 Dutch hospitals, with their outcomes displayed on the y-axis. Hospitals are ranked by their outcome of pressure ulcer prevalence, ranging in this example from ± 2% to almost 20%. This bar chart suggests an almost 17%-points difference in pressure ulcer prevalence between the best and the worst performing hospitals. It does however, not provide an insight in the statistical uncertainty of these differences. The confidence interval for small hospitals has a greater spread compared to the spread of the interval in larger hospitals.[30] The risk of 'false positive' results (indicator suggests lack of quality, but this does not withstand examination) is considerable, as a result of random variation.[15, 31, 32] Statisticians therefore often emphasize addressing random variation in reporting performance.[30, 33] It remains unclear in the Netherlands to what extent random variation influences between-hospital comparisons with commonly used outcome indicators and how to deal with this random variation.

Confounding

The confounding effect that variations in patient factors has on outcome measures is well recognised in epidemiological research.[15] Several outcome indicators are based on their use in randomised controlled trials as an unfavourable outcome of care. Examples include re-admission rates or nosocomial pressure ulcer occurrences. Findings in randomised controlled trial populations are not automatically equivalent to the groups of patients generally admitted in hospitals.[34-40] Performance indicators do not merely reflect the quality of care alone, because variations in case-mix can have a crucial influence on their values. For example, in the case of stroke patients, several studies showed that after adjustment for prognostic factors, the statistically significant differences in mortality or functional outcome between hospitals seen in the crude data became non-significant.[41, 42]

The outcome of a performance measure can be seen as the sum of factors relating to the patient, the illness, the treatment and the organisation (figure 2), whereas only the last two factors relate to quality of care that can be influenced by professionals or hospital management.[9]

Hence, adjustment for factors relating to the patient and illness (case-mix) is essential when comparing hospitals on performance.

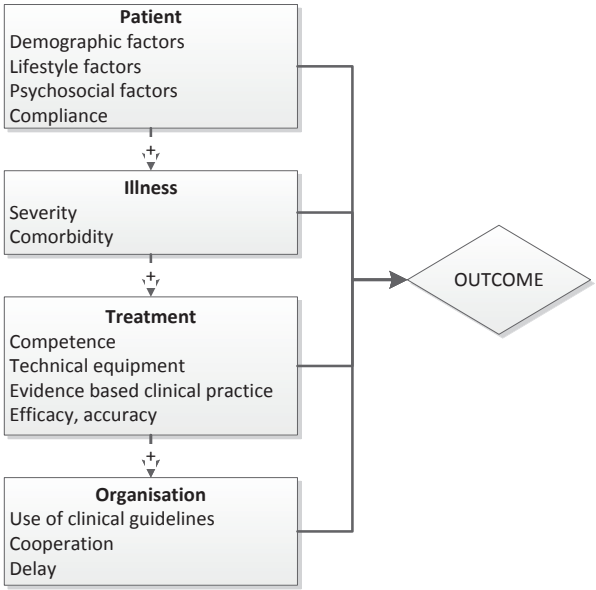


Figure 2; Conceptual framework of factors influencing an outcome[9]

Structure-process-outcome relation

Even if all requirements for reliability and validity, as described above, are met, performance indicators may still not be meaningful. The validity of a performance indicator depends on whether variations in the value of the indicator reflect variations in one or more aspects of structure or process of care, and vice versa: the validity of a structure of process indicator depends on whether variations in the value of the indicator are reflected in variations in one or more medically relevant outcomes of care. This has usually not been empirically assessed, and often evidence on validity is limited to face-validity, established by expert consensus. In other cases, the evidence is limited to construct validity based on recommendations in evidence-based guidelines, which document a link between process and outcome as found e.g. in randomized trials. A rigorous empirical examination of the causal chain from structure or process to outcome in hospital populations often lacks. It is therefore important to assess this causal link before performance indicators are used to assess quality.[7, 43]

Actionability

A basic purpose of an indicator is to improve health care. Performance indicators must thus provide clues for subsequent improvement of the quality of care delivered, so called actionability. Indicators should focus on those aspects of care in which interventions are possible and therefore have the potential for improving care. Actionability is then the degree to which a health care professional can influence the measure,

in response to an unfavourable value of the indicator.[44, 45] There is a continuous tension between the search for meaningful indicators on a national level and on their use for quality improvement within a hospital.[39] Actionability of outcome indicators is negatively influenced by the absence of information on actual care processes and subsequently performance improvement.[46, 47] On the other hand, optimal outcomes for the patient are regarded as the first purpose of health care. The purpose of performance indicators must meet the balance between the need for accountability with the need to promote quality improvement initiatives.[48]

Overall aim and research questions

This thesis aims at evaluating the usefulness of outcome indicators and process indicators in comparing hospitals and in improving the quality of hospital care. In addressing this overall aim, we distinguish between-hospital comparisons and within-hospital comparisons.

Between-hospital studies

Comparing hospitals using performance indicators for external accountability is the first topic of this thesis. In the between-hospital studies, the overall research question is: "How to interpret differences between hospitals in performance indicator measures?"

1. What is the influence of random variation in comparing quality of care between hospitals using outcome indicators (Chapter 2)?
2. How should random variation be displayed when reporting outcome indicators (Chapter 3)?
3. To what extent do random variation and case-mix influence the comparability of hospitals with respect to surgical site infections (Chapter 4)?

Within-hospital studies

The second topic of this thesis explores the use of performance indicators in within-hospital comparisons, particularly for internal quality improvement. In the within-hospital studies the overall research question is: "How strong is the relation between outcome indicators and the underlying care processes, and can performance indicators guide quality improvement?"

1. Does pressure ulcer prevalence reflect the quality of the preventive care processes in adult hospitalized patients (Chapter 5)?
2. Can the effect of governmental surveillance be quantified using performance indicators for health care institutions (Chapter 6)?
3. Does door-to-needle time reflect the effect of improvement initiatives in the care for stroke patients (Chapter 7)?
4. Does pressure ulcer prevalence reflect improvements in quality of pressure ulcer prevention in surgical patients (Chapter 8)?

REFERENCES

1. Nightingale, F., *Notes on Hospitals*. 1863, London: LONGMAN, GREEN, LONGMAN, ROBERTS, and GREEN.
2. *Kwaliteitswet Zorginstellingen*. 1996: the Netherlands.
3. Lohr, K.N., *Applications of health status assessment measures in clinical practice. Overview of the third conference on advances in health status assessment.* Med Care, 1992. **30**(5 suppl): p. MS1-14.
4. Lohr, K.N., M. Donaldson, and J. Harris-Wehling, *Medicare: a stratigy for quality assurance, V: Quality of care in a changing health care environment.* QRB Qual Rev Bull, 1992. **18**(4): p. 120-6.
5. Bengoa, R., et al., *Quality of care : a process for making strategic choices in health systems*. 2006, World Health Organization: Geneva, Switzerland. p. 9-11.
6. Donabedian, A., *Quality of care: problems of measurement. II. Some issues in evaluating the quality of nursing care.* Am J Public Health Nations Health, 1969. **59**(10): p. 1833-6.
7. Donabedian, A., *The quality of care. How can it be assessed?* JAMA, 1988. **260**(12): p. 1743-8.
8. Brook, R.H., E.A. McGlynn, and P.D. Cleary, *Quality of health care. Part 2: measuring quality of care.* N Engl J Med, 1996. **335**(13): p. 966-70.
9. Mainz, J., *Defining and classifying clinical indicators for quality improvement.* Int J Qual Health Care, 2003. **15**(6): p. 523-30.
10. Mant, J. and N. Hicks, *Detecting differences in quality of care: the sensitivity of measures of process and outcome in treating acute myocardial infarction.* Bmj, 1995. **311**(7008): p. 793-6.
11. Campbell, S.M., M.O. Roland, and S.A. Buetow, *Defining quality of care.* Soc Sci Med, 2000. **51**(11): p. 1611-25.
12. Campbell, S.M., et al., *Research methods used in developing and applying quality indicators in primary care.* Bmj, 2003. **326**(7393): p. 816-9.
13. Johnson, M.L., et al., *Effect of definition of mortality on hospital profiles.* Med Care, 2002. **40**(1): p. 7-16.
14. Leander, W.J., *Performance indicators. Defining and evolving the "right" measures of restructuring success.* Rev Patient Focus Care Assoc, 1993: p. 2-9.
15. Powell, A.E., H.T. Davies, and R.G. Thomson, *Using routine comparative data to assess the quality of health care: understanding and avoiding common pitfalls.* Qual Saf Health Care, 2003. **12**(2): p. 122-8.
16. Tsang, C., et al., *A review of patient safety measures based on routinely collected hospital data.* Am J Med Qual, 2012. **27**(2): p. 154-69.
17. Iezzoni, L.I., *Assessing quality using administrative data.* Ann Intern Med, 1997. **127**(8 Pt 2): p. 666-74.
18. Keller, D.S., et al., *The HARM Score: A Novel, Easy Measure to Evaluate Quality and Outcomes in Colorectal Surgery.* Ann Surg, 2013.
19. Breslow, M., *Assessing ICU performance using administrative data.* J Crit Care, 2001. **16**(4): p. 189-95.
20. Glance, L.G., et al., *Accuracy of hospital report cards based on administrative data.* Health Serv Res, 2006. **41**(4 Pt 1): p. 1413-37.
21. Scott, I.A. and M. Ward, *Public reporting of hospital outcomes based on administrative data: risks and opportunities.* Med J Aust, 2006. **184**(11): p. 571-5.

22. Weingart, S.N., et al., *Use of administrative data to find substandard care: validation of the complications screening program.* Med Care, 2000. **38**(8): p. 796-806.
23. Bevan, G. and C. Hood, *What's measured is what matters: Targets and gaming in the English public health care system.* Public Administration, 2006. **84**(3): p. 517-538.
24. Hood, C., *Gaming in targetworld: The targets approach to managing British public services.* Public Administration Review, 2006. **66**(4): p. 515-521.
25. Ignagni, K., *Moving beyond the blame game.* Front Health Serv Manage, 2003. **20**(1): p. 3-14.
26. Bennet, C., et al. (2008) *Beyond the blame game*. 1-52.
27. Cohen, J., *Statistical Power Analysis for the Behavioral Sciences*. second ed. 1988, Philadelphia: Lawrence Erlbaum Associates.
28. Berg, M., et al., *Feasibility first: developing public performance indicators on patient safety and clinical effectiveness for Dutch hospitals.* Health Policy, 2005. **75**(1): p. 59-73.
29. Inspectie, v.d.G., *Rapport Het Resultaat telt 2005*. 2006, Staatstoezicht op de volksgezondheid: Den Haag.
30. Adab, P., et al., *Performance league tables: the NHS deserves better.* Bmj, 2002. **324**(7329): p. 95-8.
31. Freeman, T., *Using performance indicators to improve health care quality in the public sector: a review of the literature.* Health Serv Manage Res, 2002. **15**(2): p. 126-37.
32. Gibberd, R., et al., *Using indicators to quantify the potential to improve the quality of health care.* Int J Qual Health Care, 2004. **16 Suppl 1**: p. i37-43.
33. Spiegelhalter, D., *Funnel plots for institutional comparison.* Qual Saf Health Care, 2002. **11**(4): p. 390-1.
34. Gaynes, R.P. and S. Solomon, *Improving hospital-acquired infection rates: the CDC experience.* Jt Comm J Qual Improv, 1996. **22**(7): p. 457-67.
35. Horn, S.D., *Performance measures and clinical outcomes.* Jama, 2006. **296**(22): p. 2731-2.
36. Jiang, H.J., B. Friedman, and J.W. Begun, *Sustaining and improving hospital performance: The effects of organizational and market factors.* Health Care Manage Rev, 2006. **31**(3): p. 188-96.
37. Lapierre, S.D., C. Batson, and S. McCaskey, *Improving on-time performance in health care organizations: a case study.* Health Care Manag Sci, 1999. **2**(1): p. 27-34.
38. Marciniak, T.A., et al., *Improving the quality of care for Medicare patients with acute myocardial infarction: results from the Cooperative Cardiovascular Project.* Jama, 1998. **279**(17): p. 1351-7.
39. McLoughlin, V., et al., *Improving performance using indicators. Recent experiences in the United States, the United Kingdom, and Australia.* Int J Qual Health Care, 2001. **13**(6): p. 455-62.
40. Pritchard, R., *Measuring and improving organizationalproductivity; a practical guide.* 1990, New York: Praeger.
41. Davenport, R.J., M.S. Dennis, and C.P. Warlow, *Effect of correcting outcome data for case mix: an example from stroke medicine.* Bmj, 1996. **312**(7045): p. 1503-5.
42. Lingsma, H.F., et al., *Variation between hospitals in patient outcome after stroke is only partly explained by differences in quality of care: results from the Netherlands Stroke Survey.* J Neurol Neurosurg Psychiatry, 2008. **79**(8): p. 888-94.
43. Agency for Healthcare Research and Quality (2013) *Selecting Health Outcome Measures for Clinical Quality Measurement*. Tutorials on Quality Measures >.
44. Derose, S.F. and D.B. Petitti, *Measuring quality of care and performance from a population health care perspective.* Annu Rev Public Health, 2003. **24**: p. 363-84.

45. Spertus, J.A., et al., *American College of Cardiology and American Heart Association methodology for the selection and creation of performance measures for quantifying the quality of cardiovascular care.* J Am Coll Cardiol, 2005. **45**(7): p. 1147-56.

46. Kraus, T.W., M.W. Buchler, and C. Herfarth, *Relationships between volume, efficiency, and quality in surgery--a delicate balance from managerial perspectives.* World J Surg, 2005. **29**(10): p. 1234-40.

47. Casparie, A.F., *[Postoperative wound infections: a useful indicator of quality of care?] Postoperatieve wondinfecties: een bruikbare indicator voor de kwaliteit van zorg?* Ned Tijdschr Geneeskd, 2000. **144**(10): p. 460-2.

48. Rubin, H.R., P. Pronovost, and G.B. Diette, *The advantages and disadvantages of process-based measures of health care quality.* Int J Qual Health Care, 2001. **13**(6): p. 469-74.

Chapter **2**

Random variation and rankability of hospitals using outcome indicators

van Dishoeck AM, Lingsma HF, Mackenbach JP, Steyerberg EW. Random variation and rankability of hospitals using outcome indicators. BMJ Qual Saf. 2011 Oct;20(10):869-74.

ABSTRACT

Objective

There is a growing focus on quality and safety in health care. Outcome indicators are increasingly used to compare hospital performance and to rank hospitals, but the reliability of ranking (rankability) is under debate. We aim to quantify the rankability of several outcome indicators of hospital performance currently used by the Dutch government.

Methods

From 52 indicators used by the Netherlands Inspectorate, we selected nine outcome indicators presenting a fraction and absolute numbers. Of these indicators, four were combined into two, resulting in seven indicators for analysis. We used the official data of 97 Dutch hospitals of the year 2007. We estimated uncertainty of the observed outcome within the hospitals (within hospital variance, $\sigma 2$) with fixed effect logistic regression models. We measured heterogeneity (between hospital variance, $\tau 2$) with random effect logistic regression models. Subsequently, we calculated the rankability by relating heterogeneity to uncertainty within and between hospital s ($\tau 2$ / ($\tau 2$ +median $\sigma 2$)).

Results

Sample sizes varied typically around 200 per hospital (range of median 90-277) with median 2-21 cases, causing a substantial uncertainty of outcomes per hospital. Although 4-8 fold differences between hospitals were noted, the uncertainty within the hospitals caused a poor (< 50%) rankability in 3 indicators and moderate rankability (50-75%) in the other 4 indicators.

Conclusion

The currently used Dutch outcome indicators are not suitable for ranking hospitals. When judging hospital quality the influence of random variation must be accounted for to avoid overinterpretation of the numbers in the quest for more transparency in health care. Adequate sample size is a prerequisite in attempting reliable ranking.

INTRODUCTION

There is a growing focus on quality and safety in health care. Increasingly indicators are used to assess hospital performance. In different countries nationwide systems have been set up to monitor the performance of health care institutions using a framework of structure, process and outcome indicators.(1-2) Public disclosure of the results of hospital performance leads to several inconsistent comparisons and rankings and there is concern among professionals about the value and reliability of such rankings. (3-10) Although rankings seem to be simple, they ignore the chance variability in differences between hospitals and the magnitude of differences. (11) In this research, we focus on the suitability of indicators, specifically outcome indicators, to provide reliable hospital comparisons.

Two core components determine the reliability of hospital comparisons within hospital uncertainty (how reliable are the estimates for each hospital) and between hospital heterogeneity (how large are the differences between hospitals). The amount of uncertainty in the analysis of hospital performance is higher than intuition might suggest.(12) For low-incidence outcome and for smaller subgroups in the population uncertainty can be large.(13) The smallest hospitals would likely experience five to seven times more uncertainty concerning their true performance.(14) The second component is heterogeneity between hospitals.(15) Heterogeneity relates to the true differences beyond chance between hospitals and can be estimated with random effect models. Both components determine the reliability of ranking with an indicator, the “rankability”. The term rankability was first used by van Houwelingen et al. (webpublished research (16)) it measures what part of the variation between the crude hospital effects is due to unexplained differences as opposed to uncertainty. We loosely interpret rankability as the signal to (statistical) noise ratio.

Since there are no minimal sample size requirements for the indicators used by the Dutch government, the numbers may be small, making ranking attempts less reliable. We aim to quantify the rankability of several outcome indicators of hospital performance in the Netherlands

METHODS

Data

We obtained the data from the Netherlands Inspectorate’s indicator set. The inspectorate uses this set to assess possible flaws in the quality of care in Dutch hospitals. This obligatory set includes 21 areas with 52 performance indicators (PIs), of which 14 are outcome indicators presenting both fraction and absolute numbers. Five indicators

were excluded because of clear evidence of registration bias, such as extrapolation of a limited sample in time or patient groups, leaving 9 outcome indicators (table 1). We used the data of 2007, which are publicly available at www.ziekenhuizentransparant.nl. For acute myocardial infarction (AMI), the majority of hospitals reported the in-hospital mortality instead of the 30-day mortality. Several hospitals report both. Using these data, we multiplied 0.74 to the 30-day mortality to include data for the five hospitals that only reported 30-day mortality.

Indicator	Numerator	Denominator
Nosocomial Pressure Ulcer (PU) prevalence among hospitalized patients	Number of patients with a pressure ulcer gr. 2-4	All hospitalized patients who were examined for the presence of PU
Pressure Ulcer (PU) incidence after total hip replacement	Number of patients with a pressure ulcer gr. 2-4	All total hip replacement patients
Bile duct leakage within 30 days after cholecystectomy	Number of patients with bile duct leakage within 30 days after cholecystectomy	All patients with a cholecystectomy
Unintended reoperation after colorectal surgery	Number of unintended reoperation after colorectal surgery	All colorectal operations excluding appendix
In hospital mortality after acute myocardial infarction (AMI) for patients younger than 65 years	Number of patients younger than 65 years deceased during hospitalization because of AMI	All patients younger than 65 years hospitalized because of AMI
In hospital mortality after acute myocardial infarction (AMI) for patients of 65 year and older	Number of patients 65 years and older deceased during hospitalization because of AMI	All patients 65 years and older hospitalized because of AMI
Readmission after heart failure for patients younger than 75 year	Number of readmissions after heart failure within 12 weeks after hospital discharge in patients younger than 75 years	All patients younger than 75 years admitted for heart failure.
Readmission after heart failure for patients 75 year and older	Number of readmissions after heart failure within 12 weeks after hospital discharge in patients 75 years and older	All patients younger 75 years and older admitted for heart failure.
Remaining cancer tissue after breast-conserving lumpectomy	Number of patients in whom cancer tissue is left after an initial local excision of a malignant breast tumour	All patients treated with a local excision of a malignant breast tumour

Table 1 Outcome indicators and their description

Uncertainty

We used numerators and denominator data for each hospital to create a patient level dataset. We estimated a coefficient for unfavorable outcome for each hospital and compared it to the overall average, using a fixed effect logistic regression model with an offset variable and hospital as a categorical variable. The standard error of the esti-

mated coefficient (σ^2) indicates the uncertainty of the estimate, or the within-hospital variance. We take the median σ^2 over all hospitals as a summary of the within-hospital variance. The median is used because of the skewed distribution of the σ^2.

Heterogeneity

We fitted a random effect logistic regression model to estimate unexplained heterogeneity, indicated by τ^2 (the between hospital variance). Unlike the fixed effect model, the random effect model accounts for the fact that the observed outcomes for smaller hospitals can take on extreme values because of random variation. The variance indicates the differences between hospitals beyond chance.(17)

For the interpretation of τ^2 we calculated a 95% range of odds ratios for the hospital s compared to the average as $=\exp(-1{,}96 * \tau^2);\exp(1{,}96 * \tau^2)$.(18)

Rankability

To estimate rankability, we use the following formula:

$$\rho = \tau^2 / (\tau^2 + \text{median } \sigma^2)$$

Rankability relates the heterogeneity τ^2 from the random effect logistic regression model (differences between the hospitals) to the standard error σ^2 of the individual hospitals from the fixed effect logistic regression model. Rankability can be interpreted as the part of heterogeneity between hospitals that is due to unexplained differences, and the rest is due to natural variation or chance. Therefore, rankability describes the reliability of ranking.

Case-mix adjustment

The data on performance indicators did not include patient characteristics, except for two outcomes; AMI mortality and heart failure re-admission. The original indicators are stratified by age. We combined the indicators AMI <65 years + ≥65 years; and heart failure <75 years + ≥75 years in two datasets and applied a limited age adjustment by putting age group in the fixed part of the random effect model.

The statistical analysis was performed with R statistical software (version 2.7.1, R Foundation for Statistical Computing, Vienna, Austria), using the lme4 library to fit random effect logistic regression models.

RESULTS

We studied nine outcome indicators, described in table 1, of which we combined the age groups of the myocardial infaction indicator and the heart failure indicator, resulting in two indicators in stead of four.

Within hospital uncertainty

The number of cases as well as the total number of patients per hospital varied widely for the different indicators (table 2).

Indicator	Number of hospitals	Median Cases (range)	Median N (range)	Median outcome % (range)
Nosocomial Pressure Ulcer prevalence	93	10 (0-39)	233 (59-548)	3,7 (0-11,1)
Nosocomial Pressure Ulcer incidence total hip replacement	90	2 (0-23)	197 (26-1131)	1,1 (0-8,9)
Leakage of the bile duct within 30 days after cholecystectomy	95	2 (0-7)	255 (109-625)	0,5 (0-3,63)
Unintended reoperation after colorectal surgery	94	15 (0-47)	209 (57-557)	6,9 (0-18,4)
In hospital mortality after AMI age <65 years	88	1 (0-17)	85,5 (4-720)	1,1 (0-6,8)
In hospital mortality after AMI age ≥65 years	88	10 (0-46)	117,5 (28-541)	8,6 (0-20,8)
Readmission after heart failure age <75 years	93	6 (0-30)	77 (13-389)	7,9 (0-22,6)
Readmission after heart failure age ≥75 years	93	10 (0-50)	133 (13-376)	8,0 (0-23,1)
Remaining cancer tissue after breast-saving lumpectomy	94	7 (1-46)	76 (14-300)	10,5 (1,2-35,7)

Table 2 Descriptive statistics

For instance, pressure ulcer prevalence varied from 0-39 cases, while the number of patients ranged from 59–548. For cholecystectomy, the number of cases with bile duct leakage was very small (median 2). A considerable number of hospitals reported zero cases (29 out of 97), resulting in a median incidence of leakage of the bile duct of 0,5%. The within hospital uncertainty was largest among cholecystectomy patients (σ 1,01), and pressure ulcer incidence (σ 0,85), due to small number of cases (table 3).

Indicator	sigma^2	tau2	95% range OR -	95% range OR +	rankability
Nosocomial Pressure Ulcer prevalence	0,19	0,11	0,52	1,91	37%
Nosocomial Pressure Ulcer incidence total hip replacement	0,85	0,16	0,46	2,17	38%
Leakage of the bile duct within 30 days after cholecystectomy	1,01	0,00	1	1	0%
Unintended reoperation after colorectal surgery	0,12	0,29	0,35	2,86	71%
In hospital mortality after AMI age groups combined#	0,19	0,27	0,36	2,76	58%
Readmission after heart failure age groups combined#	0,14	0,15	0,47	2,11	51%
Remaining cancer tissue after breast-saving lumpectomy	0,25	0,28	0,35	2,82	53%

Table 3 Rankability. # results for the combined age groups are adjusted for age

Between hospital heterogeneity

Heterogeneity between the hospitals varied from none (τ^2 0) for cholecystectomy, to τ^2 0,29 for colorectal surgery. The corresponding 95% range of the odds ratios was 0,35 and 2,86 for colorectal surgery, meaning that hospitals at the higher end of the distribution had a 2,86 higher chance of re-operation than in the average hospital. Similar at the lower end of the distribution patients had a 0,35 lower chance of reoperation. This was equivalent to an eight-fold difference between the hospitals for this indicator.

Rankability

Due to the large between hospital differences, rankability was the highest (71%) for colorectal surgery and the lowest (<50%) for the indicators pressure ulcer prevalence, pressure ulcer incidence, and cholecystectomy (table 3). For pressure ulcer the rankability was relatively low despite a σ^2 of 0,19 related to the small between hospital differences (τ^2). Rankability was moderate (50%-75%) for the indicators colorectal surgery, AMI, heart failure readmission, and breast saving lumpectomy.

Adjustment for case-mix revealed that a part of the heterogeneity in the AMI indicator was by age. For heart failure readmission, age was borderline significant. Rankability for the combined indicator AMI was 58% and for heart failure 51%.

DISCUSSION

We tested several outcome indicators on their reliability for ranking hospitals using the concept of rankability. Rankability indicates what part of the variation between the crude hospital effects is due to true differences (as opposed to measurement error). Combining fixed effect logistic regression models and random effect logistic regression models, we could estimate uncertainty within the individual hospitals and the unexplained heterogeneity between hospitals. We found considerable variability due to chance alone within the hospitals. On the other hand, the unexplained differences between the hospitals were small for some indicators. Both lead to low rankability.

It should be noted that ranking is a specific form of hospital comparisons. Although the amount of uncertainty is an important factor in all hospital comparisons, rankings in addition ignores the magnitude of the differences. E.g. when the random effect estimates of 10 hospitals show that they all have very similar outcomes, ranking hem from 1 to 10 ignores the similarity. Therefore, reporting rankability is even more relevant for rankings.

The indicators in our research showed substantial uncertainty that influenced rankability. For cholecystectomy, there were no differences other than those by chance alone between the hospitals. Using this indicator for ranking hospitals is useless. This adds to the criticism by de Reuver et al about this indicator.(19) Substantial heterogeneity led to larger rankability in the colorectal surgery indicator (71%). Nevertheless, for this indicator it remains unclear how much of these differences are caused by case mix. It is plausible that a different indication for surgery such as traumatic injury or colorectal cancer may play a role in reoperation rate. Case mix correction should be performed before using this indicator for ranking hospitals. The lack of heterogeneity influences the rankability of the pressure ulcer prevalence. For AMI and heart failure, we were able to perform a simple stratification for two age groups. Combining both age groups resulted in a larger number of cases and total numbers. While rankability of the group of patients younger than 65 was low due to the limited number of cases, the pooled data stratified for age had a moderate rankability (51%).

In order for rankability to be large, the between variance needs to dominate the within variance. Therefore measuring performance should be precise and with adequate sample size if we want to distinguish between hospitals. Rankability combines both the within variance and the between variance. If the between-variance (heterogeneity) is large, we can accept more within-variance to still be able to distinguish between hospitals.

The measurement of rankability provides a way of assessing reliability of ranking. We might compare rankability with the signal-to-noise ratio that is used for electrical signals and is defined as the power ratio between a signal (meaningful information) and the background noise (unwanted signal). So, an indicator provides a signal on quality of care, which is corrupted by random variation. The problem with ranking on crude hospital performance occurs when a rare event is chosen for the indicator, like mortality. Some hospitals have small sample sizes that make the statistics for the performance unstable and the rank order unlikely to replicate. One might also argue that ranking should be avoided. Furthermore, if for "pay for performance" or "quality bonus" initiatives are attempted, the signal to noise ratio should be large not to falsely accuse hospitals or individuals.

Lingsma et al used rankability to assess ranking of a small numbers of IVF clinics.(20) They found considerable heterogeneity, while uncertainty per clinic was small because of large numbers (median 654 cycles). This resulted in a substantial rankability with only 10% of the observed differences between the clinics attributed to chance. (20) Compared to this research, rankability in our data was much lower. In the Dutch outcome indicators, not only the total numbers of patients was sometimes small (median between 90 and 277) but also the outcome was frequently low. Simple rankings based on fixed effects of hospital performance disregards both the magnitude and the uncertainty of the differences between hospitals. (21) An illustrative example is the cholecystectomy indicator, where the number of cases was too low to detect any differences between hospitals. Small samples and low event rates limit the statistical power of the comparison between hospitals.(22)

This raises questions about minimal power calculations or combining indicators to provide sufficient sample size to decrease measurement error. Classical power calculation or estimating minimal cases and total numbers might be performed using Cohen's D. D is defined as the difference between two means divided by the standard deviation. Effect sizes are commonly defined as small, d =0.2, medium, d =0.5, and large, d =0.8. We might use a variant of Cohen's D for event rate. The population size for d=0,5 than is at least 200 and at least 800 for d= 0,2 for indicators with sufficient event rates.(23) These numbers can be used as "a rule of thumb" for the assessment of the reliability of ranking hospitals. Actual calculations of required sample sizes for random effect models are much more complex and theoretical work on this topic is needed. Looking at the sample sizes for the pressure ulcer indicator (59-548) in the Dutch hospitals, it is questionable if this indicator will ever be suitable for ranking hospitals. The maximal sample size is limited by the number of beds in a hospital. In case of inadequate numbers the presentation the results of a specific indicator could be done using funnel plots, since this presentation visualizes the differences between hospitals in relation to random variation.(24) In addition, crude random effect esti-

mates including a measure for rankability might be informative for stakeholders that are able to interpret them, e.g. hospitals or the government. Realistic presentation is important to avoid gaming and truly encourage actions to improve the quality of care. (25)

A categorization for rankability is yet still arbitrary. Lingsma et al suggested that > 70% rankability should be fair to rank hospitals.(20) Higgins et al assigned adjectives of low, moderate and high to the I^2 values of 25%, 50% and 75%.(26) I^2 is used to measure heterogeneity in meta-analyses(27) and is similar in nature to our rankability measure. I^2 can be interpreted as the percentage of the total variability in a set of effect sizes due to heterogeneity, that is, to between study variability. Adopting this categorization, we found that none of the outcome indicators had a high rankability. It could be argued that in case of moderate rankability, "expected ranks" should be used that take into account random variability.(13-15) This requires statistical knowledge and access to advanced statistical programs. No ranking attempt should be made when rankability is low. It might also be interesting to identify subsets of hospitals that meet or exceed a standard, fall below a standard, and a subset that cannot be classified due to sample size limitations. The random effect estimates with confidence intervals shows if a hospital significantly differs from the mean beyond statistical

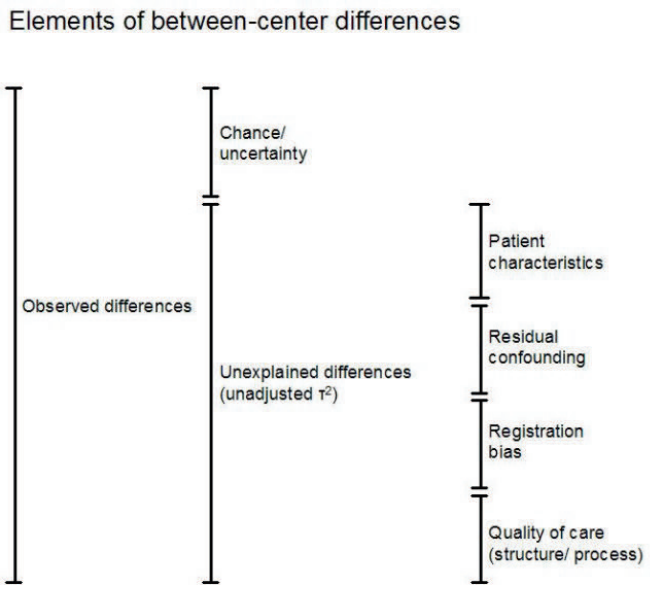


Figure 1 Conceptual framework of between- hospital differences. Observed differences can be divided in random variation and unexplained differences, which can be further attributed to patient characteristics that were not adjusted for, residual confounding because of imperfect case-mix correction, registration bias. Differences in quality of care remain as explanation for a final part of between- hospital differences.

uncertainty. In that, random effect estimates can be used to identify subsets, next to funnel plots as a graphical display of these subsets.

Reliability of ratings depends on sample size and heterogeneity, but also on biases. We can draw a conceptual framework to summarize the elements of between hospital differences (figure 1).(20)

The observed differences can be divided in unexplained differences and chance. By using random effect models chance can be corrected for, leaving patients characteristics, registration bias, quality of care and residual confounding as elements of the unexplained differences. Consequently, ranking reflects the total of unexplained differences between hospitals and not true differences in the quality of care. This is a limitation of this study, but the data as publicly reported does not provide any additional information.

We conclude that none of the currently used Dutch outcome indicators is suitable for ranking hospitals. When judging hospital quality the influence of random variation must be accounted for to avoid overinterpretation of the numbers in the quest for more transparency in health care. Adequate sample size is a prerequisite in attempting reliable ranking.

Acknowledgement; JPM and EWS had the original idea for the study, EWS developed the study design. AMvD gathered and analyzed the data with HL. AMvD wrote a first draft of the paper. All authors contributed to further drafts.

Competing interests: None

Source of Funding: Internal Erasmus MC grant for health care research (Mrace)

REFERENCES

1. Jencks SF, Cuerdon T, Burwen DR, Fleming B, Houck PM, Kussmaul AE, et al. Quality of medical care delivered to Medicare beneficiaries: A profile at state and national levels. Jama. 2000 Oct 4;284(13):1670-6.
2. Berg M, Meijerink Y, Gras M, Goossensen A, Schellekens W, Haeck J, et al. Feasibility first: developing public performance indicators on patient safety and clinical effectiveness for Dutch hospitals. Health Policy. 2005 Dec;75(1):59-73.
3. Halasyamani LK, Davis MM. Conflicting measures of hospital quality: ratings from "Hospital Compare" versus "Best Hospitals". Journal of hospital medicine (Online). 2007 May;2(3):128-34.
4. Lemmers O, Kremer JA, Borm GF. Incorporating natural variation into IVF clinic league tables. Human reproduction (Oxford, England). 2007 May;22(5):1359-62.
5. Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. Lancet. 2004 Apr 3;363(9415):1147-54.
6. Mohammed MA, Mant J, Bentham L, Raftery J. Comparing processes of stroke care in high- and low-mortality hospitals in the West Midlands, UK. Int J Qual Health Care. 2005 Feb;17(1):31-6.
7. Ranstam J, Wagner P, Robertsson O, Lidgren L. Health-care quality registers: outcome-orientated ranking of hospitals is unreliable. The Journal of bone and joint surgery. 2008 Dec;90(12):1558-61.
8. Jacobs R, Goddard M, Smith PC. How robust are hospital ranks based on composite performance measures? Med Care. 2005 Dec;43(12):1177-84.
9. Spiegelhalter D. Ranking institutions. The Journal of thoracic and cardiovascular surgery. 2003 May;125(5):1171-3; author reply 3.
10. Anderson J, Hackman M, Burnich J, Gurgiolo TR. Determining hospital performance based on rank ordering: is it appropriate? Am J Med Qual. 2007 May-Jun;22(3):177-85.
11. Adab P, Rouse AM, Mohammed MA, Marshall T. Performance league tables: the NHS deserves better. Bmj. 2002 Jan 12;324(7329):95-8.
12. Diehr P, Cain K, Connell F, Volinn E. What is too much variation? The null hypothesis in small-area analysis. Health Serv Res. 1990 Feb;24(6):741-71.
13. Robertsson O, Ranstam J, Lidgren L. Variation in outcome and ranking of hospitals: an analysis from the Swedish knee arthroplasty register. Acta orthopaedica. 2006 Jun;77(3):487-93.
14. Davidson G, Moscovice I, Remus D. Hospital size, uncertainty, and pay-for-performance. Health care financing review. 2007 Fall;29(1):45-57.
15. Normand S-LT, Shahian DM. Statistical and clinical aspects of hospital profiling. Statistical Science. 2007;22(2):206-26.
16. Houwelingen van JC, Brand R, Louis TA. Empirical bayes methods for monitoring health care quality. 2005.
17. Spiegelhalter DJ. Handling over-dispersion of performance indicators. Qual Saf Health Care. 2005 Oct;14(5):347-51.
18. Spiegelhalter D, Abrahams, Myles. Baysean approaches to clinical trials and health care evaluation2004.
19. de Reuver PR, Gouma DJ. [Bile leakage. A performance indicator with markedly different consequences for the patient, specialist and care insurer]. Nederlands tijdschrift voor geneeskunde. 2007 Aug 4;151(31):1709-12.

20. Lingsma HF, Eijkemans MJ, Steyerberg EW. Incorporating natural variation into IVF clinic league tables: The Expected Rank. BMC medical research methodology. 2009 Jul 16;9(1):53.

21. Lingsma HF, Dippel DW, Hoeks SE, Steyerberg EW, Franke CL, van Oostenbrugge RJ, et al. Variation between hospitals in patient outcome after stroke is only partly explained by differences in quality of care: results from the Netherlands Stroke Survey. J Neurol Neurosurg Psychiatry. 2008 Aug;79(8):888-94.

22. Dimick JB, Welch HG, Birkmeyer JD. Surgical mortality as an indicator of hospital quality: the problem with small sample size. Jama. 2004 Aug 18;292(7):847-51.

23. Cohen J. Statistical Power Analysis for the Behavioral Sciences. second ed. Philadelphia: Lawrence Erlbaum Associates; 1988.

24. Spiegelhalter DJ. Funnel plots for comparing institutional performance. Stat Med. 2005 Apr 30;24(8):1185-202.

25. Gibberd R, Hancock S, Howley P, Richards K. Using indicators to quantify the potential to improve the quality of health care. Int J Qual Health Care. 2004 Apr;16 Suppl 1:i37-43.

26. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Bmj. 2003 Sep 6;327(7414):557-60.

27. Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychological methods. 2006 Jun;11(2):193-206.

Chapter 3

Displaying random variation in comparing hospital performance

van Dishoeck AM, Looman CW, van der Wilden-van Lier EC, Mackenbach JP, Steyerberg EW. Displaying random variation in comparing hospital performance. BMJ Qual Saf. 2011 Aug;20(8):651-7.

ABSTRACT

Introduction

The role of transparency in quality of care is becoming ever more important. Various indicators are used to assess hospital performance. Judging hospitals using rank order takes no account of disturbing factors such as random variation and case-mix differences. The purpose of this article is to compare displays for the influence of random variation on the apparent differences in the quality of care between the Dutch hospitals.

Method

We analysed the official 2005 data of all 97 hospitals on the following performance indicators: pressure ulcer, cerebro-vascular accident and acute myocardial infarction. We calculated confidence intervals of the point estimate and the simulated confidence intervals of the ranks with bootstrap sampling. We visualized the influence of random variation with three modern graphical techniques; forest plot, funnel plot and rank plot.

Results

Statistically significant differences between hospitals were found for nearly all performance indicators ($p < 0.001$). However, the confidence intervals in the forest plot revealed that only a small number of hospitals performed significantly better or worse. The funnel plot provides a representation of differences between hospitals compared to a target value and allows for the uncertainty of these differences. Ranking hospitals was very uncertain, as well visualized by a rank plot.

Conclusion

Despite statistically significant differences between hospitals, random variation is a crucial factor that must be taken into account when judging individual hospitals. The funnel plot provides easily interpretable information on hospital performance, including the influence of random variation.

INTRODUCTION

Hospitals face increasing demands with regards to the quality, transparency and accountability of health care. Since the early 1980s interest in measuring hospital performance has led to the development of many performance indicators (PIs). A PI is a measurable element of practice performance for which there is evidence or consensus that it can be used to assess the quality of care.(1) The purpose of performance indicators must balance the need for accountability with the need to promote quality improvement initiatives (2), therefore provide the incentive to improve the quality of care.

League tables are often used for displaying hospitals performance (figure 1), suggesting a rank order.

League tables provoke concerns among health service providers for several reasons including concerns over adjustment for case-mix and the role of chance in determining their rank.(3) Because there are no minimal sample size requirements in PI measurement, random variation plays an important role in the interpretation of the results. In the Netherlands quality of care in hospitals is assessed by the Health Care Inspectorate (NHCI).(4) In 2003, the NHCI developed a public and obligatory set of PIs to guide their assessment of the quality of care delivered in hospitals.(5-6) In principle

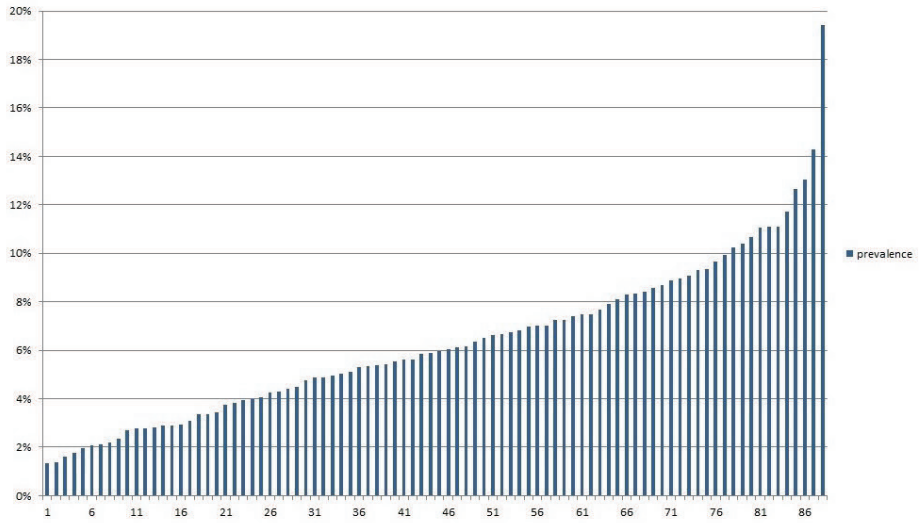


Figure 1. League table of the pressure ulcer prevalence in 89 Dutch hospitals in 2005

this set of PIs enables the Inspectorate to identify hospitals whose performance lies below the minimum standard, and guide further investigation into these hospitals. The Inspectorate publishes the anonymous data in a yearly report that presents data on more than 40 indicators concerning structure, process and health outcome using league tables of point estimates. This paper focuses on the particular aspect of random variation in comparing and ranking of institutions and explores three graphical displays describing the data.

METHODS

Data

For the analysis we used the 2005 publicly available data on three performance indicators; pressure ulcer (PU), cerebro vascular accident (CVA) and acute myocardial infarction (AMI). The indicators are selected to reflect several problems that occur with PI, like low total numbers or low number of cases. We assume hypothetically that the data provide a fair reflection of the quality of care in individual hospitals and that there is no significant effect of case-mix. The data is publicly available at www.ziekenhuizentransparant.nl.(7)

Statistical analysis

On the basis of absolute numbers (n) and the cases (y), we calculated the standard error and 95% confidence interval (CI), where y is the number of cases, and n the total number of patients in a hospital. The confidence interval was calculated using the formulas

$$CI=e^{\alpha \pm 1.96.se}, \; se=\sqrt{(1/y+(1/n-y))}, \text{ and } \alpha=\log (y/n/(1-y/n)). \qquad (8)$$

With the function Qbinom in S plus we calculated the 95% CI for the number of successes obtained in a number of binomial trials equal to the number of patients that was judged with the observed probability of being a case. These were divided again by the number of trials to obtain a CI that reflects the discrete character of the observations. CI's for the ranks were calculated by a parameterized bootstrap with the observed probability of being a case per hospital as input.(9) Differences between the hospitals were calculated using a likelihood ratio test. A p-value $P<0.05$ was considered statistically significant.

Graphical methods

We considered three techniques to visualize the influence of random variation:

1. A forest plot ranks the point estimate and the confidence interval represented by horizontal lines for each hospital in ascending order. A vertical line represents a preselected norm or standard. (10-11)
2. In a funnel plot the estimates of the hospitals are plotted together with the confidence limits of a norm or national average.(12) The confidence limits are calculated in relation to the number of patients per hospital. It is customary to plot both 95% and 99.8% confidence intervals, corresponding to approximately 2 and 3 standard errors width.(12) We calculated the confidence interval taking into account the discrete nature of the numbers. This exact calculation was necessary because the number of scores in which y = 0 was high in some indicators.
3. A rank plot uses bootstrapping to estimate the confidence interval around the rank and plots the true rank against the estimate of the bootstrap replica's and their confidence intervals.(13) Bootstrap samples are generated using a random draw with replacement to resample the individual observations from the original group. Per hospital thousand bootstrap replicas were generated from the binomial distribution. In this way, 1000 new datasets reflect what could be observed under the same circumstances. The rank numbers were determined for each new dataset. The distribution of ranking on the 1000 data sets was the basis of 95% confidence limits of the ranks.

RESULTS

The overall results are described in table 1 and the graphical display of the influence of random variation is illustrated by examples for each of the indicator areas separately. There were significant differences between the hospitals for almost all PI reported, as is summarized more in detail in table 1.

The population size varied substantially between the indicators. For both age groups of the hemorrhagic CVA indicator the mean population size was rather small (n < 40). For the AMI indicators in the age group <65 years the number of cases was small (3 and 4) leading to borderline significant difference between the hospitals.

Pressure Ulcers

The forest plot shows, that the point estimates ranged from 1.3 to 19.4% for pressure ulcer. The confidence intervals surrounding the point estimates had a wide variation (figure 2a).

For instance, the prevalence of pressure ulcers in the first hospital in figure 2a was 1.3%, but with a confidence interval ranging from 0 to 9%. The wide range was due to

Indicator	centra n	mean cases and patients	mean outcome* % (95% CI)	p†
Pressure Ulcer prevalence	89	14/238	6,0 (4,7-7,8)	< 0,0001
Ischemic Stroke; 7-day mortality				
patients< 65 year	91	2/63	3,2 (1,6-6,3)	< 0,0001
patients ≥ 65 year	91	10/178	5,6 (4,1-7,7)	< 0,0001
Hemorrhagic Stroke; 7-day mortality				
patients< 65 year	90	2/12	16,7 (8,4-30,3)	0,0005
patients ≥ 65 year	90	9/35	25,7 (19,0-33,8)	< 0,0001
Acute myocardial infarction patients ≥ 65 year				
In-hospital mortality	37	16/131	12,2 (9,6-15,4)	< 0,0001
30-day mortality	53	16/146	11,0 (8,6-13,8)	< 0,0001
Acute myocardial infarction patients < 65 year				
In-hospital mortality	37	3/102	2,9 (1,7-5,2)	0,06
30-day mortality	53	4/124	3,2 (0,2-5,3)	0,003

Table 1 Description of the performance indicators used in this comparison. Data was obtained from the public available database. *The mean outcome refers to the outcome of the mean Dutch hospital. †The p-value tests the hypothesis that there are no differences in quality of care between the Dutch hospitals.

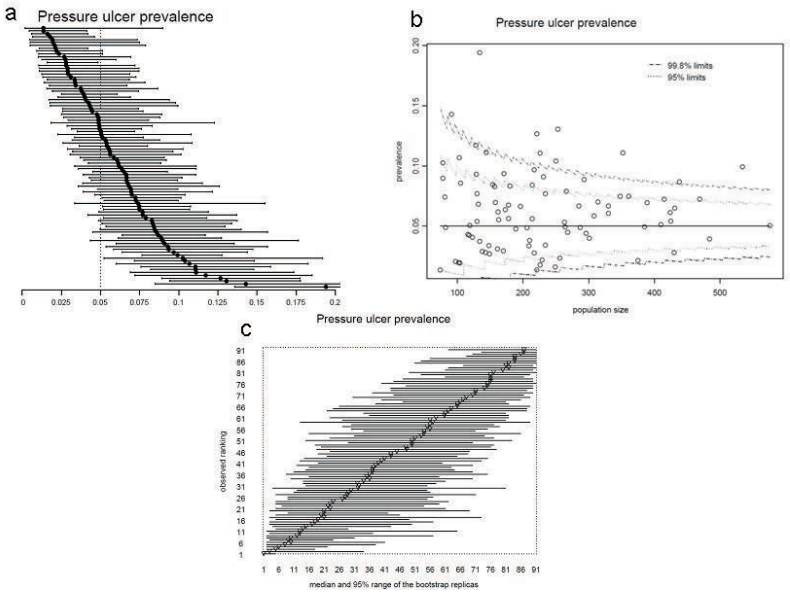


Figure 2 Three different graphical representations of the pressure ulcer prevalence in the Dutch hospitals in 2005. (a) 'Forestplot' with point estimate for each hospital and 95%-CI. The vertical line represents the Dutch norm for the pressure ulcer prevalence. (b) 'Funnelplot' in which pressure ulcer prevalence is plotted against the population measured in the hospital. The horizontal line represents the Dutch norm (5%), the funnel shaped lines are the limits of the 95%-CI (2xSE) and the 99,8%-CII (3xSE); Hospitals situated outside these limits perform significantly different from the norm. (c) 'Rank plot' showing the ranks for pressure ulcer prevalence compared with the ranking according to the median of the bootstrap replicas and its CI.

the small number of patients (74). In contrast, the second hospital listed in this graph, also scored 1.3% prevalence, but with confidence intervals of 0.5 and 4%. Despite the equal results, the second hospital performed significantly better than the national standard of the Inspectorate, which is 5%.The funnel plot shows the confidence intervals of this 5% norm (figure 2b). Hospitals situated above and below the 95% confidence intervals had a point estimate more than 2 times the standard error. Seven hospitals performed significantly better, with point estimate below 95% confidence level. Ranking hospitals on the basis of presser ulcer prevalence was very uncertain, based on the wide confidence intervals of the bootstrap samples (figure 2c).

Cerebro vascular accident

To illustrate the influence of random variation, we consider the 7-day mortality after a hemorrhagic stroke in patients younger than 65 years (figure 3).

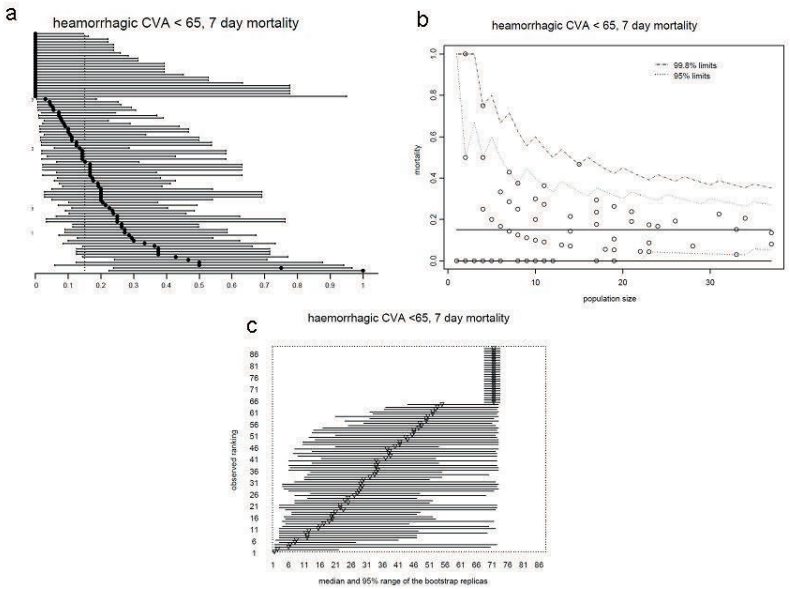


Figure 3. Three different graphical representations of the stroke indicator in the Dutch hospitals in 2005. (a) 'Forestplot' with point estimate for each hospital and 95%-CI. The vertical line represents the average for stroke mortality. (b) 'Funnelplot' in which mortality is plotted against the population measured in the hospital. The horizontal line represents the average mortality, the funnel shaped lines are the limits of the 95%-CI (2xSE) and the 99,8%-CII (3xSE); Hospitals situated outside these limits perform significantly different from the average. (c) 'Rank plot' showing the ranks for stroke mortality compared with the ranking according to the median of the bootstrap replicas and its CI.

The point estimates of hospital mortality varied, ranging from 0-100% mortality. The wide confidence intervals in the forest plot are due to the fact that small numbers of patients were admitted to the hospitals in 2005. The first 24 hospitals reported a mortality of 0%. However, hospital number 24 admitted only 2 that year, providing a confidence interval from 0-100% (figure 3a). The funnel plot shows that apart from random variation there were few differences between the hospitals (figure 3b). The rank plot reveals wide confidence intervals making the ranking attempt very uncertain (figure 3c).

Acute myocardial infarction

We illustrate the influence of random variation on 30-day hospital mortality after AMI in patients younger than 65 years. The point estimates in the forest plot ranged from 0-9,8%, with different confidence intervals based on patient numbers (figure 4a). Given a mean score of 2.5% mortality, only two hospitals performed significantly worse than the others. The funnel plot shows that it is hard to distinguish between

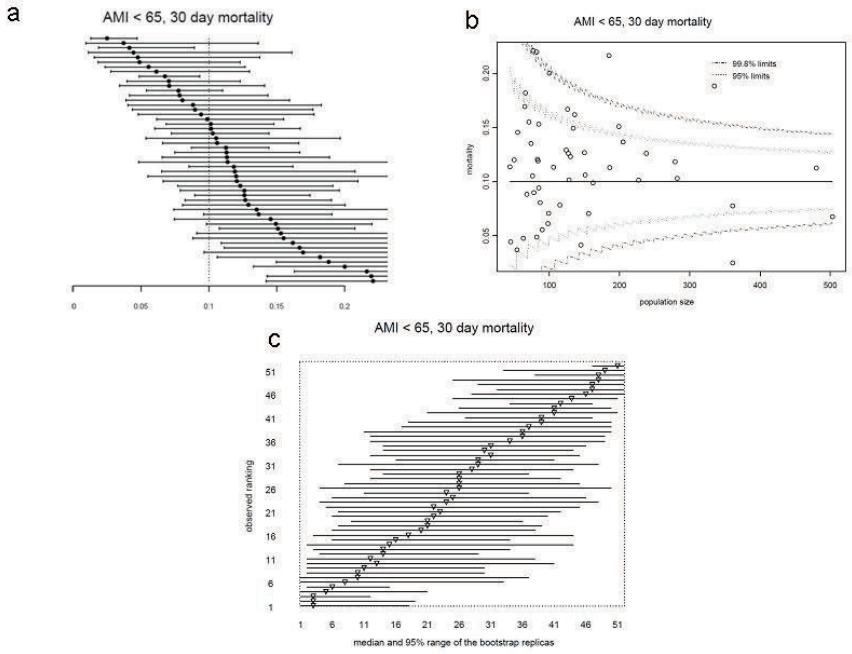


Figure 4 Three different graphical representations of the AMI mortality in the Dutch hospitals in 2005. (a) 'Forestplot' with point estimate for each hospital and 95%-CI. The vertical line represents the average for AMI mortality. (b) 'Funnelplot' in AMI mortality is plotted against the population measured in the hospital. The horizontal line represents the average AMI mortality, the funnel shaped lines are the limits of the 95%-CI (2xSE) and the 99,8%-CII (3xSE); Hospitals situated outside these limits perform significantly different from the average. (c) 'Rank plot' showing the ranks for AMI mortality compared with the ranking according to the median of the bootstrap replicas and its CI.

hospitals that performed well and those that performed poorly (figure 4b). No meaningful ranking of hospitals could be done on the basis of AMI mortality (figure 4c).

DISCUSSION

Although league tables provide a simple overview of the data, they are easy to misinterpret for this ranking of crude hospital performance does not include chance variability.(3) Graphical displays should show the data and avoid distorting what the data have to say.(14) The graphs should encourage the eye to compare different aspects of the data, such as the magnitude of differences in performance between the hospitals, as well as uncertainty. Therefore visualizing random variation is crucial. To accomplish this we chose the most common graphical display in scientific medical research; the forest plot.(11) Researchers often use the funnel plot in meta-analyses to display publication bias and other sample size effects.(15-16) In more recent research the funnel plot has been suggested for displaying data in public reporting of hospital performance.(12, 17-19). Using the rank plots we aimed at only visualizing the uncertainty in the rank order. All three displays focus on random variation, on different levels. The forest plot visualizes the CI of the individual hospital, while the funnel plots focuses on the significant differences between the hospitals. The rank plot provides insight in the chance variation of the ranks.

The forest plot ranks hospitals on the point estimate, but also provides information on the confidence intervals (CI). Our data shows that the CI varies substantially, depending on sample size. For example, in pressure ulcer two hospitals had the same point estimate, but the interpretation regarding the quality of care delivered in these hospitals differs in that only one performed significantly better. Therefore, the relative position is displayed but not easily interpretable.

In our experience, the funnel plot provides a straightforward representation of the differences between hospitals. The hospitals situated outside the 95% confidence intervals performed significantly worse or better in relation to a target or national average. The funnel plot clearly reveals that quality of care could not be measured using the stroke indicators because of the small numbers in individual hospitals. Small numbers make proper interpretation virtually impossible, because the vast majority of the apparent differences may be due to random variation.(12) The funnel plot also provides professionals the information to compare their own performance with that of other hospitals with the same volume and subsequently set their own targets. The funnel plot provides a good overview of the relative position of the individual hospitals.

Although ranking raw scores provides an easy way to compare hospitals, the variance of the original measures strongly influence the rank. This is seen in the rank plot which shows that random variation greatly influences the interpretation of what the true rank might be. Ranking may even be misleading, since random variation plays a dominant role for some indicators, such as stroke and AMI. The overview of the relative position is limited.

When a graph is constructed, information is encoded. The visual decoding is called graphical perception.(20) Judging the graphical perception of the different plots we choose the two; criteria described by Cleveland; pattern recognition (including random variation) and table look-up, relating to the accuracy of the relative position of hospital performance.(20) As summarized in table 2, we conclude that the funnel plot is the most attractive graphical display of the three techniques.

Graph	Pattern recognition	Table look-up
Forest plot	+	+
Funnel plot	++	++
Rank plot	+	+

Table 2 Comparing different plots

Several articles discuss the use of league tables in presenting the results of hospital performance.(10, 19, 21-30) They all conclude that, even when there are substantial differences between institutions, simple league tables or ranks are unreliable statistical summaries of performance. Since Spiegelhalter suggested the use of funnel plots for institutional comparison in 2002, several studies describe the usefulness of this plot. (3, 12, 19, 31-32).

This research has several limitations. We concentrated on the role of random variation and paid no attention to bias, such as registration differences, organizational differences, or the influence of case mix. With regard to the latter, it is likely that university hospitals or hospitals in urban areas have a different patient population than small hospitals or hospitals in rural areas.(33) Correction for these confounding factors with these PI is impossible because the public available data does not include patient characteristics. This requires further investigation. In our methodology we used the most common scientific approach calculating CI and using the described plots. We did not intensively search for other graphical displays to visualise the data. Also the usefulness of the different graphs was not systematically assessed. This requires a more structural approach.(20)

We conclude that despite statistically significant differences between hospitals, random variation is a crucial factor that must be taken into account when judging individual hospitals. The funnel plot provides easily interpretable information on hospital performance, including the influence of random variation.

Acknowledgement

JPM and EvdWvL had the original idea for the study, EWS developed the study design. AMvD gathered and analyzed the data with CL. AMvD wrote a first draft of the paper. All authors contributed to further drafts. Parts of this research has been published in the Dutch language in the Nederlands Tijdschrift voor Geneeskunde.(34)

Competing Interest: None declared.

REFERENCES

1. Colsen P, Casparie A. Indicatorregistratie: een model ten behoeve van integrale kwaliteitszorg in een ziekenhuis. Medisch Contact. 1995;50:297-9.
2. Rubin HR, Pronovost P, Diette GB. The advantages and disadvantages of process-based measures of health care quality. Int J Qual Health Care. 2001 Dec;13(6):469-74.
3. Adab P, Rouse AM, Mohammed MA, Marshall T. Performance league tables: the NHS deserves better. Bmj. 2002 Jan 12;324(7329):95-8.
4. IGZ IvdG. English homepage Utrecht2009 [cited 2009 11-01-09]; Available from: http://www.igz.nl/uk/.
5. Berg M, Meijerink Y, Gras M, Goossensen A, Schellekens W, Haeck J, et al. Feasibility first: developing public performance indicators on patient safety and clinical effectiveness for Dutch hospitals. Health Policy. 2005 Dec;75(1):59-73.
6. IGZ., NVZ., OMS., NFU. Basisset Prestatie-Indicatoren Ziekenhuizen 2005. Gezondheidszorg Ivd, Ziekenhuizen. NVv, Specialisten. OvM, centra. NFvUm, editors. Den Haag: Inspectie voor de Gezondheidszorg; 2005.
7. IGZ. Het Resultaat telt 2005; prestatie-indicatoren als onafhankelijke graadmeter voor de kwaliteit van in ziekenhuizen geleverde zorg. Den Haag: Staatstoezicht op de volksgezondheid2006 december 2006.
8. Altman D. Practical Statistics for Medical research. 8 ed. Roca Raton, London, New York, Washington D.C.: Chapman & hall/CRC; 1999.
9. Efron B, Tibshirani R. An introduction to the bootstrap. New York, London: Chapman & Hall; 1993.
10. Marshall EC, Spiegelhalter DJ. Reliability of league tables of in vitro fertilisation clinics: retrospective analysis of live birth rates. BMJ (Clinical research ed. 1998 Jun 6;316(7146):1701-4; discussion 5.
11. Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. BMJ (Clinical research ed. 2001 Jun 16;322(7300):1479-80.
12. Spiegelhalter DJ. Funnel plots for comparing institutional performance. Stat Med. 2005 Apr 30;24(8):1185-202.
13. Hinkley DV. Bootstrap methods (with Discussion). Journal of the Royal Statistical Society. 1988;B, 50:312-37, 55-70.
14. Tufte E. The visual display of quantitative information. 16 ed. Cheshite, Conneticut: Graphic Press; 1998.
15. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. J Clin Epidemiol. 2005 Sep;58(9):882-93.
16. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics. 2000 Jun;56(2):455-63.
17. Spiegelhalter D. Funnel plots for institutional comparison. Qual Saf Health Care. 2002 Dec;11(4):390-1.
18. Spiegelhalter DJ. Handling over-dispersion of performance indicators. Qual Saf Health Care. 2005 Oct;14(5):347-51.
19. Kirkham JJ, Bouamra O. The use of statistical process control for monitoring institutional performance in trauma care. J Trauma. 2008 Dec;65(6):1494-501.
20. Cleveland WS. The elements of graphical data. Revised edition ed. Summit, New Jersey: Hobart Press; 1994.

21. Denton AS, Bond SJ, Matthews S, Bentzen SM, Maher EJ. Treatment-related morbidity and hospital league tables: experience from a national audit of radiotherapy-induced morbidity in cervical carcinoma. Clinical oncology (Royal College of Radiologists (Great Britain)). 2002 Feb;14(1):40-2.

22. Joyce R, Webb R, Peacock JL, Stirland H. Adjusted mortality rates: a tool for creating more meaningful league tables for stillbirth and infant mortality rates. Public health. 2002 Nov;116(6):315-21.

23. Lemmers O, Broeders M, Verbeek A, Heeten G, Holland R, Borm GF. League tables of breast cancer screening units: worst-case and best-case scenario ratings helped in exposing real differences between performance ratings. Journal of medical screening. 2009;16(2):67-72.

24. Lemmers O, Kremer JA, Borm GF. Incorporating natural variation into IVF clinic league tables. Human reproduction (Oxford, England). 2007 May;22(5):1359-62.

25. Marshall T, Mohammed MA, Rouse A. A randomized controlled trial of league tables and control charts as aids to health service decision-making. Int J Qual Health Care. 2004 Aug;16(4):309-15.

26. Mohammed MA, Mant J, Bentham L, Raftery J. Comparing processes of stroke care in high- and low-mortality hospitals in the West Midlands, UK. Int J Qual Health Care. 2005 Feb;17(1):31-6.

27. Ng JQ, Morlet N, Bremner AP, Bulsara MK, Morton AP, Semmens JB. Techniques to monitor for endophthalmitis and other cataract surgery complications. Ophthalmology. 2008 Jan;115(1):3-10.

28. Nutley S, Smith PC. League tables for performance improvement in health care. J Health Serv Res Policy. 1998 Jan;3(1):50-7.

29. Shaw CD. Health-care league tables in the United Kingdom. Journal of quality in clinical practice. 1997 Dec;17(4):215-9.

30. Tu YK, Gilthorpe MS. The most dangerous hospital or the most dangerous equation? BMC Health Serv Res. 2007;7:185.

31. Schulman J, Spiegelhalter DJ, Parry G. How to interpret your dot: decoding the message of clinical performance indicators. J Perinatol. 2008 Sep;28(9):588-96.

32. Wilson J, Charlett A, Leong G, McDougall C, Duckworth G. Rates of surgical site infection after hip replacement as a hospital performance indicator: analysis of data from the English mandatory surveillance system. Infect Control Hosp Epidemiol. 2008 Mar;29(3):219-26.

33. Martens PJ, Stewart DK, Mitchell L, Black C. Assessing the performance of rural hospitals. Healthc Manage Forum. 2002 Winter;Suppl:27-34.

34. van Dishoeck AM, Looman CM, van der Wilden-van Lier EC, Mackenbach JP, Steyerberg EW. [Outcome assessment in hospitals. The influence of insecurity]. Ned Tijdschr Geneeskd. 2009 Apr 25;153(17):804-11.

Chapter **4**

Use of surgical-site infection rates to rank hospital performance across several types of surgery.

van Dishoeck AM, Koek MB, Steyerberg EW, van Benthem BH, Vos MC and Lingsma HF. Use of surgical-site infection rates to rank hospital performance across several types of surgery. Br J Surg. 2013 Apr;100(5):628-36; discussion 37.

ABSTRACT

Background; Comparing and ranking hospitals based on health outcomes is becoming increasingly popular, although case-mix differences between hospitals and random variation are known to distort interpretation. The aim of this study was to explore whether surgical-site infection (SSI) rates are suitable for comparing hospitals, taking into account case-mix differences and random variation.

Methods; Data from the national surveillance network in the Netherlands, on the eight most frequently registered types of surgery for the year 2009, were used to calculate SSI rates. The variation in SSI rate between hospitals was estimated with multivariable fixed- and random-effects logistic regression models to account for random variation and case mix. 'Rankability' (as the reliability of ranking) of the SSI rates was calculated by relating within-hospital variation to between-hospital variation.

Results; Thirty-four hospitals reported on 13 629 patients, with overall SSI rates per surgical procedure varying between 0 and 15.1 per cent. Statistically significant differences in SSI rate between hospitals were found for colonic resection, caesarean section and for all operations combined. Rankability was 80 per cent for colonic resection but 0 per cent for caesarean section. Rankability was 8 per cent in all operations combined, as the differences in SSI rates were mainly explained by case mix.

Conclusion; When comparing SSI rates in all operations, differences between hospitals were explained by case mix. For individual types of surgery, case mix varied less between hospitals, and differences were explained largely by random variation. Although SSI rates may be used for monitoring quality improvement within hospitals, they should not be used for ranking hospitals.

INTRODUCTION

Surgical-site infection (SSI) is one of the most common complications after surgery. SSIs have a major influence on mortality and morbidity, and length of hospital stay[1,2]. According to the international literature, on average 5 per cent of all surgical patients develop a wound infection, the rate varying from 1 to 15 per cent depending on the type of surgery[3,4]. Recent surveillance in Japan among patients undergoing colorectal surgery showed a SSI rate of 15 per cent[5].{Kobayashi, 2008 #653} In the Netherlands, the overall SSI rate is 3 per cent[6–9].

SSIs can partly be prevented using specific interventions before, during and after surgery[4,10–13]. The effect of these interventions is often measured using outcome indicators such as SSI rates, either prevalence or incidence[14,15]. Surveillance systems have been set up to facilitate the comparison of SSI rates between hospitals, and to stimulate improvement and compliance with guidelines[7,16]. Feedback of the results of these surveillance systems and benchmark activities have shown a reduction in SSI rates[17–20].

Outcome measures such as SSI rates are being used increasingly to compare hospitals' performance using league tables and rank orders. There is, however, doubt concerning the validity of such hospital comparisons based on outcome, because observed differences between hospitals may be partly explained by random variation and by differences in case mix[21–24].

The aim of this study was to explore the influence of random variation and case mix on SSI rates in Dutch hospitals, and to examine whether SSI rates are a suitable measure by which to rank hospitals.

METHODS

The PREZIES network database (http://www.prezies.nl/zkh/index.html) was used. PREZIES is a Dutch initiative for the nationwide surveillance of healthcare-related infections, set up in 1996. The PREZIES network is a consortium of participating hospitals and the National Institute for Health and Environment. The PREZIES initiative aims to improve the quality of care in hospitals by measuring and communicating the incidence of heathcare-related infections. PREZIES achieves this by introducing and maintaining surveillance of SSIs in hospitals[9,11,25]. To obtain the number of SSIs, standardized data collection, registration and follow-up after discharge is mandatory. The required follow-up period differs according to the type of surgery from 30 days to 1 year. The data reflect SSI rates in elective procedures and emergency surgery

combined. Because the distinction between elective and emergency surgery was not always reported, this could not be used as a case-mix variable. The PREZIES network carries out an extensive audit procedure with periodic visits to all participating hospitals.

Study period and types of surgery

Data from 2009 on the eight most frequently registered types of surgery were selected: mastectomy, laparoscopic cholecystectomy, colonic resection, hip replacement, knee replacement, abdominal hysterectomy and caesarean section. One hospital that reported fewer than 20 patients in the total group of selected operations was excluded. The primary outcome was whether the patients developed a SSI, either superficial or deep; such infections are defined in a standardized way and audited[6,25,26].

Statistical analysis

The statistical analysis was performed with R version 2.11 (R Foundation for Statistical Computing, Vienna, Austria) using the *rms* and *lme4* package. First, the variation between hospitals in SSI rates was estimated by fixed-effects logistic regression, with hospital as a categorical variable. This gives a coefficient for the odds of SSI in each hospital, compared with the average, representing only the observed differences between hospitals. The standard error (σ^2) of the individual hospital's coefficients from the fixed-effects logistic regression model represents the within-hospital uncertainty. Low-volume hospitals have larger standard errors and confidence intervals than high-volume hospitals.

Random-effects model and adjusting for case mix

To account for the expectation that part of the variation between hospitals was due to random variation, a random-effects logistic regression model was used. Random-effects models account for random variation, and estimate hospital coefficients and the total variation 'beyond chance'. The total variation is indicated by the model parameter τ^2, the variance of the random effects. For interpretation, τ^2 can be transformed into a 95 per cent range of centre differences. This 95 per cent range represent the odds of SSI of a hospital on the lower end (2.5th percentile) of the distribution and the odds of a hospital on the higher end (97.5th percentile) compared with the average[27].

To display random variation visually, funnel plots were created. In a funnel plot, the crude SSI rates are plotted against the total number of patients in each hospital. In addition, 95 and 99.8 per cent confidence limits are shown[28–30]. These are calculated in relation to the mean number of patients per hospital, taking into account the discrete nature of the numbers.

As a next step, to account for the differences in case mix between hospitals, the random-effects models were fitted, including patient and surgery characteristics that might vary between hospitals and affect the SSI rate. The case-mix variables used were: age, sex, type of surgery (mastectomy, laparoscopic cholecystectomy, colonic resection, reconstruction of abdominal vessels, hip replacement, knee replacement, abdominal hysterectomy or caesarean section), duration of operation, wound classification and American Society of Anesthesiologists (ASA) grade[31,32]. Different potential confounders were not available from the PREZIES database.

Rankability

To see whether it makes sense to rank the hospitals, the 'rankability' was calculated. Rankability relates the variance of the random-effects τ^2 (differences between hospitals) to the standard error (σ^2) of the individual hospital coefficients (differences within hospitals). Rankability can be interpreted as the part of heterogeneity between hospitals represented by unexplained differences that might be due to the quality of SSI prevention, as opposed to random variation. Rankability (ρ) is calculated as $\rho = \tau^2 / (\tau^2 + \text{median } \sigma^2)$[33,34]. The information provided by the rankability addresses the reliability of ranking hospitals by means of the SSI indicator. All analyses were performed for the individual surgery types and for all operations combined.

RESULTS

Thirty-four hospitals provided data on 13.629 patients undergoing one of the selected types of surgery. Six hospitals reported on one surgical intervention, whereas 18 hospitals reported on between two and four different operations; nine hospitals provided data on five to seven different procedures and one hospital provided data on all eight surgical interventions (*Table S1*, supporting information). The study population was predominantly female (74 per cent), owing to four surgical procedures performed mainly on women (mastectomy, abdominal hysterectomy, cholecystectomy and caesarean section). The median age was 64 years. The overall incidence of SSI was 2.8 per cent (378 SSIs in 13.629 patients). Descriptive statistics on duration of operation, ASA grade and wound class for the individual types of surgery are shown in Table 1.

Eight missing values for duration of operation were imputed with the surgery-specific mean. In a limited number of cases (less than 0.1 per cent), hospitals reported an operating time of less than 10 min. It is assumed that these are registration errors, but because of their minor impact they were not corrected. The median number of patients per hospital varied between the different types of surgery, from 12 for the reconstruction of abdominal vessels to 204 for hip replacement. The SSI rate also

	Duration of surgery (min)*	ASA grade (%)			Wound grade (%)		
		≤ II	> II	Unknown	1–2	3–4	Unknown
Mastectomy	69 (1–247)†	90	9	1	99	0	1
Laparoscopic cholecystectomy	50 (6–382)†	91	6	3	97	2	1
Colonic resection	105 (16–530)	63	32	5	85	14	1
Abdominal vessel reconstruction	163 (14–464)	51	43	6	97	0	3
Hip replacement	72 (3–960)†	87	11	2	100	0	0
Knee replacement	75 (3–920)†	87	12	1	100	0	0
Abdominal hysterectomy	75 (30–233)	94	4	2	99	1	0
Caesarean section	36 (1–209)†	93	1	6	99	0	1
Overall	68 (1–960)†	86	11	3	98	2	0

Table 1 Descriptive statistics for each type of surgery. *Values are median (range). †In a limited number of cases (less than 0.1 per cent) hospitals reported an operating time of less than 10 min; these are assumed to be registration errors, but because of their minor impact they have not been corrected. ASA, American Society of Anaesthesiologists.

Indicator	No. of procedures	No. of hospitals	Procedures per hospital*	SSIs per hospital (range)*	SSI rate (%)†
Mastectomy	1284	17	54 (9–268)	3 (0–7)	4.5 (0–14.3)
Laparoscopic cholecystectomy	1558	12	135 (43–231)	2 (0–9)	1.3 (0–5.1)
Colonic resection	965	16	61 (11–113)	9 (1–23)	15.1 (1.2–36.4)
Abdominal vessel reconstruction	271	11	12 (3–111)	0 (0–3)	0 (0–10)
Hip replacement	4199	21	204 (49–324)	2 (0–7)	1.4 (0–2.9)
Knee replacement	3404	21	140 (28–396)	2 (0–7)	1.1 (0–5.1)
Abdominal hysterectomy	284	9	32 (2–64)	1 (0–2)	2.1 (0–5.7)
Caesarean section	1664	11	140 (43–321)	0 (0–8)	0 (0–2.5)
Overall	13.629	34	326 (30–1007)	9.5 (1–35)	2.8 (0.2–17.6)

Table 2 Outcome for each type of surgery. *Values are median (range); †values in parentheses are ranges. SSI, surgical-site infection.

varied widely depending on the type of surgery, from 15.1 per cent for colonic resection to 0 per cent for reconstruction of abdominal vessels (*Table 2*).

OBSERVED SURGICAL-SITE INFECTION RATES

Observed SSI rates are visualized in funnel plots for all operations combined and after colonic resection (*Fig. 1*). Although there were outliers, the majority of hospitals had an estimate close to the average, within the confidence limits. The hospital estimates varied more widely for colonic resection.

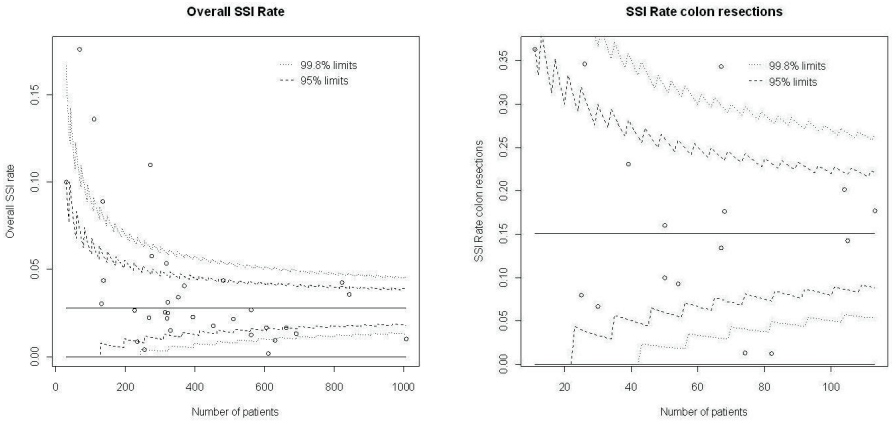


Fig. 1 Funnel plots of observed a overall surgical-site infection (SSI) rates and b SSI rate after colorectal surgery. The observed SSI rate is plotted against the total number of interventions in each hospital

FIXED-EFFECTS MODEL

The fixed-effect models, representing the observed differences between hospitals in SSI rates, showed significant differences for overall SSI rate in colonic resection and caesarean section. All remained significant when differences between hospitals were estimated using random-effect models, taking into account random variation. For example, the 95 per cent odds ratio range for colonic resection was 0.21 to 4.67 ($P < 0.001$) (*Table 3*), meaning that the odds ratio for SSI was 0.21 in a hospital at the lower end of the SSI rate distribution compared with the average; in a hospital at the higher end the odds ratio was 4.67.

	Unadjusted				Adjusted			
			95% range* of centre differences				95% range* of centre differences	
Indicator	τ^2	P†	Lower	Upper	τ^2	P†	Lower	Upper
Mastectomy	0.20	0.046	0.42	2.40	0.16	0.149	0.46	2.18
Laparoscopic cholecystectomy	0.22	0.083	0.40	2.50	0.20	0.118	0.42	2.38
Colonic resection	0.62	< 0.001	0.21	4.67	0.58	< 0.001	0.23	4.42
Abdominal vessel reconstruction	0	0.500	1.00	1.00	0	0.472	1.00	1.00
Total hip replacement	0	0.500	1.00	1.00	0	0.496	1.00	1.00
Knee replacement	0.22	0.078	0.4	2.48	0.17	0.128	0.45	2.25
Abdominal hysterectomy‡	0	0.500	1.00	1.00	0	0.481	1.00	1.00
Caesarean section§	0.81	0.032	0.17	5.82	1.10	0.024	0.13	7.59
Overall	0.66	< 0.001	0.20	4.91	0.01	< 0.001	0.42	2.33

Table 3 Between-hospital differences estimated by random-effects analysis, with or without adjustment for case mix. *This 95 per cent range represent the odds of surgical-site infection (SSI) compared with the average of a hospital on the lower end (2.5th percentile) of the distribution and the odds of a hospital on the higher end (97.5th percentile). †P value for τ^2 (variance of random effects). ‡Abdominal hysterectomy sex not in the model; §caesarean section adjusted only for age and duration of surgery.

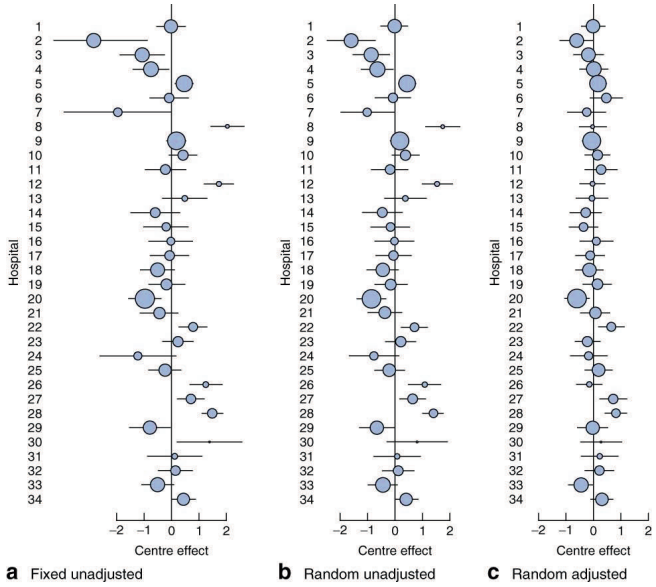


Fig. 2 Effect estimates for differences in overall surgical-site infection (SSI) rate between 34 hospitals in all surgical procedures combined: first column a fixed-effects unadjusted, second column b random-effects unadjusted and third column c random-effects adjusted models. Values are logistic regression coefficients, compared with the average outcome. In each plot, hospitals on the right side have estimated SSI rates above the average, whereas those on the left have a lower than average estimated SSI rate. Dot sizes indicate the number of patients per hospital

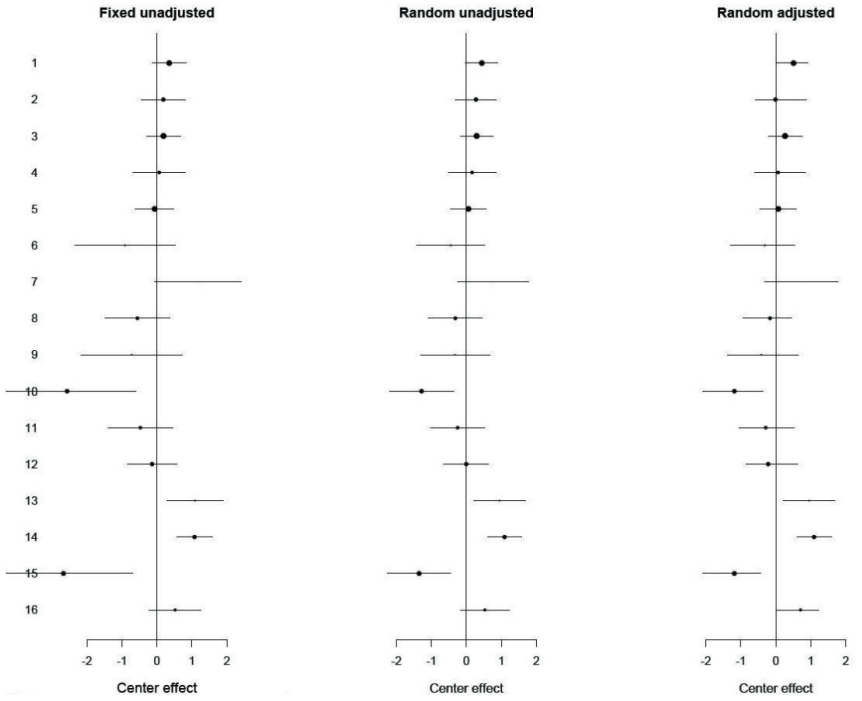


Fig. 3 Effect estimates for differences in surgical-site infection (SSI) rate for colonic resection between 16 hospitals: **a** fixed-effects unadjusted, **b** random-effects unadjusted and **c** random-effects adjusted models. Values are logistic regression coefficients, compared with the average outcome. In each plot, hospitals on the right side have estimated SSI rates above the average, whereas those on the left have a lower than average estimated SSI rate. Dot sizes indicate the number of patients per hospital

However, the differences declined between fixed- and random-effects estimates (*Figs 2a,b* and *3a,b*), indicating that part of the observed differences between hospitals was explained by random variation. This decline was most prominent in the SSI rates for colonic resection (*Fig. 3*) because of the smaller numbers per hospital.

After adjustment for case mix, the differences between hospitals declined further, especially for the overall SSI rates (*Fig. 2c*). The 95 per cent range of centre differences declined from 0.20 to 4.91 to 0.42 to 2.33 (*Table 3*), indicating that part of the differences between hospitals in SSI rate could be explained by their case mix. Within the specific procedure types case-mix difference explained less of the variation in SSI rate (*Table 3*, *Fig. 3*). Although the overall differences between hospitals in SSI rates for colonic resection and for all types of surgery remained significant, this significance was mainly attributed to a few hospitals that were notably different from the average (*Fig. 3*).

Rankability

Rankability was calculated for the surgical interventions that showed significant differences in the random-effects analysis. For caesarean section, rankability could not be calculated because the majority of hospitals reported no SSIs. For the combined SSI rates, rankability was 85 per cent without case-mix adjustment, but only 8 per cent after case-mix correction. This means that, of the observed case-mix-adjusted differences between hospitals, 8 per cent at most may have been due to the quality of prevention of SSI; 92 per cent was explained by random variation and case mix. For colonic resection, rankability before and after case-mix correction was 80 and 78 per cent respectively, meaning that a much larger part of the observed variation might have been due to quality differences.

DISCUSSION

Random variation largely explained apparent differences between hospitals in SSI rates for individual surgery types. Apparent differences between hospitals in combined SSI rates were mostly explained by case mix. The significant differences that remained after adjustment for random variation and case mix were mainly caused by a few deviating hospitals. Therefore, SSIs rate are not suitable for ranking hospitals. These findings reflect the complexity of the comparison of SSI rates between different hospitals.

Frequently, the number of procedures per hospital is low. This is often combined with low SSI rates, for instance in reconstruction of abdominal vessels or caesarean section (*Table 2*). The low SSI rates hinder the detection of differences in quality of care between hospitals. There is just not enough power. This finding is in line with other research on SSI as a performance indicator for individual operations[35–37]. The funnel plots show that around 80–90 cases per hospital might be sufficient for reliable estimation of the individual SSI rate. Rankability, however, contains a second element: the magnitude of the differences between hospitals. When differences are small but estimates very reliable, rankability will be reasonable. When differences are large, greater uncertainty can be accepted.

By combining fixed-effects logistic regression models and random-effects logistic regression models, it was possible to estimate uncertainty within the individual hospitals and the unexplained heterogeneity between hospitals. It was found that ranking hospitals, adjusted random-effect estimates, led to overinterpretation of the small and uncertain differences in SSI rates, except in colonic resection. Significant differences were found in SSI rates after colonic resection and rankability remained good after case-mix correction. It is plausible that the indication for surgery, such

as chronic bowel disease, traumatic injury or colorectal cancer, may cause residual confounding that explains part of these significant differences. Nevertheless, the finding of possible quality-of-care differences between hospitals is in line with the results of Hübner and colleagues[5].They found that the surgeon might constitute an independent risk factor for SSI after colonic surgery. Additional prospective research is needed to assess whether SSI after colonic resection can be used as a performance indicator.

There are two separate elements in the present analysis: statistical uncertainty and adjustment for case mix. Statistical uncertainty affects hospitals with small numbers most. Their estimates are too extreme (either too good or too poor) and shift towards the mean in the random-effects analysis. The estimates will, however, never change direction, such as from worse than expected to better than expected. The effect of case-mix adjustment is independent of sample size and can result in a change of direction; for example a hospital with a SSI rate below the average but also an extremely favourable case mix might go from better than expected before adjustment to worse than expected after adjustment.

Although this study is based on data from the Netherlands, the findings are generalizable to other countries. The effect of case mix is dependent on the magnitude of the differences in case mix between hospitals. These are likely to be present in any country. The role of statistical uncertainty depends on the number of patients per hospital. Although the total number of patients will be greater in larger countries, the size of the hospital is likely to be comparable. Therefore, statistical uncertainty will always be a factor to be considered, especially for smaller hospitals. This is supported by recent research showing that it is difficult to distinguish between hospitals that are performing well and those doing badly using crude SSI rates[24].

With growing pressure to report hospital performance publicly, the issue of random variation and case mix should be addressed more explicitly, and alternative methods of comparing hospitals should be used. Instead of league tables, funnel plots were used to compare outcomes between hospitals in the present study; these are useful for visualizing the influence of random variation[28–30,38,39]. Other alternatives for benchmarking include comparisons with the best-performing hospitals, a national average or a norm[17–20], or to compare rates from year to year within the same hospital.

In this study, SSI rates were adjusted for random variation and case-mix differences between hospitals. However, there may be residual confounding caused by incomplete case-mix correction and registration bias that explains the remaining differences. In fact, only a small part of observed unadjusted differences in SSI rates between hospitals was likely to be attributable to quality-of-care differences (*Fig. 4*)[33,40–42].

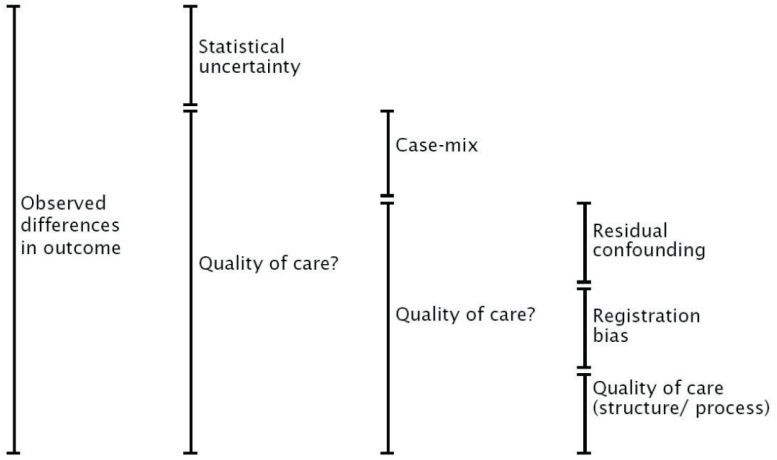


Fig. 4 Conceptual framework of between-hospital differences.

The length of the bars in this figure is arbitrary and differs for the various indicators. For example, for overall surgical-site infection (SSI) the effect of case mix is large and that of uncertainty is small, whereas for colonic resection the effect of uncertainty is large but that of case mix is limited.

With regard to case-mix correction, adjustment was made for relevant and available predictors of SSI. In addition, there are unmeasured confounders that may cause differences between hospitals in SSI rate. For hip replacement, for example, revisional surgery is associated with a significantly higher risk of SSI[43], but this was information was not available in the present data. Other examples of possible confounders that were not available for analysis include diabetes, obesity and use of immunosuppressive medication. This incomplete case-mix correction may lead to overestimation of the differences attributed to quality of care. Registration bias is probably limited in this study. The authors believe the data to be reliable and comparable because they were collected in a voluntary registry. Moreover, the hospitals contribute to the surveillance system in a standardized way and the PREZIES network has set up an extensive external audit of the data.

The study has limitations, mainly related to the database. The data include elective and emergency interventions, but there was not sufficient information for case-mix correction for urgency of surgery. Unplanned procedures are associated with a higher incidence of SSI[44], which will add to the case mix and make comparison between hospitals even more problematic. Participation in the registry is voluntary and reflected by the number of participating hospitals (34 of 94). Hospitals can contribute no data

at all, or data for some procedures only. This causes the large effect of case-mix adjustment in overall SSI rates. These differences in case mix between hospitals are not necessarily caused by differences in patient populations; they are also the result of hospitals contributing data for different types of surgery. In addition, the degree of possible case-mix adjustment is limited by variables available in the data and it is almost impossible to provide a complete case-mix correction. It could be argued that the degree of risk adjustment with the Dutch programme is restricted. In comparison, the American College of Surgeons National Surgical Quality Improvement Program has been collecting data on approximately 40 preoperative and 20 intraoperative variables[16,45]. Therefore, the present study might have overestimated the quality-of-care effects. Because of the voluntary registration, it is also possible that hospitals with a focus on quality of care are more likely participate in the PREZIES network. This seems to be reflected in the low SSI incidence, compared with international data (3 *versus* 5 per cent).

The rankability measure uses the median standard error, which is only a summary of the overall uncertainty of individual hospital estimates. Rankability as presented here is therefore a measure for a complete set of hospitals. For individual hospital comparisons, the individual standard errors might be used[33].

When comparing SSI rates for all procedures, apparent differences between hospitals were explained by case mix. For individual procedures, case mix varied less between hospitals, and differences were explained largely by random variation. Although SSI rates may be used for quality improvement within hospitals, they should not be used for ranking hospitals, especially if the number of patients per hospital is limited. Random-effects modelling and funnel plots should be used to avoid overinterpretation of apparent differences in SSI rates.

Acknowledgements

M.C.V. and H.F.L. contributed equally to this work. A.M.v.D. received an internal Erasmus MC grant for healthcare research (Mrace).

Disclosure: The authors declare no conflict of interest.

REFERENCES

1. Herwaldt LA, Cullen JJ, Scholz D, French P, Zimmerman MB, Pfaller MA *et al*. A prospective study of outcomes, healthcare resource utilization, and costs associated with postoperative nosocomial infections. *Infect Control Hosp Epidemiol* 2006; **27**: 1291–1298.
2. Sparling KW, Ryckman FC, Schoettker PJ, Byczkowski TL, Helpling A, Mandel K *et al*. Financial impact of failing to prevent surgical site infections. *Qual Manag Health Care* 2007; **16**: 219–225.
3. Leaper DJ. Risk factors for and epidemiology of surgical site infections. *Surg Infect (Larchmt)* 2010; **11**: 283–287.
4. Wilson J, Ramboer I, Suetens C; HELICS-SSI working group. Hospitals in Europe Link for Infection Control through Surveillance (HELICS). Inter-country comparison of rates of surgical site infection – opportunities and limitations. *J Hosp Infect* 2007; **65**(Suppl 2): 165–170.
5. Kobayashi, M., et al., Continuous follow-up of surgical site infections for 30 days after colorectal surgery. World J Surg, 2008. 32(6): p. 1142-6.
6. Geubbels EL, Mintjes-de Groot AJ, van den Berg JM, de Boer AS. An operating surveillance system of surgical-site infections in The Netherlands: results of the PREZIES national surveillance network. Preventie van Ziekenhuisinfecties door Surveillance. *Infect Control Hosp Epidemiol* 2000; **21**: 311–318.
7. Mannien J, van den Hof S, Muilwijk J, van den Broek PJ, van Benthem B, Wille JC. Trends in the incidence of surgical site infection in the Netherlands. *Infect Control Hosp Epidemiol* 2008; **29**: 1132–1138.
8. Mannien J, van der Zeeuw AE, Wille JC, van den Hof S. Validation of surgical site infection surveillance in the Netherlands. *Infect Control Hosp Epidemiol* 2007; **28**: 36–41.
9. Manniën J, Wille JC, Snoeren RL, van den Hof S. Impact of postdischarge surveillance on surgical site infection rates for several surgical procedures: results from the nosocomial surveillance network in The Netherlands. *Infect Control Hosp Epidemiol* 2006; **27**: 809–816.
10. Bratzler DW, Hunt DR. The surgical infection prevention and surgical care improvement projects: national initiatives to improve outcomes for patients having surgery. *Clin Infect Dis* 2006; **43**: 322–330.
11. Manniën J, van Kasteren ME, Nagelkerke NJ, Gyssens IC, Kullberg BJ, Wille JC *et al*. Effect of optimized antibiotic prophylaxis on the incidence of surgical site infection. *Infect Control Hosp Epidemiol* 2006; **27**: 1340–1346.
12. van Benthem BH, van Vliet JA. Reflections on an evaluation of the Dutch Infectious diseases Surveillance Information System. *Euro Surveill* 2008; **13**: pii: 8070.
13. Wilson J, Burman-Roy S, Leaper D. Preventing and managing surgical site infections. *Br J Hosp Med (Lond)* 2009; **70**: 184–185.
14. Bratzler DW. The Surgical Infection Prevention and Surgical Care Improvement Projects: promises and pitfalls. *Am Surg* 2006; **72**: 1010–1016.
15. Rioux C, Grandbastien B, Astagneau P. The standardized incidence ratio as a reliable tool for surgical site infection surveillance. *Infect Control Hosp Epidemiol* 2006; **27**: 817–824.
16. Itani KM. Fifteen years of the National Surgical Quality Improvement Program in review. *Am J Surg* 2009; **198**(Suppl): S9–S18.
17. Avato JL, Lai KK. Impact of postdischarge surveillance on surgical-site infection rates for coronary artery bypass procedures. *Infect Control Hosp Epidemiol* 2002; **23**: 364–367.
18. Emori TG, Culver DH, Horan TC, Jarvis WR, White JW, Olson DR *et al*. National nosocomial infections surveillance system (NNIS): description of surveillance methods. *Am J Infect Control* 1991; **19**: 19–35.

19. Horan TC, Culver DH, Gaynes RP, Jarvis WR, Edwards JR, Reid CR. Nosocomial infections in surgical patients in the United States, January 1986-June 1992. National Nosocomial Infections Surveillance (NNIS) System. *Infect Control Hosp Epidemiol* 1993; **14**: 73–80.
20. Jarvis WR. Benchmarking for prevention: the Centers for Disease Control and Prevention's National Nosocomial Infections Surveillance (NNIS) system experience. *Infection* 2003; **31**(Suppl 2): 44–48.
21. Astagneau P, L'Heriteau F. Surveillance of surgical-site infections: impact on quality of care and reporting dilemmas. *Curr Opin Infect Dis* 2010; **23**: 306–310.
22. Marshall T, Mohammed MA, Rouse A. A randomized controlled trial of league tables and control charts as aids to health service decision-making. *Int J Qual Health Care* 2004; **16**: 309–315.
23. Adab P, Rouse AM, Mohammed MA, Marshall T. Performance league tables: the NHS deserves better. *BMJ* 2002; **324**: 95–98.
24. Campbell DA Jr, Henderson WG, Englesbe MJ, Hall BL, O'Reilly M, Bratzler D *et al*. Surgical site infection prevention: the importance of operative duration and blood transfusion – results of the first American College of Surgeons–National Surgical Quality Improvement Program Best Practices Initiative. *J Am Coll Surg* 2008; **207**: 810–820.
25. Manniën J, van den Hof S, Brandt C, Behnke M, Wille JC, Gastmeier P. Comparison of the National Surgical Site Infection surveillance data between The Netherlands and Germany: PREZIES *versus* KISS. *J Hosp Infect* 2007; **66**: 224–231.
26. Muilwijk J, Walenkamp GH, Voss A, Wille JC, van den Hof S. Random effect modelling of patient-related risk factors in orthopaedic procedures: results from the Dutch nosocomial infection surveillance network 'PREZIES'. *J Hosp Infect* 2006; **62**: 319–326.
27. Lingsma HF, Roozenbeek B, Li B, Lu J, Weir J, Butcher I *et al*. Large between-center differences in outcome after moderate and severe traumatic brain injury in the international mission on prognosis and clinical trial design in traumatic brain injury (IMPACT) study. *Neurosurgery* 2011; **68**: 601–607.
28. Moloney ED, Bennett K, Silke B. Effect of an acute medical admission unit on key quality indicators assessed by funnel plots. *Postgrad Med J* 2007; **83**: 659–663.
29. Spiegelhalter D. Funnel plots for institutional comparison. *Qual Saf Health Care* 2002; **11**: 390–391.
30. Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Stat Med* 2005; **24**: 1185–1202.
31. Berard F, Gandon J. Postoperative wound infections: the influence of ultraviolet irradiation of the operating room and of various other factors. *Ann Surg* 1964; **160**(Suppl 2): 1–192.
32. Keats AS. The ASA classification of physical status – a recapitulation. *Anesthesiology* 1978; **49**: 233–236.
33. Lingsma HF, Eijkemans MJ, Steyerberg EW. Incorporating natural variation into IVF clinic league tables: The Expected Rank. *BMC Med Res Methodol* 2009; **9**: 53.
34. van Dishoeck AM, Lingsma HF, Mackenbach JP, Steyerberg EW. Random variation and rankability of hospitals using outcome indicators. *BMJ Qual Saf* 2011; **20**: 869–874.
35. Birkmeyer JD, Dimick JB. Understanding and reducing variation in surgical mortality. *Annu Rev Med* 2009; **60**: 405–415.
36. Dimick JB, Welch HG, Birkmeyer JD. Surgical mortality as an indicator of hospital quality: the problem with small sample size. *JAMA* 2004; **292**: 847–851.
37. Kao LS, Ghaferi AA, Ko CY, Dimick JB. Reliability of superficial surgical site infections as a hospital quality measure. *J Am Coll Surg* 2011; **213**: 231–235.

38. Morton A, Mengersen K, Rajmokan M, Whitby M, Playford EG, Jones M. Funnel plots and risk-adjusted count data adverse events. A limitation of indirect standardisation. *J Hosp Infect* 2011; **78**: 260–263.

39. van Dishoeck AM, Looman CW, van der Wilden-van Lier EC, Mackenbach JP, Steyerberg EW. Displaying random variation in comparing hospital performance. *BMJ Qual Saf* 2011; **20**: 651–657.

40. Lemmers O, Broeders M, Verbeek A, Heeten G, Holland R, Borm GF. League tables of breast cancer screening units: worst-case and best-case scenario ratings helped in exposing real differences between performance ratings. *J Med Screen* 2009; **16**: 67–72.

41. Lingsma HF, Dippel DW, Hoeks SE, Steyerberg EW, Franke CL, van Oostenbrugge RJ et al.; Netherlands Stroke Survey Investigators. Variation between hospitals in patient outcome after stroke is only partly explained by differences in quality of care: results from the Netherlands Stroke Survey. *J Neurol Neurosurg Psychiatry* 2008; **79**: 888–894.

42. Wilson J, Charlett A, Leong G, McDougall C, Duckworth G. Rates of surgical site infection after hip replacement as a hospital performance indicator: analysis of data from the English mandatory surveillance system. *Infect Control Hosp Epidemiol* 2008; **29**: 219–226.

43. Hubner M, Diana M, Zanetti G, Eisenring MC, Demartines N, Troillet N. Surgical site infections in colon surgery: the patient, the procedure, the hospital, and the surgeon. *Arch Surg* 2011; **146**: 1240–1245.

44. Leekha, S., et al., Should national standards for reporting surgical site infections distinguish between primary and revision orthopedic surgeries? Infect Control Hosp Epidemiol, 2010. 31(5): p. 503-8.44 Ramos-Luces O, Molina-Guillén N, Pillkahn-Díaz W, Moreno-Rodríguez J, Vieira-Rodríguez A, Gómez-León J. Surgical wound infection in general surgery. *Cir Cir* 2011; **79**: 323–329.

45. Raval MV, Cohen ME, Ingraham AM, Dimick JB, Osborne NH, Hamilton BH *et al*. Improving American College of Surgeons National Surgical Quality Improvement Program risk adjustment: incorporation of a novel procedure risk score. *J Am Coll Surg* 2010; **211**: 715–723.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Center	N	SSI % (n)	Age median years	Sex male %	Surgery types
1.	561	2,7 (15)	44	15	1,3,5,8,10
2.	612	0,2 (1)	57	17	6,7,8,10
3.	631	1,0 (6)	63	25	6,7,8,10
4.	691	1,3 (9)	69	31	6,7
5.	822	4,3 (35)	69	35	1,4,5,6,7
6.	321	2,5 (8)	31	0	10
7.	255	0,4 (1)	68	38	6,7
8.	68	17,7 (12)	70	49	4
9.	844	3,6 (30)	70	28	1,4,6,7
10.	369	4,1 (15)	72	34	4,6
11.	320	2,2 (7)	70	31	6,7
12.	110	13,6 (15)	71	46	4,5
13.	138	4,4 (6)	59	0	1
14.	329	1,5 (5)	71	23	1,4,5,6,7
15.	268	2,2 (6)	58	0	1
16.	226	2,7(6)	62	26	1,3,4,6,7,8,10
17.	314	2,6 (8)	71	32	4,6,7
18.	606	1,7 (10)	68	34	1,3,4,5,6,7
19.	395	2,3 (9)	44	11	1,3,6,7,8,10
20.	1007	1,0 (10)	61	23	1,3,4,5,6,7,8,10
21.	453	1,8(8)	60	20	1,6,7,10
22.	277	5,8 (16)	46	16	1,3,10
23.	352	3,4 (12)	56	22	1,3,4,10
24.	234	0,9 (2)	35	8	3,8,10
25.	510	2,2 (11)	64	32	6,7
26.	135	8,9 (12)	68	27	1,4,5
27.	318	5,4 (17)	70	32	4,6,7
28.	273	11,0 (30)	69	37	4,5,7
29.	562	1,3 (7)	69	32	6,7
30.	30	10,0 (3)	73	90	5
31.	132	3,0 (4)	51	26	3
32.	322	3,1 (10)	57	21	1,3,6,7,8
33.	663	1,7 (11)	67	33	1,3,4,5,6,7
34.	481	4,4 (21)	65	28	1,3,4,5,6,7,8
Total	13629	2,8% (378)	64	26%	all

Supplemental table; individual hospitals Surgery types: 1=mastectomy 3=Laparoscopic cholecystectomy 4=Colon resection 5=Reconstruction abdominal vessels 6=Hip Replacement 7=Knee replacement 8=Abdominal hysterectomy 10=Caesarean section

Appendix 2: Formulas

Fixed effect logistic regression

$$\text{Logit}(P(Y_{ij} = 1|X_{ij})) = \beta X_{ij} + \theta_i$$

with

X_{ij}:	the regression coefficients describing the effect of the covariates and the intercept,
β:	the covariates (in this case the confounders) describing the patients characteristics of patient *j* in hospital *i*, including the constant term,
$\theta_{:i}$:	the effect of hospital *i*, that is the coefficient with respect to some overall mean.

Random effect logistic regression

$$\text{Logit}(P(Y_{ij} = 1|X_{ij})) = \beta X_{ij} + \theta_i$$

with

X_{ij}:	the covariates (in this case the confounders) describing the patients characteristics of patient *j* in hospital *i*, including the constant term,
β:	the regression coefficients describing the effect of the covariates and the intercept,
θ_i:	the effect of hospital *i*, that is the coefficient with respect to some overall mean, drawn from a normal distribution with mean μ and variance

Rankability

$$\rho = \frac{\tau^2}{(\tau^2 + \text{median}(s_i^2))}$$

with

τ^2:	the variance of the random effects,
s^2_i:	the variance of the fixed effect individual hospital effect estimates.

Chapter 5

The association between quality of preventative care and hospital-acquired skin lesions in adult hospital patients; a case control study

van Dishoeck A.M., MSc RN[1,3], Looman C.W.N., PhD[2], Steyerberg E.W., Prof. PhD[1], Halfens R.J.G, PhD[4], Mackenbach J.P., Prof. PhD MD[2]

Journal of Advanced Nursing. 2015;revision submitted.

ABSTRACT

Aim. To explore the relation between the occurrence of pressure ulcers or incontinence dermatitis and the quality of the preventive care process.

Background.The prevalence of pressure ulcers with or without incontinence dermatitis is widely used as indicators for the quality of nursing care.

Design. Matched case control study

Methods. We collected information on 132 patients selected from a prevalence study (April 2010). We matched 88 controls to 44 cases, controlling for duration of hospitalization and type of nursing unit. We wrote 132 patient reports including patient factors and process criteria using chart review. Five expert teams assessed nine processes of care with guideline based review criteria. The expert teams assessed the reports blinded for outcome. The care process was assessed using a four point quality score ranging from optimal care to vital suboptimal care.

Results. In a multivariable analysis using conditional logistic regression, the pressure ulcer risk score (OR 1.3, CI 1.07-1.46, p-value 0.018) and the quality score (OR 1.87, CI 1.06 - 3.32, p-value 0.032) were independently associated with poor outcome after adjustment for type of illness, age, care needs prior to hospitalisation, stay in intensive care and the number of care problems.

Conclusion. We found that developing pressure ulcers or incontinence dermatitis was associated with the quality of the preventive care process, indicating that variation in this prevalence reflects variation in quality of care.

BACKGROUND

Pressure ulcers and incontinence dermatitis are important health care problems for hospitalized patients with reduced mobility or immobilization. A pressure ulcer (PU) is caused by uninterrupted pressure and shear on soft tissue, muscle and bone, as defined by the National Pressure Ulcer Advisory Panel (NPUAP) and the European Pressure Ulcer Advisory panel (EPUAP)(1). These ulcers decrease the quality of life for patients(2, 3) and increase morbidity and mortality(4). Incontinence associated dermatitis (IAD) is moister related skin breakdown occurring when urine or stool is left in contact with the skin(5-7).

PU patients require intensified nursing and medical care, resulting in an increased workload for healthcare workers and increased healthcare costs(8-11). PU and IAD can be avoided with adequate measures and preventive care, though not all PU's are avoidable(12-14). During the 2011 Consensus Conference in Baltimore, Maryland(15) avoidable was defined by the NPAUP as: "*An avoidable pressure ulcer can develop when the provider did not do one* or more of the following:

1. evaluate the individual's clinical condition and pressure ulcer risk factors;
2. define and implement interventions consistent with individual needs, individual goals, and recognized standards of practice;
3. monitor and evaluate the impact of the interventions;
4. or revise the interventions as appropriate".

Each of these individual concepts (evaluating, implementing prevention, monitoring, communicating) refer to specific care processes, that can be used to identify avoidable PU cases and adherence to guidelines as an important aspect of the preventive care process. Based on these factors, we drew a conceptual framework of the relation between factors relating to process quality and factors relating to the patient in the development of pressure ulcers (figure 1; conceptual framework).

In this framework, we assumed that one cannot control the health status of the patient on admission to the hospital (patient factors), nor factors relating to treatment like immobility during surgery. Nurses can plan and provide optimal care in minimising risk and preventing pressure ulcer development (process quality), including interventions to improve health status, nutrition status or mobility. Therefore, patient factors relate to the initial and evolving status of the patient.

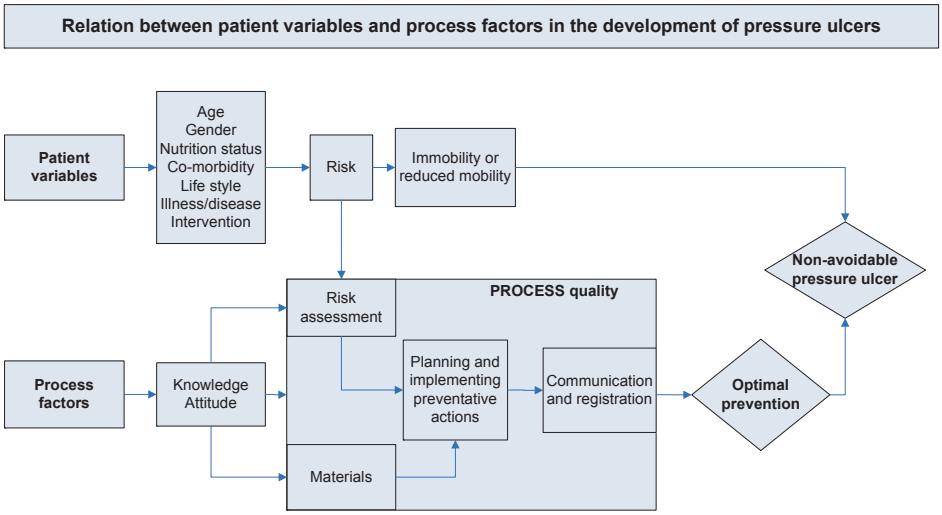


Figure 1; conceptual framework of the relation between non-modifiable variables and controllable factors in the development of pressure ulcers

PU occurrence is widely used as an indicator for the quality of care (16-18). In the Netherlands, PU prevalence, including IAD, is generally accepted and used as an indicator of the quality of nursing care, based on the assumption that PU occurrence reflects the quality of the processes of the care given to prevent PU. However, it is still unclear whether there is an association between PU/IAD prevalence and PU preventive care processes (18, 19). Brandeis et al found several risk factors associate with pressure ulcers incidence, but these could not explain the three-fold difference in the incidence rates for pressure ulcers.(20). The residual confounding in this research might be explained by unmeasured process variables. Morris et al describes an association between PU prevalence and a global quality score of PU care processes (21). Whereas, Bates-Jensen and colleagues could not relate the indicator to differences in PU care processes (22). In the latter prospective study, nurses in nursing homes observed ten preventive care processes during a 3-day data-collection period, but it is unclear if these observations interfered with normal care processes. Other studies directly applied one or a set of predefined criteria for the quality of care to PU cases (23-25), but this does not capture the complexity and time-dependence of the combined preventive care processes.

Aim

The aim of this study was to explore the relation between the occurrence of pressure ulcers or incontinence dermatitis as an outcome indicator and the quality of the preventive care processes.

METHODS

Design; We used a matched case-control design. The case-control sample was derived from the data of a cross sectional pressure ulcer prevalence measurement in April 2010 (parent study).

Setting; This study was performed in a large university hospital in the Netherlands.

Participants; The parent study, PU prevalence including IAD, is measured biannually in all adult hospitalized patients. In April 2010, data was collected in 24 units represented seven intensive care units (ICUs) or step-down units, eight medical units, and nine surgical units. Patients with PU or IAD on admission were excluded, as well as patients in day care and low risk units (psychiatry, maternity ward), in concordance with the definition of the Dutch performance indicator in 2010 (26, 27). In the case-control study patients were regarded as cases when free from PU or IAD at admission but showed a PU or IAD at the day of the data collection of the parent study.

Variables; The depended variable in this study was the occurrence of PU or IAD and we analysed the relation with the quality of the preventive care processes. Data on the preventive care processes comprised eight guideline-based criteria(1);

1. Risk assessment using risk assessment tools, clinical assessment or both
2. Patient information about PU risk and prevention (verbal or written)
3. Turning regime or repositioning for the prevention of PU
4. The use of support surfaces to prevent heel pressure ulcers
5. Alternating-pressure active support replacement mattresses
6. Skin protection from exposure to excessive moisture
7. Adequate nutrition
8. Skin assessment

Added to these eight criteria was the criterion of “non-recommended intervention”. This dichotomous variable measured any non-recommended action defined as massage, synthetic sheepskins or inserting a catheter in case of incontinence.

Although there is no reliable evidence to suggest that the use of structured, systematic pressure ulcer risk assessment tools reduces the incidence of pressure ulcers, the EPUAP/NPUAP guidelines advice a risk assessment to be carried out within 24 hour of admission in the hospital. Risk assessment is viewed as a first step in identifying patients for whom prevention should be a part of the care process.

The following variables were analysed as potential confounders; age, gender, medication, pre admission care needs, activity of daily living (ADL)(28, 29), care problems, (pain, delirium, incontinence, malnutrition), intensive care stay during admission, PU risk score and type of illness (benign/malign).

Data collection and in depth case reports

Parent study

Prevalence measurement are performed yearly since 2003 and biannual since 2008. Data was collected between 8:00 and 12:00 AM on the first Tuesday of April 2010 by one or two nurses of each clinical unit. Written instructions were provided six weeks and one week before the measurement. Each patient was examined from head to toe for signs of PU or IAD. A predefined form guided the data collection. During the April 2010 measurement, 454 patients were assessed and 44 patients with PU category 1-4 or IAD were identified. All initially reported category 1 PU were reassessed by the tissue viability nurses and only patients with true category 1 PU were considered as having PU. Category 1 PU is often excluded in research, due to high dependency on knowledge and experience of the nurses for accurate assessment(30). In this sample only ten of the fifteen reported cases of PU category 1 were marked correctly by the nurse. Because of the reassessment by the nurse consultant, we could include this category. Inclusion of IAD in light of this study is based on the research by Houwing et al finding that there is no justification for singling out moisture lesions from pressure ulcers(31), which hereafter was adopted by the Dutch Health Care Inspectorate(27). Skin protection from exposure to excessive moisture being part of proper PU prevention, the process of addressing IAD could be evaluated within the guideline criteria.

Case-control study

The quality of the preventive care processes was assessed by an expert panel using case report. An independent researcher (AMvD) collected data on patient variables and process factors between April 2010 and September 2010 using a pre-structured data collection form. The primary source of information was the patients records (paper and/or electronic) in all patients. In case of lack of information, we additionally asked the nursing staff, using closed format questions listed on the case report form on the care usually provided. Finally, we looked at on site protocols to complete the information. The data source was included in the data collection.

With this information, 132 in depth patient reports of all cases and controls were written addressing patient characteristics and medical history, current medical problem and hospital stay, aspects of care and care problems, preventive actions, and factors influencing the risk of PU development. Medical terms were explained in footnotes. The case reports were blinded for outcome by reporting only on the preventive care

until the onset of PU or the PU prevalence measurement. The time line was constructed in a way that the audit panel members could not infer the patients' outcome status (pressure ulcer or IAD present yes or no) from the time points of measurement mentioned in the case reports. For all preventive processes, the source of the information was added, thus alerting to the susceptibility of the information for possible bias.

Audit

Audit panel; We set up a 15 member expert panel including two Intensivists, one dermatologist, two scientists in the field of PU, five tissue viability nurse consultants and five nurses. The expert panel members were selected on the basis of their specific knowledge and experience with pressure ulcers and pressure ulcer prevention. Four of the 15 experts were external to the university hospital (2 scientists, 2 consultants). The expert panel member were informed about of the purpose of the study. The audit expert panel discussed the interpretation of the guidelines during a meeting prior to the start of the assessment period and agreed on the criteria determining the quality of care. At this meeting, the expert panel assessed the concept patient report and valued information on timely use of pressure releasing mattresses as vital to include in the patient reports. We split the panel into five teams of three experts, each consisting of one physician or scientist, one tissue viability nurse consultant and one nurse.

Audit procedure; We electronically sent the expert teams the case reports comprising the detailed information on the patient and on the process of care as described above. The teams each assessed 26 or 27 patient reports. There were no significant differences between the teams with regard to the number of reports for cases or controls (chi^2 2.9, p-value 0.50). The teams evaluated the patient reports on the quality of the preventive care processes. First, each team members used the nine criteria of the care process and assessed them on the severity of the shortcoming (none, minor or major) in relation to the individual patient. Based on the severity of the shortcomings and their likely relationship with an unfavourable outcome of care (i.e., occurrence of PU or IAD), the experts gave their final assessment of the care process in the form of a single quality score. We used four categories of (sub)optimal care, taken from previous audit research (32-35).

Grading of suboptimal factors for PU resulted in a quality score;

0 No suboptimal factors have been identified.
1 One or more suboptimal factors have been identified, but these are unlikely to have contributed to a failure to prevent pressure ulcer in this patient.
2 One or more suboptimal factors have been identified, and possibly have contributed to a failure to prevent pressure ulcer in this patient.

3 One or more suboptimal factors have been identified, and are likely to have contributed to a failure to prevent pressure ulcer in this patient.

In giving their final judgements, the experts took into account the specific conditions and circumstances of the patient, which sometimes justified that an element of care prescribed by the guidelines was omitted (e.g., after hip fracture regular repositioning of patients may not be feasible or even undesirable). The panel team members could reject case reports if they found the information insufficient for the assessment. This first assessment was done individually by the experts without information on the assessments of their team members. In the first evaluation, the team members used a predefined assessment form on which they could comment on the degree of the shortcoming and include the arguments underlying their final assessment on suboptimal care factors (quality score). Consensus in the first round was considered to be reached if all three team members assigned an equal quality score, which was the case in 57 patient reports. In 85 reports, there was disagreement in the first round and we sent the assessments to all three team members asking them if they would revise their quality score based on the arguments of their team members. In the second round, we considered consensus reached if the difference between the team members was not more than one point, of which the lowest score counted. The expert panel discussed six reports without consensus in the second round in a final plenary session. The physicians within the expert panel tended towards a milder judgment of care quality, loading heavier on patient factors, in relation to the nurses and the nurse consultants. In one case the discrepancy was based on different views on the mobility of the patient.

These teams each assessed a part of the patient reports, that might introduce bias if one of the teams were milder in their assessments. Therefore, each team also assessed some patient reports from the other teams for the measurement of agreement. The assessment of the patient reports of the other teams resulted in ten paired assessment of which the majority (6) was identical in the final assessment and three paired assessments differed only one point. One paired assessment differed two pint based on different views on the mobility of the patient. The agreement between the nurse and the nurse consultant was good (weighted kappa 0.74, CI 0.64-0.83) and higher than the agreement between the nurse consultant and the physician (weighted kappa 0.39, CI 0.23-0.56) or the nurse and the physician (weighted kappa 0.44, CI 0.32-0.57) in the first round of the assessment. In the second round, the agreement between the nurse and the nurse consultant was excellent (weighted kappa 0.92, CI 0.89-0.95) and again higher than the agreement between the nurse consultant and the physician (weighted kappa 0.74, CI 0.63-0.85) or the nurse and the physician (weighted kappa 0.72, CI 0.61-0.83).

Analysis

Study size; For the case-control study, we performed a power analysis using the bootstrap method. We simulated datasets with 44 cases (a number that was already known), each with two controls. The main predictor was quality of care in four categories. We assumed a phi correlation of 0.275 between quality of care and case versus control and a inter cluster correlation of 0.15 for the quality of care within the matched triple. When using these specifications, we produced 500 replica's. In each replica, we performed a binomial multilevel analysis with case as outcome, quality of care as fixed factor and triples as random factor. The result showed in 80% of the replica's a significant relation (alpha 0.05) between quality and case.

Statistical methods; Descriptive statistics; continuous variables were analysed using mean and standard deviation in case of normal distribution or median and interquartile range in case of non-parametric variables. Ordinal and nominal variables were described using numbers an percentage. The relation between PU or IAD and the quality of prevention was estimated by calculating odds ratios using conditional logistic regression. For the conditional logistic regression in SPSS, the Cox regression was used in which the paring was represented by "strada" and the outcome was PU (yes/no) using the formula; logit (p=1Ix)= B quality score + B additional variables. All cases had their event at the same time while all cases were censored represented by the grouping variable. In the multivariable analysis, a force enter method with a cut-off of 0.05 was used to include variables in the model. We tested the linearity assumptions of the ordinal variables in order to approach them as a continuous variable A two-sided p-value <0.05 was considered statistically significant. Possible interaction was expected and tested between the risk score and care needs at admission, age and gender. We calculated agreement between the different experts within the teams using a weighted Kappa analysis in Medcalc vesion 11.6.

Matching procedure; For every case, two controls were selected matching on type of nursing unit and length of stay, expecting these factors to be confounders, and because matching on (instead of statistically controlling for) confounders increases the precision of the study, given the limited number of cases(36, 37).

Ethical considerations

In this study, we used routinely collected data and anonymized the patient characteristics that could lead to recognition of an individual. This project was approved by the Erasmus MC Medical Ethics Committee.

RESULTS

Patient characteristics

We included 132 patients (44 cases and 88 controls) in our study sample, of whom 61% was male and the mean age was 60 years (Table 1). More than half of these patients (64%) lived with a partner. Thirty-one percent of the patients received some kind of help for ADL. The majority of the cases (66%) had PU category 1 or 2 and 10 cases (23%) suffered from incontinence dermatitis.

Variable	Cases, n=44	Controls n=88
Gender male % (n)	50% (22)	66% (58)
Age (years), mean (sd)	64.5 (13,0)	54,6 (15.3)
Ethnicity, Caucasian % (n)	84% (37)	75% (66)
Living with partner, % (n)	61% (27)	65% (57)
Care needs before admission (ADL) % (n)		
Independent	54% (24)	77% (68)
Needs help	32% (14)	18% (16)
Dependent	14% (6)	5% (4)
Type of illness, malignancy	43% (19)	25% (22)
Number of care problems mean (sd)	3.05 (1.6)	1.4 (1.4)
ICU admission during hospital stay, % (n)	55% (24)	35% (31)
PU risk score		
no risk	39% (17)	86% (76)
risk	45% (20)	14% (12)
high risk	16% (7)	0% (0)
PU or IAD		
PU category 1	32% (14)	
PU category 2	34% (15)	
PU category 3	9% (5)	
PU category 4	2%(1)	
IAD	23% (10)	

Table 1 Demographic variables. ADL- activities of daily life, ICU- intensive care unit, PU- pressure ulcer, IAD-incontinence associated dermatitis.

Audit procedure

The audit teams, blinded for outcome, provided 132 assessments of patients. The assessments of the nine criteria (Figure 2) showed differences between the cases and the controls for risk assessment, prevention of heel PU and the use of alternating

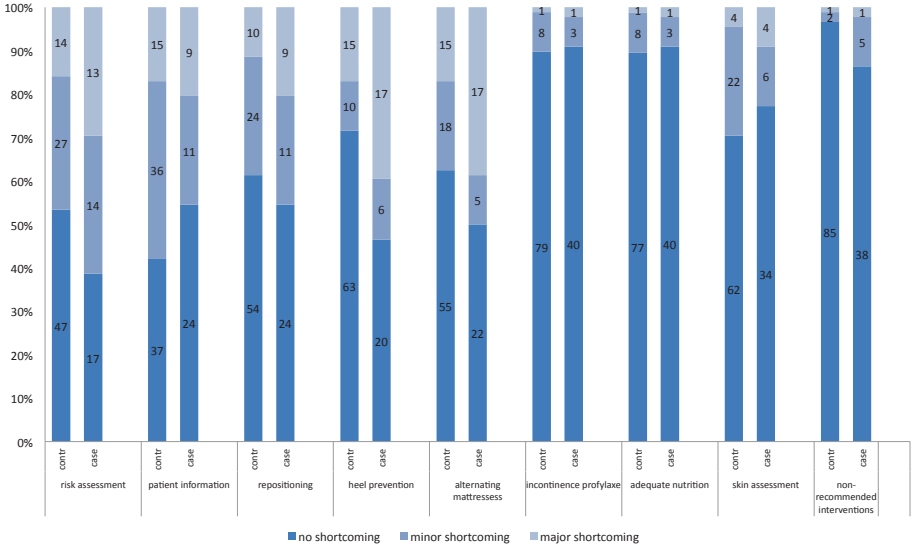


Figure 2: assessment of the individual care processes; differences between cases and controls. The raw number of patients are listed per category within the bars.

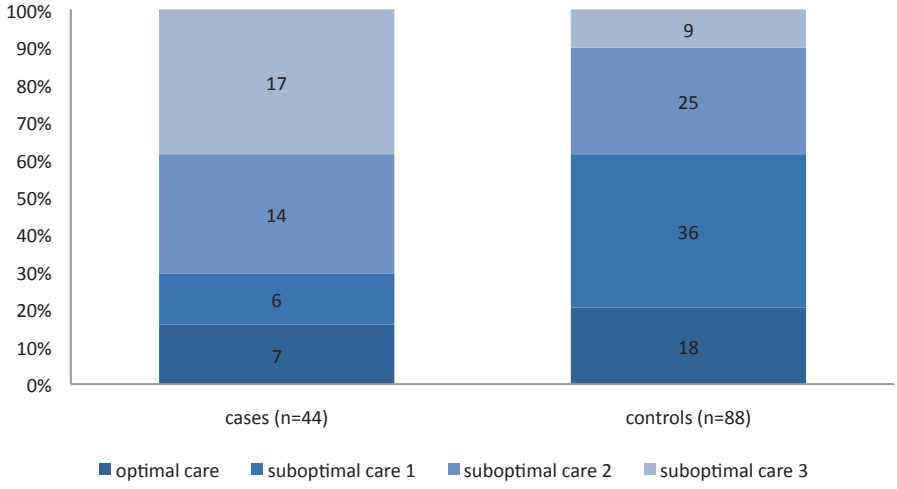


Figure 3: quality score distribution between cases and controls. The raw number of patients are listed per category within the bars.

pressure relieving mattresses.We observed differences between cases and controls in suboptimal care score. (Figure 3)

Case control analysis

In the univariate analysis of the conditional logistic regression showed that associated with PU/IAD were the factors quality score, PU risk score, age, care needs before admission, type of illness, number of care problems and IC admission during stay (table 2).

	Univariable			Multivariable		
Variable	**OR**	**CI**	**p-value**	**OR**	**CI**	**p-value**
Quality score	2.0	1.3-3.0	0.001	1.9	1.1-3.3	0.032
PU risk score	1.3	1.2-1.5	<0.001	1.3	1.0-1.4	0.018
Type of Illness, malignancy	3.3	1.2-9.3	0.024	4.3	0.9-20.1	0.067
Care needs before admission	2.6	1.2-5.6	0.014	2.3	0.7-7.1	0.153
Number of care problems	1.6	1.2-2.1	0.003	1.2	0.8-1.8	0.338
Age per decade	1.6	1.2-2.0	0.001	1.2	0.8-1.7	0.511
ICU admission during hospital stay	3.9	1.4-11.0	0.011	1.4	0.3-6.7	0.708
Gender	1.9	0.9-4.0	0.089			
Ethnicity (autochthonous/immigrant)	0.5	0.2-1.5	0.213			
Living with partner	1.2	0.5-2.5	0.694			

Table 2 Logistic regression analysis. ADL-activities of daily living, PU- pressure Ulcer, ICU, intensive care unit

In the multivariable analysis, the final quality score (p 0,032), and the PU risk score (p 0,018) were significantly associated with the occurrence of PU/IAD, indicating that variations in the prevalence are a reflection of variations in the quality of care. We examined interaction terms between predictors, and found none was sufficient relevant to extend the model beyond the main effect for each predictor.

To determine if care processes differ in importance for category 1 or category 2 and greater ulcers, the OR for the cases of pressure ulcer without category 1 PU was calculated separately. We found a comparable association between these cases and the quality score (OR 1.9 CI 1.03-3.72).

DISCUSSION

Main findings

The aim of this study was to explore the relation between the occurrence of pressure ulcers or incontinence dermatitis as an outcome indicator and the quality of the preventive care processes. We found a significant relation between the quality of the preventative care processes and the occurrence of hospital-acquired skin lesions in adult hospital patients. This finding indicates that variation in pressure ulcer prevalence reflects variation in quality of care. Furthermore, we found a significant role of patients' pressure ulcer risk as independent predictor, showing that pressure ulcer prevalence provides an indication of the quality of nursing care, not an absolute measure. Given an optimal process quality, there will always be unavoidable PU. This bias in the outcome indicator might influence the actionability of the indicator. Actionability relates to the clues for subsequent improvement of the quality of care that the performance indicators provides. Indicators should focus on those aspects of care in which interventions are possible and therefore have the potential for improving care. Actionability is then the degree to which a health care professional can influence the measure, in response to an unfavourable value of the indicator(38, 39). There is a continuous tension between the search for meaningful indicators on a national level and on their use for quality improvement within a hospital(40). Actionability of outcome indicators is described to be negatively influenced by the absence of information on actual care processes and subsequently performance improvement(41, 42). Although we found a significant relation between outcome and process, our finding might suggests that process indicators are more actionable than outcome indicators.

Both in cases as well as in controls the optimal care percentage was low. This might be explained by our strict assessment by agreeing not to use the optimal care assessment if one of the criteria was not up to the international standard, even if the consequences for the patient were negligible. But this finding is in line with recent literature. Beeckman, et al, assessed the process quality according to the international guidelines and found optimal care in only 14% of the patients (23). Gunningberg found that in less than 20% of the high risk patients a PU prevention protocol was present at the time of the survey (24). It is unclear why patients received sub-optimal care in PU prevention. Nurses' knowledge and attitude play an important role in their capacity in preventing PU. Several studies detected a lack of knowledge in nurses towards pressure ulcers prevention (43-46). Others found attitude to be significantly influencing PU prevention, underpinning the complex nature of behavioural change(47). Recent findings from the RN4Cast study showed nursing care left undone was associated with nurse-related organisational factors(48, 49). Our model, hypothesizing which factors influence the occurrence of PU might warrant inclusion. Further research is needed and important for inquiries into the major factors contributing to the sub-optimal

delivery of preventative care. The findings of this study add to the growing body of evidence indicating significant room of improvement in the daily implementation of pressure ulcer prevention care.

On exploring the inclusion of IAD to this study, it must be said that there is a dispute among professionals on how to value the influence of moisture on the development of PU. In the pathology study by Houwing et al, no justification was found for singling out moisture lesions from pressure ulcers by indicating that the distinction may even be dangerous when proper preventive measures for the development of pressure ulcers are not undertaken.(31) However, other studies do not include this skin condition, based on the causal relation between IAD and the exposure to urine and stool. IAD being a risk factor for de development of PU(50) will result in wounds that are a combination of both. Since skin protection from exposure to excessive moisture is part of proper PU prevention, including IAD does not make the relation between care processes and the outcome indicator less relevant.

Using an audit study design, we focussed on all processes preventing PU/IAD and the audit panel could assess the preventive actions in light of the complexity and time dependence of the combined care processes. In assessing the care processes, the expert teams looked only at preventable flaws. If patient factors hinder a preventive action, such as repositioning and hip replacement surgery, this was not seen as inadequate care delivery, as long as other preventive actions were undertaken. To our knowledge, we are the first to explore the relation between PU/IAD and process quality this way. Amir et al investigated the decline in PU prevalence in the Netherlands and related this decline to the use of special beds/mattresses and special cushions in wheelchairs, as well as repositioning, dehydration/malnutrition prevention and PU information (51). They used the Dutch National Prevalence Survey of Care Problems database from 2001 to 2008 and the process indicators available in that database. Other methods address the compliance of professionals to PU prevention guidelines (23, 24, 52, 53). Because we aimed at addressing all PU/IAD prevention processes, we were faced with accompanying complexity. Nurses take different preventive actions and these actions need to be applied in the specific timeframe in which the patient is at risk, sometimes for days or weeks. To measure the exact compliance to guidelines, researchers can prospectively observe the whole timeframe to assess whether prevention was sufficient and timely. On the other hand, this approach is enormously time consuming and may interfere with normal care process. Our audit design proved to be a feasible and practical approach in addressing the quality of all processes in preventing PU.

In nursing research, the case control design is sparsely used in prediction studies and explanatory studies (54, 55). Our study illustrates the potential usefulness of this design for studies of process-outcome relationships in nursing care.

Limitations

Limitations of the study were found in underreporting of preventive measures in patients records, thereby limiting available data. To expand the reported information, we added available information through nursing staff recall and adhered on site protocols or standard care of PU prevention on the different nursing units. Since the nursing staff recall was part of preparation of the in depth case reports for the retrospective audit, this method of data collection gives way to bias on recall and whether the on-site protocol was on the actually followed up. Here for, we added the source of the information to the case reports, thus providing the expert team members insight in these probable biases. The experts had the possibility to reject a case report if they thought the information was insufficient for the assessment. This possibility of rejection was not enforced in any of the case reports. Despite the systematic approach of data collection and reporting, the lack of exact information of what was done within the timeframe can be addressed in future research by accompanying the parent study with a questionnaire on the care processes.

In the parent study, true cases of category 1 PU might have been missed, because they were overlooked by nurses in the initial assessment.(56-58) This misclassification could have influenced the control group, making the difference more difficult to prove.

Due to limited resources, the data collection could not be performed by two independent researchers. Instead, an independent researcher performed the extraction process of the data collection directed by a predefined protocol and guided by a data collection form. Therefore, no inter-rater reliability was performed on the extraction process in relation to the case reports. However, a small part of the case reports were submitted to the nurse who was familiar with the specific patient. In all cases, the nurse evaluated the description of the patient's factors and process factors to be correct.

Despite these limitations, we have found a high predictive value of the quality of the care processes on the occurrence of PU.

Conclusion

We found a significant association between the quality of the preventative care processes and the occurrence of hospital-acquired skin lesions in adult hospital patients. This finding indicates that variation in pressure ulcer prevalence reflects variation in quality of care. In clinical practice, measuring both pressure ulcer occurrence as well as process-indicators will give viable information for improvement. Future research should aim at understanding the major factors contributing to the sub-optimal delivery of preventative care.

ACKNOWLEDGEMENT

We would like to mention a special thanks to our expert panel; dr. F. Heule, dr. J. van Bommel, dr. L. Schoonhoven, E. Strippe, I.M. van den Berg, W. Dekker-Verdoorn, B. den Boogert, J. van Boekel, B. Ruit, S. De Mar, J. Brugman, J.M.C. Blom and S. Theuns.

Source of Funding: Internal Erasmus MC grant for health care research (Mrace)

Participation of authors; JPM and EWS had the original idea for the study and developed the study design. AMvD gathered the data, wrote the patient reports and analysed the data with CWNL. RJGH participated and commented the assessment process as expert in the field of PU, AMvD wrote a first draft of the paper. All authors contributed to further drafts.

Competing interests: None

Impact statement

Pressure ulcer prevalence is widely used as a performance indicator for the quality of nursing care but its association with the preventive care process is still unclear and frequently debated. We aimed at exploring the total bundle of preventive care processes and its relation with the outcome indicator. We found a significant association between the quality of the preventative care processes and the occurrence of hospital-acquired skin lesions in adult hospital patients. This finding indicates that variation in pressure ulcer prevalence reflects variation in quality of care.

Key words; quality of care, performance indicator, nursing, quality improvement, process indicators, pressure ulcer

Summary statement

Why is this research or review needed?

- Although pressure ulcer prevalence is widely used as a performance indicator for the quality of nursing care, its association with the preventive care process is still unclear and frequently debated.
- To explore the relation between the occurrence of pressure ulcers or incontinence dermatitis and the quality of the preventive care process.

What are the key findings?

- We found a significant association between pressure ulcer prevalence and the quality of preventive care, indicating that the performance indicator does reflect the quality of nursing care.

- A significant part of the prevalence reflects patient factors, showing that pressure ulcer prevalence provides an indication of the quality of nursing care, not an absolute measure.

How should the findings be used to influence policy/practice/research/education?

- Pressure ulcer prevalence should be used as an indicator of the quality of nursing care.
- Measuring the different process criteria identifies quality improvement opportunities.
- The case control design is useful for nursing studies on process-outcome relationships.

REFERENCES

1. National Pressure Ulcer Advisory Panel, European Pressure Ulcer Advisory Panel. Prevention and treatment of pressure ulcer: quick reference guide. Pressure Ulcer Prevention. Washington DC: National Pressure Ulcer Advisory Panel; 2009.
2. Bale S, Dealey C, Defloor T, Hopkins A, Worboys F. The experience of living with a pressure ulcer. Nurs Times. 2007 Apr 10-16;103(15):42-3.
3. Hopkins A, Dealey C, Bale S, Defloor T, Worboys F. Patient stories of living with a pressure ulcer. J Adv Nurs. 2006 Nov;56(4):345-53.
4. Landi F, Onder G, Russo A, Bernabei R. Pressure ulcer and mortality in frail elderly people living in community. Arch Gerontol Geriatr. 2007;44 Suppl 1:217-23.
5. Chatman N, Carls C. How to manage incontinence-associated dermatitis; Prevention and prompt action are keys to managing this painful and costly complication. Wound Care Advisor 2012;May/June(1):7-10.
6. Gray M. Optimal management of incontinence-associated dermatitis in the elderly. Am J Clin Dermatol. 2010;11(3):201-10.
7. Gray M, Bliss DZ, Doughty DB, Ermer-Seltun J, Kennedy-Evans KL, Palmer MH. Incontinence-associated dermatitis: a consensus. J Wound Ostomy Continence Nurs. 2007 Jan-Feb;34(1):45-54; quiz 5-6.
8. Bennett G, Dealey C, Posnett J. The cost of pressure ulcers in the UK. Age Ageing. 2004 May;33(3):230-5.
9. Schuurman JP, Schoonhoven L, Defloor T, van Engelshoven I, van Ramshorst B, Buskens E. Economic evaluation of pressure ulcer care: a cost minimization analysis of preventive strategies. Nurs Econ. 2009 Nov-Dec;27(6):390-400, 15.
10. Padula WV, Mishra MK, Makic MB, Sullivan PW. Improving the quality of pressure ulcer care with prevention: a cost-effectiveness analysis. Med Care. 2011 Apr;49(4):385-92.
11. Reddy M, Gill SS, Rochon PA. Preventing pressure ulcers: a systematic review. Jama. 2006 Aug 23;296(8):974-84.
12. Meesterberends E, Halfens R, Lohrmann C, de Wit R. Pressure ulcer guideline development and dissemination in Europe. Journal of clinical nursing. 2010 Jun;19(11-12):1495-503.
13. de Laat EH, Pickkers P, Schoonhoven L, Verbeek AL, Feuth T, van Achterberg T. Guideline implementation results in a decrease of pressure ulcer incidence in critically ill patients. Crit Care Med. 2007 Mar;35(3):815-20.
14. NPUAP, EPUAP. Prevention and treatment of pressure ulcer: quick reference guide. Pressure Ulcer Prevention. In: European Pressure Ulcer Advisory Panel, National Pressure Ulcer Advisory Panel, editors. Washington DC2009.
15. Black JM, Edsberg LE, Baharestani MM, Langemo D, Goldberg M, McNichol L, et al. Pressure ulcers: avoidable or unavoidable? Results of the National Pressure Ulcer Advisory Panel Consensus Conference. Ostomy Wound Manage. 2011 Feb;57(2):24-37.
16. Beeckman D, Defloor T, Demarre L, Van Hecke A, Vanderwee K. Pressure ulcers: Development and psychometric evaluation of the Attitude towards Pressure ulcer Prevention instrument (APuP). International Journal of Nursing Studies. 2010;47(11):1432-41.
17. McGlynn EA, Cassel CK, Leatherman ST, DeCristofaro A, Smits HL. Establishing national goals for quality improvement. Med Care. 2003 Jan;41(1 Suppl):I16-29.

18. Simon M, Bergquist-Beringer S, Gajewski B, Dunton N. Pressure ulcer prevention as an indicator of nursing care quality: a comment on Jull and Griffiths (2010). International journal of nursing studies. 2010 Sep;47(9):1194-5.

19. Jull A, Griffiths P. Is pressure sore prevention a sensitive indicator of the quality of nursing care? A cautionary note. International journal of nursing studies. 2010 May;47(5):531-3.

20. Brandeis GH, Ooi WL, Hossain M, Morris JN, Lipsitz LA. A longitudinal study of risk factors associated with the formation of pressure ulcers in nursing homes. J Am Geriatr Soc. 1994 Apr;42(4):388-93.

21. Morris J, Moore T, Jones R, Mor V, Angelelli J, Berg K, et al. Validation of Long-Term and Post-Acute Care Quality Indicators, final report. Baltimore, Maryland 21244: Centers for Medicare and Medicaid Services2003. Report No.: CMS Contract No: 500-95-0062/T. O. #4.

22. Bates-Jensen BM, Cadogan M, Osterweil D, Levy-Storms L, Jorge J, Al-Samarrai N, et al. The minimum data set pressure ulcer indicator: does it reflect differences in care processes related to pressure ulcer prevention and treatment in nursing homes? J Am Geriatr Soc. 2003 Sep;51(9):1203-12.

23. Beeckman D, Defloor T, Schoonhoven L, Vanderwee K. Knowledge and attitudes of nurses on pressure ulcer prevention: a cross-sectional multicenter study in Belgian hospitals. Worldviews Evid Based Nurs. 2011 Sep;8(3):166-76.

24. Gunningberg L, Stotts NA, Idvall E. Hospital-acquired pressure ulcers in two Swedish County Councils: cross-sectional data as the foundation for future quality improvement. Int Wound J. 2011 Oct;8(5):465-73.

25. Vanderwee K, Clark M, Dealey C, Gunningberg L, Defloor T. Pressure ulcer prevalence in Europe: a pilot study. J Eval Clin Pract. 2007 Apr;13(2):227-35.

26. IGZ IvdG. English homepage Utrecht2009 [cited 2009 11-01-09]; Available from: http://www.igz.nl/uk/.

27. Inspectie vd, Gezondheidszorg, NVZ, NFU, OMS. Basisset prestatie-indicatoren ziekenhuizen 2010. Inspectie vd, Gezondheidszorg, editor. Utrecht2009.

28. Roper N, Logan W, Tierney A. The Elements of Nursing. Churchill Livingstone1980.

29. Roper N. LWWTAJ. The Roper-Logan-Tierney Model of Nursing: Based on Activities of Living. Edinburgh: Elsevier Health Sciences; 2000.

30. Bethell E. Controversies in classifying and assessing grade I pressure ulcers. Journal of wound care. 2003 Jan;12(1):33-6.

31. Houwing RH, Arends JW, Canninga-van Dijk MR, Koopman E, Haalboom JR. Is the distinction between superficial pressure ulcers and moisture lesions justifiable? A clinical-pathologic study. Skinmed. 2007 May-Jun;6(3):113-7.

32. Vredevoogd CB, Wolleswinkel-van den Bosch JH, Amelink-Verburg MP, Verloove-Vanhorick SP, Mackenbach JP. [Perinatal mortality assessed: results of a regional audit] Perinatale sterfte getoetst: resultaten van een regionale audit. Ned Tijdschr Geneeskd. 2001 Mar 10;145(10):482-7.

33. Richardus JH, Graafmans WC, Verloove-Vanhorick SP, Mackenbach JP. Differences in perinatal mortality and suboptimal care between 10 European regions: results of an international audit. Bjog. 2003 Feb;110(2):97-105.

34. de Koning JS, Klazinga NS, Koudstaal PJ, Prins A, Dippel DW, Heeringa J, et al. Quality of care in stroke prevention: results of an audit study among general practitioners. Prev Med. 2004 Feb;38(2):129-36.

35. de Koning JS, Klazinga N, Koudstaal PJ, Prins AD, Borsboom GJ, Mackenbach JP. Quality of stroke prevention in general practice: relationship with practice organization. Int J Qual Health Care. 2005 Feb;17(1):59-65.

36. Greenland S, Morgenstern H, Thomas DC. Considerations in determining matching criteria and stratum sizes for case-control studies. Int J Epidemiol. 1981 Dec;10(4):389-92.

37. Breslow NE. Statistics in epidemiology: the case-control study. J Am Stat Assoc. 1996 Mar;91(433):14-28.

38. Derose SF, Petitti DB. Measuring quality of care and performance from a population health care perspective. Annu Rev Public Health. 2003;24:363-84.

39. Spertus JA, Eagle KA, Krumholz HM, Mitchell KR, Normand SL, American College of Cardiology/American Heart Association Task Force on Performance M. American College of Cardiology and American Heart Association methodology for the selection and creation of performance measures for quantifying the quality of cardiovascular care. Journal of the American College of Cardiology. 2005 Apr 5;45(7):1147-56.

40. McLoughlin V, Leatherman S, Fletcher M, Owen JW. Improving performance using indicators. Recent experiences in the United States, the United Kingdom, and Australia. Int J Qual Health Care. 2001 Dec;13(6):455-62.

41. Kraus TW, Buchler MW, Herfarth C. Relationships between volume, efficiency, and quality in surgery--a delicate balance from managerial perspectives. World J Surg. 2005 Oct;29(10):1234-40.

42. Casparie AF. [Postoperative wound infections: a useful indicator of quality of care?] Postoperatieve wondinfecties: een bruikbare indicator voor de kwaliteit van zorg? Ned Tijdschr Geneeskd. 2000 Mar 4;144(10):460-2.

43. Cox J, Roche S, Gandhi N. Critical care physicians: attitudes, beliefs, and knowledge about pressure ulcers. Advances in skin & wound care. 2013 Apr;26(4):168-76.

44. Strand T, Lindgren M. Knowledge, attitudes and barriers towards prevention of pressure ulcers in intensive care units: a descriptive cross-sectional study. Intensive Crit Care Nurs. 2010 Dec;26(6):335-42.

45. El Enein NY, Zaghloul AA. Nurses' knowledge of prevention and management of pressure ulcer at a health insurance hospital in Alexandria. Int J Nurs Pract. 2011 Jun;17(3):262-8.

46. Ham WH, Schoonhoven L, Schuurmans MJ, Veugelers R, Leenen LP. Pressure ulcer education improves interrater reliability, identification, and classification skills by emergency nurses and physicians. J Emerg Nurs. 2015 Jan;41(1):43-51.

47. Demarre L, Vanderwee K, Defloor T, Verhaeghe S, Schoonhoven L, Beeckman D. Pressure ulcers: knowledge and attitude of nurses and nursing assistants in Belgian nursing homes. J Clin Nurs. 2012 May;21(9-10):1425-34.

48. Ausserhofer D, Zander B, Busse R, Schubert M, De Geest S, Rafferty AM, et al. Prevalence, patterns and predictors of nursing care left undone in European hospitals: results from the multicountry cross-sectional RN4CAST study. BMJ Qual Saf. 2014 Feb;23(2):126-35.

49. Sasso L, Bagnasco A, Zanini M, Catania G, Aleo G, Santullo A, et al. RN4CAST@IT: why is it important for Italy to take part in the RN4CAST project? J Adv Nurs. 2015 Jun 23.

50. Demarre L, Verhaeghe S, Van Hecke A, Clays E, Grypdonck M, Beeckman D. Factors predicting the development of pressure ulcers in an at-risk population who receive standardized preventive care: secondary analyses of a multicentre randomised controlled trial. J Adv Nurs. 2015 Feb;71(2):391-403.

51. Amir Y, Meijers J, Halfens R. Retrospective study of pressure ulcer prevalence in Dutch general hospitals since 2001. Journal of wound care. 2011 Jan;20(1):18, 20-5.

52. van Gaal BG, Schoonhoven L, Mintjes JA, Borm GF, Koopmans RT, van Achterberg T. The SAFE or SORRY? programme. part II: effect on preventive care. Int J Nurs Stud. 2011 Sep;48(9):1049-57.

53. Voz A, Williams C, Wilson M. Who is turning the patients? A survey study. J Wound Ostomy Continence Nurs. 2011 Jul-Aug;38(4):413-8.

54. Shyu YI, Lee HC. Predictors of nursing home placement and home nursing services utilization by elderly patients after hospital discharge in Taiwan. J Adv Nurs. 2002 May;38(4):398-406.

55. Smith CE, Curtas S, Werkowitch M, Kleinbeck SV, Howard L. Home parenteral nutrition: does affiliation with a national support and educational organization improve patient outcomes? JPEN J Parenter Enteral Nutr. 2002 May-Jun;26(3):159-63.

56. Beeckman D, Schoonhoven L, Fletcher J, Furtado K, Gunningberg L, Heyman H, et al. EPUAP classification system for pressure ulcers: European reliability study. J Adv Nurs. 2007 Dec;60(6):682-91.

57. Vanderwee K, Grypdonck MH, De Bacquer D, Defloor T. The reliability of two observation methods of nonblanchable erythema, Grade 1 pressure ulcer. Appl Nurs Res. 2006 Aug;19(3):156-62.

58. Defloor T, Schoonhoven L, Katrien V, Weststrate J, Myny D. Reliability of the European Pressure Ulcer Advisory Panel classification system. J Adv Nurs. 2006 Apr;54(2):189-98.

ADDITIONAL MATERIAL

Case report

Casusbeschrijving auditonderzoek patiënt 10613

Algemene gegevens en voorgeschiedenis

Nederlandse vrouw, 74 jaar en alleenstaande weduwe. Voor opname zelfstandig met hulp. Bekend met reumatoïde artritis, waarvoor biologicals[1] en corticosteroïden. Mw. ontwikkelde een spondylodiscitis[2] niveau thoracaal 9 t/m 12, met epidurale abcessen, rond de wervels en de dorsale rugspier, waarvoor opname in een ziekenhuis in Rotterdam.

Medische gegevens

Opname vanuit een ander ziekenhuis voor een second opinion en behandelopties op een niet-chirurgische risicoafdeling. Vanaf opname strikte bedrust en fysiotherapie voor oefeningen in bed t.a.v. de zithouding. Er werd gekozen voor een conservatief medicamenteus beleid, waarbij de MRI verbetering liet zien. Tijdens de opname ontwikkelde Mw. dyspnoeklachten t.g.v. hartfalen. Ligduur tot prevalentiemeting 13 dagen. Bij de behandeling waren reumatoloog, revalidatiearts, cardioloog, en fysiotherapeut betrokken.

Zorgaspecten

Mw. is zeer angstig. Algeheel zieke vrouw, met zeer geringe inspanningstolerantie. Bij vlagen verward en onrustig. Bedgebonden en ADL afhankelijk.

Zorgproblemen:

Decubitus	Ja, geen systematische risico-inventarisatie. Wel zorgplanning en registratie. Verpleegkundigen bespreken het risico van decubitus. Risicoplaatsen stuit, hielen, ellebogen
Ondervoeding	Thuis normaal gewicht (BMI 22). Voor opname 9 kg afgevallen bij koorts misselijkheid en braken; ondervoeding. Tijdens opname slechte eetlust en geringe intake. Na 8 dagen bengmarksonde[3] ingebracht om adequaat te kunnen voeden.
Delirium	Ja, medicatie en vrijheidsbeperkende maatregelen
Pijn	Nee
Incontinentie en diarree	Urinecatheter, incontinent van faeces en diarree

Preventieve maatregelen

Patiëntinformatie	Nee
Wisselligging	Ja, moeizaam bij onrust.
Preventie hieldecubitus	Ja
Inspectie van de huid	Ja, tijdens verzorging
Anti-decubitus matras	Ja, volgens planning bij opname, uitvoering 3 dagen later, matras voor hoogrisico patiënten.
Bescherming tegen incontinentie	Ja, barrièrespray
Voedingsinterventies	Diëtist i.c. vanaf opname. Mw. heeft voorkeur voor vloeibare producten, zoals yoghurt en kwark en wordt gestimuleerd meer te eten. De aangeboden drinkvoeding wil zij niet proberen. Mw. wil ook geen sonde. Energieverrijkt dieet. Bij duodenemsonde verzelverrijkte sondevoeding 1500 kcal.
Zinloze preventieve maatregelen	Geen

Decubitus beoordeling

1. Druk- en schuifkrachten[4]	Aanwezig door immobiliteit en onrust
2. Weefseltolerantie[5]	Beperkt door temp, voedingstoestand, corticosteroïden en hartfalen
Retrospectief decubitusrisico	Waterlowscore basisrisico 22, max. risico 24

1 Reumatoïde artritis is een auto-immuunziekte waarbij het afweersysteem zich keert tegen het lichaam door het produceren van een teveel aan ontstekingsstimulerende stoffen. Het gevolg is een chronische ontsteking van de gewrichten. Biologicals behoren tot een nieuwe generatie geneesmiddelen die de werking van lichaamseigen stoffen of cellen van het immuunsysteem beïnvloeden of nabootsen. Het effect van biologicals is het remmen van het ontstekingsproces en het voorkomen van gewrichtsschade.

2 Spondylodiscitis; ontsteking van een tussenwervelschijf met een ontsteking van de aanliggende wervels en een botvliesontsteking van de wervels

3 Bengmarksonde, deze transnasale microsonde heeft, na verwijdering van de voerdraad, een krul aan het einde die, bij voldoende maagmotiliteit, naar het jejunum migreert.

4 Oorzaak, Intensiteit; onderlaag, houding, lichaamsbouw, interventies als OK/CT e.d., duur, pijn en gevoeligheid (denk ook aan pijnmedicatie en sedatie)

5 Weefselmassa, Doorbloeding van de huid, Vochtletsel, verweking van de huid (maceratie)

Chapter **6**

Transparency: can the effect of governmental surveillance be quantified?

van Dishoeck AM, Oude Wesselink SF, Lingsma HF, Steyerberg E, Robben PB, Mackenbach JP. [Transparency: can the effect of governmental surveillance be quantified?] Transparantie: is het effect van toezicht te meten? Ned Tijdschr Geneeskd. 2013;157(16):A1676.

ABSTRACT

Purpose

The Netherlands Health Care Inspectorate (NHCI) protects and promotes health and healthcare by ensuring that care providers, care institutions and companies comply with laws and regulations. The effects of these actions on public health are not quantified. The objective of this study is exploring the feasibility of measuring the impact of surveillance actions on public health.

Method

We examined the magnitude of the health problems suicide, pressure ulcers and medication errors before and after the surveillance by the NHCI. In addition, we estimated if the reports were likely to express complete coverage or major underreporting or overreporting. Finally, we determined the effect of the surveillance initiated by the NHCI to avoid health damage.

Results

Medication errors are not clearly defined to measure the magnitude of the health problem or the effect of surveillance. Pressure ulcers and suicide can be quantified using the data from the inspectorate. Using a time series design, the trend before and after the surveillance can be made transparent. The exact impact of the effect of the surveillance in both pressure ulcer and suicide cannot yet be quantified.

Conclusion

Currently, it is not possible to quantity the impact of the surveillance on public health. In case of clearly defined health problems, it is possible to quantify the extent of the problem using the data at the Inspectorate or from external data and time series analysis. Establishing a causal relation between supervision and observed time trends, however, requires an experimental research design, including a prospective randomized or a stepped wedge design.

INTRODUCTION

Each healthcare professional in the Netherlands comes across the Dutch Healthcare Inspectorate (*Inspectie voor de Gezondheidszorg*, IGZ) sooner or later, since the Inspectorate keeps a critical eye on the activities of all care providers. The IGZ has also been subjecting its own activities to an increasing level of critical scrutiny in recent years. This is necessary because the social importance of surveillance is increasing, due among other things to the effect of the market in healthcare and the increasing need for transparency. These trends have led to a greater need for insights into the effectiveness of surveillance, both within the Inspectorate and externally.[1]

The IGZ started a surveillance evaluation programme in 2008. The aim of this programme is "to evaluate and improve the methods and instruments (see box on "Surveillance instruments") used for the purposes of surveillance, and to document the contribution surveillance makes to the safety and quality of healthcare".[2] Data on the results of interventions before and after surveillance can be used to measure the impact of the surveillance. Thanks to the improvement of the data collection methods used by the IGZ, such as redesign of the reports registration system or the performance indicator database, the IGZ has the part of data it needs for evaluation purposes in-house.

Surveillance instruments

Incident surveillance (IS) deals with the reporting of incidents and calamities in practice. The relevant data sources are individual incident reports and the Incident Surveillance reporting system.

Themed surveillance (TS) deals with high-risk aspects or areas within the healthcare field. The relevant data come from reports and databases from a selection of the institutions involved.

Risk-indicator surveillance (RS) is periodic surveillance of the risks and quality of the care provided in healthcare institutions on the basis of performance indicators (PI). The data in question are derived from reports and databases from all healthcare institutions covered by the surveillance.

The objective of the present study was to investigate the possibility of using the Inspectorate's own data as a basis for surveillance impact measurements. Two questions may be distinguished in this connection:

1. Do IGZ data sources contain enough information to permit reliable estimation of healthcare outcomes, or are these sources characterised by over- or underreporting compared with external data sources?
2. Is the impact of IGZ interventions measurable?

Three different health problems subject to IGZ surveillance were selected for the purposes of this feasibility study; these represent three different areas of healthcare, and three health problems on which the Inspectorate may be expected to have adequate

data, namely suicide in mental healthcare, pressure ulcers in hospitals and medication errors in nursing homes and care homes.

METHOD

To investigate the first question mentioned above, we determined whether it was possible to measure the extent of the three health problems in question in the Netherlands with the aid of IGZ data. We estimated the extent of possible over- or underreporting by comparison with external data. We looked for information on the selected health problems in external data sources. Sources were considered to be comparable when there was a clear resemblance between them in population and measuring method used, or when there was an explicable and quantifiable difference between them. We compared IGZ data with independent data both before and after an intervention (surveillance).

As regards the second question, we investigated whether it was possible to measure the impact of surveillance by the IGZ and we also analysed the impact of the intervention (that is, the surveillance). Since we were interested in effects that had taken place in the past, we used an interrupted time series design or time series analysis (see box on "Research methods") for this purpose. This approach allowed us to examine trends before and after the introduction of certain IGZ activities. The difference in the extent of the health problem before and after the intervention reflects the possible impact of surveillance.[3] In order to gain an insight into the extent to which this effect was actually due to the surveillance, we searched external data sources and the literature for indications that other factors could have influenced the health outcomes.

The IGZ makes a distinction between different forms of surveillance, and uses different data sources for each type of surveillance (see Table 1). We used the following IGZ data sources for the present study, listed below by health problem involved.

- Suicide: the IGZ reporting system used for Incident surveillance. Mental healthcare institutions are obliged to report all patient suicides. All these suicide reports are stored in this system.

- Pressure ulcers: risk indicator database. Annual records of the prevalence of pressure ulcers are stored in the basic data set "hospital performance indicators", for the purposes of risk-indicator surveillance (see box on 'Surveillance instruments').[4]

- Medication errors: risk indicator database. The number of medication errors forms part of the basic data set "nursing- and care-home performance indicators", which is also used for the purposes of risk-indicator surveillance.

The selection of these health problems was based on estimates of the amount of information available, and the number of years the IGZ has been collecting data on them.

Type of surveillance	Instrument	Data sources
Incident surveillance (IS) (reports)	Compulsory and voluntary reports	Reporting system
Themed surveillance (TS)	Structured and/or unstructured questionnaires	Data from healthcare institutions, stored in databases and reports.
Risk-indicator surveillance (RS)	Performance indicators (PI)	PI databases

Table 1 Surveillance by type, instrument and data source

RESULTS

Suicide

The IGZ has been receiving about 600 reports each year of suicides occurring in Dutch mental healthcare institutions since 1984. Statistics Netherlands (known as Centraal Bureau voor de Statistiek, CBS, in Dutch) keeps an annual register of all suicides occurring in the Netherlands, including those outside mental healthcare institutions. The Statistics Netherlands data show an annual total of about 1550 suicides in the Netherlands. This number is larger than the number of suicides reported to the IGZ (Fig. 1a), since the Inspectorate only receives reports of suicides occurring in a mental healthcare institution.[5] Experts indicate that mental healthcare institutions comply well with the obligation to report suicides.[6] It follows that there is no reason to believe that the data on suicides in mental healthcare institutions collected by the IGZ are underreported.

Mental healthcare institutions are obliged not only to report all cases of suicide to the IGZ but also to give full details of each case together with an analysis of the incident and suggestions for avoiding such incidents in the future that could be included in a suicide prevention policy. The impact of the obligation to report suicides could not be measured on the basis of the IGZ data alone since they only date back to 1984, when the obligation to report suicides was introduced. Statistics Netherlands does have data on suicides from before 1984, however.[7] This shows that the annual number of suicides in the Netherlands had been rising since the 1950s, flattened off around the

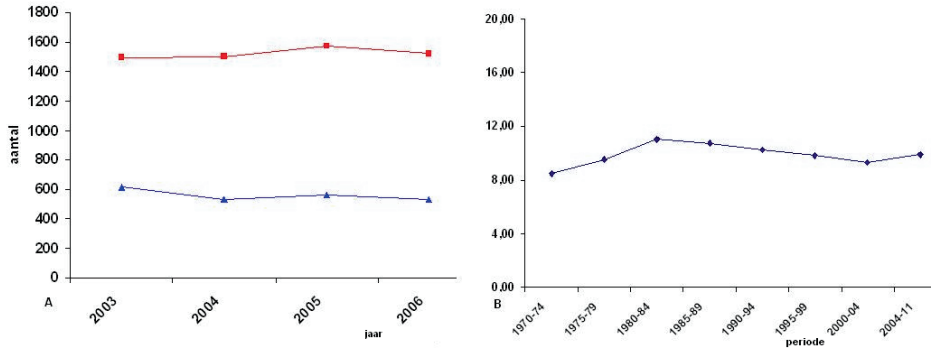


Figure 1 (a) Annual number of suicides in the Netherlands, according to the data of Statistics Netherlands (—■—) and the IGZ (—▲—); only suicides occurring in mental healthcare institutions are reported to the IGZ. (b) Number of suicides per 100,000 head of population (—◆—) in the Netherlands in the period 1970-2011, according to the data of Statistics Netherlands.

mid-1980s and has remained fairly stable since then (see Fig. 1b). A similar trend is also observed in other European countries.[8]

Pressure ulcers

The prevalence of pressure ulcers in Dutch hospitals was 9.8% in 2003, according to the first IGZ report on this indicator.[9] This figure had fallen to 3.8% by 2007.[10] The University of Maastricht has also been collecting pressure ulcer data since 1998, through the programme National Prevalence Measurements on Healthcare Problems (*Landelijke Prevalentiemeting Zorgproblemen*, LPZ). Hospitals throughout the Netherlands participate in this programme on a voluntary basis.[11]

The national average prevalence of nosocomial pressure ulcers in the Netherlands was about 7.5% up to 2004, with a slight tendency to fall that was enhanced after introduction of surveillance by the IGZ. The prevalence of pressure ulcers was over-reported in the IGZ data for 2003, because some hospitals incorrectly included stage 1 pressure ulcers in their reports.[9] The LPZ databases did not include data on these early-stage pressure ulcers. IGZ and LPZ data are comparable after 2003.(Fig. 2).

The prevalence of pressure ulcers has fallen (3.7% in 2008) since this condition has been included in the IGZ basic set of hospital performance indicators in 2003.[12,13] However, this intervention by the IGZ was not the only measure that could have reduced the prevalence of pressure ulcers. Other nationwide initiatives that were introduced, such as the "Get better quicker" (*Sneller Beter*) project, also led to an improvement in the situation.[14, 15]

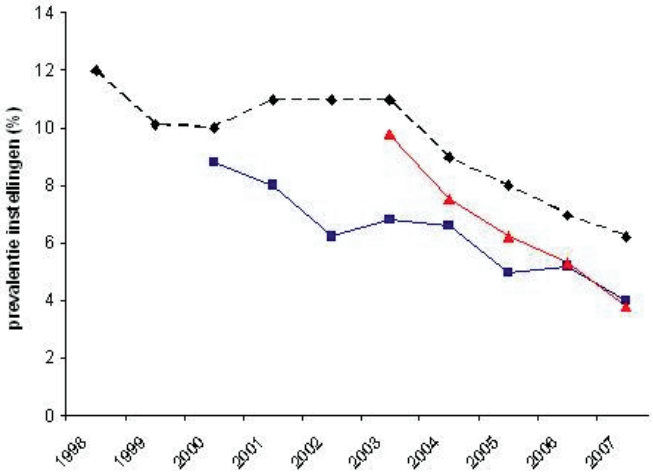


Figure 2 Comparison of data on the prevalence of pressure ulcers in the Netherlands from two data sources: the National Prevalence Measurements on Healthcare Problems programme (Landelijke Prevalentiemeting Zorgproblemen, LPZ), all healthcare institutions (- — ◆ — — -) and only hospitals (——▲——); and the IGZ hospital performance indicators data set (——■——).

Medication errors

"Medication errors" is the term covering all mistakes healthcare professionals can make in connection with medication.[16] They can occur at all stages of the medication process, from prescription through preparation and delivery to administration.[17] There is no uniform definition of medication errors, and the quality of the data collected on this point is not good enough to permit good interpretation.

The most important source of information on medication errors is the "Hospital admissions related to medication" (HARM) study.[18] This study estimates the proportion of medication-related hospital admissions in the Netherlands to be 2.4% of all hospital admissions, or 5.6% of all acute admissions; it also found that 7% of all medication-related hospital admissions had a fatal outcome, and that 28% of all patients admitted because of avoidable medication errors came from nursing or care homes.[18] This one-off study was performed in 2006, and no time trend is available. Studies from other countries, in particular America, show a rise in the number of medication errors due to an increase in the complexity of medication and the ways it can be administered.[19-20]

DISCUSSION

The aim of this study was to investigate whether it is possible to measure the impact of surveillance with the aid of the IGZ's own data. In order to answer the first question we posed – concerning the reliability of the Inspectorate's own data – we compared IGZ data with external data. This comparison showed that the IGZ data on suicide and pressure ulcers are reliable. Medication errors, on the other hand, are not defined clearly enough to permit insights to be gained on the scope of this health problem. As a result, it was not possible to compare IGZ data on this point with that from external sources.

The second question we asked, about the measurability of the impact of surveillance on these health problems, was much more difficult to answer for a number of reasons. It was impossible to determine whether the observed trends in the health problems investigated were due to IGZ intervention. It is conceivable, for example, that the obligation on mental healthcare institutions to report to the IGZ suicides among their patients may have led to a flattening off of the trend in the number of suicides. A similar flattening off was however found in other European countries,[8] and since none of the mental healthcare institutions in these countries were under any obligation to report suicides to the Dutch Healthcare Inspectorate it seems likely that the observed trend was due to some other factor.

Surveillance by the IGZ was not the only intervention that might have had an impact on the prevalence of pressure ulcers. Other initiatives aimed at improving the quality of healthcare, such as the "Get better quicker" (*Sneller Beter*) project, also led to an improvement in the situation. While the IGZ did highlight the problem of pressure ulcers,[13] it is impossible to draw any conclusions about the actual cause of the observed drop in the frequency of this problem.

Measuring the impact of surveillance

As Table 2 shows, quantification of the impact of surveillance by the IGZ is not directly possible with the aid of existing data sources. The Inspectorate does regularly highlight health problems, which causes other actors in the healthcare field to deploy various initiatives to deal with them. Greater insight into the influence of external factors on the health problems in question is needed before a causal relationship can be established for the observed trends.

The complexity of this issue is illustrated by the schematic "impact chain" shown in Fig. 3. The people and resources deployed by the IGZ for the purposes of surveillance form the input to this chain. The IGZ uses various methods to perform this surveillance (see Table 1). The "products" shown as the next link in the chain are the results of the

Investigative step	Suicide	Pressure ulcers	Medication errors
1a Survey of available IGZ data and reliability	– TS – RS (PI) + Reports	– TS + RS (PI) – Reports	– TS – RS – Reports
1b Survey of external data and comparability	+ CBS data known	+ LPZ pressure ulcer prevalence data known	– lack of uniform definitions
2a Study of trends around intervention	Internal and external	Internal and external	None
2b Estimation of impact of surveillance	An impression can be gained, but impact is not quantifiable	An impression can be gained, but impact is not quantifiable	Not possible

Table 2; Summary of analysis of selected health problems and feasibility of quantification of impact of surveillance TS= thematic surveillance, RS= risk-indicator surveillance, Reports= registration of reports CBS= Statistics Netherlands (Centraal bureau voor de Statistiek), LPZ= National Prevalence Measurements on Healthcare Problems programme (landelijke prevalentiemeting zorgproblemen)

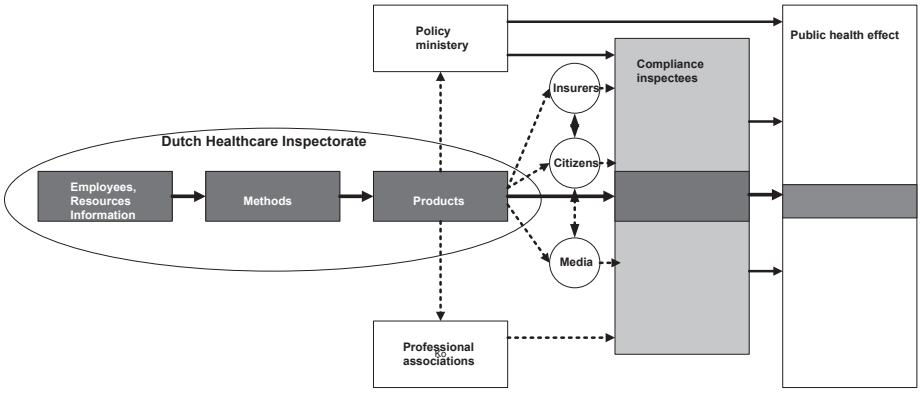


Figure 3 Schematic representation of the "impact chain", the relationship between IGZ activities and changes in public health in the Netherlands. The input to this chain is formed by the people and resources deployed by the IGZ for the purposes of surveillance. The IGZ uses various methods and activities to perform this surveillance. The "products" shown as the next link in the chain are the results of these activities. These products have an impact on healthcare insurers, members of the public, the media and the policy of the authorities and professional associations. Government policy also has an impact on healthcare providers and public health, but indirect effects due to activities of professional associations, healthcare insurers, members of the public and the media also play a role. All these factors influence the compliance of healthcare providers and healthcare organizations with the duties they are supposed to perform, and thus make a contribution to public health.

Inspectorate's activities, for example the number of visits performed by inspectors in a given year. These products have an impact on healthcare insurers, members of the public, the media and the policy of the authorities and professional associations.

All these factors influence the compliance of healthcare providers and healthcare organisations with the duties they are supposed to perform, and thus make a contribution to public health. Surveillance not only makes a limited direct contribution to the quality of public health, but also affects health indirectly in a number of ways.

Comparison with the literature

This feasibility study is unique in its kind. No previous attempts have been made to quantify the impact of government surveillance. Previous studies have however shown that it is possible to quantify the effect on compliance behaviour. For example, the sale of tobacco to minors can be reduced by providing information to retailers found guilty of illegal sales or by fining them.[21] These authors found that fining the retailers in question led to a drop of 58 percentage points in illegal sales to minors (from 76% to 18%). It is not known, however, to what extent this intervention actually reduced smoking among minors.

The Health Council of the Netherlands (*Gezondheidsraad*) suggested in its report on the development of evidence-based surveillance that the IGZ should make its work more evidence-based.[1] While the Inspectorate investigates a wide range of different topics, there is no clear line in the development of evidence-based knowledge. Studies are currently under way on IGZ activities, for example on the reliability and validity of the surveillance instruments used by the Inspectorate.[22]

Limitations of this study

The retrospective nature of the present study limits the conclusions that may be drawn from it. Data collected at different moments in time need to be uniform. Changes in the definition of an indicator make comparison between different times unreliable. For example, the IGZ added "lesion due to incontinence" to the indicators for stage 2-4 pressure ulcers in 2008. As a result, data collected after 2008 can no longer be compared with those from an earlier date. It was found that medication errors were not defined clearly enough to permit estimation of the scope of the problem on the basis of IGZ data. The situation was further complicated by the fact that external sources made use of different definitions of medication errors.

We use an interrupted time series design in the present study. This allows trends around the time of an intervention to be analysed. In principle, the IGZ can apply this time series approach to its own data. If however the data set used by the IGZ forms

part of the intervention under investigation, external data on the extent of the health problem before the IGZ intervention are also required.

Implications for further research and policy

An experimental approach in study design is an alternative to the use of time series analysis. Randomised controlled trials are the gold standard for the evaluation of interventions, since it can be assumed in such cases that any external effects are the same for both groups so that any health differences found can be ascribed to the intervention. The Health Council of the Netherlands has advised the IGZ to use a prospective randomised trial for the evaluation of surveillance. However, a randomised trial approach can have an undesirable effect on surveillance strategy since it can require one institution to be under surveillance while another is not. Phased introduction of surveillance, for example on a regional basis, might provide a solution to this problem. The "stepped wedge" approach (see box) is one form of this design that has been enjoying increasing popularity in recent years.[23]

Research methods

Interrupted time series (ITS) design

ITS is a quasi-experimental investigative design in which a group of participants are repeatedly tested both before and after a manipulation or a natural event. The repeated measurements make it possible to observe a trend in the results.

Stepped wedge design

Stepped wedge design is an experimental design in which the intervention in question is gradually extended over for example a number of regions or GP practices. It involves sequential roll-out of the intervention to participants (individuals or clusters) over a number of time periods. By the end of the study, all participants will have received the intervention, though the order in which they receive it is determined at random. They act as controls while waiting to receive the intervention.

Recently, a follow-up study has been performed, investigating the use of a prospective randomised trial to quantify the health effect of the surveillance of multidisciplinary care teams treating patients with diabetes mellitus, and a retrospective study of the impact of surveillance on support by midwives who are acting as primary care providers in helping pregnant women to stop smoking.[24]

Conclusions and recommendations

When health problems are clearly defined, it is possible to quantify their scope with the aid of data available to the IGZ and to observe trends with the aid of time series analysis. However, an experimental study design is required to establish a causal relationship between surveillance and the observed trends.

In general, only clearly defined health problems permit measurement of the impact of their surveillance. In addition, data must be available from the literature on the scope of the problem, the health effect of surveillance and external effects.

The data collected by the IGZ must be clearly defined. Possible questions that might be asked about the data must be taken into account during its collection. The data collection and the interventions must be designed to provide an effective basis for studies.

A prospective randomised trial or a stepped wedge design is one alternative approach that should be taken into consideration for measurement of the impact of surveillance activities.

Lessons learnt from this study

- The Dutch Healthcare Inspectorate (IGZ) collects information about various activities and their health implications in order to determine whether they comply with statutory requirements.
- The IGZ can perform interventions on the basis of its findings.
- The result of these interventions can be determined at an individual level, but it is not possible to acquire insights into the impact of these activities on public health.
- Comparison of IGZ data with external data makes it possible to gain an impression of the scope of health problems and improvements that may be achieved.
- It is difficult to quantify the exact contribution to these improvements without extensive information on external effects and autonomous developments.

REFERENCES

1 Mackenbach J, et al. Op weg naar evidence-based toezicht. Het onderzoek naar effecten van toezicht door de Inspectie voor de Gezondheidszorg. Den Haag: Gezondheidsraad; 2011. p. 102.

2 Van der Wal G. Meerjarenbeleidsplan 2012-2015. Voor gerechtvaardig vertrouwen in verantwoorde zorg. Utrecht: Inspectie voor de Gezondheidszorg; 2011. p. 1-33.

3 Armitage P, Berry G, Matthews J. Statistical methods in medical research. 4e dr. Malden, MA: Blackwell Publishing; 2005. p. 430-54.

4 Prestatie-indicatoren ziekenhuizen. Basisset 2008. Den Haag: Inspectie voor de Gezondheidszorg; 2007.

5 Gezondheid and zorg in cijfers 2008. Den Haag/Heerlen: Centraal Bureau voor de Statistiek; 2009.

6 Huisman A, Robben PB, Kerkhof AJ. An examination of the Dutch Health Care Inspectorate's supervision system for suicides of mental healthcare users. Psychiatr Serv. 2009;60:80-5 Medline. doi:10.1176/appi.ps.60.1.80

7 CBS. Bron: CBS, http://statline.cbs.nl/statweb/, geraadpleegd op 27 februari 2013.

8 Bool M, et al. Verminderen van suicidaliteit. Beleidsadvies. Utrecht: Trimbos instituut; 2007.

9 Het resultaat telt 2003. Utrecht: Inspectie voor de Gezondheidszorg; 2005.

10 Het resultaat telt 2006. Utrecht: Inspectie voor de Gezondheidszorg; 2007.

11 Bours GJ, Halfens RJ en Wansink SW. Landelijk Prevalentie Onderzoek Decubitus: resultaten zesde jaarlijkse meting 2003. Maastricht: Universiteit Maastricht; 2003.

12 Decubitus doorgelicht: richtlijn onvoldoende in de praktijk toegepast. Den Haag: Inspectie voor de Gezondheidszorg; 2004.

13 Van Woensel M, Van der Valk PHM, Te Velde L. Effecten van de ziekenhuisranglijsten. ZM Magazine. 2007;23 (9):2-.

14 Sneller Beter. Tien geleerde lessen uit de praktijk. Rotterdam/Utrecht/Den Haag: iBMG/CBO/OMS/ZonMw; 2008.

15 Oosterhof F. Het decubitusproject in het Jeroen Bosch ziekenhuis. Best Practices Zorg. 2008;4:18-21.

16 Dovey SM, et al. A preliminary taxonomy of medical errors in family practice. Qual Saf Healthcare. 2002;11:233-8 Medline. doi:10.1136/qhc.11.3.233

17 Barker KN, et al. Medication errors observed in 36 healthcare facilities. Arch Intern Med. 2002;162:1897-903 Medline. doi:10.1001/archinte.162.16.1897

18 Van den Bemt PM, Egberts TC. Hospital admissions related to medication (HARM) [eindrapport]. Utrecht: Nederlandse Vereniging van Ziekenhuisapothekers/Orde van Medisch Specialisten; 2006.

19 Handler SM, et al. Epidemiology of medication-related adverse events in nursing homes. Am J Geriatr Pharmacother. 2006;4:264-72 Medline. doi:10.1016/j.amjopharm.2006.09.011

20 Phillips DP, Bredder CC. Morbidity and mortality from medical errors: an increasingly serious public health problem. Annu Rev Public Health. 2002;23:135-50 Medline. doi:10.1146/annurev.publhealth.23.100201.133505

21 Gemson DH, et al. Laying down the law: reducing illegal tobacco sales to minors in central Harlem. Am J Public Health. 1998;88:936-9 Medline. doi:10.2105/AJPH.88.6.936

22 Tuijn SM, et al. Evaluating instruments for regulation of healthcare in the Netherlands. J Eval Clin Pract. 2011;17:411-9 Medline. doi:10.1111/j.1365-2753.2010.01431.x

23 Dekkers OM. Het 'stepped wedge'-design. Ned Tijdschr Geneeskd. 2012;156:A4069 Medline.

24 Oude Wesselink, S.F. Towards evidence-based government supervision in health care. Thesis Erasmus University Rotterdam, 2015. ISBN 978-94-62330-39-9. www.oudewesselink.net/proefschrift

Chapter 7

Measuring quality improvement in acute ischemic stroke care; interrupted time series analysis of door-to-needle time

van Dishoeck AM, Dippel DW, Dirks M, Looman CM, Mackenbach JP, Steyerberg E. Measuring quality improvement in acute ischemic stroke care; interrupted time series analysis of door-to-needle time. Cerebrovascular Diseases Extra. 2014(4):149-55

ABSTRACT

Background; Timely thrombolysis is a vital aspect of acute stroke treatment, and is reflected in the widely used performance indicator “door-to-needle time” (DNT). We aimed to study quality improvement from the first implementation of thrombolysis in stroke patients in a university hospital in the Netherlands. We further aimed to identify specific interventions that affected the door-to-needle time.

Methods; We included all consecutive patients admitted with acute ischemic stroke in a large university hospital in the Netherlands between January 2006 and December 2012. We used an interrupted time series design to study explanations for a trend in time between emergency room entry and treatment with thrombolytic therapy, analyzed by means of segmented regression.

Results; Between 2006 and 2012, 1703 patients were admitted with an ischemic stroke, of whom 262 received thrombolytic therapy. The percentage of patient treated with thrombolysis increased from 5 to 22%. Door-to-needle time decreased significantly (1.0% per month, CI 0.7-1.4%). In 2006, the median door-to-needle time was 75 minutes and none of the patients were treated within one hour. In 2012, these numbers had improved to 45 minutes and 81% treated within one hour. We could not find a significant association between any specific intervention and the door-to-needle time.

Conclusion; The door-to-needle time steadily improved from the first implementation of thrombolysis, Specific explanations for this improvement require further study, and may relate to the combined impact of a series of structural and logistic interventions.

BACKGROUND

In patients with acute ischemic stroke, early treatment with recombinant tissue plasminogen activator (rtPA) improves functional outcome by effectively reducing disability and dependency.(1, 2) Recent guidelines for the treatment of ischemic stroke recommend that the time from arrival at the hospital to the initiation of the thrombolytic treatment should be 60 minutes or less.(3) The quality of the in-hospital care pathway is often measured by means of the time from entrance in the emergency department (ED), until the patient receives intravenous rtPA: the door-to-needle time (DNT). Intra-organizational barriers to timely thrombolysis relate to the availability of the neurologist, blood drawing and measurements, computed tomographic (CT) imaging and skilled nursing staff.(4, 5) In the Netherlands, the percentage of patients receiving thrombolysis within one hour is an obligatory indicator of hospital performance for external accountability.

In a large university hospital in the Netherlands, stroke care was guided by a hospital wide protocol since 2001. The neurology department implemented several quality initiatives to improve the care for acute stroke patients, especially focusing on the percentage of patients receiving thrombolysis.(6, 7) Improving door-to-needle time started with a yearly training of residents and nursing staff since 2005, including "dummy runs". Pocket flow-charts with protocol summaries were first handed out in 2006 and were updated regularly. In July 2007, the ED initiated the use of Manchester triage system (MTS) protocol. The MTS is a sensitive tool for marking those who need critical care on arrival in the ED. Stroke patients obtain the highest emergency code red.(8) Since October 2007, treatment was started in the CT-room, and DNT was reported for every patient at the morning report. Individual feedback was given to all doctors who exceeded the 1 hour time threshold. In October 2009 a CT scanner was placed in the ED and treatment started immediately after non-contrast computed CT was done, but before CT angiography. In November 2010, a pre-notification single call activation system was put in place, alarming the neurology resident, radiologist, radiology laboratory personal and the emergency department nurse. Since May 2011, a second neurology resident was on duty in the weekends to ensure the availability of a doctor at any time.

We aimed to study quality improvement from the first implementation of thrombolysis in this university hospital. We further aimed to identify specific interventions that affected the door-to-needle time.

METHODS

We included all acute ischemic stroke patients admitted to a large university hospital in the Netherlands between 2006 and 2012. We focused on those treated with thrombolytic therapy on admission. We used a retrospective interrupted time series design to evaluate longitudinal effects.(9) Segmented regression analysis of the interrupted time series data allowed us to assess how much an intervention changed the DNT. The time series experiment is a reasonable alternative when the condition of a true experiment cannot be met.(9)

Data was collected routinely for research purposes and internal quality measurement (the "Erasmus Stroke Study") and for the reporting of the performance indicator "timely thrombolysis" in ischemic stroke patients. All patients with acute stroke admitted to the neurology department were entered into the registry. Completeness was cross-checked with hospital administrative systems. Data was entered and checked by medical researchers. The data collection did not change over time. The data was anonymized for analysis and could not be related to individual patients. The Erasmus Stroke Study has been approved for use in scientific medical studies by the institutional review board of Erasmus MC. We selected quality interventions that had a fixed starting point in time to include in the model. Selected interventions (i) were the start of the educational program (i1), MST-protocol (i2), CT-scanner at ED (i3), pre-notification system (i4) and second neurology resident (i5).

We report descriptive statistics using percentages, mean and standard deviation or median and interquartile range. We estimated the trend in DNT from the start of the measurement in 2006 and tested for changes in dependent variable pre and post intervention with a segmented regression analysis. We considered 2 models. The first model was: $\log(dtn)=\alpha + \beta_T T$. The second model was: $\log(dtn)=\alpha + \beta_T T + \sum_{1\text{-}5} [\beta_i I_i] + \sum_{1\text{-}5} [\beta_{i*T} I_T * T]$, where *T* (time) represents the time form the start of the measurement period (continuous variable, months starting at 1), and β_T expresses the overall trend before the interventions. *I* (intervention) represents the difference in pre and post intervention *i*, coded 0 prior to the intervention, 1 post intervention, β_i expresses the drop in DNT immediately after an intervention, and β_{i*T} expresses the change in trend over time. Both models were fitted with and without inclusion of potential confounders (age and sex). The estimate $(e^{\beta}-1)*100$ represents the percentage change in DNT. The confidence interval was calculated as $100*(e^{\beta}\pm(1.96*se)-1)$, where se is the standard error for the β parameter considered. We additionally performed a logistic regression analysis to estimate change over time in the percentage of patients receiving thrombolysis within one hour. Statistical analysis was done with IBM SPSS statistics v20 and R v3.0.1. (R foundation for statistical computing, Vienna, Austria).

RESULTS

Between January 2006 and December 2012, 1703 patients with ischemic stroke were admitted and 285 (17%) were treated with rtPA. We excluded 17 patients because of referral from another hospital for intra-arterial thrombolysis, 3 patients because of an in-hospital event and 3 patients because of missing data, leaving 262 patients. Patients treated with thrombolysis were on average 63 years old at the time of the stroke and 52% were male (Table 1).

Year	Hospital admissions for ischemic stroke N	Thrombolysis after ED admission n (%)	Age* mean (sd)	Gender* male n (%)	Door-to-needle time, median (IQR)
2006	262	12 (5)	60 (14.9)	6 (50)	75 (70-100)
2007	266	13 (5)	60 (17.2)	8 (62)	65 (45-85)
2008	200	31 (16)	67 (16.5)	15 (48)	60 (42-90)
2009	232	45 (19)	62 (17.4)	25 (56)	50 (36-72)
2010	235	54 (23)	62 (16.1)	27 (51)	40 (27-68)
2011	242	49 (20)	65 (14.7)	26 (53)	40 (25-55)
2012	261	58 (22)	63 (15.8)	29 (50)	35 (23-56)
Total	1703	262	63 (16.1)	136 (52)	45 (30-70)

Table 1; Demographic characteristics. *Age and gender are related to the rtPA patients

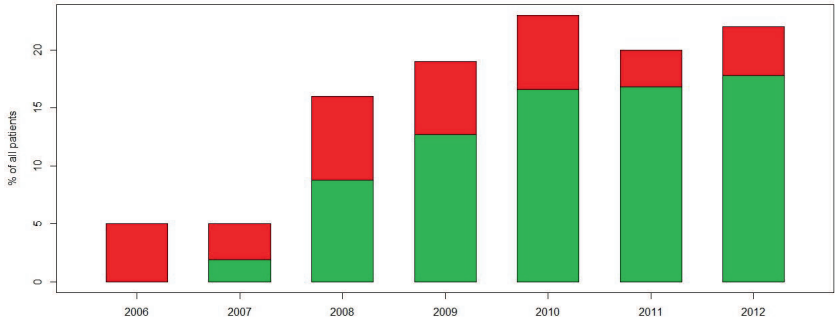


Figure 1; Percentage of patients with acute ischemic stroke treated with thrombolysis (red/green bars) and the fraction treated within one hour (green) per year from 2006 to 2012.

Mean age (p=0.58) and sex distribution (p=0.98) did not change over the years. The proportion treated with thrombolysis increased from 5% in 2006 to 22% in 2012 (figure 1). In 2006, none of the patients were treated within one hour. In 2012, this had increased to 81% (figure 1).

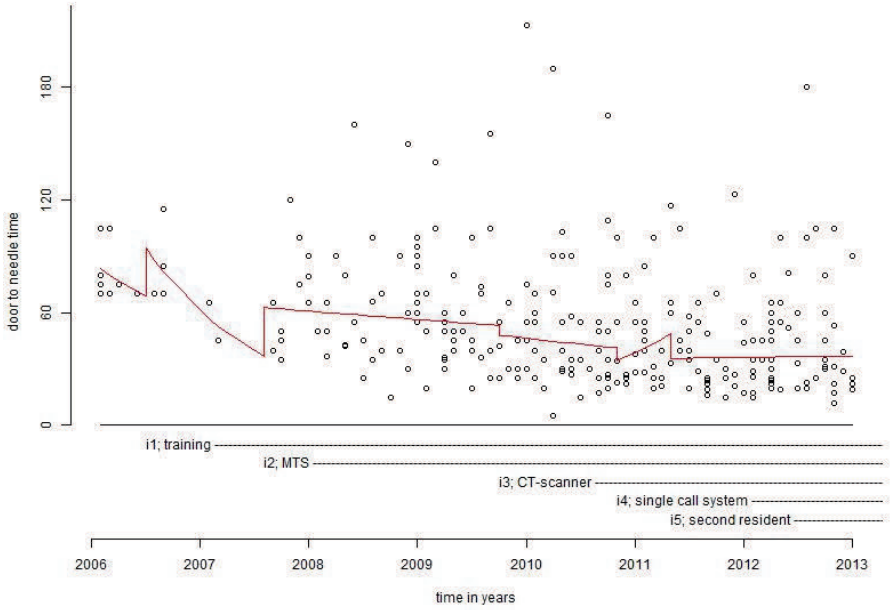


Figure 2; Scatter plot with regression lines indicating effects of interventions aimed at improving door-to-needle time for acute ischemic stroke treatment with rtPA.

In a logistic regression analysis this trend was significant (OR 1.6 per year CI 1.4-1.8). Since 2006, the median door-to-needle time was reduced from 75 minutes to 45 minutes in 2012 ($p<.001$ in a linear regression model). In this period a 12% annual decrease in door-to-needle time was achieved (CI 16%-8%). We could not find a significant association between any specific intervention (figure 2) and the trend in the DNT.

DISCUSSION

We found that median DNT was successfully reduced by 30 minutes between 2006 and 2012. The percentage of patients treated within 60 minutes increased from 0 to 81%. Although DNT improved significantly, we could attribute this trend to one or more specific interventions.

We note that all implemented interventions have been proven effective in the literature.(3, 5, 10-17) An explanation for the lack of significance in our analysis may lie in a slow and gradual effect of our interventions. We selected only those interventions with a fixed starting point in time to include in our analysis. Other initiatives,

such as discussing DNT for every patient at the morning report, could also explain the reduction in DNT (residual confounding). We hypothesize that the cumulative effect of various interventions lowered the DNT. The constant and increasing focus on improvement will have steered the perception of urgency among physicians and ED personnel. Such a perception may translate in faster action to initiate treatment. (5) This highlights the complexity of quality improvement within a single center setting and of relating the results to a single measurement. A recent review evaluated the effectiveness of improvements from quality collaboratives, especially feed-back systems.(18) It concluded that although the evidence of the impact of quality collaboratives is positive, it is also limited because of the complex nature of improvements and the different ways they are applied. Our results resemble those of Meretoja et al, (19) who reduced the median in-hospital delay to 20 minutes with multiple concurrent strategies, but they did not relate the decrease of DNT to any single intervention. It hence remains unclear what specific mechanisms or interventions are responsible for the quality improvement.

We assumed that the groups of patients were similar every year. We could check this assumption for age and gender and stroke severity, which were similar over the years and did not impact on the findings. Other, unmeasured, confounders may however have influenced the results. We speculate that the increase in proportion of patients treated with rtPA over the years means that more complicated patients were also treated, for example those with not readily available information on contra-indicated medication, or with fluctuating symptoms or high blood pressure. This implies that the observed trend in reducing DNT would even have been stronger if the same selection of patients had been made as in the early years of rtPA treatment.

Our results support the use of performance measures for internal communication. Median DNT should be used on a monthly or quarterly basis to inform all professionals treating stroke patient of their achievements. Measuring and reporting DNT could be helpful in keeping professionals focused and in improving performance.(12)

Limitations of our research are the single center design without control group and relatively small sample size. The small sample size may explain our lack of statistically significant results for specific interventions. It implies that scientifically valid evaluations of local implementation are only possible in large centers with large caseloads. The lack of effect of specific interventions may be explained by type 2 error (lack of power) or by a true absence. It remains unclear if the performance indicator is not suitable for explaining the individual interventions (type 2 error or lack of power) in the single centre setting or if the intervention itself did not have a major impact. Furthermore, we did not focus on measuring more specific parts of the care processes, like “onset to door time”, “door-to-CT time” or “door-to-neurologist time”,

while this could be beneficial in guiding future improvements. Recent findings suggest a more comprehensive approach to the total chain of care enabling rtPA treatment to eliminate bottlenecks in the entire pre- and intra hospital care pathway.(20)

In conclusion, both door-to-needle time and the percentage of patients treated within 60 minutes after ED admission, improved significantly, presumably through the combined impact of a series of structural and logistic interventions.

Acknowledgment

We would like to thank M. Scheele, A. van der Lugt and M. Eesteren for their contribution in finding the dates of all interventions.

Source of Funding: Internal Erasmus MC grant for health care research (Mrace)

Disclosure statement: Participation of authors; JPM had the original idea for the study and EWS and AMvD developed the study design. DD and MD provided the data base. AMvD analyzed the data with CWNL. AMvD wrote a first draft of the paper. All authors contributed to further drafts.

Competing interests: None

REFERENCES

1. Hacke W, Donnan G, Fieschi C, et al. Association of outcome with early stroke treatment: pooled analysis of ATLANTIS, ECASS, and NINDS rt-PA stroke trials. Lancet 2004;363:768-774.
2. Wardlaw JM, Murray V, Berge E, et al. Recombinant tissue plasminogen activator for acute ischaemic stroke: an updated systematic review and meta-analysis. Lancet 2012;379:2364-2372.
3. Jauch EC, Saver JL, Adams HP, Jr., et al. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/ American Stroke Association. Stroke 2013;44:870-947.
4. Desai JA, Smith EE. Prenotification and Other Factors Involved in Rapid tPA Administration. Curr Atheroscler Rep 2013;15:337.
5. Mikulik R, Kadlecova P, Czlonkowska A, et al. Factors influencing in-hospital delay in treatment with intravenous thrombolysis. Stroke 2012;43:1578-1583.
6. Dirks M, Niessen LW, Huijsman R, et al. Promoting Acute Thrombolysis for Ischaemic Stroke (PRACTISE). Int J Stroke 2007;2:151-159.
7. Dirks M, Niessen LW, van Wijngaarden JD, et al. Promoting thrombolysis in acute ischemic stroke. Stroke 2011;42:1325-1330.
8. van der Wulp I, van Baar ME, Schrijvers AJ. Reliability and validity of the Manchester Triage System in a general emergency department patient population in the Netherlands: results of a simulation study. Emerg Med J 2008;25:431-434.
9. Glass G. Interrupted time series quasi-experiments. In: Complementary methods for research in education, 2nd ed. Washington D.C.: American Educational Research Association, 1997: 589-609.
10. Bae HJ, Kim DH, Yoo NT, et al. Prehospital notification from the emergency medical service reduces the transfer and intra-hospital processing times for acute stroke patients. J Clin Neurol 2010;6:138-142.
11. Casolla B, Bodenant M, Girot M, et al. Intra-hospital delays in stroke patients treated with rt-PA: impact of preadmission notification. J Neurol 2013;260:635-639.
12. Fassbender K, Balucani C, Walter S, Levine SR, Haass A, Grotta J. Streamlining of prehospital stroke management: the golden hour. Lancet neurol. 2013;12:585-596.
13. Helsinki Stroke Thrombolysis Registry G. Does time of day or physician experience affect outcome of acute ischemic stroke patients treated with thrombolysis? A study from Finland. Int J Stroke 2012;7:511-516.
14. Kim SK, Lee SY, Bae HJ, et al. Pre-hospital notification reduced the door-to-needle time for iv t-PA in acute ischaemic stroke. Eur J Neurol 2009;16:1331-1335.
15. Lau AY, Soo YO, Graham CA, et al. An expedited stroke triage pathway: the key to shortening the door-to-needle time in delivery of thrombolysis. Hong Kong Med J 2010;16:455-462.
16. Lindsberg PJ, Happola O, Kallela M, Valanne L, Kuisma M, Kaste M. Door to thrombolysis: ER reorganization and reduced delays to acute stroke treatment. Neurology 2006;67:334-336.
17. Saver JL. Time is brain--quantified. Stroke 2006;37:263-266.
18. Schouten LM, Hulscher ME, van Everdingen JJ, Huijsman R, Grol RP. Evidence for the impact of quality improvement collaboratives: systematic review. BMJ 2008;336:1491-1494.
19. Meretoja A, Strbian D, Mustanoja S, Tatlisumak T, Lindsberg PJ, Kaste M. Reducing in-hospital delay to 20 minutes in stroke thrombolysis. Neurology 2012;79:306-313.

20. Lahr MM, Luijckx GJ, Vroomen PC, van der Zee DJ, Buskens E. The chain of care enabling tPA treatment in acute ischemic stroke: a comprehensive review of organisational models. J Neurol 2013;260:960-968.

Chapter **8**

Measuring quality improvement using pressure ulcer prevalence; segmented regression analysis of an interrupted time series

van Dishoeck AM, Steyerberg EW, van Lanschot JJB, Hovius SER, Mackenbach JP. Measuring quality improvement using pressure ulcer prevalence; segmented regression analysis of an interrupted time series. to be submitted. 2015.

ABSTRACT

Background; A basic purpose of an indicator is to improve health care. Pressure ulcer prevalence is widely used as quality indicator of nursing care. A quality improvement program was initiated in two surgical units after the increase in pressure ulcer prevalence over a 2-year period. The aim of this research was measuring the effect of a trainings program improving knowledge towards pressure ulcer prevention using the outcome (pressure ulcer prevalence) combined with several process indicators.

Methods; Design; quasi-experimental interrupted time series design. We used the outcome indicator, pressure ulcer prevalence, and several process indicators in a time series design in two surgical wards to monitor the effect of a quality improvement program. The prevalence of these indicators was measured monthly in a 3 and 5 months period before and after the intervention. We estimated the trend in the pressure ulcer prevalence and process indicator prevalence (dependent variables) prior to the intervention and after the intervention. We tested for changes in dependent variable pre intervention and post intervention and tested for changes in the slope of the trend pre and post intervention.

Results; We performed eight prevalence measures in which 299 patients (120 pre-intervention, 179 post-intervention) were included. The pressure ulcer prevalence prior to the intervention varied between 5 and 14%. After the training, we observe a drop in pressure ulcer prevalence to 3 and 0%. In the last two measurements, the prevalence rises to 15%. These differences were not significant. The trend in risk assessment improved significantly (β -0.7 before β 8.4 (p-value <0.01))

Conclusion; The process indicators provide insight in the daily practice and offer opportunities for further improvement of process quality. The outcome measure presents only an indication of the quality of the preventive care process.

BACKGROUND

A basic purpose of measuring quality is to improve health care. Performance indicators must thus provide clues for subsequent improvement of the quality of care delivered, so called actionability. Actionability is then the degree to which a health care professional can influence the measure, in response to an unfavourable value of the indicator.[1, 2]

Pressure ulcers (PU) in hospitals occurs in patients who are bedridden or less mobile [3]. Nurses can avoid pressure ulcers in many cases with early identification of risk, planning and implementation of preventive actions and the systematic registration and communication of the effects of these actions [4]. The quality of these processes (timely and correct application) can be summarized in the term process quality. Does the patient develop a pressure ulcer despite optimal process quality; this was unavoidable [5]. The quality of the preventive process can be measured using outcome indicators such as incidence or prevalence of pressure ulcers [6, 7]. The Dutch Healthcare Inspectorate considered the pressure ulcer prevalence as an indicator of the quality of nursing care [8, 9]. In this context, all clinical units in a large university hospital measure the pressure ulcer prevalence twice a year since 2008. Although the individual measurements are largely influenced by random variation, the historical data provides trend information on pressure ulcers prevalence. A quality project was carried out after an increase of the pressure ulcers prevalence on two surgical units over a two-year period. The aim of the project was to improve the care processes preventing the development of pressure ulcers. Since it is still unclear if pressure ulcer prevalence can be used to monitor quality improvement within a hospital, the aim of this research was measuring the effect of a trainings program improving knowledge towards pressure ulcer prevention using the outcome (pressure ulcer prevalence) combined with several process indicators.

METHOD

Theoretical framework

For this quality project, we set up a theoretical framework based on controllable and non-modifiable factors to give insight in their relationships (Figure 1).

The patient characteristic's and disease or treatment factors determine the non-modifiable factors influencing the risk of pressure ulcer development. Combined with immobility and / or limited mobility they provide the extent of the pressure ulcer risk and the need for prevention. Optimal prevention is essential in avoiding pressure ulcers. The knowledge and attitude of caregivers affect the assessment of the risks to

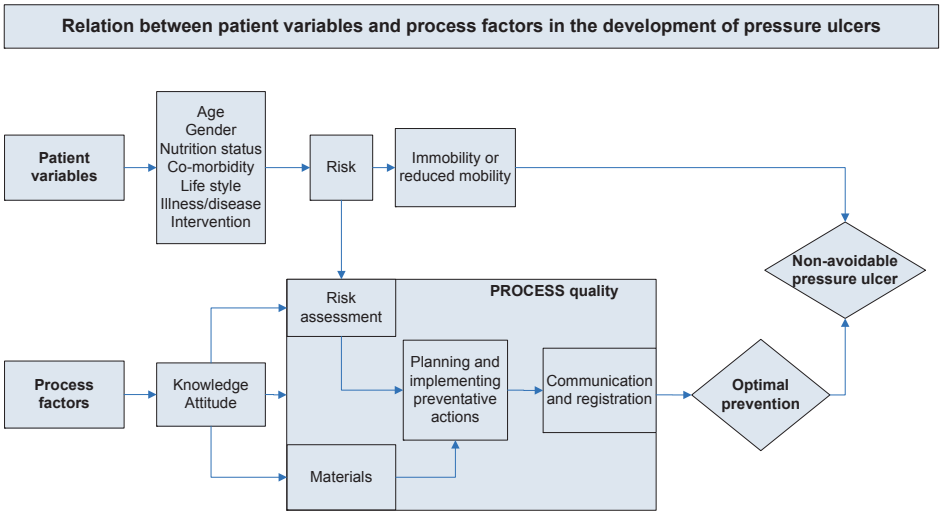


Figure 1 Theoretical framework of the relation between modifiable and non-modifiable factors in the development of pressure ulcers.

the patient, the planning and implementation of prevention and the available pressure reducing materials. Skin observations, communication and registration of prevention and pressure ulcers are just as important as the foregoing processes. Accordingly, all these controllable factors determine the quality of the preventive processes.

Population

The population consisted of all consecutive patients with admitted at two surgical units on the days that the prevalence measurements took place. On these two nursing units, patients were surgically treated for gastrointestinal diseases or after trauma. For the measurement of pressure ulcers of risk assessment and registration, we included all eligible patients. For the preventive measures, we selected patients with an increased risk of developing pressure ulcers. We excluded patients with pressure ulcers on admission.

Design

We chose an interrupted time series design with serial prevalence measurements of pressure ulcers and process measurements before and after a training program for nurses. Interrupted time series is a quasi-experimental design to evaluate longitudinal effects. Next to the observation of the occurrence of pressure ulcers and incontinence-associated dermatitis (IAD), we also examined the preventive measures and risk assessment in the patient files. Furthermore, we collected patient data concerning non-modifiable factors. After three prevalence measurements, the intervention took place. The intervention consisted of a comprehensive training program for

all nurses of the clinical units. With mini posters, we drew extra attention to various preventive measures throughout the intervention period. After the intervention, five serial prevalence measures took place.

Statistical analysis

Segmented regression analysis of interrupted time series data allowed us to assess how much the intervention changed the outcome, both immediately and over time. [10, 11] We estimated the trend in the dependent variable prior to the intervention, after intervention and tested for changes in the slope of the trend pre and post intervention;

$$\textbf{Model PU } = b0 + b1T + b2I + b3P$$

T (time) represent the time form the start of the measurement period starting at 1 (continuous variable in months). Beta T expresses the overall trend. If not significant, the trend was flat (not changing). I (intervention) represent the difference in pre and post intervention, coded 0 prior to the intervention, 1 post intervention. Beta I expresses the drop in pressure ulcer prevalence and process measure prevalence's after an intervention. P (post) represent the time since intervention, coded 0 prior to the intervention and post intervention starting at 1 (continuous variable in months). Beta P expresses the change in trend after the intervention.

Graphical and statistical analyses were performed with SPSS 20.0 and Excel 2010.

Results

During the project period, we performed eight prevalence measurements in which 299 patients were involved. The three pre-intervention measurements contained 120 patients. The five post-intervention measurements contained 179 patients. The majority of the patients were of the male sex (61%) and the mean age was 55 years with ranging from 19-94 years. The median length of hospital stay was 7 days (inter quartile range (IQR) 3-15). The cause of hospital admission was predominantly gastric-intestinal diseases (54%) and trauma or multi-trauma (33%) (table 1).

In 30% of patients, the disease involved a malignancy. The treatment was in most cases surgery for primary disease (54%) and surgery for complications or additional surgery (19%). An increased risk of developing pressure ulcers was present in 201 patients (67%). Twenty-one patients had pressure ulcers on admission at the Erasmus MC or developed pressure ulcers on another unit in the Erasmus MC during the hospitalization and were excluded. Twenty-two patients developed pressure ulcers category 2-4 on the units during the project period. The outcome measures are presented in table 2 and figure 2.

Variable	Total N=299	Pre N=121	Post N=178	p-value
Age, mean (SD)	60(14)	53(17)	56(17)	0.18
Gender, male n (%)	182(61)	69(58)	113(63)	0.33
Increased PU risk	201(67)	71(59)	130(73)	0.15
Length of stay, mean (SD)	13(16)	14(19)	11(14)	0.13
Disease n (%)				0.04
oesophagus and stomach	38 (13)	14 (12)	24 (13)	
small intestine	22 (7)	11 (9)	11 (6)	
colon	42 (14)	9 (7)	33 (19)	
liver and gallbladder	34 (11)	22 (18)	12 (7)	
pancreas	25 (8)	9 (7)	16 (9)	
trauma	71 (24)	28 (23)	43 (24)	
multi-trauma	27 (9)	11 (9)	16 (9)	
endocrine system	1 (0.3)	1 (1)	0 (0)	
bones and joints (no trauma)	8 (3)	5 (4)	3 (2)	
inguinal, umbilical, incisional hernia	13 (4)	6 (5)	7 (4)	
renal diseases	1 (0.3)	0 (0)	1 (1)	
skin and soft tissue	1 (0.3)	0 (0)	1 (1)	
chronic pain syndrome	8 (3)	2 (2)	6 (3)	
other	8 (2)	3 (2)	5 (3)	
Malignancy n (%)	90.(30)	41.(34)	49 (27)	0.21
Treatment n (%)				0.94
surgery for primary disease	160 (54)	63.(53)	97.(54)	
surgery for complications or reconstruction	58 (19)	25.(21)	33.(18)	
drain of stent	27 (9)	12.(10)	15.(8)	
“wait and see”	20 (67)	8 (7)	12.(7)	
other	34 (11)	12.(10)	22.(12)	
Intensive Care during admission	97 (32)	42.(35)	55 (31)	0.44

Table 1 Descriptive variables (p-values; chi^2 nominal and ordinal variables, t-test continuous variables)

Pressure ulcer occurrence	Total N=299	Pre n=120	Post n=179	p-value
Pressure ulcers before unit admission n (%)	21 (7,0)	7 (5,8)	14 (7,8)	0.49
Nosocomial pressure ulcers				0.16
Pressure ulcer cat 1 n (%)	28 (9,4)	16 (13,3)	12 (6,7)	
Pressure ulcer cat. 2 n (%)	29 (9,7)	11 (9,2)	18 (10,1)	
Pressure ulcer cat. 3 n (%)	4 (1,3)	3 (2,5)	1 (<1)	
Pressure ulcer cat. 4 n (%)	1 (<1)	0 (<1)	1 (<1)	

Table 2; pressure ulcer occurrence during prevalence measures

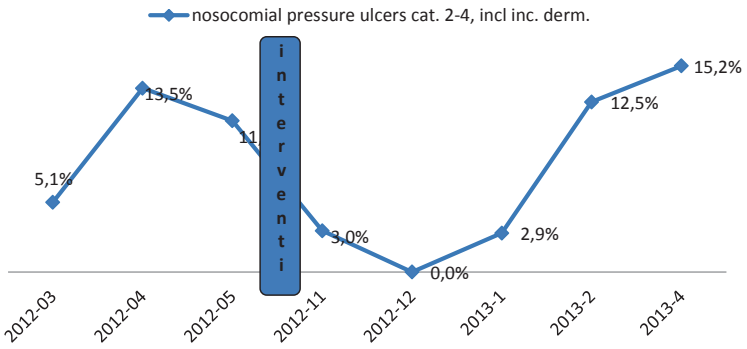


Figure 2; PU prevalence before and after an intervention aiming at improving the knowledge on pressure ulcer prevention among nursing staff of two surgical units.

The trend prior to the intervention varies between 5 and 14%. After the training, we observe a drop in pressure ulcer prevalence to 3 and 0%, which is a clinically relevant effect. This effect is long lasting, since in the last two measurements, the prevalence rises again to 15%. In segmented logistic regression analysis, this outcome was not statistically significant due to the small numbers (Table 3).

Variable	Trend before intervention B (p-value)	Direct effect of the intervention B (p-value)	Trend after intervention B (p-value)
PU cat. 2-4	0.2 (0.51)	-5.0 (0.15)	0,4 (0.51)
Risk assessment	-0.7 (0.03)	8.4 (<0.01)	0.8 (0.09)
Patient information	0.3 (0.65)	4.0 (0.41)	0.4 (0.51)
Alternating mattresses	-0.4 (0.2)	3.8 (0.26)	0.5 (0.30)
Heel prevention	0.1 (0.83)	-0.2 (0.96)	-0,1 (0.76)
Repositioning	0.1 (0.77)	1.3 (0.66)	0.2 (0.55)

Table 3 Trend analyses. PU = pressure ulcers PU and Risk assessment was measured among all patients; prevention was measured among risk patients.

The graphical display of the process measures (figure 3) shows that particularly in risk assessment we were able to improve the quality of care, which remained stable over the follow-up measurements.

At the last measurement after the intervention, risk assessment was performed in 97% of the patients.

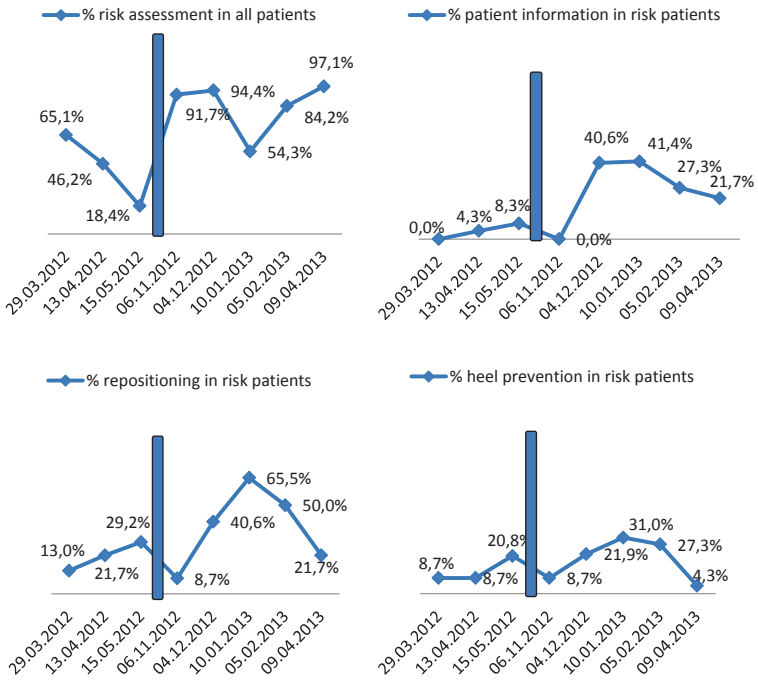


Figure 3 Process indicators before and after the intervention

The trend in this process indicators was statistically significant with a downward trend prior to the intervention, a significant increase directly after the intervention followed by an increasing trend (Table 2). Patient information, heel prevention and repositioning show some variability, but not directly after the training, and there was no permanent effect observed. The use of alternating matrasses remained unchanged. In the other process indicators we did not find an significant effect of the intervention.

DISCUSSION

The aim of this research was exploring the use of quality indicators in measuring the effect of an intervention targeting at lowering pressure ulcer occurrence on two surgical units. In the outcome indicator (pressure ulcer prevalence), we achieved a clinically relevant improvement after the training, however this was not permanent. The same pattern was seen in the process indicators in which we observed an initial and temporary improvement. Thus, the process measures provided insight in the cause of the relatively high PU prevalence. None of these results were significant in the interrupted time series analyses. Since the majority of process indicators was measured

in risk patients only, the numbers were small, making significant changes difficult to assess. Thus it is unclear what contributed to these non-significant results. We did achieve a significant improvement in the process indicator risk assessment, measured in all patients. The lack of effect of the intervention is consistent with research on the effectiveness of training as a single intervention.[12]. Also, van Gaal et al found no overall difference in preventive pressure ulcer measures in hospitals as effect of a multifaceted implementation strategy.[13]. Research by Beeckman et al [14] showed that attitude towards prevention was significantly correlated with pressure ulcer occurrence. Other research shows successful interventions. Research by Uzun et al showed education regarding preventive care to be effective in reducing the incidence of PUs in an ICU setting.[15] Anderson et al implemented a pressure ulcer prevention bundle using frequent nurse rounds by a tissue viability nurse.[16] This resulted in a statistically significant and clinically relevant reduction in the incidence of pressure ulcers. These findings confirm the complexity of improvement initiatives in guideline adherence, as well as measuring improvement.

Limitations of our research are the single centre pre-post intervention design without control group and relatively small sample sizes. The small sample size may explain our lack of statistically significant results for the process measures, or this can caused by a true absence of effect of the intervention.

We used a time series analysis, a well-known and widely used design.[17-19] We did not explore the use of other approaches to the analysis, such as control charts. The control chart combines time series analysis with a graphical presentation of the data. [20] Presenting limits of one, two and three standard deviations (SD) might be more informative in showing diverging outcomes.

Generalizability to quality improvements is limited for only two surgical wards in a specific setting were included in this quality project. Generalizability to other indicators however, is valid in case of small number. Despite these limitations, the findings of this research contribute to the understanding of the value and constraints of performance indicators in measuring and improving quality of care.

We conclude that the process measurements provide insight in the daily practice and offers opportunities for further improvement of process quality. The outcome indicator pressure ulcer prevalence presents an indication of the quality of the preventive care process.

Source of Funding: Internal Erasmus MC grant for health care research (Mrace)

Participation of authors; JPM had the original idea for the study and EWS and AMvD developed the study design. JL and SH initiated the quality improvement. AMvD coordinated the project and analyzed the data. AMvD wrote a first draft of the paper. All authors contributed to further drafts.

Competing interests: None

REFERENCES

1. Derose, S.F. and D.B. Petitti, *Measuring quality of care and performance from a population health care perspective.* Annu Rev Public Health, 2003. **24**: p. 363-84.
2. Spertus, J.A., et al., *American College of Cardiology and American Heart Association methodology for the selection and creation of performance measures for quantifying the quality of cardiovascular care.* J Am Coll Cardiol, 2005. **45**(7): p. 1147-56.
3. Allman, R.M., *Pressure ulcer prevalence, incidence, risk factors, and impact.* Clin Geriatr Med, 1997. **13**(3): p. 421-36.
4. European Pressure Ulcer Advisory Panel and National Pressure Ulcer Advisory Panel. *Prevention of pressure ulcers: quick reference guide* 2009 [cited 2013 08-06-2013]; Available from: http://www.epuap.org/guidelines/QRG_prevention_in_Belgian.pdf.
5. Black, J.M., et al., *Pressure ulcers: avoidable or unavoidable? Results of the National Pressure Ulcer Advisory Panel Consensus Conference.* Ostomy Wound Manage, 2011. **57**(2): p. 24-37.
6. Berlowitz, D., VanDeusen Lukas, C., Niederhauser, A., Parker,V., Silver, J., Ayello, E., Zulkowski, K. *How Do We Measure Our Pressure Ulcer Rates and Practices?* [Publication # 11-0053-EF] 2011 [cited 2013 08-06-2013]; Available from: http://www.ahrq.gov/professionals/systems/long-term-care/resources/pressure-ulcers/pressureulcertoolkit/putool5.html.
7. Defloor, T., et al., *EPUAP statement on prevalence and incidence monitoring of pressure ulcer occurrence.* J Tissue Viability, 2005. **15**(3): p. 20-7.
8. IGZ, et al., *Basisset prestatie-indicatoren ziekenhuizen* . ed. I.v.d. Gezondheidszorg. 2009, Utrecht
9. IGZ, I.v.d.G. *English homepage* 2009 [cited 2009 11-01-09]; Available from: http://www.igz.nl/uk/.
10. Aiken, A.M., et al., *Changing use of surgical antibiotic prophylaxis in Thika Hospital, Kenya: a quality improvement intervention with an interrupted time series design.* PLoS One, 2013. **8**(11): p. e78942.
11. Yu, S.C., et al., *Using interrupted time series design to analyze changes in hand, foot, and mouth disease incidence during the declining incidence periods of 2008-2010 in China.* Biomed Environ Sci, 2012. **25**(6): p. 645-52.
12. Thompson, C. and S. Stapley, *Do educational interventions improve nurses' clinical decision making and judgement? A systematic review.* Int J Nurs Stud, 2011. **48**(7): p. 881-93.
13. van Gaal, B.G., et al., *The SAFE or SORRY? programme. part II: effect on preventive care.* Int J Nurs Stud, 2011. **48**(9): p. 1049-57.
14. Beeckman, D., et al., *Knowledge and attitudes of nurses on pressure ulcer prevention: a cross-sectional multicenter study in Belgian hospitals.* Worldviews Evid Based Nurs, 2011. **8**(3): p. 166-76.
15. Uzun, O., R. Aylaz, and E. Karadag, *Prospective study: reducing pressure ulcers in intensive care units at a Turkish medical center.* J Wound Ostomy Continence Nurs, 2009. **36**(4): p. 404-11.
16. Anderson, M., et al., *Universal Pressure Ulcer Prevention Bundle With WOC Nurse Support.* J Wound Ostomy Continence Nurs, 2015. **42**(3): p. 217-25.
17. Anaby, D., et al., *Interrupted Time Series Design: A Useful Approach for Studying Interventions Targeting Participation.* Phys Occup Ther Pediatr, 2013.
18. Hanbury, A., L. Wallace, and M. Clark, *Use of a time series design to test effectiveness of a theory-based intervention targeting adherence of health professionals to a clinical guideline.* Br J Health Psychol, 2009. **14**(Pt 3): p. 505-18.
19. Harlow, L.L. and M.D. Anglin, *Time series design to evaluate effectiveness of methadone maintenance intervention.* J Drug Educ, 1984. **14**(1): p. 53-72.

20. Duclos, A. and N. Voirin, *The p-control chart: a tool for care improvement.* Int J Qual Health Care, 2010. **22**(5): p. 402-7.

Chapter 9

General Discussion

Although Florence Nightingale had high expectations of the possibilities of performance indicators ("hospital statistics") it took more than a century before her ideas were carried out on a large scale with the purpose of "*enabling us to ascertain the mortality in different hospitals, as well as from different diseases and in different districts of the same country"[1].* This goal has now largely been met, thanks to extensive efforts of data collection. However, we are still far from her ideal that performance indicators will "*improve the treatment and management of the sick and maimed poor*". We found that performance indicators often provide only a crude and potentially misleading indication of the quality of care.

MAIN FINDINGS

The aim of this research was to study the value of performance indicators in comparing the quality of care between hospitals and their usefulness for the improvement of the quality of care within hospitals.

1. *"How to interpret differences between hospitals in performance indicator measures?"*

We found considerable influence of random variation when we compared hospitals using the outcome indicators of the Dutch Health Care Inspectorate. In the ranking of hospitals, the uncertainty of the estimates led to unreliable positioning. Therefore, none of the tested indicators could be used for the ranking of hospitals (chapter 2). Furthermore, the graphical displays in which indicators are presented must include information on random variation. The funnel plot provided a representation of differences between hospitals compared to a target value, therewith allowing simple interpretation of the uncertainty of these differences. A forest plot gave appropriate insight in the number of Dutch hospitals that actually significantly deviated from the average by presenting the hospital estimates and their confidence intervals. We also used rank plots and showed that the substantial uncertainty made rankings with these outcome indicators unreliable. Of all three graphical displays used, the funnel plots provided most the valuable insight in the magnitude of random variation and is therefore best used for the overall interpretation of differences between hospitals

(chapter 3). In addition, random effects analysis provided means to correct the effect size of the observed differences for chance. For surgical site infections, we found that the apparent differences between Dutch hospitals were predominantly attributable to random variation and case-mix. This case study provided a clear illustration that both random variation and case-mix must be addressed systematically in performance measurement before conclusions can be drawn on the quality of hospital care (chapter 4).

2. *"How strong is the relation between outcome indicators and the underlying care processes, and can the performance indicator be used for quality improvement?"*

Exploring the process-outcome relation, we found that the outcome indicator 'pressure ulcer occurrence' reflected differences in the quality of the bundle of preventive care processes provided by nurses. This significant relation between outcome and process in pressure ulcer care supports the usefulness of this indicator in assessing the quality of nursing care (chapter 5). Addressing the process-outcome relation in performance measurement from another angle, we explored the feasibility of measuring the effect of surveillance by the Dutch Health Care Inspectorate using retrospective data on health outcomes in three health problems. We found that in case of clearly defined health problems, such as pressure ulcers and suicide, the frequency of these outcomes could be measured using Inspectorate data and trends could be analysed using an interrupted time series design. However, support of a causal relationship between supervision and observed trends can only be derived with data on external factors that influenced this trend (chapter 6). In a process measure of acute stroke care, we found a significant improvement in "door-to-needle time" (DNT) measured over several consecutive years. We could not attribute this trend to one or more specific interventions. We hypothesised that the combined effect of various interventions together and the constant focus of care-givers on quality improvement explained the significant improvement of the indicator DNT (chapter 7). In a quality project that aimed at decreasing pressure ulcer occurrence, we found a significant improvement in the process measure risk assessment However, we found no statistically significant decrease of pressure ulcer occurrence using an interrupted time series design (chapter 8).

LIMITATIONS

Scope and generalizability

Our research centred on performance indicators and quality of care in Dutch hospitals. Using the yearly published hospital data, we explored the influence of random

variation and case-mix. We did not investigate these factors in other areas of health care, such as general practice or long-term care. We do not know whether the influence of random variation and case-mix are more or less prominent in other health care sectors. Therefore, our findings cannot be generalized to other sectors, and our conclusions will only address the use of performance indicators in hospital settings.

Of all available hospital indicators, we selected in our research only a limited number of indicators for specific conditions, that is, cardiology, neurology, surgery and nursing. Our findings on outcome measures likely generalise to other outcome indicators in this field, because random variation and case-mix are known distorting factors. Process measures are commonly considered to be less influenced by random variation and differences in case-mix [2, 3]. This is because process measures such as door-to-needle-time are less rare compared to outcome measures such as mortality. Moreover, many care processes need to be followed in all patients irrespective of their specific risk profile. However for door-to-needle-time it could be argued that in high comorbidity patients, additional tests need to be performed which can prolong the door-to-needle time. Therefore, the impact of case-mix should be explored further for specific process measures.

We used databases containing information on hospitals in the Netherlands. Although the total number of patients will be greater in larger countries, the sizes of the hospitals are likely to be similar, making our findings on the role of random error generalizable to other western countries. The effect of case-mix depends on the magnitude of the differences in case-mix between hospitals. This is expected to be present in any country, and even more so in settings with higher specialized centres.

Internal validity

Our study designs were mostly retrospective and observational in nature. As a consequence thereof, our analyses were limited to the available data. For the Dutch Health Care Inspectorate's indicators (chapter 2 and chapter 3), the data did not include case-mix variables. Thus, we were unable to explore the influence of case-mix in these outcome measures, while case-mix differences likely impact on the comparisons of the individual hospitals. This limitation in the data restricts interpretations of the Dutch Health Care Inspectorate's indicators, in addition to the statistical uncertainty in many of the indicators. The PREZIES-database (chapter 4) did include case-mix variables but unmeasured aspects could have caused residual confounding that might have explained part of the statistically significant differences. More valid interpretations may be possible with analyses of individual patient data, such as initiated by the Dutch Institute for Clinical Auditing (DICA).

We concentrated on the role of random variation and case-mix and paid no attention to bias, such as registration errors or differences in operationalization of indicators. In consequence, the statistically significant differences that we found in almost all indicators comparing hospitals cannot be attributed with confidence to the quality of delivered care. Therefore, we might have overestimated or underestimated the quality-of-care effects.

In our quality improvement research, we used a time series analysis, a well-known and widely used design [4-6]. We did not explore the use of other approaches to the analysis, such as control charts. The control chart combines time series analysis with a graphical presentation of the data.[7] Presenting limits of one, two and three standard deviations (SD) might be informative in showing diverging outcomes.

Despite these limitations, the findings of our project contribute to the understanding of the value and constraints of performance indicators in measuring and improving quality of care.

INTERPRETATION

The use of performance indicators has become popular in the last decades based on the belief that achieving good health outcomes for patients is the fundamental purpose of healthcare.[8] In light of this study project, we review what these measurements tell us currently in comparison to that what is known in the field.

Comparing hospitals

In comparing and ranking hospitals random variation and case-mix are the two major obstacles to reliable ranking.

Random variation

We found that ranking hospitals using outcome indicators issued by the Dutch Health Inspectorate did not give a trustworthy picture of the quality differences between hospitals due to inherent random variation. This unreliability is caused by small sample sizes and/or rare event rates[9]. Other researchers also found that inadequate sample sizes were influencing the reliable assessment of performance when performance was assessed in specific patient subpopulations.[10, 11] Methodological research confirmed that uncertainty affects hospitals with small numbers the most, making it difficult to distinguish between hospitals that are performing well and those doing badly.[12, 13] The hospital estimates are too extreme (either too good or too poor) and shift towards the mean in the random-effects analysis. The outcome measure for performance evaluation is often the same as used in clinical trials. In these trials,

a formal power calculation is used to determine a sufficient sample size. In performance measures, the sample size is determined by the number of patients treated in a hospital in the given timeframe for which the indicator is reported. The variability between hospitals is also influenced by the event rates of the selected outcome or process. Rare outcomes, such as reoperations or mortality, are found to have inadequate reliability for the comparisons of hospitals.[14, 15] Recently, the percentage of limb amputation in patients with diabetic foot ulcers is suggested as an indicator for out-patient diabetic foot clinics in the Netherlands. Although lower limb amputation is regarded as an unfavourable outcome of diabetic foot ulcer care, both the number of patients yearly treated in an out-patient clinic (between 50 and 150 in 2013 [16]), as well as the amputation rate in this population (3,8-4,6%.[17]) are small. These low numbers will lead to an indicator score in which differences between hospitals do not overcome random variation. Combining years of observation may increase the total number of patients, but this time frame may be too broad for quality improvement purposes. For quality improvement, an adequate report frequency is ideally monthly or quarterly or at best yearly. Therefore, estimates of sample size and event rate should be a major topic in the development of indicators.

In order to assess reliability of ranking, we used the concept of rankability in chapter 2. We can compare rankability with the signal-to-noise ratio that is used for electrical signals, defined as the power ratio between a signal (meaningful information) and the background noise (in this case statistical noise or random variation). So, a performance indicator provides a signal on quality of care, which is distorted by random variation. Our research showed that none of the tested indicators are suitable for ranking of hospitals.[9, 18, 19] Using the rankability concept to evaluate the reliability of ranking hospitals on mortality after colorectal surgery, research showed that only 38% of the differences between hospitals were due to true differences after correcting for random variation and case-mix.[20] Similar results were found in stroke patients.[21] Hospitals with small sample sizes make the rank order unlikely to be replicated. On the other hand in IVF patients a high rankability was found in comparing IVF clinics on the number of treatment cycles and the number of pregnancies.[22] The difference between these applications lies predominantly in frequency of the outcome. Interpretation of ranks should be avoided in case of low rankability, or ranks should be shown as an expected rather than observed ranks [42, 71].

For the reader, graphs are often more transparent and give a clearer picture of the situation than tables with numbers. Hospital outcomes are often published graphically in league tables. Although these displays provide a simple overview of the performance, they do not give insight in the underlying numbers.[19, 23, 24] Since Spiegelhalter in 2002 suggested the use of funnel plots for institutional comparisons, several studies described the usefulness of this plot [24-28], although some commented on their

limitations in reports on standardized mortality ratios (SMRs). [29] Funnel plots for SMRs should be used with caution either when the expected number of events is small or when the expected number of events is large. Although more research is needed to clarify the desired graphical display in specific conditions, addressing random variation is indispensable in hospital assessments.

Case-mix

Next to the importance of random variation, our research showed the importance of correcting for case-mix. The effect of case-mix adjustment is independent of sample size and can result in a change of direction, such as from worse than expected to better than expected. This finding is very common in this area of research, where it has often been found that correcting for patient factors changed the perspective on hospital performance from bad to good, or vice versa in case of a favourable case-mix.[30-36] Other research concentrated on the development of case-mix models to correct for patient factors in performance measures.[30, 37-41] It is hence indisputable that correction for patient factors should be part of the assessment of hospital performance.

Improving quality

Process-outcome relation

In our study, we explored the process-outcome interaction and found a significant relationship between the preventive care processes and nosocomial pressure ulcer prevalence.[42] We found that the odds of developing a pressure ulcer was related to the quality of the preventive care processes, indicating that variation in the prevalence reflects variation in quality of care.[42] The process-outcome relationship is not always confirmed in observational research. Tillman et al found that the implementation of a surgical safety checklist improved the compliance to prevention strategies, but it did not affect the overall surgical site infection rate.[43] A systematic review into process indicators for diabetes care showed that process indicators focusing on drug treatment were significantly associated with outcomes, while process indicators measuring numbers of tests or visits were not related to outcomes.[44] In a cohort study, treatment indicators measuring lipid-lowering and albuminuria-lowering status were valid quality measures, but the indicators for blood pressure-lowering treatment did not predict patient outcomes.[45] Although the relationship between process and outcome may seem straightforward and applicable, this needs further research.

Quality improvement projects

One of the aims of a performance indicator is that it provides clues for subsequent improvement of the quality of care, so called “actionability”. For this purpose, we investigated two quality improvement projects. In our pressure ulcers quality project, we measured both process and outcome variables and the intervention significantly

improved the pressure ulcer risk assessment in surgical patients.[46] Despite the fact that the pressure ulcers prevalence decreased from 14% to 3%, this was not statistically significant. In stroke patients, we demonstrated an improvement in door-to-needle time, but could not relate this finding to any of the structural or logistic interventions. The process measure improved steadily, but it was unclear to exactly what improvement this could be contributed.[47] For internal quality improvement, process indicators seem to be more informative than outcome indicators.[42, 47] More research is needed on the use of process indicators, outcome indicators or a combination of these two as a tool for quality improvement and whether studies reporting on multi-centre cohorts give a clear picture on the effect in the individual hospitals.

RECOMMENDATIONS

Two important factors should always be addressed in performance measures: random variation and case mix. Observed differences may be corrected using statistical methods for shrinkage to the mean. Several methods have been suggested [9, 21, 22, 48, 49]. The concept of rankability provides a method for assessing the consistency of ranking and should therefore be considered when ranks are presented. Case-mix differences distort hospital comparisons. Careful registration of relevant characteristics is therefore essential to allow for statistical corrections.

Furthermore, we recommend that the process-outcome relation should be addressed and explored in existing indicators as well as in the development of new quality indicators. All performance indicators should be scrutinised carefully according to their aim for bringing transparency, accountability or improving quality of care.

OVERALL CONCLUSIONS; BEYOND THE NUMBERS

Hospital performance is more than reflected in the ratio of two numbers: a numerator and a denominator. When judging hospital quality, in the quest for more transparency in health care, the influence of random variation and case-mix must be dealt with to avoid over-interpretation of the numbers.[9, 19] Random variation can basically be addressed with larger sample sizes, and respecting uncertainty when small sample size is at stake. Case-mix correction should be applied notwithstanding the extra burden on data collection.[18] When developing performance indicators, one must not only consider the interpretability of indicators used for quality measurement or quality improvement, but also their statistical properties: availability of adequate sample size, and insensitivity to case-mix difference. The structure-process-outcome relation

of indicators should be explored beyond the level of the expert opinion or guideline directed level.[42] Outcome measurement does not provide sufficient information for improving the quality of care.[46, 47] Outcome indicators should be paired with process indicators to gain insight for the improvement of quality of care processes.

The measurement of quality of care is a multidimensional and complex process. We must be aware that a performance indicator may offer an uncertain and invalid signal on quality and is by no means an absolute measure. At best we have a measure like a one hand clock, which indicates roughly what time it is.

REFERENCES

1. Nightingale, F., *Notes on Hospitals*. 1863, London: LONGMAN, GREEN, LONGMAN, ROBERTS, and GREEN.
2. Palmer, R.H., *Process-based measures of quality: the need for detailed clinical data in large health care databases.* Ann Intern Med, 1997. 127(8 Pt 2): p. 733-8.
3. Shojania, K.G., J. Showstack, and R.M. Wachter, *Assessing hospital quality: a review for clinicians.* Eff Clin Pract, 2001. 4(2): p. 82-90.
4. Anaby, D., et al., *Interrupted Time Series Design: A Useful Approach for Studying Interventions Targeting Participation.* Phys Occup Ther Pediatr, 2013.
5. Hanbury, A., L. Wallace, and M. Clark, *Use of a time series design to test effectiveness of a theory-based intervention targeting adherence of health professionals to a clinical guideline.* Br J Health Psychol, 2009. 14(Pt 3): p. 505-18.
6. Harlow, L.L. and M.D. Anglin, *Time series design to evaluate effectiveness of methadone maintenance intervention.* J Drug Educ, 1984. 14(1): p. 53-72.
7. Duclos, A. and N. Voirin, *The p-control chart: a tool for care improvement.* Int J Qual Health Care, 2010. 22(5): p. 402-7.
8. Murphy, P.J., *Measuring and recording outcome.* Br J Anaesth, 2012. 109(1): p. 92-8.
9. van Dishoeck, A.M., et al., *Random variation and rankability of hospitals using outcome indicators.* BMJ Qual Saf, 2011. 20(10): p. 869-74.
10. Eijkenaar, F. and R.C. van Vliet, *Profiling individual physicians using administrative data from a single insurer: variance components, reliability, and implications for performance improvement efforts.* Med Care, 2013. 51(8): p. 731-9.
11. Fung, V., et al., *Meaningful variation in performance: a systematic literature review.* Med Care, 2010. 48(2): p. 140-8.
12. Campbell, D.A., Jr., et al., *Surgical site infection prevention: the importance of operative duration and blood transfusion--results of the first American College of Surgeons-National Surgical Quality Improvement Program Best Practices Initiative.* J Am Coll Surg, 2008. 207(6): p. 810-20.
13. Davidson, G., I. Moscovice, and D. Remus, *Hospital size, uncertainty, and pay-for-performance.* Health Care Financ Rev, 2007. 29(1): p. 45-57.
14. Krell, R.W., et al., *Profiling hospitals on bariatric surgery quality: which outcomes are most reliable?* J Am Coll Surg, 2014. 219(4): p. 725-34 e3.
15. Krell, R.W., et al., *Reliability of risk-adjusted outcomes for profiling hospital surgical quality.* JAMA Surg, 2014. 149(5): p. 467-74.
16. Dutch Hospital Data. *DHD-databank Kwaliteit*. 26-04-2015]; Available from: http://www.ziekenhuizentransparant.nl/.
17. Doggen, K., et al., *Implementation of a quality improvement initiative in Belgian diabetic foot clinics: feasibility and initial results.* Diabetes Metab Res Rev, 2014. 30(5): p. 435-43.
18. van Dishoeck, A.M., et al., *Use of surgical-site infection rates to rank hospital performance across several types of surgery.* Br J Surg, 2013. 100(5): p. 628-36; discussion 637.
19. van Dishoeck, A.M., et al., *Displaying random variation in comparing hospital performance.* BMJ Qual Saf, 2011. 20(8): p. 651-7.
20. Henneman, D., et al., *Ranking and rankability of hospital postoperative mortality rates in colorectal cancer surgery.* Ann Surg, 2014. 259(5): p. 844-9.

21. Lingsma, H.F., et al., *Variation between hospitals in patient outcome after stroke is only partly explained by differences in quality of care: results from the Netherlands Stroke Survey.* J Neurol Neurosurg Psychiatry, 2008. 79(8): p. 888-94.
22. Lingsma, H.F., M.J. Eijkemans, and E.W. Steyerberg, *Incorporating natural variation into IVF clinic league tables: The Expected Rank.* BMC Med Res Methodol, 2009. 9(1): p. 53.
23. van Dishoeck, A.M., et al., *[Outcome assessment in hospitals. The influence of insecurity].* Ned Tijdschr Geneeskd, 2009. 153(17): p. 804-11.
24. Adab, P., et al., *Performance league tables: the NHS deserves better.* Bmj, 2002. 324(7329): p. 95-8.
25. Kirkham, J.J. and O. Bouamra, *The use of statistical process control for monitoring institutional performance in trauma care.* J Trauma, 2008. 65(6): p. 1494-501.
26. Schulman, J., D.J. Spiegelhalter, and G. Parry, *How to interpret your dot: decoding the message of clinical performance indicators.* J Perinatol, 2008. 28(9): p. 588-96.
27. Spiegelhalter, D.J., *Funnel plots for comparing institutional performance.* Stat Med, 2005. 24(8): p. 1185-202.
28. Wilson, J., et al., *Rates of surgical site infection after hip replacement as a hospital performance indicator: analysis of data from the English mandatory surveillance system.* Infect Control Hosp Epidemiol, 2008. 29(3): p. 219-26.
29. Seaton, S.E., et al., *What is the probability of detecting poorly performing hospitals using funnel plots?* BMJ Qual Saf, 2013. 22(10): p. 870-6.
30. Bateman, P., et al., *BDA special care case mix model.* Br Dent J, 2010. 208(7): p. 291-6.
31. Talsma, A.K., et al., *Re-resection rates after breast-conserving surgery as a performance indicator: introduction of a case-mix model to allow comparison between Dutch hospitals.* Eur J Surg Oncol, 2011. 37(4): p. 357-63.
32. Wouters, M.W., et al., *High-volume versus low-volume for esophageal resections for cancer: the essential role of case-mix adjustments based on clinical data.* Ann Surg Oncol, 2008. 15(1): p. 80-7.
33. Davenport, R.J., M.S. Dennis, and C.P. Warlow, *Effect of correcting outcome data for case mix: an example from stroke medicine.* Bmj, 1996. 312(7045): p. 1503-5.
34. DuBard, C.A., J.C. Jacobson Vann, and C.T. Jackson, *Conflicting Readmission Rate Trends in a High-Risk Population: Implications for Performance Measurement.* Popul Health Manag, 2015.
35. Hollis, S., et al., *Standardized comparison of performance indicators in trauma: a new approach to case-mix variation.* J Trauma, 1995. 38(5): p. 763-6.
36. van der Veer, S.N., et al., *Measuring the quality of renal care: things to keep in mind when selecting and using quality indicators.* Nephrol Dial Transplant, 2014. 29(8): p. 1460-7.
37. Bours, G.J., et al., *Development of a model for case-mix adjustment of pressure ulcer prevalence rates.* Med Care, 2003. 41(1): p. 45-55.
38. Damhuis, R., et al., *A case-mix model for monitoring of postoperative mortality after surgery for lung cancer.* Lung Cancer, 2006. 51(1): p. 123-9.
39. Watret, L., *Using a case-mix-adjusted pressure sore incidence study in a surgical directorate to improve patient outcomes in pressure ulcer prevention.* J Tissue Viability, 1999. 9(4): p. 121-5.
40. Neumann, A., et al., *Measuring performance in health care: case-mix adjustment by boosted decision trees.* Artif Intell Med, 2004. 32(2): p. 97-113.
41. Timmers, T.K., et al., *Intensive care performance: How should we monitor performance in the future?* World J Crit Care Med, 2014. 3(4): p. 74-9.

42. van Dishoeck, A.M., et al., *The prevalence of pressure ulcers or incontinence dermatitis reflects the poor quality of care in adult hospitalized patients. An audit case control study.* Journal of Advanced Nursing, 2015. submitted.

43. Tillman, M., et al., *Surgical care improvement project and surgical site infections: can integration in the surgical safety checklist improve quality performance and clinical outcomes?* J Surg Res, 2013. 184(1): p. 150-6.

44. Sidorenkov, G., et al., *Review: relation between quality-of-care indicators for diabetes and patient outcomes: a systematic literature review.* Med Care Res Rev, 2011. 68(3): p. 263-89.

45. Sidorenkov, G., et al., *Do treatment quality indicators predict cardiovascular outcomes in patients with diabetes?* PLoS One, 2013. 8(10): p. e78821.

46. van Dishoeck, A.M., et al., *Measuring quality improvement using pressure ulcer prevalence; segmented regression analysis of an interrupted time series.* to be submitted, 2015.

47. van Dishoeck, A.M., et al., *Measuring quality improvement in acute ischemic stroke care; interrupted time series analysis of door-to-needle time.* Cerebrovascular Diseases Extra, 2014(4): p. 149-155.

48. Anderson, J., et al., *Determining hospital performance based on rank ordering: is it appropriate?* Am J Med Qual, 2007. 22(3): p. 177-85.

49. Ash, A., Fienberg, SE., Louis, TA.,Normand, S-LT., Stukel, Utts, J., *Statistical issues in assessing hospital perfromance.* Commissioned by the Committee of Presidents of Statistical Societies, 2012(The COPSS-CMS White Paper Committee): p. 1-70.

Chapter **10**

Summary

SUMMARY

Quality of care is a broad and abstract concept and the attempts of measuring quality places constrains on the interpretability of the outcomes. The aim of this research was to study the value of performance indicators in comparing the quality of care between hospitals and their usefulness for the improvement of the quality of care within hospitals.

Comparing hospitals

We found considerable influence of random variation when we compared hospitals using the outcome indicators of the Dutch Health Care Inspectorate. Using the concept of rankability we found that in the ranking of hospitals, both the between hospital uncertainty and the uncertainty of the within hospitals estimates led to unreliable positioning. Therefore, none of the tested indicators could be used for the ranking of hospitals (chapter 2). Low numbers in sample size or event rates lead to an indicator score in which differences between hospitals do not overcome random variation. Both sample size and event rates need to be addressed in the development of indicators. Furthermore, we found that a forest plot gave appropriate insight in the number of Dutch hospitals that actually significantly deviated from the average. The funnel plot provided a visual representation of differences between hospitals therewith allowing simple interpretation of the uncertainty of these differences. We also used rank plots and showed that the substantial uncertainty makes current rankings with these outcome indicators unreliable. Of all three graphical displays used, the funnel plots provided most valuable insight in the magnitude of random variation and is therefore best used for the interpretation of differences between hospitals (chapter 3). Although more research is needed to clarify the desired graphical display in specific conditions, not addressing random variation in graphical displays potentially misleads hospital assessments.

For surgical site infections, we found that the apparent differences between Dutch hospitals in this specific outcome indicator were predominantly attributable to random variation and case-mix. This case study provided a clear illustration that both random variation and case-mix must be addressed systematically in performance measure-

ment before conclusions can be drawn on the quality of hospital care (chapter 4). It is indisputable that correction for patient factors should be part of the assessment of hospital performance.

STRUCTURE-PROCESS-OUTCOME RELATION

Exploring the process-outcome relation, we found that the outcome indicator 'pressure ulcer occurrence' reflected differences in the quality of the bundle of preventive care processes provided by nurses. This significant relation between outcome and process in pressure ulcer care, supports the usefulness of this indicator in assessing the quality of nursing care. We confirmed that the pressure ulcer prevalence was also determined by several patient factors that cannot be influenced (chapter 5). Addressing the process-outcome relation in performance measurement from another angle, we explored the feasibility of measuring the effect of surveillance by the Dutch Health Care Inspectorate using retrospective data on health outcomes in three health problems: pressure ulcers, suicide and medication errors. We found that in case of clearly defined health problems, such as pressure ulcers and suicide, the frequency of these outcomes could be measured using Inspectorate data and trends could be analysed using an interrupted time series design. However, support of a causal relationship between supervision and observed trends could only be derived with data on external factors that influenced this trend (chapter 6). Although the relationship between process and outcome may seem straight forward and applicable, this is not always confirmed in research. We recommend, that the process-outcome relation should be addressed and explored in existing indicators as well as in the development of new quality indicators.

MEASURING IMPROVEMENT OF THE QUALITY OF CARE

In a process measure of acute stroke care, we found a significant improvement in "door-to-needle time" (DNT) over recent years. We could not attribute this trend to one or more specific interventions. We hypothesised that the combined effect of various interventions together and the constant focus of care-givers on quality improvement explained the significant improvement of the indicator DNT, chapter 7). In a quality project that aimed to improve pressure ulcer prevalence, we did not find a statistically significant decrease of nosocomial pressure ulcer occurrence using an interrupted time series design. However, we did see a significant improvement in the process measure risk assessment (chapter 8). Outcome indicators should be paired with process indicators to gain insight for the improvement of quality of care.

CONCLUSION

The measure of quality of care is a multidimensional and complex process. We must be aware that a performance indicator offers only a certain signal on quality and is by no means an absolute measure. Like a one hand clock, we roughly know what time it is.

SAMENVATTING

Kwaliteit van zorg is een breed en abstract begrip en de pogingen kwaliteit te meten stelt eisen aan de interpreteerbaarheid van de uitkomsten. Het doel van dit project was om de waarde van prestatie-indicatoren in de vergelijking van de kwaliteit van zorg tussen ziekenhuizen en hun nut voor de verbetering van de kwaliteit van de zorg binnen het ziekenhuiste te onderzoeken.

VERGELIJKING ZIEKENHUIZEN; TOEVALSVARIATIE EN PATIËNTFACTOREN

We vonden een aanzienlijke invloed van toevalsvariatie in de vergelijking en rangordening van ziekenhuizen met behulp van de uitkomstindicatoren van de Nederlandse Inspectie voor de Gezondheidszorg. Met behulp van het concept van rankability werd aangetoond dat een rangordening van de ziekenhuizen met behulp van deze prestatie-indicatoren onbetrouwbaar is. Toevalsvariatie, zowel tussen ziekenhuizen als binnen ziekenhuizen, beïnvloedt de onzekerheid van de rangordening. Derhalve kan geen van de onderzochte prestatie-indicatoren worden gebruikt voor de rangordening van ziekenhuizen. Lage aantallen in de noemer van de indicator (steekproef) en/of weinig voorkomende uitkomsten/events (teller) leiden tot een indicatorscore waarvan de verschillen tussen ziekenhuizen het toeval niet overstijgen. (hoofdstuk 2) De grootte van zowel de teller als de noemer moeten in de ontwikkeling van indicatoren worden meegewogen.

We onderzochten in hoeverre de grafische weergave van de indicatoruitkomsten inzicht gaf in deze toevalsvariatie. Daarbij bleek een “forrest plot” inzicht te geven in de ziekenhuizen die aanzienlijk afweken van het gemiddelde. De “funnel plot” gaf een visuele weergave van de verschillen tussen ziekenhuizen en een eenvoudige interpretatie van de onzekerheid van deze verschillen. Het “rank plot” toonde aan dat ranglijsten gebaseerd op uitkomstindicatoren onbetrouwbaar zijn. Van alle drie de gebruikte grafische displays bleek de funnel plot waardevolle inzichten te geven in de omvang van de toevalsvariatie zonder de suggestie van een rangordening. Daarom is deze het best te gebruikt voor de interpretatie van verschillen tussen ziekenhuizen

(hoofdstuk 3). Hoewel meer onderzoek nodig is naar de gewenste grafische weergaven, leidt het niet weergeven van toevalsvariatie in grafische displays tot misleidende ziekenhuis beoordelingen.

Voor de uitkomstindicator postoperatieve wondinfectie vonden we dat de schijnbare verschillen tussen de Nederlandse ziekenhuizen voornamelijk toe te schrijven waren aan toevalsvariatie en patiëntfactoren (case-mix). Deze case studie is een duidelijke illustratie dat correctie voor zowel toevalsvariatie als patiëntfactoren systematisch moet worden uitgevoerd bij het meten van zorgprestaties voordat conclusies kunnen worden getrokken over de kwaliteit van de ziekenhuiszorg (hoofdstuk 4). Het staat buiten kijf dat de correctie voor de patiëntfactoren onderdeel van de beoordeling van de prestaties van een ziekenhuis moet zijn.

PROCES-UITKOMST RELATIE

In de exploratie van de proces-uitkomst relatie vonden we dat de uitkomstindicator 'decubitusprevalentie' de verschillen in de kwaliteit van de bundel van preventieve zorgprocessen weerspiegeld. Deze significante relatie tussen uitkomst (decubitus) en proces (preventie van decubitus) onderbouwt de waarde van deze indicator bij de beoordeling van de kwaliteit van de verpleegkundige zorg. Uit het onderzoek bleek dat de prevalentie van decubitus ook significant werd beïnvloed door patiëntfactoren (hoofdstuk 5).

Met het onderzoek naar de haalbaarheid van het meten van effect van het toezicht door de Nederlandse Inspectie voor de Gezondheidszorg benaderden wij de proces-uitkomst relatie vanuit een andere hoek. Daarbij onderzochten wij de invloed van het toezicht op drie gezondheidsproblemen; decubitus, suïcide en medicatiefouten. Bij duidelijk gedefinieerde gezondheidsproblemen, zoals decubitus en suïcide, bleek dat de omvang van deze gezondheidsproblemen kon worden gemeten met behulp van retrospectieve gegevens (waaronder prestatie-indicatoren) van de Inspectie. Daarnaast bleken de trends kunnen worden geanalyseerd met een "interrupted time series design" voor en na het instellen van het toezicht. Echter, de onderbouwing van een oorzakelijk verband tussen het toezicht en de waargenomen trends kon daarmee niet worden vastgesteld. Hiervoor misten de gegevens over externe factoren die deze trend ook beïnvloeden (hoofdstuk 6).

Hoewel de relatie tussen structuur, proces en uitkomst ongecompliceerd en toepasbaar lijkt, wordt deze niet altijd bevestigd met onderzoek. Nader onderzoek naar de structuur-proces-uitkomst relatie moeten is wenselijk voor zowel bestaande indicatoren als bij de ontwikkeling van nieuwe kwaliteitsindicatoren.

METEN VAN VERBETERING VAN DE KWALITEIT VAN DE ZORG

In een procesindicator van acute zorg na een cerebro vasculair accident (beroerte of CVA) vonden we een significante verbetering van de “door-to-needle tijd” (DNT) over de gemeten jaren. We konden echter deze trend niet toeschrijven aan één of meer specifieke interventies (hoofdstuk 7). Het is mogelijk dat het gecombineerde effect van verschillende interventies en de constante focus van de zorgverleners op kwaliteitsverbetering deze aanzienlijke verbetering van de indicator DNT kan verklaren.

In een kwaliteit project dat gericht was op de verbetering van decubitusprevalentie onder chirurgische patiënten, kon er geen statistisch significante daling van nosocomiale decubitus worden aangetoond. Wel bleek er een aanzienlijke verbetering van de screening op de kans op het ontwikkelen van decubitus gemeten met de procesindicator risico-inventarisatie (hoofdstuk 8).

Uitkomstindicatoren moeten worden gecombineerd met procesindicatoren om inzicht te krijgen in de verbetering van de kwaliteit van de zorg.

CONCLUSIE

De meting van de kwaliteit van zorg is een multidimensionaal en complex proces. We moeten ons ervan bewust zijn dat een indicator slechts een beperkt signaal geeft over de zorgkwaliteit en geenszins absolute maat is. Net als een klok met één wijzer weten we ongeveer hoe laat het is.

DANKWOORD

Na 26 jaar het mooiste beroep, dat van Intensive Care verpleegkunde, te hebben uitgeoefend ging ik ging de weg van het wetenschappelijk onderzoek naast een baan als stafadviseur. Er bleek geen groter contrast denkbaar. Ik heb die weg bewandeld, bewonderd en aanschouwd, en ja, het heeft een hart (Castenada, Teachings of Don Juan, 1961). Voor u ligt het resultaat van dat besluit en dit was nooit mogelijk geweest zonder de hulp en inzet van velen aan wie ik dank wil zeggen.

Allereerst drie bijzondere mensen zonder wie ik nooit aan dit project was begonnen. Juist zij hebben de basis gelegd voor deze stap. Als eerste mijn vader aan wie ik, postuum, dit proefschrift opdraag; “Pappie, ik begrijp nu pas wat je me toen vertelde”. Atie Immink, een *hoofdzuster* uit duizenden; “dank dat je mij steeds mogelijkheden voor verdere ontwikkeling voorhield en mij leerde mijn twijfels te overwinnen”. Dr. Wilma Scholte op Reimer, verplegingswetenschapper, epidemioloog en verpleegkundige; “dank dat je mij de liefde voor het onderzoek leerde”.

Mijn promotoren, prof. dr. J.P. Mackenbach en prof. dr. E.W. Steyerberg zeg ik dank. Johan; ik vond het een eer om bij jou te promoveren. Dank voor alles wat je me leerde. Ewout; dank voor je begeleiding en speciale dank voor je vertrouwen in mijn onderzoekskwaliteiten, vooral aan het begin van dit project. Dat zorgde ervoor dat ik doorging en het afmaakte. Ik heb bewondering voor je snelle denken. Meestal begreep ik een dag later wat je bedoelde.

Dr. C.W.N. Looman, Casper, dank voor je bijdrage aan de statistische bewerking van mijn data in meerdere onderzoeken, je geduld en gezellige overleggen. Ik heb laatst nog een biertje uit de Aveyron voor je gevonden. Ik hoop dat je hem nog niet hebt.

Dr. H.L. Lingsma; Hester, ”Dank voor de samenwerking bij een groot deel van de projecten binnen dit proefschrift. Ik ben er trots op dat ik de eerste auteur was van jouw eerste artikel als laatste auteur. Ik voel me vereerd dat jij mijn paranimf bent.

Kamergenoten bij MGZ; Corine, Yvonne, Ida en Judith; dank voor gezelligheid, gesprekken en jullie goede raad.

Mijn auditpanel; dr. F. Heule, dr. J. van Bommel, dr. L. Schoonhoven, E. Strippe, I.M. van den Berg, W. Dekker-Verdoorn, B. den Boogert, J. van Boekel, B. Ruit, S. De Mar, J. Brugman, J.M.C. Blom en S. Theuns “dank voor alle tijd en moeite die jullie stopten in 132 casusbeschrijvingen. Dank voor jullie professionele oordeel”.

Joke Brugman en Astrid Paul, verpleegkundig aandachtsvelder decubitus en wondzorg, jullie wil ik danken voor de samenwerking bij het kwaliteitsproject en jullie niet aflatende doorzettingsvermogen om de decubituszorg te verbeteren. Het succes wat jullie maakten van dit project legde voor mij de basis van een onderzoek. Joke, ik ben er trots op dat jij na dit project ook de kwaliteit van de wondzorg bent gaan aanpakken. Ik help je erg graag mee.

Drs. G. Venderbos-Smith; Georgina dank voor je tijd en geduldige beoordeling en bijdrage aan mijn Engelstalige teksten. Als geen ander kan jij de essentie op de juiste plaats in een zin zetten.

Wondexpertise Erasmus MC; Prof. dr. S.E.R. Hovius, dr. C. Dekker-van Doorn en drs. D. van Duijn; Steven, Connie en Dick; “jullie wil ik dank zeggen voor de veilige haven van de Werkgroep decubitus- en wondzorg, van waaruit ik dit onderzoeksproject heb kunnen doen. Ook dank aan mijn collega’s en in het bijzonder Dymmie Landa; Dymmie dank voor de samenwerking in de opzet van jouw onderzoek naar wondanalyse en wondregistratie. Jij liet mij kennis maken met de wondere wereld van de wondzorg. Ik bouw nog steeds voort op jouw kennis daarvan.

Lieve vrienden en vriendinnen; de laatste jaren is er te weinig tijd geweest om met iedereen contact te houden. Ik dank jullie voor alle contact die jullie telkens weer met mij opnamen en jullie interesse in dit project.

Choco+, mijn bijzondere groep van studievrienden met chocola als basis; Dank voor jullie steun en interesse. Ik had te weinig tijd de laatste paar jaar en hoop jullie nu weer vaker te zien. Als laatste in de lijn van Choco+ PhD’s kan ik alleen nog vragen: “Wie volgt?”

Dank aan mijn buren en oud-buren voor jullie interesse en bemoediging tijdens dit project. Jullie maken (maakten) de Spoorsingel een mooi plekje om te wonen.

Al varend leert men bijzondere mensen kennen, die in de kuip luisterden naar mijn verhalen over onderzoek. Arie en Gudi, Georg en Ulrike, Rick en Ineke, Cor en Betty, Just en Marjolein dank voor jullie aandacht.

Wat is een dankwoord toch moeilijk! Ik blijf veranderen en ben zelfs nu vast nog mensen vergeten. Kan het eerste deel van deze zin nog gewijzigd worden?

Vijf geweldige vrouwen wil ik noemen omdat zij mij bij de gewone dingen van het leven hielden. Margreet, Henny, Wanny (en Roland), Marjolein en Antoinette,, dank voor koffietijd, ontbijtjes, glaasjes en dinertjes kletsend over kinderen, kleinkinderen, kleren naaien, poppen en beren maken en alle andere alledaagse dingen van het leven. Jullie gaven mij de broodnodige afleiding.

Speciale dank aan Robert en Fredy Pleijsier voor heerlijke vakanties samen en glaasjes achter een anker. Bij de volgende vakantie laat ik mijn laptop thuis.

En dan mijn familie aan wie ik dank wil zeggen;

Dudok van Heel-en, ooms, tantes, neven nichten; wie weet heb ik nu weer eens tijd om langs te komen.

Visée-s; Mo, Peta, Pierre, Thomas, Antoine, Olivier, Axel, Isabelle, Niels en Fleur; dank voor gezelligheid en het goede Franse leven.

Boelens-en; de hele Boel, Marian, Joop, Maryciel, Marcel, Laìs, Laszlo en Charlotte; ik hoop op nog vele mooie Boelensweekeinden.

Van Dishoeck-en; Nelleke, Jaap, Florentien, Daniël, Matthijs, Pieter, Anja, Jasper, Tijmen, Rozemarijn en Caroline, dank voor jullie meeleven met mijn ambities en mijn project. Lieve Mammie; ergens op een wolkje ben jij trots op mij.

Een dankwoord voor een schip; het is misschien niet gebruikelijk, maar ik doe het toch. Mijn Groot Frisia; twee mastjes, een paar zeiltjes. Jij maakt het leven eenvoudig en genietbaar. Met jou komt een mens tot rust.

Lieve Pien en Naud, als ik had geweten dat kleinkinderen zo leuk waren, dan was ik daar eerst aan begonnen (onbekend). Jullie laten mij opnieuw zien hoe men met verwondering naar de eenvoudige dingen van het leven moet kijken. Ama heeft nu eindelijk meer tijd voor jullie.

Kim en Pauline, mijn schoondochters, ik ben gelukkig dat juist jullie mijn gezin versterken en er een hele nieuwe (vrouwelijke) dimensie aan geven. Jullie zijn, elk op eigen wijze, bijzondere vrouwen en ik dank jullie voor jullie betrokkenheid en aanmoediging bij dit project.

Ewout en Folkert, jullie zijn het beste wat mij in mijn leven is overkomen. Klein waren jullie “mijn jongens” en eenmaal groot en tijdens Paps ziekte waren jullie de hoeksteen van mijn bestaan. Jullie maken mij een rijk mens. Ewout, dank voor je hulp bij mijn opleiding epidemiologie en je geduldig lezen van mijn teksten (“*Mam: wat wil je eigenlijk zeggen?* en “*doet het pijn om een punt te zetten?*”). Jij naast mij tijdens mijn verdediging betekent veel voor mij.

Lieve Frits; jij was mijn grootste criticus en mijn grootste bewonderaar tijdens dit project. Je plande en organiseerde mijn vakanties en lag geduldig voor anker als ik wilde “werken”. Met jou kan ik blij zijn, boos zijn, gelukkig zijn, verdrietig zijn, mijzelf zijn. Je kunt niets zeker weten en alles gaat voorbij, maar ik geloof in jou en mij (B de Groot). Eindelijk tijd om samen te reizen, te vliegen, te zeilen of toch een huisje in Frankrijk? Ik hou van je.

ABOUT THE AUTHOR

Adriana Margaretha (Anne-Margreet) van Dishoeck (The Hague, 1956) studied Health Science at the University Utrecht and graduated in 2005 on *the reliable assessment of sedation level in routine clinical practice by adding an instruction to the Ramsay Scale*. In addition to her function as intensive care nurse, she performed research into quality of care for two intensive care units from 2002-2006.

In 2006 she started her PhD project at the department of Public Health at the of the Erasmus MC Medical Center Rotterdam and studied Epidemiology at the Netherlands Institute for Health Sciences (NIHES) Erasmus MC Rotterdam. She graduated in 2009 on *displaying random variation in comparing hospital performance*. Next to her PhD project, she works at the department of Plastic & Reconstructive Surgery and Hand Surgery of the Erasmus MC Medical Center Rotterdam as a staff advisor and board member of the Wound Care Center since 2007. Additional to her advisory work, she performs and coordinates research into the quality of wound care.

She started career as a nurse in 1975 at the oldest Dutch hospital, the Oude en Nieuwe Gasthuis in Delft, the Netherlands and worked as a nurse at the Hippolytes Ziekenhuis in Delft. From 1980 after an intensive training program, she continued her professional line of work as an Intensive Care nurse in the Surgical Intensive Care unit, Intensive Coronary Care Unit and the Intensive Care Thorax surgery at the Erasmus MC Medical Center in Rotterdam until 2006.

Between 1990 and 2006, she also lectured hemodynamics and left ventricle assist devices at the Health Academy of the Erasmus MC University Medical Centre in Rotterdam at the training for intensive care nurse. Since 2014, she lectures evidence based practice (EBP) at the Health Care Academy. From 1992 to 1999 she was editor of the Dutch nursing journal Cordiaal, initially published by the Dutch Heart Foundation and later by the Dutch Association for Cardiovascular Nurses.

Anne-Margreet is married to Frits Boelens, mother of Ewout and Folkert Boelens, mother in law of Kim Semeijn and Pauline van Leeuwen-Boelens and grandmother (Ama) of Pien and Naud Boelens.

PUBLICATIONS

Publications in scientific peer-reviewed journals

1. van Dishoeck A.M., Steyerberg E.W. Ingezonden brief, Ziekenhuizen top 100; wisselende ranglijsten, wisselende reputatie. Ned Tijdschr Geneeskd 2007;151(4):265-6.
2. van Dishoeck, A. M., T. van der Hooft, et al. (2009). "Reliable assessment of sedation level in routine clinical practice by adding an instruction to the Ramsay Scale." Eur J Cardiovasc Nurs **8**(2): 125-8. Poster presentation: Acute Cardiac Care 2006;vol 8 supplement 2:59-60.
3. van Dishoeck, A. M., C. M. Looman, et al. (2009). "[Outcome assessment in hospitals. The influence of insecurity]Prestatie-indicatoren voor ziekenhuizen. De invloed van onzekerheid." Ned Tijdschr Geneeskd **153**(17): 804-11
4. van Dishoeck, A.M., Lingsma H.F., Steyerberg E.W. (2010) Prestatie-indicatoren; de invloed van taakonzekerheid. Artikelbespreking Nederlands Tijdschrift voor Geneeskunde. [Performance indicators: the role of 'task uncertainty'] Ned Tijdschr Geneeskd. 2010;154:A1775.
5. van Dishoeck, A.M., Looman, C.M.N., van der Wilden-van Lier, E.C.M., Mackenbach, J.P. Steyerberg, E.W. Displaying random variation in comparing hospital performance. BMJ Quality & Safety in Health Care, 2011;20:651e657. doi:10.1136/bmjqs.2009.035881 Poster presentation International Forum on Quality & Safety in Health Care, Nice apr. 2011
6. van Dishoeck, A.M., Lingsma, H.F., Mackenbach, J.P., Steyerberg, E.W. Random variation and rankability of hospitals using outcome indicators. BMJ Qual Saf 2011;20:869-874. Poster presentation International Forum on Quality & Safety in Health Care, apr. 2010
7. van Dishoeck AM, Koek MB, Steyerberg EW, van Benthem BH, Vos MC, Lingsma HF. Use of surgical-site infection rates to rank hospital performance across several types of surgery. Br J Surg. 2013;100:628-636
8. van Dishoeck AM, Oude Wesselink SF, Lingsma HF, Steyerberg E, Robben PB, Mackenbach JP. [transparency: Can the effect of governmental surveillance be quantified?] transparantie: Is het effect van toezicht te meten? Ned Tijdschr Geneeskd. 2013;157:A1676
9. van Dishoeck, A.M., Dippel, D., Dirks M., Looman C., Mackenbach, J.P., Steyerberg, E.W. Measuring Quality Improvement in Acute Ischemic Stroke Care: Interrupted Time Series Analysis of Door-to-Needle Time. Cerebrovasc Dis Extra 2014;4:149–155
10. Landa DLC, van Dishoeck AM, Steyerberg EW, Hovius SER. Quality of measurements of acute surgical and traumatic wounds using a digital wound-analysing tool. International Wound Journal 2014; doi: 10.1111/iwj.12330

OTHER PUBLICATIONS

1. van Dishoeck AM, Reanimatielessen aan huisgenoten van hartpatiënten. Cordiaal 1995;4:141-4
2. van Dishoeck AM, CPAP, een zuurstof masker met iets meer. Cordiaal 1995; 5:168-71
3. van Dishoeck AM, Acute Pericarditis. Cordiaal 1999;3: 82-7
4. van Dishoeck AM, Linker Ventrikel Assist Devices (LVAD). Cordiaal 1999; 5:158-62
5. Van Dishoeck AM, Eenkhoorn E., Non invasief beademen; een duik in de literatuur. Cordiaal 2000;375-6
6. van Dishoeck AM, Het Percutane Linker Ventrikel Assist Device. Cordiaal 2002; 2:22-4

7. van Dishoeck AM, Frankhuizen P., Giljam J., Scholte op Reimer WJM. Aanwezigheid van Familie bij een Reanimatie; een case-report. Cordiaal 2004;4:
8. van Dishoeck AM, Het acute myocardinfarct en de psychosociale concepten binnen deze crisis: een literatuuronderzoek. Cordiaal 2005;3:
9. van Dishoeck AM, Koppelaar C., Bevochtigen van zuurstof. Cordiaal 2006;1:40-3
10. van Dishoeck AM, Luttik M.L. De theorie van Kwalitatief onderzoek Cordiaal 2007;3:98
11. van Dishoeck AM, van 't Verlaat P.G. Power of onderscheidend vermogen van een onderzoek. Cordiaal 2008; 4: 132-4.
12. van Dishoeck, AM, Bibollet-Ruche, E., Buijl-Nout, H., Verhoef, K., Hovius, S.E.R. Wondzorg na vulvectomie. Oncologica 2010
13. van Dishoeck AM, Van 't Verlaat P.G. Casestudy; een onderzoeksmethode voor verpleegkundigen. Cordiaal 2011; 4: 134-5.
14. van Dishoeck AM, Lenzen ML, Milde hypothermie na een out-of hopital reanimatie; wetenschappelijk bewijs nader bekeken. Cordiaal 2012;5:152-4.
15. van Dishoeck AM, Hovius SER. Valkuilen van decubitusprevalentie in ziekenhuizen als maat voor de verpleegkundige zorgverlening. WCS nieuws 2013;29(1):58-61
16. de Waard BE, van Dishoeck AM, Hovius SER. Behandeling van Hypergranulatie in de wondzorg. WCS nieuws 2014;30(2):10-4
17. van Dishoeck AM, Nijssen WE, Hovius SER. Veranderende wetgeving en financiering in de zorg. Cordiaal 2014; 3:*pag*
18. van Dishoeck AM, Snaterse M, van der Wetering H, Een succesvolle abstract schrijven; de eerste stap naar een posterpresentatie! Cordiaal 2015; 3:*pag*

PHD PORTFOLIO

Name PhD student: A.M. van Dishoeck, MSc.	PhD period: 2006-2015
Erasmus MC Department: Public Health	Promotor 1: prof. dr. J.P. Mackenbach
Research School: NIHES	Promotor 2: prof. dr. E.W. Steyerberg

Subject	Year	ECTS
1. PhD training		
NIHES Summer course	2006	5,6
Study design (ESPCC01)		
o Follow-up studies and case control studies	2006	4,3
o Methodological topics in Epidemiologic research	2006	1,4
Data-analysis		
o Classical methods CC02	2007	5,7
o Modern Statistical Methods (EP03)	2007	4,3
Program-specific courses		
o Analysis of Population Health HS02a	2006	1,4
o Analysis of Determinants HS02b	2006	1,4
o Intervention Development and Evaluation HS02c	2006	1,4
Elective courses 2		
o Introduction to clinical research	2007	0,7
o Ethnicity, Health and Health care	2007	1,1
o Health Services: Research and Practice	2007	0,9
o Analysis of time-varying exposures	2008	0,7
Skills courses		
o Working with SPSS (SC04)	2007	0,15
o Good clinical practice (BROK)	2008	1
o Re-certification BROK	2013	
o Biomedical English Writing and Communication	2009	4
o Research Integrity	2010	1
Department journal club	2007-2014	6
Presentations at seminars and workshops;		
o Symposium Dutch Inspectorates. Workshop; Kwantificeren van effect van toezicht, is het haalbaar?	2009	1
o		
International conferences		
o Oral presentation . Meeting the Reference group of the Dutch and Norwegian Board of Health Supervision Rotterdam, January 22th, 2010. Institute of Health Policy and Management (iBMG); Quantifying the effect of governmental surveillance.	2010	1,0

Subject	Year	ECTS
o Poster presentation; International Forum on Quality & Safety in Health Care, apr 2010. Random variation and rankability of hospitals using outcome indicators.	2010	1,0
o Poster presentation; International Forum on Quality & Safety in Health Care, Nice apr. 2011 Displaying random variation in comparing hospital performance.	2011	1,0
o Oral presentation. Fifteenth Annual European Pressure Ulcer Advisory Panel Meeting; The prevalence of pressure ulcers or incontinence dermatitis reflects the poor quality of care in adult hospitalized patients.	2012	1,0
o Poster presentation; Can surgical site infections reflect hospital performance? Statistical analysis of the influence of random variation and case mix	2012	1,0
2. Teaching		
Master class EBP for Nurse Practitioners	2012 - 2013	2,0
Module EBP (42 hours per module twice a year)	Since2014	3,0
Continuous Nursing Education CNE's Werkgroep Wetenschappelijk Onderzoek van de Nederlandse Vereniging van Hart- en Vaatverpleegkundigen	2009-2015	6,0
Supervising medical students in 4th year research projects; • Surgical Site Infection after gastrointestinal surgery, influence of factors concerning quality of care	2014	1,0
• Feasibility of tele monitoring in wound care	2015	1,0
Supervising Master's theses		

This thesis addresses two major topics in measuring, comparing and improving quality of care. We found considerable influence of random variation and case-mix in comparing hospitals using performance indicators. Although we found a significant relation between outcome and care processes, chance variation is the major limitation for the interpretability of indicators used for quality measurement or quality improvement. Like a one hand clock, we roughly know what time it is.